Chapter 0

Preliminaries

Notation: In what follows, we will use the following notation:

- $\mathbb{N}$ is the set of all natural numbers (positive integers);
- $\mathbb{N}_0$ is the set of all non-negative integers;
- $\mathbb{Z}$ is the set of all integers;
- $\mathbb{Q}$ is the set of all rational numbers;
- $\mathbb{R}$ is the set of all real numbers;
- $\mathbb{C}$ is the set of all complex numbers.

0.1 Mathematical induction

Mathematical induction is a proof technique that can be used to prove that a certain statement holds for all positive integers n.

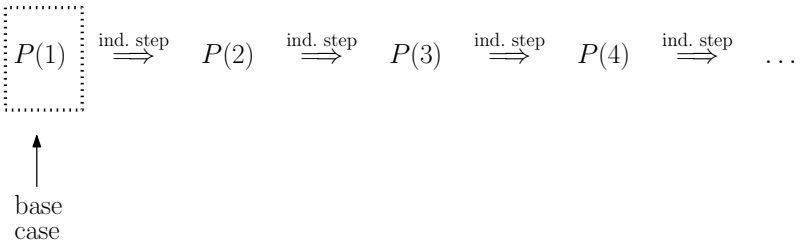
Let $P(n)$ be a statement about the number n. In order to prove that $P(n)$ holds for every positive integer n, it suffices to prove the following two statements:

- **Base case:** $P(1)$ is true;
- **Induction step:** for every positive integer n,
  if $\underbrace{P(n) \text{ is true}}_{\text{“induction hypothesis”}}$, then $P(n+1)$ is true.

Why does this work? Here is the intuition: We are trying to prove an infinite sequence of statements, namely,

$$P(1),\ P(2),\ P(3),\ P(4),\ \ldots$$

By the base case, $P(1)$ is true. By the induction step, since $P(1)$ is true, $P(2)$ is also true. Again by the induction step, since $P(2)$ is true, so is $P(3)$. Once again by the induction step, since $P(3)$ is true, so is $P(4)$. And so on! Thus, $P(n)$ is true for all positive integers n. Schematically, this is shown in the diagram below.



Example 0.1.1. *Prove that* $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$ *for all positive integers* n.

Solution. Let $P(n)$ be the statement that $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. Thus:

- $P(1)$ is the statement that $1 = \frac{1\cdot(1+1)}{2}$;
- $P(2)$ is the statement that $1 + 2 = \frac{2\cdot(2+1)}{2}$;
- $P(3)$ is the statement that $1 + 2 + 3 = \frac{3\cdot(3+1)}{2}$;
- etc.

We need to prove that the statement $P(n)$ is true for all positive integers n.

Base case: $n = 1$. Obviously, $1 = \frac{1\cdot(1+1)}{2}$. Thus, $P(1)$ is true.

Induction step: Fix a positive integer n, and assume inductively that $P(n)$ is true. We must show that $P(n+1)$ is true.

The induction hypothesis states that $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. Using this, we must prove that $1 + 2 + \cdots + n + (n+1) = \frac{(n+1)\big((n+1)+1\big)}{2}$. We compute:

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= (1 + 2 + \cdots + n) + (n+1) \\ &= \frac{n(n+1)}{2} + (n+1) && \text{by the induction hypothesis} \\ &= (n+1)(\tfrac{n}{2} + 1) \\ &= \frac{(n+1)\big((n+1)+1\big)}{2}. \end{aligned}$$

Thus, $P(n+1)$ is true. This completes the induction. □

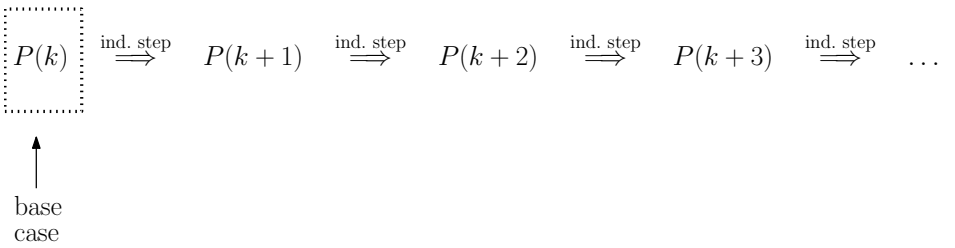
Sometimes, the base case may be different from $n = 1$. This may happen if we need to prove that a statement $P(n)$ is true for all $n \geq k$, where k is an integer other than 1. (Typically, we will have $k = 0$ or $k > 1$. However, in principle, k may even be a negative integer.) In this case, the base case will be $n = k$, i.e. we will need to prove the following two statements:

- **Base case:** $P(k)$ is true;
- **Induction step:** for every integer $n \geq k$,
  if $\underbrace{P(n) \text{ is true}}_{\text{"induction hypothesis"}}$, then $P(n+1)$ is true.

Here, the intuition is similar to what we saw above. We are trying to prove an infinite sequence of statements, namely

$$P(k),\ P(k+1),\ P(k+2),\ P(k+3),\ \ldots$$

By the base case, $P(k)$ is true. By the induction step, since $P(k)$ is true, $P(k+1)$ is also true. Again by the induction step, since $P(k+1)$ is true, so is $P(k+2)$. Once again by the induction step, since $P(k+2)$ is true, so is $P(k+3)$. And so on! Thus, $P(n)$ is true for all integers $n \geq k$. Schematically, this is shown in the diagram below.



Example 0.1.2. *Prove that $3n < 2^n$ for all integers $n \geq 4$.*

Proof. Since we are proving the statement for integers $n \geq 4$, our base case is $n = 4$.

Base case: $n = 4$. Clearly, $3 \cdot 4 = 12 < 16 = 2^4$.

Induction step: Fix an integer $n \geq 4$, and assume inductively that $3n < 2^n$. We must show that $3(n+1) < 2^{n+1}$. We observe the following:

$$\begin{aligned} 3(n+1) &= 3n + 3 \\ &< 2^n + 3 && \text{by the induction hypothesis} \\ &< 2^n + 2^2 \\ &< 2^n + 2^n && \text{because } n > 2 \\ &= 2^{n+1} \end{aligned}$$

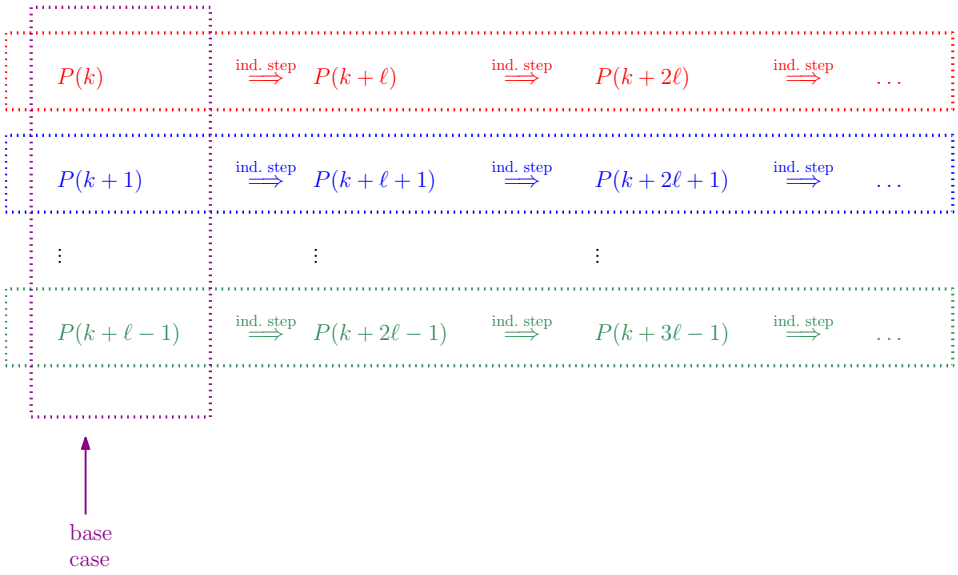
Thus, the statement is true for $n + 1$. This completes the induction. □

0.1.1 Induction with more than one base case

Suppose that k is an integer, and that we wish to prove inductively that $P(n)$ holds for all integers $n \geq k$. Ordinarily, we would expect $n = k$ to be the base case. However, suppose that we do not know how to prove the implication "$P(n) \implies P(n+1)$," but we do know how to prove that "$P(n) \implies P(n+\ell)$," where ℓ is some positive integer (other than 1). In this case, we will have a slightly modified induction step ("$P(n) \implies P(n+\ell)$" instead of "$P(n) \implies P(n+1)$"), and we will have ℓ base cases, namely, $P(k), P(k+1), \ldots, P(k+\ell-1)$. More precisely, we will need to prove the following:

- **Base case:** $P(k), P(k+1), \ldots, P(k+\ell-1)$ are true;
- **Induction step:** for every integer $n \geq k$,
  if $\underbrace{P(n) \text{ is true}}_{\text{"induction hypothesis"}}$, then $P(n+\ell)$ is true.

Indeed, this is enough to show that $P(n)$ holds for all integers $n \geq k$. The intuition behind this is given in the diagram below.



Example 0.1.3. *Suppose you have an unlimited number of 3 Kč stamps and 5 Kč stamps (and no other stamps). Show that you can pay any amount of postage greater or equal to 8 Kč (as long as it is in whole Kč).*

Solution. We need to show that any integer $n \geq 8$ (our postage in Kč) can be expressed in the form

$$n = 3a + 5b,$$

where a and b are non-negative integers (the number of 3 Kč and 5 Kč stamps, respectively, that we can use to pay our n Kč postage). We will prove this by

induction on n. Obviously, if we can pay n Kč using our stamps, then we can also pay $(n+3)$ Kč: we simply use one 3 Kč stamp more. In other words, if the statement is true for n, then it is also true for $n+3$. This means that we will need three base cases: $n = 8$, $n = 9$, and $n = 10$. Let us give the details.

Base case: We must show that for each $n \in \{8, 9, 10\}$, there exist non-negative integers a and b such that $n = 3a + 5b$. But this is clearly true:

- $8 = 3 \cdot 1 + 5 \cdot 1$;
- $9 = 3 \cdot 3 + 5 \cdot 0$;
- $10 = 3 \cdot 0 + 5 \cdot 2$.

Induction step: Fix an integer $n \geq 8$, and assume inductively that the statement is true for n. We must show that it is true for $n+3$. By the induction hypothesis, these exist non-negative integers a and b such that $n = 3a+5b$. But then $n+3 = 3(a+1)+5b$, and so the statement holds for $n+3$. This completes the induction. □

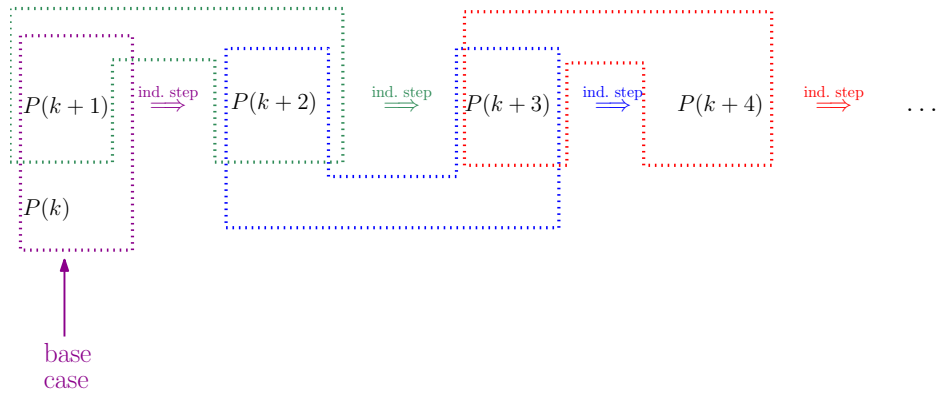
Suppose, again, that k is an integer, and that we wish to prove inductively that $P(n)$ holds for all integers $n \geq k$. However, suppose that we are not able to prove the implication "$P(n) \implies P(n+1)$," but that we are able to prove that $P(n), P(n+1), \ldots, P(n+\ell-1)$ together imply $P(n+\ell)$, where ℓ is some positive integer (other than 1). In this case, we will again have ℓ base cases, namely, $P(k), P(k+1), \ldots, P(k+\ell-1)$. More precisely, we will need to prove the following:

- **Base case:** $P(k), P(k+1), \ldots, P(k+\ell-1)$ are true;
- **Induction step:** for every integer $n \geq k$,
  if $\underbrace{P(n), P(n+1), \ldots, P(n+\ell-1) \text{ are all true}}_{\text{"induction hypothesis"}}$, then $P(n+\ell)$ is true.

Once again, this is enough to show that $P(n)$ holds for all integers $n \geq k$. The intuition behind this is as follows:

- $P(k), P(k+1), \ldots, P(k+\ell-1)$ hold by the base case;
- since $P(k), P(k+1), \ldots, P(k+\ell-1)$ hold, the induction hypothesis guarantees that $P(k+\ell)$ holds;
- now $P(k+1), P(k+2), \ldots, P(k+\ell)$ hold, and so by the induction hypothesis, $P(k+\ell+1)$ holds;
- now $P(k+2), P(k+3), \ldots, P(k+\ell+1)$ hold, and so by the induction hypothesis, $P(k+\ell+2)$ holds;
- and so on!

For the case when $\ell = 2$, the idea behind this is illustrated in the diagram below.



Example 0.1.4. *The* Fibonacci numbers *are defined as follows:*

- $F(1) = F(2) = 1$*;*
- $F(n+2) = F(n) + F(n+1)$ *for all positive intgers* n*.*

Prove that $F(n) = \frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\sqrt{5}}$ *for all positive integers* n*.*

Solution. The general term is defined in terms of the previous two terms. Thus, instead of one base case, we have two: $n = 1$ and $n = 2$.

Remark: If the general term were defined in terms of, say, the previous fifteen terms, then we would have fifteen base cases!

Base case: For $n = 1$, we have:

$$\frac{(1+\sqrt{5})^1 - (1-\sqrt{5})^1}{2^1\sqrt{5}} = \frac{2\sqrt{5}}{2\sqrt{5}} = 1 = F(1).$$

For $n = 2$, we have:

$$\frac{(1+\sqrt{5})^2 - (1-\sqrt{5})^2}{2^2\sqrt{5}} = \frac{(1+2\sqrt{5}+5) - (1-2\sqrt{5}+5)}{4\sqrt{5}} = \frac{4\sqrt{5}}{4\sqrt{5}} = 1 = F(2).$$

Thus, the statement is true for $n = 1$ and $n = 2$.

Induction step: Fix a positive integer n, and assume inductively that the statement is true for n and $n+1$. We must show that it is true for $n+2$.

By the induction hypothesis, we have that

- $F(n) = \frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\sqrt{5}}$;
- $F(n+1) = \frac{(1+\sqrt{5})^{n+1} - (1-\sqrt{5})^{n+1}}{2^{n+1}\sqrt{5}}$.

We must show that $F(n+2) = \frac{(1+\sqrt{5})^{n+2} - (1-\sqrt{5})^{n+2}}{2^{n+2}\sqrt{5}}$. We compute:

$$\begin{aligned}
F(n+2) &\overset{(*)}{=} F(n) + F(n+1) \\
&\overset{(**)}{=} \tfrac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\sqrt{5}} + \tfrac{(1+\sqrt{5})^{n+1} - (1-\sqrt{5})^{n+1}}{2^{n+1}\sqrt{5}} \\
&= \tfrac{4(1+\sqrt{5})^n - 4(1-\sqrt{5})^n}{2^{n+2}\sqrt{5}} + \tfrac{2(1+\sqrt{5})(1+\sqrt{5})^n - 2(1-\sqrt{5})(1-\sqrt{5})^n}{2^{n+2}\sqrt{5}} \\
&= \tfrac{(6+2\sqrt{5})(1+\sqrt{5})^n - (6-2\sqrt{5})(1-\sqrt{5})^n}{2^{n+2}\sqrt{5}} \\
&= \tfrac{(1+\sqrt{5})^2(1+\sqrt{5})^n - (1-\sqrt{5})^2(1-\sqrt{5})^n}{2^{n+2}\sqrt{5}} \\
&= \tfrac{(1+\sqrt{5})^{n+2} - (1-\sqrt{5})^{n+2}}{2^{n+2}\sqrt{5}},
\end{aligned}$$

where (*) follows from the definition of Fibonacci numbers, and (**) follows from the induction hypothesis. This completes the induction. □

0.1.2 Strong induction

We now discuss a type of induction (sometimes called "strong induction") that lacks a base case. Again, let $P(n)$ be a statement about the number n. In order to prove that $P(n)$ holds for every positive integer n, it suffices to prove the following:

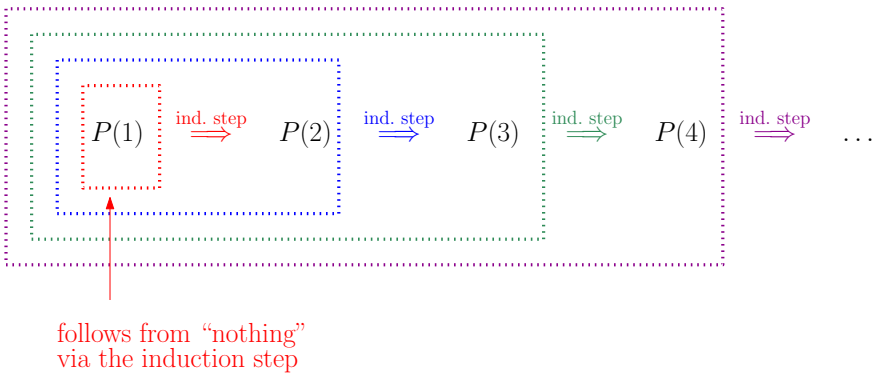
- **Induction step:** for every positive integer n,
  if $\underbrace{P(1), \ldots, P(n-1) \text{ are all true}}_{\text{"induction hypothesis"}}$, then $P(n)$ is true.

Here is a slightly different way of writing the same thing:

- **Induction step:** for every positive integer n,
  if $\underbrace{P(i) \text{ is true for all positive integers } i < n}_{\text{"induction hypothesis"}}$, then $P(n)$ is true.

Why does this make sense? Here is the intuition. Suppose that we have proven the induction step above. For $n = 1$, the induction hypothesis is vacuously true,[1] and so it follows that $P(1)$ is true. Put in another way, for $n = 1$, the induction step essentially says "if 'nothing,' then $P(1)$ is true," which is the same as "$P(1)$ is true." What about $P(2)$, $P(3)$, $P(4)$, etc.? Since $P(1)$ is true, the induction step guarantees that $P(2)$ is true. Now $P(1), P(2)$ are true; so, by the induction step, $P(3)$ is true. Now $P(1), P(2), P(3)$ are true; so, by the induction step, $P(4)$ is true. And so on! The intuition behind this is summarized in the diagram below.

[1] This is because there are no positive integers $i < 1$.



As before, slight variations on the theme are possible. In particular, for a fixed integer k, we may wish to prove by strong induction that $P(n)$ holds for all integers $n \geq k$. In this case, it is enough to prove the following:

- **Induction step:** for every integer $n \geq k$,
  if $\underbrace{P(k), \ldots, P(n-1) \text{ are all true}}_{\text{"induction hypothesis"}}$, then $P(n)$ is true.

Another way of writing the same thing is as follows:

- **Induction step:** for every integer $n \geq k$,
  if $\underbrace{P(i) \text{ is true for all integers } i \text{ such that } k \leq i < n}_{\text{"induction hypothesis"}}$, then $P(n)$ is true.

Example 0.1.5. *Prove that every integer $n \geq 2$ can be written as a product of one or more prime numbers.*

Proof. Fix an integer $n \geq 2$, and assume inductively that each of $2, \ldots, n-1$ can be written as a product of primes.[2] We must show that n can be written as a product of primes.

Clearly, n is either prime or composite.

Suppose first that n is prime. Then, obviously, n can be written as a product of primes, namely

$$n = \underbrace{n}_{\text{prime}}.$$

Suppose now that n is composite. Then there exist integers n_1, n_2 such that $2 \leq n_1, n_2 < n$ and $n = n_1 n_2$. By the induction hypothesis, n_1 and n_2 can be written as products of primes. Set $n_1 = p_1 \cdot \cdots \cdot p_k$ and $n_2 = q_1 \cdot \cdots \cdot q_\ell$, where $p_1, \ldots, p_k, q_1, \ldots, q_\ell$ are prime numbers. Then $n = n_1 n_2 = p_1 \cdot \cdots \cdot p_k \cdot q_1 \cdot \cdots \cdot q_\ell$. Thus, n is a product of primes. This completes the induction. □

[2] In other words, we are assuming that for all integers m such that $2 \leq m < n$, m can be written as a product of primes. Note that if $n = 2$, then we are in fact not assuming anything because there are no integers m satisfying $2 \leq m < 2$.

0.2 Modular arithmetic. Arithmetic in $\mathbb{Z}_n$

0.2.1 Modular arithmetic

Given $n \in \mathbb{N}$ and $m \in \mathbb{Z}$, we write $n \mid m$ if m is divisible by n, that is, if there exists some $k \in \mathbb{Z}$ such that $m = kn$.

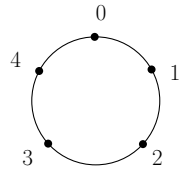
Given $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$, we say that a and b are *congruent modulo* n, and we write $a \equiv b \pmod{n}$ or $a \equiv_n b$, provided that $n|(a-b)$, i.e. $a - b = kn$ for some $k \in \mathbb{Z}$; equivalently, we have that $a \equiv b \pmod{n}$ provided that a and b leave the same remainder when divided by n (where the remainder is required to be one of the integers $0, 1, \ldots, n-1$). Note that for a positive integer n and an integer a, we have that a is divisible by n (equivalently: a is a multiple of n) if and only if $a \equiv 0 \pmod{n}$.

Example 0.2.1. *All the following hold:*

- $2 \equiv 17$ *(mod* 3*);*
- $-13 \equiv 8$ *(mod* 7*);*
- $-1 \equiv 7$ *(mod* 4*);*
- $2 \not\equiv 17$ *(mod* 2*);*
- $-13 \not\equiv 8$ *(mod* 5*);*
- $-1 \not\equiv 7$ *(mod* 6*).*

Remarks:

- For fixed $n \in \mathbb{N}$, every integer is congruent modulo n to exactly one of the following n integers: $0, \ldots, n-1$. As we shall see, doing arithmetic modulo n essentially boils down to doing arithmetic with only n values (namely $0, \ldots, n-1$), as opposed to infinitely many. This is quite useful for certain applications.
- Congruence modulo n can be visualized in terms of an "n-hour clock" (see the picture below for the case $n = 5$). Suppose we are given an integer a, and we wish to determine which of $0, 1, \ldots, n-1$ it is congruent to modulo n. Obviously, if $a = 0$, then $a \equiv 0 \pmod{n}$. If a is positive, then we start at 0 and make n clockwise steps; the number we finish at is the number we need. For example, we have that $14 \equiv 4 \pmod{5}$ because if we start at 0 and then make 14 steps clockwise on the 5-hour clock, we finish at 4. On the other hand, if a is negative, then we make $|a| = -a$ many counterclockwise steps. For example, we have that $-7 \equiv 3 \pmod{5}$ because if we start at 0 and then make 7 steps counterclockwise on the 5-hour clock, then we finish at 3.



Proposition 0.2.2. *Let $n \in \mathbb{N}$ and $a, b, c \in \mathbb{Z}$. Then the following hold:*

(a) $a \equiv a \pmod{n}$;

(b) if $a \equiv b \pmod{n}$, then $b \equiv a \pmod{n}$;

(c) if $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $a \equiv c \pmod{n}$.

Proof. (a) and (b) are obvious. For (c), assume that $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$. Then $n \mid (a-b)$ and $n \mid (b-c)$, i.e. there exist $k, \ell \in \mathbb{Z}$ such that $a - b = kn$ and $b - c = \ell n$. But then

$$a - c \;=\; (a-b) + (b-c) \;=\; kn + \ell n \;=\; (k+\ell)n,$$

i.e. $n \mid (a-c)$. Thus, $a \equiv c \pmod{n}$. □

Remark: Proposition 0.2.2 states that congruence modulo n is an "equivalence relation" on $\mathbb{Z}$. (If you are not yet familiar with equivalence relations, you will soon learn about them in Discrete Math.)

Proposition 0.2.3. *Let $n \in \mathbb{N}$ and $a, b, c, d \in \mathbb{Z}$, and assume that $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$. Then:*

(a) $a + c \equiv b + d \pmod{n}$;

(b) $a - c \equiv b - d \pmod{n}$;

(c) $ac \equiv bd \pmod{n}$.

Proof. Since $a \equiv b \pmod{n}$, we have that $n|(a-b)$, and so there exists some $k \in \mathbb{Z}$ such that $a - b = kn$. Similarly, since $c \equiv d \pmod{n}$, there exists some $\ell \in \mathbb{Z}$ such that $c - d = \ell n$.

To prove (a), we observe that

$$(a+c) - (b+d) \;=\; (a-b) + (c-d) \;=\; kn + \ell n \;=\; (k+\ell)n,$$

and so $n \mid \big((a+c) - (b+d)\big)$. Thus, $a + c \equiv b + d \pmod{n}$. This proves (a).

For (b), we observe that

$$(a-c) - (b-d) \;=\; (a-b) - (c-d) \;=\; kn - \ell n \;=\; (k-\ell)n,$$

and so $n \mid \big((a-c) - (b-d)\big)$. Thus, $a - c \equiv b - d \pmod{n}$. This proves (b).

Finally, for (c), we have that

$$\begin{aligned} ac - bd &= ac - ad + ad - bd \\ &= a(c-d) + (a-b)d \\ &= a\ell n + knd \\ &= (a\ell + dk)n, \end{aligned}$$

and so $n \mid (ac - bd)$. Thus, $ac \equiv bd \pmod{n}$. This proves (c). □

Warning: Do not divide!!! For example, we have that $4 \equiv 8 \pmod{4}$, but if we divide both sides by 2, we get $2 \not\equiv 4 \pmod{4}$.

Proposition 0.2.4. *Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. Assume that $a \equiv b \pmod{n}$. Then $a^t \equiv b^t \pmod{n}$ for all integers $t \geq 0$.*

Proof. We proceed by induction on t.

Base case: $t = 0$. By definition, $r^0 = 1$ for all integers r.[3] So, $a^0 = 1 = b^0$, and so by Proposition 0.2.2(a), we have that $a^0 \equiv b^0 \pmod{n}$.

Induction case: Fix a non-negative integer t, and assume inductively that $a^t \equiv b^t \pmod{n}$. Since we also have that $a \equiv b \pmod{n}$, Proposition 0.2.3(c) implies that $a^t a \equiv b^t b \pmod{n}$, i.e. that $a^{t+1} \equiv b^{t+1} \pmod{n}$. This completes the induction. □

Remark: In what follows, we will repeatedly use Propositions 0.2.2, 0.2.3, and 0.2.4 without explicitly stating this.

Example 0.2.5. *Compute the last digit of 2018^{2019}.*

Solution. In principle, we could compute the value of 2018^{2019}, and then simply check what its last digit is. However, 2018^{2019} is an enormous number, and so this is impractical (even with the help of a computer). However, note that the last digit of a non-negative integer is simply its remainder when divided by 10. So, we need only figure out which of $0, 1, \ldots, 9$ the number 2018^{2019} is congruent to modulo 10.[4]

Clearly, $2018 \equiv 8 \pmod{10}$, and so $2018^{2019} \equiv 8^{2019} \pmod{10}$. Now, note the following:

- $8^1 \equiv 8 \pmod{10}$;
- $8^2 \equiv 4 \pmod{10}$;

[3] In fact, $r^0 = 1$ for all real numbers r.

[4] If we were looking for the last two digits, then we would be considering congruence modulo 100; for the last three digits, we would need congruence modulo 1000, etc.

- $8^3 \equiv 2 \pmod{10}$;
- $8^4 \equiv 6 \pmod{10}$;
- $8^5 \equiv 8 \pmod{10}$.

This looks like a periodic pattern! The general formula will be as in the Claim below (and we prove the Claim by mathematical induction).

> **Claim.** For all integers $k \geq 0$, we have the following:
>
> - $8^{4k+1} \equiv 8 \pmod{10}$;
> - $8^{4k+2} \equiv 4 \pmod{10}$;
> - $8^{4k+3} \equiv 2 \pmod{10}$;
> - $8^{4k+4} \equiv 6 \pmod{10}$.

Proof of the Claim. We proceed by induction on k.

Base case: For $k = 0$, we have:

- $8^{4\cdot 0+1} = 8 \equiv_{10} 8$;
- $8^{4\cdot 0+2} = 8^2 = 64 \equiv_{10} 4$;
- $8^{4\cdot 0+3} = 8^3 = 8 \cdot 8^2 \overset{(*)}{\equiv}_{10} 8 \cdot 4 = 32 \equiv_{10} 2$, where for (*), we used the fact that $8^2 \equiv_{10} 4$ (proven above);
- $8^{4\cdot 0+4} = 8^4 = 8 \cdot 8^3 \overset{(*)}{\equiv}_{10} 8 \cdot 2 = 16 \equiv_{10} 6$, where for (*) we used the fact that $8^3 \equiv_{10} 2$ (proven above).

Thus, the claim is true for $k = 0$.

Induction step: Fix a non-negative integer k, and assume inductively that the statement is true for k.[5] We must show that it is true for $k + 1$.[6] We saw in the

[5] So, we are assuming that all the following hold:

- $8^{4k+1} \equiv 8 \pmod{10}$;
- $8^{4k+2} \equiv 4 \pmod{10}$;
- $8^{4k+3} \equiv 2 \pmod{10}$;
- $8^{4k+4} \equiv 6 \pmod{10}$.

[6] So, we must prove all the following:

- $8^{4(k+1)+1} \equiv 8 \pmod{10}$;
- $8^{4(k+1)+2} \equiv 4 \pmod{10}$;
- $8^{4(k+1)+3} \equiv 2 \pmod{10}$;
- $8^{4(k+1)+4} \equiv 6 \pmod{10}$.

base case that $8^4 \equiv 6 \pmod{10}$, and consequently, for all non-negative integers ℓ, we have that

$$8^{4(k+1)+\ell} \quad = \quad 8^4 \cdot 8^{4k+\ell} \quad \equiv_{10} \quad 6 \cdot 8^{4k+\ell}.$$

In the following calculations, (*) follows from what we just showed,[7] and (**) follows from the induction hypothesis. We compute:

- $8^{4(k+1)+1} \overset{(*)}{\equiv}_{10} 6 \cdot 8^{4k+1} \overset{(**)}{\equiv}_{10} 6 \cdot 8 = 48 \equiv_{10} 8$;
- $8^{4(k+1)+2} \overset{(*)}{\equiv}_{10} 6 \cdot 8^{4k+2} \overset{(**)}{\equiv}_{10} 6 \cdot 4 = 24 \equiv_{10} 4$;
- $8^{4(k+1)+3} \overset{(*)}{\equiv}_{10} 6 \cdot 8^{4k+3} \overset{(**)}{\equiv}_{10} 6 \cdot 2 = 12 \equiv_{10} 2$;
- $8^{4(k+1)+4} \overset{(*)}{\equiv}_{10} 6 \cdot 8^{4k+4} \overset{(**)}{\equiv}_{10} 6 \cdot 6 = 36 \equiv_{10} 6$.

This completes the induction. ♦

Since $2019 = 4 \cdot 504 + 3$, the Claim guarantees that $8^{2019} \equiv 2 \pmod{10}$. Consequently,

$$2018^{2019} \quad \equiv \quad 8^{2019} \quad \equiv \quad 2 \pmod{10},$$

and it follows that the last digit of 2018^{2019} is 2. □

Notation: For $a_n, a_{n-1}, \ldots, a_0 \in \{0, 1, \ldots, 9\}$, we define:

$$\overline{a_n a_{n-1} \ldots a_0} \quad := \quad \sum_{k=0}^{n} a_k 10^k.$$

Thus, $\overline{a_n a_{n-1} \ldots a_0}$ is the number whose first digit is a_n,[8] whose second digit is a_{n-1}, and so on.

Proposition 0.2.6. *Let $a = \overline{a_n a_{n-1} \ldots a_0}$. Then $a \equiv a_n + a_{n-1} + \cdots + a_0$ (mod 9). Therefore, a positive integer is divisible by 9 if and only if the sum of its digits is divisible by 9.*

Proof. By definition, an integer is divisible by 9 if and only if it is congruent to 0 modulo 9. So, the second statement of the proposition follows immediately from the first. It remains to prove the first statement. Note that $10 \equiv 1 \pmod{9}$. So, by Proposition 0.2.4, we have that $10^k \equiv 1 \pmod{9}$ for all non-negative integers k. It follows that for all $k \in \{0, \ldots, n\}$, we have that $a_k \cdot 10^k \equiv a_k \pmod{9}$. Consequently,

$$a \quad = \quad \overline{a_n a_{n-1} \ldots a_0} \quad = \quad \sum_{k=0}^{n} a_k 10^k \quad \equiv_9 \quad \sum_{k=0}^{n} a_k \quad = \quad a_n + a_{n-1} + \cdots + a_0,$$

which is what we needed to show. □

[7] That is, from the fact that $8^{4(k+1)+\ell} \equiv_{10} 6 \cdot 8^{4k+\ell}$ for all non-negative integers ℓ.

[8] It is possible that this first digit is zero. We could eliminate this possibility, but that would result in a messier definition.

Proposition 0.2.7. *Let $a = \overline{a_n a_{n-1} \dots a_0}$. Then $a \equiv a_n + a_{n-1} + \dots + a_0 \pmod 3$. Therefore, a positive integer is divisible by 3 if and only if the sum of its digits is divisible by 3.*

Proof. The proof is completely analogous to that of Proposition 0.2.6: just replace 9 with 3 throughout. □

Example 0.2.8. *Show that the equation*

$$x^2 + y^2 \quad = \quad 10^{z+2} - 1$$

has no non-negative integer solutions.

Solution. We will show that for all non-negative integers x, y, z, we have that

$$x^2 + y^2 \quad \not\equiv \quad 10^{z+2} - 1 \pmod 4.$$

This will immediately imply that the equation $x^2+y^2 = 10^{z+2}-1$ has no non-negative integer solutions.

First, note that $100 \equiv 0 \pmod 4$. So, for a non-negative integer z, we have that

$$10^{z+2} - 1 \quad = \quad 100 \cdot 10^z - 1 \quad \equiv \quad 0 \cdot 10^z - 1 \quad \equiv \quad -1 \pmod 4.$$

On the other hand:

- $0^2 \equiv 0 \pmod 4$;
- $1^4 \equiv 1 \pmod 4$;
- $2^2 \equiv 0 \pmod 4$;
- $3^2 \equiv 1 \pmod 4$.

Since every integer is congruent to one of $0, 1, 2, 3$ modulo 4, it follows that the square of any integer is congruent to either 0 or 1 modulo 4. It follows that the sum of two squares is congruent to 0, 1, or 2 modulo 4, and none of these three numbers (0 , 1, or 2) is congruent to -1 modulo 4. Consequently, for integers x and y, we have that

$$x^2 + y^2 \quad \not\equiv \quad -1 \pmod 4.$$

Thus, for non-negative integers x, y, z, we have that

$$x^2 + y^2 \quad \not\equiv \quad 10^{z+2} - 1 \pmod 4,$$

and we are done. □

0.2.2 Arithmetic in $\mathbb{Z}_n$. Fermat's Little Theorem

Given $n \in \mathbb{N}$ and $a \in \mathbb{Z}$, we set

$$[a]_n \quad := \quad \big\{x \in \mathbb{Z} \mid x \equiv a \pmod{n}\big\};$$

note that $[a]_n = \big\{a + kn \mid k \in \mathbb{Z}\big\}$.[9] Note also that $a \in [a]_n$, since $a \equiv a \pmod{n}$. We define

$$\mathbb{Z}_n \quad := \quad \big\{[a]_n \mid a \in \mathbb{Z}\big\}.$$

Proposition 0.2.9. *Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. Then:*

(a) if $a \equiv b \pmod{n}$, then $[a]_n = [b]_n$;

(b) if $a \not\equiv b \pmod{n}$, then $[a]_n \cap [b]_n = \emptyset$.

Proof. This follows from the fact that, by Proposition 0.2.2, congruence modulo n is an equivalence relation on $\mathbb{Z}$. If you are not familiar with the theory of equivalence relations, here is a detailed proof.

We first prove (a). Suppose that $a \equiv b \pmod{n}$. It suffices to show that $[a]_n \subseteq [b]_n$ (the proof of the reverse inclusion is analogous). Fix $x \in [a]_n$. Then $x \equiv a \pmod{n}$. Since $a \equiv b \pmod{n}$, Proposition 0.2.2 guarantees that $x \equiv b \pmod{n}$. Consequently, $x \in [b]_n$, and we deduce that $[a]_n \subseteq [b]_n$. This proves (a).

It remains to prove (b). We prove the contrapositive: if $[a]_n \cap [b]_n \neq \emptyset$, then $a \equiv b \pmod{n}$. So, assume that $[a]_n \cap [b]_n \neq \emptyset$, and fix some $x \in [a]_n \cap [b]_n$. Since $x \in [a]_n$, we have that $x \equiv a \pmod{n}$, and since $x \in [b]_n$, we have that $x \equiv b \pmod{n}$. But now by Proposition 0.2.2, we have that $a \equiv b \pmod{n}$. This proves (b). □

Note that for $n \in \mathbb{N}$, every integer is congruent to exactly one of $0, \ldots, n-1$ modulo n; by Proposition 0.2.9, it follows that for all $x \in \mathbb{Z}$, the set $[x]_n$ is equal to exactly one of the following: $[0]_n, \ldots, [n-1]_n$. This implies that, in fact:

$$\mathbb{Z}_n \quad = \quad \big\{[0]_n, \ldots, [n-1]_n\big\}.$$

Moreover, by Proposition 0.2.9, no two of $0, \ldots, n-1$ are congruent to each other modulo n, and consequently, $[0]_n, \ldots, [n-1]_n$ are pairwise disjoint. We now deduce that the sets $[0]_n, \ldots, [n-1]_n$ form a "partition" of $\mathbb{Z}$, that is:

[9] For example:

- $[0]_2 = \{\ldots, -4, -2, 0, 2, 4, \ldots\}$;
- $[1]_2 = \{\ldots, -3, -1, 1, 3, 5, \ldots\}$;
- $[0]_3 = \{\ldots, -6, -3, 0, 3, 6, \ldots\}$;
- $[1]_3 = \{\ldots, -5, -2, 1, 4, 7, \ldots\}$;
- $[2]_3 = \{\ldots, -4, -1, 2, 5, 8, \ldots\}$.

- $\mathbb{Z} = [0]_n \cup \cdots \cup [n-1]_n$, and
- the sets $[0]_n, \ldots, [n-1]_n$ are pairwise disjoint.[10]

If you are familiar with "equivalence relations," then note that congruence modulo n is an equivalence relation on $\mathbb{Z}$ (by Proposition 0.2.2), and the sets $[0]_n, \ldots, [n-1]_n$ are the associated equivalence classes.

Notation: When working in $\mathbb{Z}_n$, we often write simply $0, \ldots, n-1$ instead of $[0]_n, \ldots, [n-1]_n$, respectively. We may do this **only** if we have previously made it clear that our numbers (which are technically sets of integers) are in $\mathbb{Z}_n$.

Example 0.2.10. *For $n = 2$, $[0]_2 = \{2t \mid t \in \mathbb{Z}\}$ and $[1]_2 = \{1 + 2t \mid t \in \mathbb{Z}\}$[11], and we have that $\mathbb{Z}_2 = \{[0]_2, [1]_2\}$. Typically, we write simply $\mathbb{Z}_2 = \{0, 1\}$, but technically, 0 stands for the set $[0]_2$, and 1 stands for $[1]_2$.*

Recall that, by Proposition 0.2.3, for all $n \in \mathbb{N}$ and $a, a', b, b' \in \mathbb{Z}$, if $a \equiv a' \pmod{n}$ and $b \equiv b' \pmod{n}$, then $a + b \equiv a' + b' \pmod{n}$, $a - b \equiv a' - b' \pmod{n}$, and $ab \equiv a'b' \pmod{n}$; equivalently, if $[a]_n = [a']_n$ and $[b]_n = [b']_n$, then $[a+b]_n = [a'+b']_n$, $[a - b]_n = [a' - b']_n$, and $[ab]_n = [a'b']_n$. Thus, we may define addition, subtraction, and multiplication in $\mathbb{Z}_n$ as follows. For $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$, we define

- $[a]_n + [b]_n = [a + b]_n$;
- $[a]_n - [b]_n = [a - b]_n$;
- $[a]_n[b]_n = [ab]_n$.

As we would expect (and as our next proposition summarizes), addition and multiplication in $\mathbb{Z}_n$ are commutative and associative, and multiplication is distributive over addition in $\mathbb{Z}_n$.

Proposition 0.2.11. *Let $n \in \mathbb{N}$. Then all the following hold:*

(a) addition and multiplication are commutative in $\mathbb{Z}_n$, that is, for all $a, b \in \mathbb{Z}_n$, we have that $a + b = b + a$ and $ab = ba$;

(b) addition and multiplication are associative in $\mathbb{Z}_n$, that is, for all $a, b, c \in \mathbb{Z}_n$, we have that $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$;

(c) multiplication is distributive over addition in $\mathbb{Z}_n$, that is, for all $a, b, c \in \mathbb{Z}_n$, we have that $a(b + c) = ab + ac$.

[10] This means that no two of $[0]_n, \ldots, [n-1]_n$ have an element in common. In other words, for all distinct $i, j \in \{0, \ldots, n-1\}$, we have $[i]_n \cap [j]_n = \emptyset$.

[11] In other words, $[0]_2$ is the set of all even numbers, and $[1]_2$ is the set of all odd numbers.

Proof. This essentially follows from the definition of $\mathbb{Z}_n$, from the fact that addition and multiplication are commutative and associative in $\mathbb{Z}$, and from the fact that multiplication is distributive over addition in $\mathbb{Z}$. We give the details for the commutativity of addition in $\mathbb{Z}_n$; the rest is left as an easy exercise.

Fix $a, b \in \mathbb{Z}_n$. Then there exist $a', b' \in \mathbb{Z}$ such that $a = [a']_n$ and $b = [b']_n$. We now have that

$$\begin{aligned} a + b &= [a']_n + [b']_n \\ &= [a' + b']_n && \text{by the definition of addition in } \mathbb{Z}_n \\ &= [b' + a']_n && \text{by the commutativity of addition in } \mathbb{Z} \\ &= [b']_n + [a']_n && \text{by the definition of addition in } \mathbb{Z}_n \\ &= b + a. \end{aligned}$$

This proves that addition is commutative in $\mathbb{Z}_n$. □

Let us now take a look at the addition and multiplication tables for $\mathbb{Z}_n$, for a few small values of n.

Example 0.2.12. *Below are the addition and multiplication tables for $\mathbb{Z}_2$.*

$+$	$[0]_2$	$[1]_2$
$[0]_2$	$[0]_2$	$[1]_2$
$[1]_2$	$[1]_2$	$[0]_2$

$\cdot$	$[0]_2$	$[1]_2$
$[0]_2$	$[0]_2$	$[0]_2$
$[1]_2$	$[0]_2$	$[1]_2$

If we omit square brackets and subscripts (as we usually do), we obtain the addition and multiplication tables for $\mathbb{Z}_2$ shown below.

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Example 0.2.13. *Below are the addition and multiplication tables for $\mathbb{Z}_3$.*[12]

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

[12]Remember, in this context, 0 stands for $[0]_3$, 1 stands for $[1]_3$, and 2 stands for $[2]_3$.

Example 0.2.14. *Below are the addition and multiplication tables for* $\mathbb{Z}_4$.[13]

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Example 0.2.15. *Below are the addition and multiplication tables for* $\mathbb{Z}_5$.[14]

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Remark/Notation: Note that for all positive integers n, each number a in $\mathbb{Z}_n$ has a unique "additive inverse," denoted by $-a$, i.e. the number (element of $\mathbb{Z}_n$) that we need to add to a in order to obtain 0 (here, $0 = [0]_n$). When using square brackets and subscripts, we do, of course, get $-[a]_n = [-a]_n = [n-a]_n$ for all positive integers n and all integers a. However, we will usually work in $\mathbb{Z}_n$ **without** such brackets. For small values of n, we get the following:

- in $\mathbb{Z}_2$: $-0 = 0, \quad -1 = 1;$
- in $\mathbb{Z}_3$: $-0 = 0, \quad -1 = 2, \quad -2 = 1;$
- in $\mathbb{Z}_4$: $-0 = 0, \quad -1 = 3, \quad -2 = 2, \quad -3 = 1;$
- in $\mathbb{Z}_5$: $-0 = 0, \quad -1 = 4, \quad -2 = 3, \quad -3 = 2, \quad -4 = 1.$

Remark: Note that for $n = 2, 3, 5$, every non-zero member of $\mathbb{Z}_n$ has a "multiplicative inverse," i.e. a number that we can multiply it by to get 1. However, for $n = 4$, this is not the case. As Theorem 0.2.16 and Corollary 0.2.17 (see below) show, this is not an accident.

Theorem 0.2.16. *Let* $n \in \mathbb{N}$ *and* $a \in \mathbb{Z}$ *be relatively prime.*[15] *Then there exists some* $b \in \mathbb{Z}$ *such that* $ab \equiv 1$ *(mod* n*), and therefore,* $[a]_n[b]_n = [1]_n$.

[13] Remember, in this context, 0 stands for $[0]_4$, 1 stands for $[1]_4$, 2 stands for $[2]_4$, and 3 stands for $[3]_4$.

[14] Remember, in this context, 0 stands for $[0]_5$, 1 stands for $[1]_5$, 2 stands for $[2]_5$, 3 stands for $[3]_5$, and 4 stands for $[4]_5$.

[15] This means that the greatest common divisor of n and a, denoted by $\gcd(n, a)$, is 1. In other words, the only positive integer that divides both n and a is 1.

Proof. Let us show that no two of $0, a, 2a, \ldots, (n-1)a$ are congruent modulo n.[16] Suppose otherwise, and fix distinct $i, j \in \{0, \ldots, n-1\}$ such that $ia \equiv ja \pmod{n}$. Then $(i-j)a \equiv 0 \pmod{n}$, that is, $n|(i-j)a$. Since n and a are relatively prime, it follows that $n|(i-j)$. But this is impossible because $i, j \in \{0, \ldots, n-1\}$ and $i \neq j$, and so $0 < |i-j| < n$. Thus, no two of $0, a, 2a, \ldots, (n-1)a$ are congruent modulo n.

We know that every integer is congruent modulo n to one of the following n integers: $0, 1, 2, \ldots, n-1$. We showed above that no two of the following n integers are congruent to each other modulo n: $0, a, 2a, \ldots, (n-1)a$. It follows that (exactly) one of $0, a, 2a, \ldots, (n-1)a$ is congruent to 1 modulo n. In other words, for exactly one value of $b \in \{0, 1, 2, \ldots, n-1\}$, we have that $ba \equiv 1 \pmod{n}$. For this b, we have that $ab \equiv 1 \pmod{n}$, and therefore, $[a]_n[b]_n = [1]_n$. This completes the argument. □

Corollary 0.2.17. *Let $p \in \mathbb{N}$ be a prime number. Then:*

(a) for all $a \in \mathbb{Z}$ such that a is not a multiple of p, there exists some $b \in \mathbb{Z}$ such that $ab \equiv 1 \pmod{p}$, and consequently, $[a]_p[b]_p = [1]_p$;

(b) for all $a \in \mathbb{Z}_p \setminus \{0\}$, there exists some $b \in \mathbb{Z}_p \setminus \{0\}$ such that $ab = 1$.[17]

Proof. We first prove (a). Since p is a prime number, every integer that is not a multiple of p is relatively prime to p; (a) now follows from Theorem 0.2.16.

Statement (b) immediately follows from (a). Indeed, fix $a \in \mathbb{Z}_p \setminus \{0\}$. Then there exists an integer $a' \in \{1, \ldots, p-1\}$ such that $a = [a']_p$. By (a), there exists an integer b' such that $a'b' \equiv 1 \pmod{p}$. We now set $b := [b']_p$, and we see that $ab = [a']_p[b']_p = [a'b']_p = [1]_p$. Moreover, $b \neq 0$, since (in $\mathbb{Z}_p$) we have that $a \cdot 0 = 0 \neq 1 = ab$. This proves (b). □

Corollary 0.2.17(b) states that, for a prime number p, every number in $\mathbb{Z}_p \setminus \{0\}$ has a multiplicative inverse. Fermat's Little Theorem (below) is a strengthening of Corollary 0.2.17 in that it gives an actual formula for this multiplicative inverse. However, before stating and proving Fermat's Little Theorem, we need some notation (which will be used in the proof). For non-negative integers n, we define $n!$ (read "n factorial") recursively, as follows:

- $0! := 1$;
- $(n+1)! := n! \cdot (n+1)$ for all non-negative integers n.

So, for a positive integer n, we have $n! = 1 \cdot 2 \cdot \cdots \cdot n$.

Fermat's Little Theorem. *If $p \in \mathbb{N}$ is a prime number, and $a \in \mathbb{Z}$ is not a multiple of p, then $a^{p-1} \equiv 1 \pmod{p}$.*

[16] Note that this implies that $[a]_n, [2a]_n, \ldots, [(n-1)a]_n$ are pairwise distinct.

[17] Here, $0 = [0]_p$ and $1 = [1]_p$.

Proof. Fix a prime number $p \in \mathbb{N}$. Let $a \in \mathbb{Z}$, and assume that a is not a multiple of p. As in the proof of Theorem 0.2.16, no two of $0, a, 2a, \ldots, (p-1)a$ are congruent modulo p.[18] Since every integer is congruent to exactly one of $0, 1, \ldots, p-1$ modulo p, it follows that there exists some rearrangement (i.e. permutation) $r_1, \ldots, r_{p-1}$ of the sequence $1, \ldots, p-1$ such that

- $a \equiv r_1 \pmod{p}$;
- $2a \equiv r_2 \pmod{p}$;

  $\vdots$

- $(p-1)a \equiv r_{p-1} \pmod{p}$.

It now follows that

$$\underbrace{a \cdot 2a \cdot \cdots \cdot (p-1)a}_{=(p-1)!a^{p-1}} \equiv \underbrace{r_1 r_2 \ldots r_{p-1}}_{=(p-1)!} \pmod{p},$$

and so $(p-1)!a^{p-1} \equiv (p-1)! \pmod{p}$. But now

$$(a^{p-1} - 1)(p-1)! \equiv 0 \pmod{p},$$

that is, $p \mid \Big((a^{p-1} - 1)(p-1)!\Big)$. Since p is prime, we see that p and $(p-1)!$ are relatively prime. It follows that $p \mid (a^{p-1} - 1)$, and consequently, $a^{p-1} \equiv 1 \pmod{p}$, which is what we needed to show. □

For a positive integer n and for $a \in \mathbb{Z}_n$, we define powers of a recursively, as follows:

- $a^0 = 1$ (where $1 := [1]_n$);
- $a^{m+1} = a^m a$ for all non-negative integers m.

So, for a positive integer m, we have the familiar formula

$$a^m = \underbrace{a \cdot \cdots \cdot a}_{m},$$

where it is understood that the multiplication on the right-hand-side is in $\mathbb{Z}_n$. With this set-up, we can restate Fermat's Little Theorem in two ways, as follows.

[18] This is exactly the same as in the proof of Theorem 0.2.16, but for the sake of completeness, here is the full proof. Suppose that some two of $0, a, \ldots, (p-1)a$ are congruent modulo p. Fix distinct $i, j \in \{0, 1, \ldots, p-1\}$ such that $ia \equiv ja \pmod{p}$. Then $(i-j)a \equiv 0 \pmod{p}$, that is, $p|(i-j)a$. Since p is prime and does not divide a, we see that $p|(i-j)$. But this is impossible because $i, j \in \{0, \ldots, p-1\}$ and $i \neq j$, and so $0 < |i-j| < p$. Thus, no two of $0, a, 2a, \ldots, (p-1)a$ are congruent modulo p.

Fermat's Little Theorem. *If $p \in \mathbb{N}$ is a prime number, and $a \in \mathbb{Z}$ is not a multiple of p, then $\big([a]_p\big)^{p-1} = [1]_p$.*

Fermat's Little Theorem. *If $p \in \mathbb{N}$ is a prime number and $a \in \mathbb{Z}_p \setminus \{0\}$, then $a^{p-1} = 1$.*

Multiplicative inverses. Suppose that p is a **prime** number and that $a \in \mathbb{Z}_p \setminus \{0\}$. By Fermat's Little Theorem, a^{p-2} is a "multiplicative inverse" of a, i.e. if we multiply a by a^{p-2} (on either side), we obtain 1.[19] Moreover, it is easy to see that a^{p-2} is the **only** multiplicative inverse of a in $\mathbb{Z}_p$. Indeed, if $b \in \mathbb{Z}_p$ satisfies $ab = 1$ (which is equivalent to $ba = 1$, by Proposition 0.2.11), then by multiplying both sides by a^{p-2}, we obtain

$$\underbrace{a^{p-2} \cdot a}_{=a^{p-1}=1} b \quad = \quad a^{p-2} \cdot 1,$$

and consequently, $b = a^{p-2}$. So, we can say that a^{p-2} is **the** multiplicative inverse of a (denoted by a^{-1}), and we write

$$\underbrace{a^{-1}}_{\substack{\text{multiplicative} \\ \text{inverse of } a}} \quad = \quad a^{p-2}$$

Note, however, that for small values of the prime number p, it is easier to read off the multiplicative inverses of non-zero numbers in $\mathbb{Z}_p$ from the multiplication table for $\mathbb{Z}_p$ than it is to compute the $(p-2)$-th powers of those numbers. Thus, by taking a quick look at the multiplication tables for $\mathbb{Z}_2$, $\mathbb{Z}_3$, and $\mathbb{Z}_5$, we get the following:

- in $\mathbb{Z}_2$: $1^{-1} = 1$;
- in $\mathbb{Z}_3$: $1^{-1} = 1$, $2^{-1} = 2$;
- in $\mathbb{Z}_5$: $1^{-1} = 1$, $2^{-1} = 3$, $3^{-1} = 2$, $4^{-1} = 4$.

0.3 Complex numbers

Remark: This section is written for students who have already studied complex numbers, but need a bit of a refresher. For this reason, most of the proofs are omitted, and more importantly, there are relatively few examples. If you have never seen complex numbers before, you might want to learn about them from a high school algebra textbook. After that, you can read this section to check your understanding.

[19] That is: $a \cdot a^{p-2} = a^{p-2} \cdot a = 1$.

0.3.1 Complex numbers: definition, basic properties, and examples

To define complex numbers, we first introduce the *imaginary unit number*, denoted by i, which satisfies

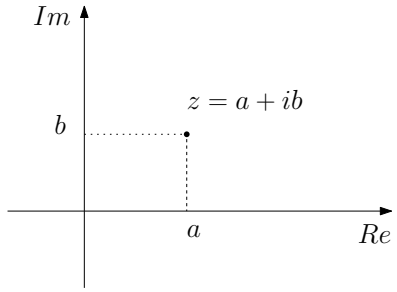
$$i^2 = -1.$$

A *complex number* is any number of the form $z = a + bi$, where a and b are real numbers; the *real part* of the complex number z is the real number a, and the *imaginary part* of z is the real number b. The real and imaginary part of a complex number z are denoted by $Re(z)$ and $Im(z)$, respectively. For example, we have the following:

- $Re(2 + i) = 2$ and $Im(2 + i) = 1$;
- $Re(-3i) = 0$ and $Im(-3i) = -3$;
- $Re(7) = 7$ and $Im(7) = 0$.

Note that real numbers are precisely those complex numbers whose imaginary part is zero.

The set of all complex numbers is denoted by $\mathbb{C}$. Complex numbers can be visualized in the "complex plane." This plane has two axes: the *real axis* (denoted by Re) and the *imaginary axis* (denoted by Im). A complex number $z = a + bi$ (where $a, b \in \mathbb{R}$) can be visualized in the complex plane as in the picture below.



Note that real numbers are precisely those complex numbers that lie on the real axis.

We define addition and multiplication of complex numbers as follows. Given complex numbers $z_1 = a_1 + b_1 i$ and $z_2 = a_2 + b_2 i$ (where $a_1, b_1, a_2, b_2 \in \mathbb{R}$), we define

- $z_1 + z_2 = (a_1 + b_1 i) + (a_2 + b_2 i) := (a_1 + a_2) + (b_1 + b_2)i$;
- $z_1 z_2 = (a_1 + b_1 i)(a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2)i$.

The definition of addition is natural. The idea behind the definition of multiplication is that we are supposed to get something like this:

$$\begin{aligned}(a_1 + b_1 i)(a_2 + b_2 i) &\overset{(*)}{=} a_1 a_2 + a_1 b_2 i + b_1 a_2 i + b_1 b_2 \underbrace{i^2}_{=-1} \\ &= (a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2)i,\end{aligned}$$

where in (*), we assumed that the addition and multiplication of complex numbers are commutative and associative, and that multiplication is distributive over addition for complex numbers. However, this is not something that we can **assume**; it is something we would need to **prove**, using the appropriate definitions.

Proposition 0.3.1. *All the following hold:*

(a) addition and multiplication in $\mathbb{C}$ are commutative, that is, for all $z_1, z_2 \in \mathbb{C}$, we have that $z_1 + z_2 = z_2 + z_1$ and $z_1 z_2 = z_2 z_1$;

(b) addition and multiplication in $\mathbb{C}$ are associative, that is, for all $z_1, z_2, z_3 \in \mathbb{C}$, we have that $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$ and $(z_1 z_2) z_3 = z_1 (z_2 z_3)$;

(c) multiplication is distributive over addition in $\mathbb{C}$, that is, for all $z_1, z_2, z_3 \in \mathbb{C}$, we have that $z_1(z_2 + z_3) = z_1 z_2 + z_1 z_3$.

Proof. This essentially follows from the definition of addition and multiplication in $\mathbb{C}$, and from the fact that addition and multiplication are commutative and associative in $\mathbb{R}$, and multiplication is distributive over addition in $\mathbb{R}$. Let us prove in detail that addition in $\mathbb{C}$ is associative; the rest is left as a straightforward exercise.

Fix $z_1, z_2, z_3 \in \mathbb{C}$; we must show that $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$. Set $z_1 = a_1 + b_1 i$, $z_2 = a_2 + b_2 i$, and $z_3 = a_3 + b_3 i$, where $a_1, b_1, a_2, b_2, a_3, b_3 \in \mathbb{R}$. We now compute:

$$\begin{aligned}
(z_1 + z_2) + z_3 &= \Big((a_1 + b_1 i) + (a_2 + b_2 i)\Big) + (a_3 + b_3 i) \\
&\overset{(*)}{=} \Big((a_1 + a_2) + (b_1 + b_2) i\Big) + (a_3 + b_3 i) \\
&\overset{(*)}{=} \Big((a_1 + a_2) + a_3\Big) + \Big((b_1 + b_2) + b_3\Big) i \\
&\overset{(**)}{=} \Big(a_1 + (a_2 + a_3)\Big) + \Big(b_1 + (b_2 + b_3)\Big) i \\
&\overset{(*)}{=} (a_1 + b_1 i) + \Big((a_2 + a_3) + (b_2 + b_3) i\Big) \\
&\overset{(*)}{=} (a_1 + b_1 i) + \Big((a_2 + b_2 i) + (a_3 + b_3 i)\Big) \\
&= z_1 + (z_2 + z_3),
\end{aligned}$$

where each instance of (*) follows from the definition of addition in $\mathbb{C}$, and (**) follows from the fact that addition in $\mathbb{R}$ is associative. This proves that addition in $\mathbb{C}$ is associative. □

Powers of complex numbers are defined in the usual way. For a complex number z, we define

- $z^0 := 1$;
- $z^{m+1} := z^m z$ for all non-negative integers m.

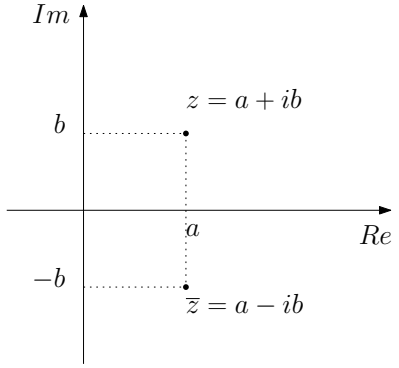
So, for all positive integers m, we have the familiar expression

$$z^m = \underbrace{z \dots z}_{m}.$$

Next, for a complex number $z = a + bi$ (where $a, b \in \mathbb{R}$), we define the following:

- the *complex conjugate* of z is $\overline{z} := a - bi$;
- the modulus (or absolute value) of z is $|z| := \sqrt{a^2 + b^2}$.

Geometrically, the complex conjugate of a complex number z is obtained by reflecting z about the Re axis, as shown in the picture below. Obviously, $\overline{\overline{z}} = z$. Further, note that $\overline{z} = z$ if and only if the complex number z is in fact a real number, i.e. $Im(z) = 0$. We also note that the modulus of a complex number is the usual Pythagorean distance between that complex number and the origin in the complex plane. Note that the modulus of a complex number z is a non-negative real number, and moreover, we have that $|z| = 0$ if and only if $z = 0$.



Proposition 0.3.2. *For all complex numbers $z = a + bi$ (with $a, b \in \mathbb{R}$), we have that*

$$z\overline{z} = a^2 + b^2 = |z|^2.$$

Proof. Fix a complex number $z = a + bi$, where $a, b \in \mathbb{R}$. We then have that

$$\begin{aligned} z\overline{z} &= (a + bi)(a - bi) \\ &= (a^2 - b(-b)) + (a(-b) + ba)i \\ &= a^2 + b^2 \\ &= |z|^2, \end{aligned}$$

which is what we needed to show. □

Note that Proposition 0.3.2, in particular, establishes that multiplying a complex number z by its conjugate produces a real number; that real number is zero if and only if $z = 0$.

Let us now explain how division works in $\mathbb{C}$. First of all, given a complex number $z = a + bi$ (with $a, b \in \mathbb{R}$) and a real number $r \neq 0$, we have

$$\frac{z}{r} = \frac{a}{r} + \frac{b}{r}i.$$

Now suppose that z_1 and $z_2 \neq 0$ are complex numbers. To compute $\frac{z_1}{z_2}$, we need to transform the denominator into a non-zero real number. We do this by multiplying both the numerator and the denominator by $\overline{z_2}$, at which point (by Proposition 0.3.2) the denominator becomes $|z_2|^2$, which is a non-zero real number, and we can divide as above. Let us take a look at an example.

Example 0.3.3. *Compute the following quotients:*

(a) $\frac{7-6i}{3+2i}$; *(b)* $\frac{1}{2-i}$; *(c)* $\frac{2-3i}{5}$; *(d)* $\frac{4-2i}{2-i}$.

Solution. (a) We multiply both the numerator and the denominator by $\overline{3+2i} = 3-2i$, and we obtain

$$\frac{7-6i}{3+2i} = \frac{(7-6i)(3-2i)}{(3+2i)(3-2i)} = \frac{9-32i}{9+4} = \frac{9}{13} - \frac{32}{13}i.$$

(b) We multiply both the numerator and the denominator by $\overline{2-i} = 2+i$, and we obtain

$$\frac{1}{2-i} = \frac{2+i}{(2-i)(2+i)} = \frac{2+i}{4+1} = \frac{2}{5} + \frac{1}{5}i.$$

(c) The denominator is a real number, and so we have

$$\frac{2-3i}{5} = \frac{2}{5} - \frac{3}{5}i.$$

(d) We could multiply both the numerator and the denominator by $\overline{2-i} = 2+i$. However, in this particular case, it is easier to compute as follows:

$$\frac{4-2i}{2-i} = \frac{2(2-i)}{2-i} \overset{(*)}{=} 2,$$

where (*) was obtained by canceling out the common factor $2 - i$ in the numerator and the denominator. □

Finally, we give some properties of the complex conjugate and the modulus of a complex number (see Propositions 0.3.4 and 0.3.5 below).

Proposition 0.3.4. *For all $z_1, z_2 \in \mathbb{C}$, the following hold:*

(a) $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$;

(b) $\overline{z_1 - z_2} = \overline{z_1} - \overline{z_2}$;

(c) $\overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}$;

(d) if $z_2 \neq 0$, *then* $\overline{z_1/z_2} = \overline{z_1}/\overline{z_2}$.

Moreover, for all $z \in \mathbb{C}$ *and non-negative integers* m, *we have that*

(e) $\overline{z^m} = (\overline{z})^m$.

Proof. We prove (c). The rest is left as an exercise. Fix a complex numbers $z_1 = a_1 + b_1 i$ and $z_2 = a_2 + b_2 i$ (with $a_1, b_1, a_2, b_2 \in \mathbb{R}$). We then have that

$$\begin{aligned} \overline{z_1 z_2} &= \overline{(a_1 + b_1 i)(a_2 + b_2 i)} \\ &= \overline{(a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2) i} \\ &= (a_1 a_2 - b_1 b_2) - (a_1 b_2 + b_1 a_2) i. \end{aligned}$$

On the other hand,

$$\begin{aligned} \overline{z_1}\ \overline{z_2} &= \left(\overline{a_1 + b_1 i}\right)\left(\overline{a_2 + b_2 i}\right) \\ &= (a_1 - b_1 i)(a_2 - b_2 i) \\ &= (a_1 a_2 - b_1 b_2) + (-a_1 b_2 - b_1 a_2) i \\ &= (a_1 a_2 - b_1 b_2) - (a_1 b_2 + b_1 a_2) i. \end{aligned}$$

So, $\overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}$. This proves (c). □

Proposition 0.3.5. *For all* $z_1, z_2 \in \mathbb{C}$, *the following hold:*

(a) $|z_1 z_2| = |z_1||z_2|$;

(b) if $z_2 \neq 0$, *then* $|z_1/z_2| = |z_1|/|z_2|$.

Moreover, for all $z \in \mathbb{C}$, *the following hold:*

(c) $|-z| = |z|$;

(d) for all non-negative integers m, *we have* $|z^m| = |z|^m$.

Proof. We prove (a); the rest is left as an exercise. Fix complex numbers $z_1 = a_1 + b_1 i$ and $z_2 = a_2 + b_2 i$ (where $a_1, b_1, a_2, b_2 \in \mathbb{R}$). We then have that

$$\begin{aligned}
|z_1 z_2| &= |(a_1 + b_1 i)(a_2 + b_2 i)| \\
&= |(a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2) i| \\
&= \sqrt{(a_1 a_2 - b_1 b_2)^2 + (a_1 b_2 + b_1 a_2)^2} \\
&= \sqrt{a_1^2 a_2^2 + 2 a_1 a_2 b_1 b_2 + b_1^2 b_2^2 + a_1^2 b_2^2 - 2 a_1 b_2 b_1 a_2 + b_1^2 a_2^2} \\
&= \sqrt{a_1^2 a_2^2 + b_1^2 b_2^2 + a_1^2 b_2^2 + b_1^2 a_2^2}.
\end{aligned}$$

On the other hand, we have that

$$\begin{aligned}
|z_1||z_2| &= |a_1 + b_1 i||a_2 + b_2 i| \\
&= \sqrt{a_1^2 + b_1^2}\sqrt{a_2^2 + b_2^2} \\
&= \sqrt{a_1^2 a_2^2 + a_1^2 b_2^2 + b_1^2 a_2^2 + b_1^2 b_2^2} \\
&= \sqrt{a_1^2 a_2^2 + b_1^2 b_2^2 + a_1^2 b_2^2 + b_1^2 a_2^2}.
\end{aligned}$$

It follows that $|z_1 z_2| = |z_1||z_2|$. This proves (a). □

0.3.2 The Fundamental Theorem of Algebra

A *root* of a polynomial $p(x)$ with complex coefficients is a complex number c such that $p(c) = 0$. For example, $1 + i$ is a root of the polynomial $p(x) = x^2 - 2x + 2$ because

$$p(1+i) = (1+i)^2 - 2(1+i) + 2 = 0.$$

In the particular case of $p(x) = x^2 - 2x + 2$, the roots could have been found via the familiar quadratic equation. There exist formulas for finding the complex roots of all third and fourth degree polynomials with complex coefficients, but no such formula exists for polynomials of degree five or more (although in some special cases, we may be able to use various tricks to find the roots of these higher-degree polynomials). Nevertheless, we do have the following **existence** result. (A *constant* polynomial is a polynomial of the form $p(x) = c$, where c is a fixed constant/number.)

The Fundamental Theorem of Algebra. *Any non-constant polynomial with complex coefficients has a complex root.*

Remark: The Fundamental Theorem of Algebra is an existence result in the sense that it guarantees the **existence** of a complex root for any non-constant polynomial with complex coefficients, even though we might not be able to actually **compute** this root. Of course, every real number is complex. So, the Fundamental Theorem

of Algebra, in particular, implies that every non-constant polynomial with real coefficients has a complex root (which may or may not be a real number). For instance, the polynomial $p(x) = x^2 + 1$ is a non-constant polynomial with real (in fact, rational) coefficients, but it has no real roots. It does, of course, have two complex roots, namely i and $-i$.

We will not prove the Fundamental Theorem of Algebra here. There are no known elementary proofs of this theorem: all the known proofs of the Fundamental Theorem of Algebra rely on advanced mathematics, such as complex analysis or topology.

The Fundamental Theorem of Algebra implies that any polynomial $p(x)$ with complex coefficients and of degree $n \geq 1$ can be factored into n linear factors. More precisely, for such a polynomial $p(x)$, there exist complex numbers $a, \alpha_1, \ldots, \alpha_\ell$ such that $a \neq 0$ and such that $\alpha_1, \ldots, \alpha_\ell$ are pairwise distinct, and positive integers $n_1, \ldots, n_\ell$ satisfying $n_1 + \cdots + n_\ell = n$, such that

$$p(x) = a(x - \alpha_1)^{n_1} \ldots (x - \alpha_\ell)^{n_\ell},$$

and moreover, this factorization into linear factors is unique up a permutation of the α_i's and the corresponding n_i's. (We omit the proof.) Here, a is the leading coefficient of $p(x)$, i.e. the coefficient in front of x^n. Complex numbers $\alpha_1, \ldots, \alpha_\ell$ are the roots of $p(x)$ with *multiplicities* $n_1, \ldots, n_\ell$, respectively. If we think of each α_i as being a root "n_i times" (due to its multiplicity), then we see that the n-th degree polynomial $p(x)$ has exactly n complex roots. This is often summarized as follows: "every n-th degree polynomial (with $n \geq 1$) with complex coefficients has exactly n complex roots, when multiplicities are taken into account."

As we already mentioned, there are formulas that allow us to compute the roots of polynomials with complex coefficients of degree at most four. However, no such formulas exist for polynomials (with complex coefficients) of degree $n \geq 5$: we know that all such polynomials have n complex roots (when multiplicities are taken into account), but in general, there is no formula for computing these roots. In fact, not only is no such formula known, but using Galois theory, one can show that no such formula can exist for polynomials of degree at least five. (Once again, we may be able to use various tricks to compute the roots of some special high-degree polynomials. However, none of these tricks will work in the general case.)

Complex roots of polynomials with real coefficients. Every real number can be seen as a complex number. So, it makes sense to speak of complex roots of polynomials with real coefficients, as in the case of the theorem below. Recall that, geometrically, the complex conjugate of a complex number z is obtained by reflecting z about the Re axis in the complex plane. Moreover, for a complex number z, we have that $\overline{z} = z$ if and only if z is a real number.

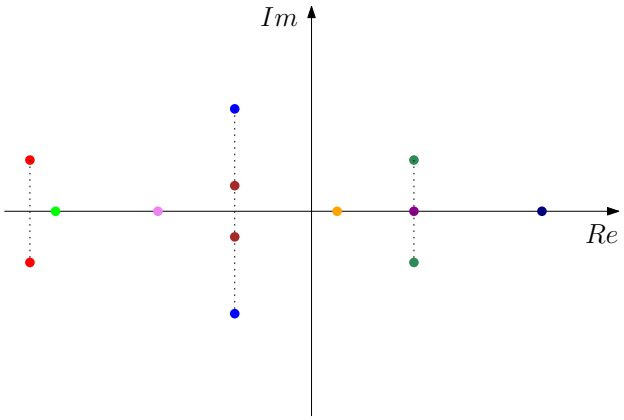
Theorem 0.3.6. *Let $p(x)$ be any polynomial with **real** coefficients, and let $z \in \mathbb{C}$. Then z is a root of $p(x)$ if and only if its complex conjugate $\overline{z}$ is a root of $p(x)$.*

Proof. Set $p(x) = a_n x^n + \cdots + a_1 x + a_0$, where $a_0, a_1, \ldots, a_n \in \mathbb{R}$. Then we have the following sequence of equivalences:

$$\begin{aligned}
p(z) = 0 &\iff \overline{p(z)} = \overline{0} \\
&\iff \overline{a_n z^n + \cdots + a_1 z + a_0} = \overline{0} \\
&\overset{(*)}{\iff} \overline{a_n}(\overline{z})^n + \cdots + \overline{a_1}(\overline{z}) + \overline{a_0} = \overline{0} \\
&\overset{(**)}{\iff} a_n(\overline{z})^n + \cdots + a_1\overline{z} + a_0 = 0 \\
&\iff p(\overline{z}) = 0,
\end{aligned}$$

where (*) follows from Proposition 0.3.4, and (**) follows from the fact that $a_0, a_1, \ldots, a_n$ and 0 are real numbers. □

Remark: Note that Theorem 0.3.6 implies that the complex roots of a non-constant polynomial are symmetric about the Re axis in the complex plane. Some (or perhaps all) of those roots may lie on the Re axis, i.e. they may be real numbers. So, if we plot the roots of a non-constant polynomial with real coefficients as points in the complex plane, we may obtain a picture that looks something like this (see below; complex conjugate pairs are colored with the same color).



Chapter 1

Systems of linear equations. Vectors and matrices

1.1 An informal introduction to fields

In this chapter, we will assume that $\mathbb{F}$ is a fixed "field." A formal definition of a field will be given in chapter 2 (see section 2.4). For now, we note that all the following are fields:

- the field $\mathbb{Q}$ of rational numbers;
- the field $\mathbb{R}$ of real numbers;
- the field $\mathbb{C}$ of complex numbers;
- the field $\mathbb{Z}_p$, where p is a **prime** number.[1]

Each field is equipped with two operations: addition and multiplication. These two operations are commutative and associative, and multiplication is distributive over addition.[2] It is also relevant that every field has an "additive identity" 0 and a "multiplicative identity" 1, which satisfy $a + 0 = 0 + a = a$ and $a \cdot 1 = 1 \cdot a = a$ for all elements a of the field. Every element a of a field has a corresponding "additive inverse," denoted by $-a$, which is a number that we can add to a in order to obtain 0. For example:

- the additive inverse of $\sqrt{17}$ in $\mathbb{R}$ is $-\sqrt{17}$, since $\sqrt{17} + (-\sqrt{17}) = 0$ in $\mathbb{R}$.

[1] If $n \in \mathbb{N}$ is not prime, then $\mathbb{Z}_n$ is **not** a field.

[2] This means that for all elements a, b, c of the field, the following are satisfied:

- $a + b = b + a$ and $ab = ba$ (commutativity of addition and multiplication);
- $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$ (associativity of addition and multiplication);
- $a(b + c) = ab + ac$ (distributive property of multiplicaiton over addition).

- the additive inverse of $2-i$ in $\mathbb{C}$ is $-2+i$, since $(2-i)+(-2+i)=0$ in $\mathbb{C}$;
- the additive inverse of 3 in $\mathbb{Z}_5$ is 2 (and we write $-3=2$), since $3+2=0$ in $\mathbb{Z}_5$;
- the additive inverse of 4 in $\mathbb{Z}_5$ is 1 (and we write $-4=1$), since $4+1=0$ in $\mathbb{Z}_5$;
- the additive inverse of 2 in $\mathbb{Z}_3$ is 1 (and we write $-2=1$), since $2+1=0$ in $\mathbb{Z}_3$.

Equally importantly, every **non-zero** element a of a field has a "multiplicative inverse," denoted by a^{-1}, which is a number we can multiply a by in order to obtain 1. For example:

- the multiplicative inverse of $\sqrt{17}$ in $\mathbb{R}$ is $\frac{1}{\sqrt{17}}$, because $\sqrt{17}\cdot\frac{1}{\sqrt{17}}=1$ in $\mathbb{R}$;
- the multiplicative inverse of $2-i$ is $\frac{2}{5}+\frac{1}{5}i$, because $(2-i)(\frac{2}{5}+\frac{1}{5}i)=1$ in $\mathbb{C}$;
  - this is obtained by computing:
$$\frac{1}{2-i} \;=\; \frac{2+i}{(2-i)(2+i)} \;=\; \frac{2+i}{5} \;=\; \frac{2}{5}+\frac{1}{5}i;$$
- the multiplicative inverse of 3 in $\mathbb{Z}_5$ is 2 (and we write $3^{-1}=2$), since $3\cdot 2=1$ in $\mathbb{Z}_5$;
- the multiplicative inverse of 4 in $\mathbb{Z}_5$ is 4 (and we write $4^{-1}=4$), since $4\cdot 4=1$ in $\mathbb{Z}_5$;
- the multiplicative inverse of 2 in $\mathbb{Z}_3$ is 2 (and we write $2^{-1}=2$), since $2\cdot 2=1$ in $\mathbb{Z}_3$.

Remark: When working over $\mathbb{Z}_p$ (for a prime number p), it is a good idea to first write out the addition and multiplication tables for $\mathbb{Z}_p$, because this allows us to easily identify additive and multiplicative inverses: for a given $a\in\mathbb{Z}_p$, we simply read off from the tables what number we need to add to a to get zero, and (assuming $a\neq 0$) what number we need to multiply it by to get 1.

Warning: The following are **not** fields: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}_n$ (where n is a positive integer that is **not** prime).

For the remainder of this chapter, you may assume that the field $\mathbb{F}$ in question is one of the following: $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p$ (where p is a prime number). However, everything that we prove in this chapter does in fact hold for general fields $\mathbb{F}$, not just the ones listed above.

1.2 Vectors and matrices

1.2.1 Matrices

A *matrix* is a rectangular array of numbers (typically, elements of some field). An $n \times m$ *matrix* (read "n by m matrix") is a matrix with n rows and m columns. Consider, for example, the following matrices:

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 3 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 & 1 \\ 2 & 5 \\ 1 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 3 & 3 & 2 \\ 1 & -1 & -5 \\ -2 & 2 & 3 \end{bmatrix}.$$

A is a 2×3 matrix, B is a 3×2 matrix, and C is a 3×3 matrix. A *square matrix* is one that has the same number of rows and columns. So, C is a square matrix, but A and B are not square matrices. The *main diagonal* of a square matrix is the diagonal between the upper left corner and the bottom right corner. For example, the main diagonal of the square matrix C is colored red (below).

$$C = \begin{bmatrix} 3 & 3 & 2 \\ 1 & -1 & -5 \\ -2 & 2 & 3 \end{bmatrix}$$

The rows of a matrix are enumerated from top to bottom, whereas the columns are enumerated from left to right. The i,j-th entry of a matrix is the entry that appears in the i-th row (from the top) and j-th column (from the left) in the matrix. A matrix A can be specified as follows:

$$A = \left[\ a_{i,j}\ \right]_{n \times m}.$$

This notation indicates that the matrix A is of size $n \times m$ (i.e. has n rows and m columns), and the i,j-th entry (i.e. the entry in the i-th row and j-th column) is $a_{i,j}$. So, if $A = \left[\ a_{i,j}\ \right]_{n \times m}$, then we have that

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{bmatrix}.$$

A *zero matrix* is a matrix all of whose entries are 0 (where the 0 comes from the field that we are working with). The zero matrix of size $n \times m$ is denoted by $O_{n \times m}$. For example,

$$O_{2 \times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

A *non-zero matrix* is a matrix that has at least one non-zero entry.

Notation: If $\mathbb{F}$ is a field, then the set of all $n \times m$ matrices with entries in $\mathbb{F}$ is denoted by $\mathbb{F}^{n\times m}$.

Terminology: A *real matrix* is a matrix whose entries are real numbers, whereas a *complex matrix* is a matrix whose entries are complex numbers.

1.2.2 Column vectors (or simply vectors)

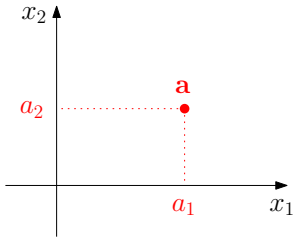
A *column vector*, or simply *vector*, is a matrix with just one column. Here are some examples of vectors (in this case, vector entries are real numbers):

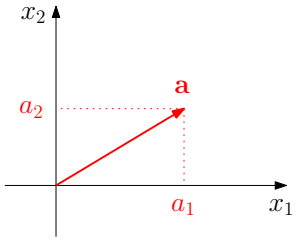
$$\mathbf{a} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} -13 \\ 0 \\ 0 \\ \pi \end{bmatrix}, \qquad \mathbf{c} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

Vectors are typically denoted by bold letters (e.g. $\mathbf{a}$, $\mathbf{u}$, $\mathbf{x}$) or by letters with an arrow on top (e.g. $\vec{a}$, $\vec{u}$, $\vec{x}$). The zero vector (i.e. vector $\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$) is denoted by $\mathbf{0}$ or $\vec{0}$. (The number of entries in a zero vector should either be made explicit or be clear from context.) A *non-zero vector* is a vector that has at least one non-zero entry.

Notation: If $\mathbb{F}$ is a field, then the set of all (column) vectors with n entries, all of them in $\mathbb{F}$, is denoted by $\mathbb{F}^n$. (Thus, $\mathbb{F}^n = \mathbb{F}^{n\times 1}$.)

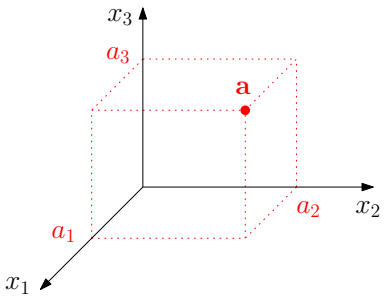
Geometric interpretation of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$. A vector $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ in $\mathbb{R}^2$ can be represented in the two-dimensional Euclidean space either as a point (see the picture below, on the left) or as a line segment with an arrow starting at the origin (see the picture below, on the right).

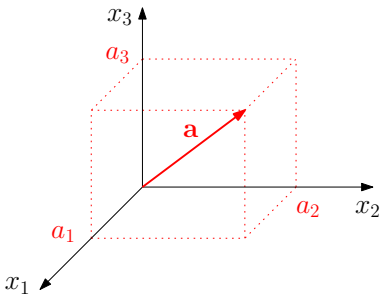




The zero vector $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is simply the origin.

A vector $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$ in $\mathbb{R}^3$ has a similar geometric interpretation in the three-dimensional Euclidean space (see the picture below).





Once again, the zero vector $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ is simply the origin.

Vectors in $\mathbb{R}^n$ for $n \geq 4$ are higher-dimensional analogs of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

1.2.3 Row vectors

A *row vector* is a matrix with only one row. For example, the following are row vectors (in this case, vector entries are real numbers):

- $\mathbf{a} = \begin{bmatrix} 1 & -3 \end{bmatrix}$;
- $\mathbf{b} = \begin{bmatrix} -13 & 0 & 0 & \pi \end{bmatrix}$;
- $\mathbf{c} = \begin{bmatrix} 1 & 2 & 0 & -1 & 1 \end{bmatrix}$.

In these lecture notes, row vectors will appear less commonly than column vectors. The set of all row vectors with n entries, all of them in some field $\mathbb{F}$, is denoted by $\mathbb{F}^{1\times n}$ (i.e. exactly the same way as the set of all $1 \times n$ matrices with entries in $\mathbb{F}$).

1.2.4 Specifying matrices in terms of their rows or columns

The columns of a matrix can be seen as (column) vectors, and matrices can be specified in terms of their columns. When we specify a matrix $A \in \mathbb{F}^{n\times m}$ (where $\mathbb{F}$ is some field) in the form

$$A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix},$$

we mean that $\mathbf{a}_1, \dots, \mathbf{a}_m$ are the columns of A (appearing in that order from left to right in the matrix A), and moreover, $\mathbf{a}_1, \dots, \mathbf{a}_m$ are vectors in $\mathbb{F}^n$. For example, if $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$, where $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $\mathbf{a}_3 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, then $A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 0 & 4 \end{bmatrix}$.

Similarly, the rows of a matrix can be seen as row vectors, and matrices can be specified in terms of their rows. When we specify a matrix $A \in \mathbb{F}^{n\times m}$ (where $\mathbb{F}$ is

some field) in the form

$$A = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix},$$

we mean that $\mathbf{r}_1, \ldots, \mathbf{r}_n$ are the rows of A (appearing in that order from top to bottom in the matrix A), and moreover, $\mathbf{r}_1, \ldots, \mathbf{r}_n$ are row vectors in $\mathbb{F}^{1\times m}$.

For example, if $A = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}$, where $\mathbf{r}_1 = \begin{bmatrix} 1 & 2 & 1 & 3 \end{bmatrix}$ and $\mathbf{r}_2 = \begin{bmatrix} 3 & 4 & 4 & 3 \end{bmatrix}$, then $A = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 3 & 4 & 4 & 3 \end{bmatrix}$.

1.3 Systems of linear equations

A *linear equation* in the variables $x_1, \ldots, x_m$ is an equation that can be written in the form

$$a_1x_1 + \cdots + a_mx_m = b,$$

where b and the coefficients $a_1, \ldots, a_n$ are elements of some field $\mathbb{F}$. For example, the equation

$$x_1 - 3(x_2 - x_1) = 7x_3 - 4,$$

with coefficients understood to be in $\mathbb{R}$, is a linear equation because it can be algebraically rearranged to have the following form:

$$4x_1 - 3x_2 - 7x_3 = -4,$$

which is obviously a linear equation. On the other hand, equations

$$x_1^3 + x_2 = 17 \quad \text{and} \quad x_1 - \sqrt{x_2} = 5$$

are **not** linear because of x_1^3 and $\sqrt{x_2}$.

A *system of linear equations*, or a *linear system*, is a collection of one or more linear equations involving the same variables, say $x_1, \ldots, x_m$ (and with coefficients from the same field). For example, the following is a linear system (here, the coefficients are assumed to be in $\mathbb{R}$):

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 7x_2 & & & - & \pi x_4 & = & -\sqrt{3} \\ & & -3x_2 & + & 17x_3 & - & 3x_4 & = & 2 \\ x_1 & + & x_2 & - & 2x_3 & + & 7x_4 & = & \frac{11}{2} \end{array}$$

Remark: Typographically, we normally arrange equations in our system so that the terms involving the same variable are below each other (i.e. visually in the same column).

A *solution* of a linear system in variables $x_1, \ldots, x_m$ is a list $s_1, \ldots, s_m$ of numbers (from the same field as the coefficients of the system) such that each equation becomes a true statement when $s_1, \ldots, s_m$ are substituted for $x_1, \ldots, x_m$, respectively.

Example 1.3.1. *Consider the linear system*

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 9 \\ & & 2x_2 & + & 3x_3 & = & 16 \\ x_1 & + & x_2 & - & x_3 & = & 4 \end{array}$$

with coefficients in $\mathbb{R}$*. Then*

$$\begin{array}{rcl} x_1 & = & 1 \\ x_2 & = & 5 \\ x_3 & = & 2 \end{array}$$

is a solution of the system above.

Example 1.3.2. *Consider the linear system*

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & 0 \\ 2x_1 & + & x_2 & = & 1 \end{array}$$

with coefficients in $\mathbb{Z}_3$*. Then*

$$\begin{array}{rcl} x_1 & = & 1 \\ x_2 & = & 2 \end{array}$$

is a solution of the system above.

The *set of solutions* or *solution set* of a linear system is the set of all solutions of that system. Our main goal in this section is to describe a procedure for finding the solution set of any linear system.

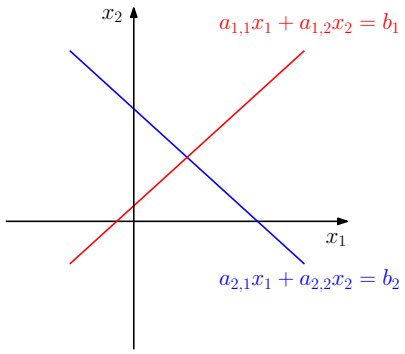
A linear system may have no solutions, may have a unique solution (i.e. exactly one solution), or may have more than one solution. A system that has at least one solution is called *consistent*; a system that has no solutions is said to be *inconsistent*.

Linear systems with real coefficients. Consider the following system of two linear equations in two variables, with coefficients in $\mathbb{R}$.

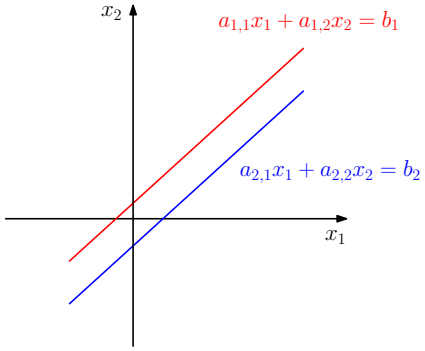
$$\begin{array}{rcrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & = & b_2 \end{array}$$

Let us assume that at least one of the coefficients $a_{1,1}, a_{1,2}$ is non-zero, and similarly, that at least one of the coefficients $a_{2,1}, a_{2,2}$ is non-zero. Then each of the two equations above defines a line in the plane. There are three possibilities for these two lines:

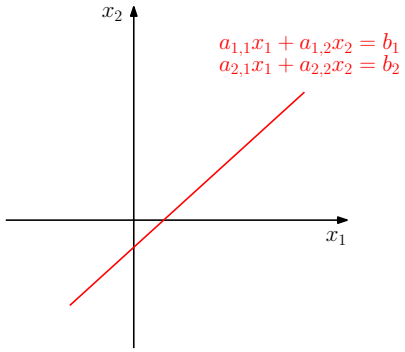
1. the two lines may intersect in one point (in this case, the system has a unique solution, and in particular, it is consistent);



2. the two lines may be distinct, parallel lines (in this case, the system has no solutions, i.e. it is inconsistent);



3. the two lines may be identical (in this case, the system has infinitely many solutions, and in particular, the system is consistent).[3]



On the other hand, suppose that we have a system of two linear equations in three variables (with coefficients in $\mathbb{R}$).

$$\begin{array}{ccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & a_{1,3}x_3 & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & a_{2,3}x_3 & = & b_2 \end{array}$$

[3]Note that the two lines may be identical even if the two equations are different. For instance, $x_1 + x_2 = 1$ and $2x_1 + 2x_2 = 2$ define the same line.

Similarly to the above, let us assume that at least one of the coefficients $a_{1,1}, a_{1,2}, a_{1,3}$ is non-zero, and that at least one of the coefficients $a_{2,1}, a_{2,2}, a_{2,3}$ is non-zero. Then each of the two equations above defines a plane in the three-dimensional Euclidean space. Those two planes may intersect in a line (in which case the system has infinitely many solutions, and in particular, the system is consistent); or the two planes may be distinct and parallel (in which case, the system has no solutions, i.e. it is inconsistent); or the two planes may be identical (in which case the system has infinitely many solutions, and in particular, the system is consistent).

1.3.1 The augmented matrix and the coefficient matrix of a linear system

Suppose we are given a system of n linear equations in m variables, as follows.

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \dots & + & a_{1,m}x_m & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \dots & + & a_{2,m}x_m & = & b_2 \\ & & & & & & & \vdots & \\ a_{n,1}x_1 & + & a_{n,2}x_2 & + & \dots & + & a_{n,m}x_m & = & b_n \end{array}$$

The *coefficient matrix* of this system is the $n \times m$ matrix

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{bmatrix}.$$

To fully capture our linear system, we need a bigger matrix, called the "augmented matrix" of the linear system. We start with the coefficient matrix A, and then we form the vector whose entries are the numbers to the right of the equality sign:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

The *augmented matrix* of our linear system is the $n \times (m+1)$ matrix

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,m} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} & b_n \end{array}\right].$$

Obviously, a linear system is fully "encoded" by its augmented matrix. The vertical dotted line is optional, but visually, it helps separate the coefficients to the left of the equality sign (those that form the coefficient matrix) from the numbers to the right of the equality sign.

Example 1.3.3. *Find the coefficient matrix and the augmented matrix of the linear system below (with coefficients understood to be in $\mathbb{R}$).*

$$\begin{array}{rcrcrcl} 3x_1 & + & 2x_2 & + & 5x_3 & = & 7 \\ & & 3x_2 & - & x_3 & = & 0 \end{array}$$

Solution. The coefficient matrix of the linear system is $\left[\begin{array}{rrr} 3 & 2 & 5 \\ 0 & 3 & -1 \end{array}\right]$, whereas the augmented matrix is $\left[\begin{array}{rrr|r} 3 & 2 & 5 & 7 \\ 0 & 3 & -1 & 0 \end{array}\right]$. □

Example 1.3.4. *Find the coefficient matrix and the augmented matrix of the linear system below (with coefficients understood to be in $\mathbb{Z}_3$).*

$$\begin{array}{rcl} 2x_1 + x_3 + 2 & = & x_2 \\ x_2 + x_3 & = & 2x_1 \end{array}$$

Solution. We first algebraically rearrange the system above to get it into standard form (below).[4]

$$\begin{array}{rcrcrcl} 2x_1 & + & 2x_2 & + & x_3 & = & 1 \\ x_1 & + & x_2 & + & x_3 & = & 0 \end{array}$$

We can now easily read off the two matrices that we need. The coefficient matrix of the linear system is $\left[\begin{array}{rrr} 2 & 2 & 1 \\ 1 & 1 & 1 \end{array}\right]$, whereas the augmented matrix is $\left[\begin{array}{rrr|r} 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{array}\right]$. □

1.3.2 Elementary row operations

Two linear systems (with the same variables) are *equivalent* if they have exactly the same solution set. Now, suppose we are given a system of linear equations such as the one below (with coefficients understood to be in some field $\mathbb{F}$).

$$\begin{array}{rcrcrcrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \ldots & + & a_{1,m}x_m & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \ldots & + & a_{2,m}x_m & = & b_2 \\ & & & & & & & \vdots & \\ a_{n,1}x_1 & + & a_{n,2}x_2 & + & \ldots & + & a_{n,m}x_m & = & b_n \end{array}$$

We would like to manipulate this system in a way that allows us to "read off" the solution set of the system. There are three basic ways that we can manipulate the

[4] Remember: We are working in $\mathbb{Z}_3$! We manipulated the first equation as follows.

1. First, we added $2x_2$ to both sides of the equation to obtain $2x_1 + 2x_2 + x_3 + 2 = 0$. (We used the fact that $x_2 + 2x_2 = (1+2)x_2 = 0x_2 = 0$.)
2. Then, we added 1 to both sides of the equation to obtain $2x_1 + 2x_2 + x_3 = 1$. (We used the fact that $2 + 1 = 0$.)

We manipulated the second equation by adding x_1 to both sides to obtain $x_1 + x_2 + x_3 = 0$. (We used the fact that $2x_1 + x_2 = (2+1)x_2 = 0x_2 = 0$.)

system in a way that does not change the solution set (i.e. in a way that produces an equivalent linear system). We list these three operations/manipulations below (the scalars that we mention always belong to the same field as the coefficients of the linear system in question), and we illustrate each operation with an example (in each example, the coefficients are understood to be in $\mathbb{R}$.)

1. Swap (interchange) two equations.

   - For example, by swapping the first and third equation of the linear system on the left, we obtain the linear system on the right.

$$\begin{array}{rcrcrcr} x_1 & + & 3x_2 & - & 2x_3 & = & -1 \\ \frac{1}{2}x_1 & & & + & 2x_3 & = & 0 \\ x_1 & + & x_2 & + & 2x_3 & = & 2 \end{array} \quad \longrightarrow \quad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 2 \\ \frac{1}{2}x_1 & & & + & 2x_3 & = & 0 \\ x_1 & + & 3x_2 & - & 2x_3 & = & -1 \end{array}$$

   It is obvious that this operation does not alter the solution set.

2. Multiply one equation by a **non-zero** scalar.

   - For example, by multiplying the second equation of the linear system on the left by 2, we obtain the linear system on the right.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 2 \\ \frac{1}{2}x_1 & & & + & 2x_3 & = & 0 \\ x_1 & + & 3x_2 & - & 2x_3 & = & -1 \end{array} \quad \longrightarrow \quad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 2 \\ x_1 & & & + & 4x_3 & = & 0 \\ x_1 & + & 3x_2 & - & 2x_3 & = & -1 \end{array}$$

   Let us explain why this does not alter the solution set. Suppose we have multiplied the i-th equation of our linear system by some scalar $\alpha \neq 0$. Obviously, all solutions of the old system are still solutions of the new system. On the other hand, by multiplying the i-th equation of the new system by α^{-1} (the multiplicative inverse of α), we get the old system back.[5] So, any solution of the new system is a solution of the old system as well.

   Warning: Do **not** multiply an equation by 0, since that "kills" the equation!

3. Add a scalar multiple of one equation to another equation.

   - For example, by adding (-1) times the second equation to the third equation of the linear system on the left, we obtain the linear system on the right.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 2 \\ x_1 & & & + & 4x_3 & = & 0 \\ x_1 & + & 3x_2 & - & 2x_3 & = & -1 \end{array} \quad \longrightarrow \quad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 2 \\ x_1 & & & + & 4x_3 & = & 0 \\ & & 3x_2 & - & 6x_3 & = & -1 \end{array}$$

[5] In the example above, we would multiply the second equation of the linear system on the right by $\frac{1}{2}$ in order to obtain the linear system on the left.

Let us explain why this does not alter the solution set. Suppose we have added α times the i-th equation to the j-th equation (where $i \neq j$). Obviously, any solution of the old system is also a solution of the new system. On the other hand, if we start with the new system, then add $-\alpha$ times the i-th equation to the j-th equation, we get the old system back.[6] So, any solution of the new system is a solution of the old system as well.

Instead of manipulating systems linear systems in this way, we can manipulate their augmented matrices. There are three types of "elementary row operations" on matrices (with entries understood to be in some field $\mathbb{F}$), which we list below. (The scalars that we mention always belong to the same field as the entries of the matrix in question.) We illustrate each type of elementary row operation with an example (in our examples, the matrix entries are assumed to be in $\mathbb{R}$).

1. Swap (interchange) two rows.
   - We denote the operation of swapping rows i and j ($i \neq j$) by "$R_i \leftrightarrow R_j$."
   - For example, we can swap the first and third row of the matrix on the left to obtain the matrix on the right.

$$\left[\begin{array}{cccc} 1 & 3 & -2 & -1 \\ \frac{1}{2} & 0 & 2 & 0 \\ 1 & 1 & 2 & 2 \end{array}\right] \overset{R_1 \leftrightarrow R_3}{\sim} \left[\begin{array}{cccc} 1 & 1 & 2 & 2 \\ \frac{1}{2} & 0 & 2 & 0 \\ 1 & 3 & -2 & -1 \end{array}\right]$$

2. Multiply one row by a **non-zero** scalar.
   - We denote the operation of multiplying row i by a scalar $\alpha \neq 0$ by "$R_i \to \alpha R_i$."
   - For instance, we can multiply the second row of the matrix on the left by 2 to obtain the matrix on the right.

$$\left[\begin{array}{cccc} 1 & 1 & 2 & 2 \\ \frac{1}{2} & 0 & 2 & 0 \\ 1 & 3 & -2 & -1 \end{array}\right] \overset{R_2 \to 2R_2}{\sim} \left[\begin{array}{cccc} 1 & 1 & 2 & 2 \\ 1 & 0 & 4 & 0 \\ 1 & 3 & -2 & -1 \end{array}\right]$$

3. Add a scalar multiple of one row to another row.
   - We denote the operation of adding scalar α times row i to row j ($i \neq j$) by "$R_j \to R_j + \alpha R_i$."
   - For example, we can add (-1) times the second row to the third row of the matrix on the left to obtain the matrix on the right.

$$\left[\begin{array}{cccc} 1 & 1 & 2 & 2 \\ 1 & 0 & 4 & 0 \\ 1 & 3 & -2 & -1 \end{array}\right] \overset{R_3 \to R_3 + (-1)R_2}{\sim} \left[\begin{array}{cccc} 1 & 1 & 2 & 2 \\ 1 & 0 & 4 & 0 \\ 0 & 3 & -6 & -1 \end{array}\right]$$

[6] In the example above, we would add 1 times the second equation to the third equation of the system on the right to obtain the system on the left.

Note: Instead of "$R_3 \to R_3 + (-1)R_2$," we could also have written (and we typically do write) just "$R_3 \to R_3 - R_2$."

Importantly, all elementary row operations are reversible:

1. we can undo (reverse) the operation of swapping two rows ("$R_i \leftrightarrow R_j$") by applying the same operation again;

2. we can undo (reverse) the operation of multiplying row i by a scalar $\alpha \neq 0$ ("$R_i \to \alpha R_i$") by multiplying row i by α^{-1} ("$R_i \to \alpha^{-1} R_i$");

3. we can undo (reverse) the operation of adding scalar α times row i to another row j ("$R_j \to R_j + \alpha R_i$") by adding $-\alpha$ times row i to row j ("$R_j \to R_j - \alpha R_i$").

Remark: Solving systems of linear equations is our primary motivation for introducing elementary row operations. However, we can, in principle, perform elementary row operations on **any** matrix (with entries in some field), even one that was not obtained as an augmented matrix of a linear system. We will, indeed, do this at various points in these lecture notes. However, for now, it is useful to think of elementary row operations on matrices as a more compact way of performing the corresponding operations on linear systems.

Terminology/Notation: If one matrix can be obtained from another via some sequence of elementary row operations, then the two matrices are said to be *row equivalent*. If matrices A and B are row equivalent, then we write $A \sim B$. Note that any two row equivalent matrices are of the same size (i.e. have the same number of rows and the same number of columns), and their entries belong to the same field.

Remark: Clearly, if two matrices with at least two columns (and with entries in some field $\mathbb{F}$) are row equivalent, then they encode equivalent linear systems (as augmented matrices).[7]

The following proposition states that, for a field $\mathbb{F}$, row equivalence is an equivalence relation on the set $\mathbb{F}^{n \times m}$.

Proposition 1.3.5. *Let $\mathbb{F}$ be a field. Then all the following hold:*

(a) for all $A \in \mathbb{F}^{n \times m}$, $A \sim A$;

(b) for all $A, B \in \mathbb{F}^{n \times m}$, if $A \sim B$, then $B \sim A$;

(c) for all $A, B, C \in \mathbb{F}^{n \times m}$, if $A \sim B$ and $B \sim C$, then $A \sim C$.

[7] A matrix that only has one column is not the augmented matrix of any linear system. That said, according to our definition, two one-column matrices (i.e. two column vectors) can be row equivalent.

Proof. (a) Fix $A \in \mathbb{F}^{n\times m}$. By, for example, multiplying the first row of A by 1 (i.e. by applying the elementary row operation "$R_1 \to 1R_1$"), we obtain the original matrix A; so, $A \sim A$.

(b) Fix $A, B \in \mathbb{F}^{n\times m}$, and assume that $A \sim B$. Then by applying some sequence $R_1, \dots, R_k$ of elementary row operations to A, we obtain the matrix B. But we know that elementary row operations are reversible! For each $i \in \{1, \dots, k\}$, let R_i' be the elementary row operation that reverses (undoes) the elementary row operation R_i. If we apply the sequence $R_k', \dots, R_1'$ of elementary row operations to B, we obtain the matrix A. So, $B \sim A$.

(c) Fix $A, B, C \in \mathbb{F}^{n\times m}$, and assume that $A \sim B$ and $B \sim C$. Since $A \sim B$, we know that B can be obtained by applying some sequence $R_1, \dots, R_k$ of elementary row operations to A. Similarly, since $B \sim C$, we know that B can be obtained by applying some sequence $R_{k+1}, \dots, R_{k+\ell}$ of elementary row operations to B. But now if we apply the sequence $R_1, \dots, R_k, R_{k+1}, \dots, R_{k+\ell}$ to A, we get C. □

1.3.3 Row reduction

A *zero row* of a matrix is a row in which all entries are zero, and a *non-zero row* is a row that has at least one non-zero entry. (*Zero* and *non-zero* columns are defined analogously.) The *leading entry* of a non-zero row is the leftmost non-zero entry of that row.

A matrix is in *row echelon form* (or simply *echelon form*), abbreviated *REF*, if it satisfies the following two conditions:

1. all non-zero rows are above any zero rows;
2. each leading entry of a non-zero row (other than the top row) is in a column strictly to the right of the column containing the leading entry of the row right above.[8]

If, in addition, the matrix satisfies the following two conditions, then it is in *reduced row echelon form* (or simply *reduced echelon form*), abbreviated RREF:

3. the leading entry in each non-zero row is 1;
4. each leading 1 is the only non-zero entry in its column.

If a matrix is in row echelon form (resp. reduced row echelon form), then we also say that the matrix is a *row echelon matrix* (resp. *reduced row echelon matrix*).

[8] So, all entries in a column below a leading entry of a row are zeros.

Schematically, a matrix in row echelon form looks like this (here, ■'s represent non-zero numbers, and $*$'s represent arbitrary numbers):

$$\begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

On the other hand, a matrix in **reduced** row echelon form schematically looks like this ($*$'s represent arbitrary numbers):

$$\begin{bmatrix} 0 & 1 & * & 0 & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A *pivot position* of a matrix in row echelon form is the position of a leading entry of a non-zero row, and a *pivot column* of a matrix in row echelon form is a column that contains a pivot position. In our diagram representing a matrix in row echelon form, the pivot positions are the positions of the black squares, and the pivot columns are the columns containing those black squares. In the special case of matrices in reduced row echelon form, the pivot positions are the positions of the leading 1's of the non-zero rows, and the pivot columns are the columns containing those leading 1's. For example, the matrix below is in reduced row echelon form (the $*$'s are arbitrary numbers from the field in question); its pivot positions are boxed, and the pivot columns (four of them) are the ones with the boxed entries.

$$\begin{bmatrix} 0 & \boxed{1} & * & 0 & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & \boxed{1} & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & \boxed{1} & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Theorem 1.3.6. *Every matrix (with entries in some field) is row equivalent to a* ***unique*** *matrix in reduced row echelon form.*

Corollary 1.3.7. *If two row equivalent matrices (with entries in some field) are both in row echelon form, then they have exactly the same pivot positions and exactly the same pivot columns.*

We postpone the proof of Theorem 1.3.6 and Corollary 1.3.7 to subsection 1.3.7. For now, we introduce some terminology and notation. By Theorem 1.3.6, every matrix A (with entries in some field) is row equivalent to a unique matrix in reduced row echelon form, which we call the *reduced row echelon form* of A, denoted by $\mathrm{RREF}(A)$. A *row echelon form* of a matrix A is any matrix that is in row echelon form and is row equivalent to A. A matrix may have more than one row echelon form (i.e. it may be row equivalent to more than one matrix in row echelon form), but by Corollary 1.3.7, all row echelon matrices of a given matrix have the same "shape," i.e. their "black squares" are in the same place. The *pivot positions* and the *pivot columns* of an arbitrary matrix A (with entries from some field) are the pivot positions and the pivot columns, respectively, of any matrix in row echelon form that is row equivalent to A; by Corollary 1.3.7, this is well defined. In particular, if we have computed the reduced row echelon form of a matrix A, then we can immediately identify the pivot positions and the pivot columns of A.

Remark/Terminology: The number of pivot columns of a matrix A (equivalently: the number of pivot positions of A, or the number of non-zero rows in a row echelon form of A) has a special name: it is called the *rank* of A and is denoted by $\mathrm{rank}(A)$. We will study rank in more detail in section 1.6, and it will subsequently reappear in various contexts throughout these lecture notes.

The following corollary of Theorem 1.3.6 is also postponed to subsection 1.3.7.

Corollary 1.3.8. *Two matrices (with entries in some field) are row equivalent if and only if they have the same reduced row echelon form.*

The row reduction algorithm. We now describe an algorithm, called the *row reduction* algorithm, that transforms any matrix (with entries in some field) into a row equivalent matrix that is in reduced row echelon form.[9] The algorithm has two parts: the "forward phase" and the "backward phase." The forward phase transforms the matrix into one in row echelon form. The backward phase transforms a matrix in row echelon form into one in **reduced** row echelon form. The forward phase of the row reduction algorithm is also called "Gaussian elimination." The entire row reduction algorithm (with both the forward and the backward phase) is also called the "Gauss-Jordan elimination." In the description of the algorithm, we will use the word "pivot" to mean the actual number that is in the pivot position in question (or that we intend to move into the pivot position).[10] We now describe the row reduction algorithm.

[9] We note this algorithm proves the existence part of Theorem 1.3.6, but not the uniqueness part.

[10] Most texts never actually define the word (noun) "pivot." Instead, they only define "pivot positions" and "pivot columns," as we did above. If they use the word "pivot" (as a noun), they do so informally. We also use it slightly informally, but the advantage is that we get a simpler description of the row reduction algorithm.

Forward phase:

1. Begin with the leftmost non-zero column. This is a pivot column. The pivot position is at the top of the column.[11]

2. Select a non-zero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

3. Use elementary row operations of the form "$R_j \to R_j + \alpha R_i$" (where row i contains the pivot position in question, row j is below row i,[12] and α is a suitable scalar) to create zeros in all positions below the pivot position.

4. Cover (or ignore) the row containing the pivot position, as well as all the rows (if any) above it. Apply steps 1-4 to the submatrix that remains. Repeat the process until there are no more non-zero rows to modify.

Backward phase:

5. Beginning with the rightmost pivot column and working upward and to the left, create zeros above each pivot position. If a pivot is not 1, make it 1 by a scaling operation ("$R_i \to \alpha R_i$," for a suitable scalar $\alpha \neq 0$).

A couple of implementations of the row reduction algorithm are given below (see Examples 1.3.9 and 1.3.10). In each case, we first implement the forward phase, and then we implement the backward phase. In the forward phase, we use a horizontal dotted line as a visual aid: it separates the rows that have already been processed (those are the ones above the horizontal dotted line) from the ones that have not yet been processed (those are the ones below the horizontal dotted line). Moreover, the pivot column that we have identified and are currently processing (as per step 1 or step 5) is colored red.

Example 1.3.9. *Apply the row reduction algorithm to the matrix A below (with entries understood to be in $\mathbb{R}$) in order to compute its reduced row echelon form.*

$$A := \begin{bmatrix} 0 & -3 & -6 & 3 & 4 & -1 \\ 2 & 1 & -4 & 13 & -4 & 3 \\ 2 & 3 & 0 & 11 & -6 & 5 \end{bmatrix}$$

Solution. We first implement the forward phase of the algorithm in order to transform the matrix into one in row echelon form, as follows.

[11] In the initial iteration, this means that the pivot position is in the top row. However, in subsequent iterations, it will mean that the pivot position is "in the top row if we ignore the rows that we have already processed and are done with."

[12] So, $j > i$.

$$
\begin{aligned}
A &= \begin{bmatrix} 0 & -3 & -6 & 3 & 4 & -1 \\ 2 & 1 & -4 & 13 & -4 & 3 \\ 2 & 3 & 0 & 11 & -6 & 5 \end{bmatrix} \\
&\underset{\sim}{R_1 \leftrightarrow R_3} \begin{bmatrix} 2 & 3 & 0 & 11 & -6 & 5 \\ 2 & 1 & -4 & 13 & -4 & 3 \\ 0 & -3 & -6 & 3 & 4 & -1 \end{bmatrix} \\
&\underset{\sim}{R_2 \to R_2 - R_1} \begin{bmatrix} 2 & 3 & 0 & 11 & -6 & 5 \\ \hdashline 0 & -2 & -4 & 2 & 2 & -2 \\ 0 & -3 & -6 & 3 & 4 & -1 \end{bmatrix} \\
&\underset{\sim}{R_3 \to R_3 - \frac{3}{2}R_2} \begin{bmatrix} 2 & 3 & 0 & 11 & -6 & 5 \\ 0 & -2 & -4 & 2 & 2 & -2 \\ \hdashline 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.
\end{aligned}
$$

The forward phase of the row reduction algorithm is now complete: our matrix is in row echelon form. It remains to implement the backward phase in order to transform the matrix into one in **reduced** row echelon form. We compute:

$$
\begin{aligned}
A &\sim \begin{bmatrix} 2 & 3 & 0 & 11 & -6 & 5 \\ 0 & -2 & -4 & 2 & 2 & -2 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \quad \text{by the forward phase} \\
&\underset{\sim}{\substack{R_1 \to R_1 + 6R_3 \\ R_2 \to R_2 - 2R_3}} \begin{bmatrix} 2 & 3 & 0 & 11 & 0 & 17 \\ 0 & -2 & -4 & 2 & 0 & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \\
&\underset{\sim}{R_2 \to -\frac{1}{2}R_2} \begin{bmatrix} 2 & 3 & 0 & 11 & 0 & 17 \\ 0 & 1 & 2 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \\
&\underset{\sim}{R_1 \to R_1 - 3R_2} \begin{bmatrix} 2 & 0 & -6 & 14 & 0 & 8 \\ 0 & 1 & 2 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \\
&\underset{\sim}{R_1 \to \frac{1}{2}R_1} \begin{bmatrix} 1 & 0 & -3 & 7 & 0 & 4 \\ 0 & 1 & 2 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.
\end{aligned}
$$

The backward phase of row reduction is now complete: our matrix is in reduced row echelon form. Thus,

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & 0 & -3 & 7 & 0 & 4 \\ 0 & 1 & 2 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

□

Example 1.3.10. *Apply the row reduction algorithm to the matrix B below (with entries understood to be in $\mathbb{Z}_3$) in order to compute its reduced row echelon form.*

$$B := \begin{bmatrix} 0 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 1 \end{bmatrix}$$

Solution. We will compute keeping the addition and multiplication tables for $\mathbb{Z}_3$ (below) in mind.

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

We first implement the forward phase of the algorithm in order to transform the matrix into one in row echelon form, as follows.

$$B = \begin{bmatrix} 0 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 1 \end{bmatrix}$$

$$\underset{R_1 \leftrightarrow R_4}{\sim} \begin{bmatrix} 1 & 0 & 2 & 2 & 1 \\ 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 2 \end{bmatrix}$$

$$\underset{\substack{R_2 \to R_2 + R_1 \\ R_3 \to R_3 + R_1}}{\sim} \left[\begin{array}{ccccc} 1 & 0 & 2 & 2 & 1 \\ \hdashline 0 & 1 & 2 & 0 & 2 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 0 & 2 \end{array}\right]$$

$$\underset{\substack{R_3 \to R_3 + 2R_2 \\ R_4 \to R_4 + 2R_2}}{\sim} \left[\begin{array}{ccccc} 1 & 0 & 2 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 \\ \hdashline 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{array}\right]$$

$$\overset{R_4\to R_4+R_3}{\sim} \left[\begin{array}{ccccc} 1 & 0 & 2 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The forward phase of our row reduction algorithm is now complete: our matrix is in row echelon form. It remains to implement the backward phase in order to transform the matrix into one in **reduced** row echelon form.

$$B \overset{(*)}{\sim} \left[\begin{array}{ccccc} 1 & 0 & 2 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \overset{\substack{R_1\to R_1+R_3 \\ R_2\to R_2+R_3}}{\sim} \left[\begin{array}{ccccc} 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

where (*) follows from the forward phase (above). The backward phase of row reduction is now complete: our matrix is in reduced row echelon form. Thus,

$$\mathrm{RREF}(B) = \left[\begin{array}{ccccc} 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

□

When row reducing, we do not normally draw the horizontal dotted line and color the pivot column that we are processing (as we did in Examples 1.3.9 and 1.3.10). It is, however, good practice to indicate which elementary row operations are being performed at each stage. Let us take a look at a couple of additional examples (Examples 1.3.11 and 1.3.12 below) in which we omit the horizontal dotted line and pivot column coloring, but carefully indicate which elementary row operation(s) we are performing. For extra clarity, we also indicate the beginning and end of the forward and backward phase of the row reduction algorithm (though this is not strictly necessary).

Example 1.3.11. *Apply the row reduction algorithm to the matrices C_1 and C_2 below (with entries understood to be in $\mathbb{Z}_2$) in order to compute their reduced row echelon form.*

$$C_1 := \left[\begin{array}{cccccc} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{array}\right] \qquad C_2 := \left[\begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right]$$

Solution. We will compute keeping the addition and multiplication tables for $\mathbb{Z}_2$ (below) in mind.

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

We first row reduce the matrix C_1, as follows:

$$C_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$ start of the forward phase

$$\overset{\substack{R_2 \to R_2 + R_1 \\ R_3 \to R_3 + R_1}}{\sim} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\overset{R_2 \leftrightarrow R_3}{\sim} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$ end of the forward phase (the matrix is in row echelon form)

$$\overset{\substack{R_1 \to R_1 + R_3 \\ R_2 \to R_2 + R_3}}{\sim} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$ start of the backward phase

$$\overset{R_1 \to R_1 + R_2}{\sim} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$ end of the backward phase (the matrix is in reduced row echelon form).

The last matrix from the calculation above is in reduced row echelon form, and we conclude that

$$\mathrm{RREF}(C_1) = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

We now row reduce C_2, as follows:

$$C_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$ start of the forward phase

$$\overset{R_2\to R_2+R_1}{\underset{R_4\to R_4+R_1}{\sim}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\overset{R_3\to R_3+R_2}{\underset{R_4\to R_4+R_2}{\sim}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\overset{R_3\leftrightarrow R_4}{\sim} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ end of the forward phase (the matrix is in row echelon form)

$$\overset{R_1\to R_1+R_3}{\underset{R_2\to R_2+R_3}{\sim}} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ start of the backward phase

$$\overset{R_1\to R_1+R_2}{\sim} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ end of the backward phase (the matrix is in reduced row echelon form).

The last matrix from the calculation above is in reduced row echelon form, and we conclude that

$$\mathrm{RREF}(C_2) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

□

Remark: Note that in our solution of Example 1.3.11, we applied elementary row operations of only two types: "$R_i \leftrightarrow R_j$" and "$R_i \to R_i + R_j$" (where $i \neq j$ in both cases). This is not an accident! It is because we were working over $\mathbb{Z}_2$, and $\mathbb{Z}_2$ contains only one non-zero element (number), namely 1. Of course, elementary row operations of the type "$R_i \to 1R_i$" and "$R_i \to R_i + 0R_j$" (for $i \neq j$) are legal, but they leave the matrix unchanged.

Example 1.3.12. *Apply the row reduction algorithm to the matrices D_1 and D_2 below (with entries understood to be in $\mathbb{Z}_5$) in order to compute their reduced row echelon form.*

$$D_1 := \begin{bmatrix} 2 & 1 & 0 & 2 & 3 \\ 4 & 2 & 2 & 1 & 2 \\ 3 & 4 & 1 & 2 & 2 \end{bmatrix} \qquad D_2 := \begin{bmatrix} 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & 3 \\ 1 & 2 & 1 & 3 \\ 2 & 1 & 3 & 3 \end{bmatrix}$$

Solution. We will compute keeping the addition and multiplication tables for $\mathbb{Z}_5$ (below) in mind.

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

We first row reduce the matrix D_1, as follows:

$$D_1 = \begin{bmatrix} 2 & 1 & 0 & 2 & 3 \\ 4 & 2 & 2 & 1 & 2 \\ 3 & 4 & 1 & 2 & 2 \end{bmatrix}$$ start of the forward phase

$$\overset{\substack{R_2 \to R_2 + 3R_1 \\ R_3 \to R_3 + R_1 \\ \sim \\ (*)}}{} \begin{bmatrix} 2 & 1 & 0 & 2 & 3 \\ 0 & 0 & 2 & 2 & 1 \\ 0 & 0 & 1 & 4 & 0 \end{bmatrix}$$

$$\overset{\substack{R_3 \to R_3 + 2R_2 \\ \sim}}{} \begin{bmatrix} 2 & 1 & 0 & 2 & 3 \\ 0 & 0 & 2 & 2 & 1 \\ 0 & 0 & 0 & 3 & 2 \end{bmatrix}$$ end of the forward phase

$$\overset{\substack{R_1 \to 3R_1 \\ R_2 \to 3R_2 \\ R_3 \to 2R_3 \\ \sim \\ (**)}}{} \begin{bmatrix} 1 & 3 & 0 & 1 & 4 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$ start of the backward phase

$$\overset{\substack{R_1 \to R_1 + 4R_3 \\ R_2 \to R_2 + 4R_3 \\ \sim \\ (***)}}{} \begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$ end of the backward phase.

The last matrix from the calculation above is in reduced row echelon form, and we conclude that

$$\mathrm{RREF}(D_1) = \begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}.$$

Before moving on to the matrix D_2, let us make some comments on the calculation above. In (*), we observe that $3 = -2$ in $\mathbb{Z}_5$, and so the elementary row operation "$R_2 \to R_2+3R_1$" is exactly the same as the elementary row operation "$R_2 \to R_2-2R_1$" (and indeed, we could have written "$R_2 \to R_2 - 2R_1$" instead of "$R_2 \to R_2 + 3R_1$"). In any case, the goal was to turn the first entry of the second row into 0. We can accomplish this using the fact that $4 + 3 \cdot 2 = 0$ (in $\mathbb{Z}_5$), or using the fact that $4-2\cdot 2 = 0$ (again, in $\mathbb{Z}_5$). Similar remarks apply to (***): instead of "$R_1 \to R_1+4R_3$" and "$R_2 \to R_2+4R_3$," we could have written "$R_1 \to R_1 - R_3$" and "$R_2 \to R_2 - R_3$," respectively, and we would have obtained the same result. Finally, in (**), we were turning all pivots into 1's, using the fact that, in $\mathbb{Z}_5$, we have that $2^{-1} = 3$ and $3^{-1} = 2$.

We now row reduce the matrix D_2, as follows:

$$D_2 = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & 3 \\ 1 & 2 & 1 & 3 \\ 2 & 1 & 3 & 3 \end{bmatrix}$$ start of the forward phase

$$\underset{\substack{R_3\to R_3+R_1\\ R_4\to R_4+2R_1}}{\sim} \begin{bmatrix} 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 3 & 4 \\ 0 & 2 & 2 & 0 \end{bmatrix}$$

$$\underset{R_4\to R_4-2R_2}{\sim} \begin{bmatrix} 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 3 & 4 \end{bmatrix}$$ this is the same as "$R_4 \to R_4 + 3R_2$"

$$\underset{R_4\to R_4-R_3}{\sim} \begin{bmatrix} 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ this is the same as "$R_4 \to R_4 + 4R_3$"

end of the forward phase

$$\underset{\substack{R_1\to 4R_1\\ R_3\to 2R_3}}{\sim} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ turning the leading entry of each non-zero row into 1

start of the backward phase

$$\underset{\substack{R_1\to R_1+2R_3\\ R_2\to R_2+3R_3}}{\sim} \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\overset{R_1 \to R_1 + 3R_2}{\sim} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{end of the backward phase.}$$

The last matrix from the calculation above is in reduced row echelon form, and we conclude that

$$\mathrm{RREF}(D_2) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

□

Warning: When working over $\mathbb{Z}_p$ (where p is a prime number), all the entries inside of your matrices should **only** be the numbers $0, 1, \ldots, p-1$. As you process your matrix, you may possibly get negative numbers $-1, \ldots, -(p-1)$ inside of your matrix, but your **final answer** should **never** contain negative numbers. For example, if you are working over $\mathbb{Z}_3$, you should turn any -1's into 2's, and any -2's into 1's. In any case, when row reducing a matrix with entries in $\mathbb{Z}_p$ (for a prime number p), you should **never** (i.e. at no stage of the algorithm) have any numbers greater than $p-1$, and you should **never** have any fractions inside of your matrix!

Reasonable deviations from the row reduction algorithm. When computing the reduced row echelon form of a matrix, it is in principle legal to apply **any** elementary row operation at any stage (since this always produces a row equivalent matrix). However, to efficiently turn a matrix into one in reduced row echelon form, we should more or less follow the row reduction algorithm as described, because otherwise, our calculation may become very long and very messy. That said, slight deviations from the algorithm are sometimes a good idea. In particular, it is often a good idea to rescale one or more rows at the beginning or in the middle of the algorithm in order to eliminate fractions (when working over $\mathbb{R}$), or perhaps to turn the leading entry of one or more rows into 1's. For instance, in Example 1.3.12, we could have started our row reduction algorithm for D_1 by rescaling the first row so that the leading entry would become 1 (and then proceeding from there):

$$D_1 = \begin{bmatrix} 2 & 1 & 0 & 2 & 3 \\ 4 & 2 & 2 & 1 & 2 \\ 3 & 4 & 1 & 2 & 2 \end{bmatrix} \overset{R_1 \to 3R_1}{\sim} \begin{bmatrix} 1 & 3 & 0 & 1 & 4 \\ 4 & 2 & 2 & 1 & 2 \\ 3 & 4 & 1 & 2 & 2 \end{bmatrix}.$$

We could also have chosen to rescale all three rows so that the leading entries in all of them become 1:

$$D_1 = \begin{bmatrix} 2 & 1 & 0 & 2 & 3 \\ 4 & 2 & 2 & 1 & 2 \\ 3 & 4 & 1 & 2 & 2 \end{bmatrix} \overset{\substack{R_1 \to 3R_1 \\ R_2 \to 4R_2 \\ R_3 \to 2R_3}}{\sim} \begin{bmatrix} 1 & 3 & 0 & 1 & 4 \\ 1 & 3 & 3 & 4 & 3 \\ 1 & 3 & 2 & 4 & 4 \end{bmatrix}.$$

In any case, the basic idea of the row reduction algorithm is that, in the forward phase, we identify pivot columns from left to right and we systematically "clean them up" downward (i.e. all the entries below the pivot position that we are processing get turned into 0), whereas in the backward phase, we identify pivot columns from right to left and we "clean them up" upward (i.e. all the entries above the pivot position that we are processing get turned into 0). This basic procedure should be respected, since significant deviations from it may lengthen the procedure very considerably.

1.3.4 Solving linear systems

To find the solution set of a linear system, we proceed as follows. First, we form the augmented matrix of our linear system, and using row reduction, we find the reduced row echelon form of that matrix. Then, we "translate" this matrix (in reduced row echelon form) into the linear system that it encodes. The linear system that we obtain is equivalent to the one that we started with, that is, the two systems have exactly the same solution set. We now read off the solution set as follows.

1. If the rightmost column of the augmented matrix (the one to the right of the vertical dotted line) is a pivot column, then the system is inconsistent, i.e. it has no solutions.

   - For example, suppose that by row reduction, we obtained the following matrix (say, with coefficients in $\mathbb{R}$).

     $$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

     This matrix encodes the following linear system:

     $$\begin{aligned} x_1 \quad - \quad x_3 &= 0 \\ x_2 + 5x_3 &= 0 \\ 0 &= 1 \\ 0 &= 0 \end{aligned}$$

     Because of the equation "$0 = 1$," the system is inconsistent (i.e. it has no solutions).

2. If the rightmost column of the augmented matrix (the one to the right of the vertical dotted line) is **not** a pivot column, but all the other columns **are** pivot columns, then the system has a unique solution.

- For example, suppose that by row reduction, we obtained the following matrix (say, with coefficients in $\mathbb{R}$).

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This matrix encodes the following linear system:

$$\begin{array}{rcrcl} x_1 & & & = & -5 \\ & x_2 & & = & 0 \\ & & x_3 & = & 3 \\ & & 0 & = & 0 \end{array}$$

This system is consistent and has a unique solution, which we can immediately read off, as follows.

$$\begin{array}{rcl} x_1 & = & -5 \\ x_2 & = & 0 \\ x_3 & = & 3 \end{array}$$

3. If the rightmost column of the augmented matrix (the one to the right of the vertical dotted line) is **not** a pivot column, and at least one of the other columns is also **not** a pivot column, then the system has more than one solution, which we read off as follows. The variables that correspond to the **non-pivot** columns (we call these variables *free variables*) may take **any** value; these values (called *parameters*) are denoted by letters such as r, s, t. The variables that correspond to the pivot columns are called *basic*, and we solve for them in terms of our parameters. This form of solution is called the *parametric form of the solution*; we will also refer to it as the *general solution*.

- For example, suppose that by row reduction, we obtained the following matrix (say, with coefficients in $\mathbb{R}$).

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 5 & 6 & 0 \\ 0 & 0 & 1 & -1 & 7 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

This matrix encodes the linear system below.

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & & & + & 5x_4 & + & 6x_5 & = & 0 \\ & & & x_3 & - & & x_4 & + & 7x_5 & = & -3 \\ & & & & & & & & 0 & = & 0 \end{array}$$

The system is consistent and has more than one solution. The variables variables x_2, x_4, x_5 are free (because the non-pivot columns of the augmented matrix to the left of the vertical dotted line are columns $2, 4, 5$). The remaining variables are basic. We now read off the solutions as follows:

$$\begin{array}{rcll} x_1 &=& -2r - 5s - 6t & \\ x_2 &=& r & \\ x_3 &=& s - 7t - 3 & \\ x_4 &=& s & \\ x_5 &=& t & \text{where } r, s, t \in \mathbb{R}. \end{array}$$

Remark: Do not forget to specify which field your parameters come from! Here, we have "$r, s, t \in \mathbb{R}$" because the coefficients of our system are in $\mathbb{R}$.

Specifying the number of solutions of a linear system. An inconsistent linear system has zero solutions. A consistent system may have a unique solution (i.e. exactly one solution), or it may have more than one solution. A consistent system with no free variables has a unique solution. A consistent system that has at least one free variable has more than one solution, since each free variable can take an arbitrary value from the field $\mathbb{F}$ in question. If our field is infinite (for example, if it is $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$), then a consistent system with at least one free variable has infinitely many solutions. On the other hand, if our field $\mathbb{F}$ is finite, and our linear system is consistent with exactly k free variables, then the number of solutions of our system is precisely $|\mathbb{F}|^k$ (where $|\mathbb{F}|$ is the cardinality of $\mathbb{F}$, i.e. the number of elements in $\mathbb{F}$). In particular, if $\mathbb{F} = \mathbb{Z}_p$ for some prime number p, then a consistent system with exactly k free variables has exactly p^k solutions.

Example 1.3.13. *Solve the linear system below (with coefficients understood to be in $\mathbb{R}$), and specify how many solutions it has.*

$$\begin{array}{rcrcrcrcrcr} & & -3x_2 & - & 6x_3 & + & 3x_4 & + & 4x_5 & = & -1 \\ 2x_1 & + & x_2 & - & 4x_3 & + & 13x_4 & - & 4x_5 & = & 3 \\ 2x_1 & + & 3x_2 & & & + & 11x_4 & - & 6x_5 & = & 5 \end{array}$$

Solution. The augmented matrix of this linear system is the matrix A below.

$$A = \left[\begin{array}{ccccc:c} 0 & -3 & -6 & 3 & 4 & -1 \\ 2 & 1 & -4 & 13 & -4 & 3 \\ 2 & 3 & 0 & 11 & -6 & 5 \end{array}\right]$$

This is precisely the matrix from Example 1.3.9. The reduced row echelon form of this matrix (computed in Example 1.3.9) is

$$\mathrm{RREF}(A) = \left[\begin{array}{ccccc:c} 1 & 0 & -3 & 7 & 0 & 4 \\ 0 & 1 & 2 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right].$$

We see from RREF(A) that the rightmost column (the one to the right of the vertical dotted line) of the augmented matrix A of our linear system is **not** a pivot column; therefore, our linear system is consistent. We further see from RREF(A) that the pivot columns of A are its first, second and fifth column; so, the basic variables of our linear system are x_1, x_2, x_5, whereas the remaining variables (namely, x_3, x_4) are free. Now, RREF(A) is the augmented matrix of the linear system below, which is equivalent to our original linear system.

$$\begin{array}{rcrcrcrcl} x_1 & & - & 3x_3 & + & 7x_4 & & = & 4 \\ & x_2 & + & 2x_3 & - & x_4 & & = & 3 \\ & & & & & & x_5 & = & 2 \end{array}$$

We read off the solutions as follows:

$$\begin{array}{rcll} x_1 & = & 3s - 7t + 4 & \\ x_2 & = & -2s + t + 3 & \\ x_3 & = & s & \\ x_4 & = & t & \\ x_5 & = & 2 & \text{where } s, t \in \mathbb{R}. \end{array}$$

Our linear system is consistent and has two free variables. Since the field $\mathbb{R}$ is infinite, it follows that the number of solutions is infinite.

Optional: It is is easy to make mistakes when row reducing, and so it is generally a good idea to check our solutions. We do this by plugging in our general solution into the original system and checking that we get true statements. In this example, we get the following.

$$\begin{array}{rcrcrcrcrcr} & & -3(-2s+t+3) & - & 6s & + & 3t & + & 4 \cdot 2 & = & -1 \\ 2(3s-7t+4) & + & (-2s+t+3) & - & 4s & + & 13t & - & 4 \cdot 2 & = & 3 \\ 2(3s-7t+4) & + & 3(-2s+t+3) & & & + & 11t & - & 6 \cdot 2 & = & 5 \end{array}$$

By simplifying the left-hand-side, we see that all the equalities above are correct.

Remark: When checking solutions, all the parameters should cancel out! If, after simplifying, one of our equations became something like $2s + 7 = -2$ or $-t = 1$, this would tell us that we miscomputed somewhere. It does not matter that equalities such as $2s + 7 = -2$ or $-t = 1$ work for **some** values of s and t. They are supposed to work for **all** possible values of the parameters. If they fail to work for some values, then we know that we made a mistake somewhere and need to compute again. □

Example 1.3.14. *Solve the linear system below (with coefficients understood to be in $\mathbb{Z}_3$), and specify how many solutions it has.*

$$\begin{array}{rcrcrcrcl} & & x_2 & + & x_3 & & & = & 2 \\ 2x_1 & + & x_2 & & & + & x_4 & = & 1 \\ 2x_1 & + & x_2 & + & x_3 & + & x_4 & = & 1 \\ x_1 & & & + & 2x_3 & + & 2x_4 & = & 1 \end{array}$$

Solution. The augmented matrix of this linear system is the matrix B below.

$$B = \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 1 \end{array}\right]$$

This is precisely the matrix from Example 1.3.10. The reduced row echelon form of this matrix (computed in Example 1.3.10) is

$$\mathrm{RREF}(B) = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We see from $\mathrm{RREF}(B)$ that the rightmost column (the one to the right of the vertical dotted line) of the augmented matrix B of our linear system is **not** a pivot column; therefore, our linear system is consistent. We further see from $\mathrm{RREF}(B)$ that the pivot columns of B are its first, second, and third column; so, the basic variables of our linear system are x_1, x_2, x_3, whereas the remaining variable (namely, x_4) is free. Now, $\mathrm{RREF}(B)$ is the augmented matrix of the linear system below, which is equivalent to our original linear system.

$$\begin{array}{rcrcrcl} x_1 & & & & + & 2x_4 & = & 1 \\ & x_2 & & & & & = & 2 \\ & & x_3 & & & & = & 0 \\ & & & & & 0 & = & 0 \end{array}$$

We read off the solutions as follows:

$$\begin{array}{rcll} x_1 & = & t+1 & \\ x_2 & = & 2 & \\ x_3 & = & 0 & \\ x_4 & = & t & \text{where } t \in \mathbb{Z}_3. \end{array}$$

Our linear system is consistent and has one free variable. Since the field $\mathbb{Z}_3$ has three elements, the number of solutions is $3^1 = 3$.

Remark: To get x_1, we computed $x_1 = -2x_4 + 1 = x_4 + 1 = t + 1$, where we used the fact that in $\mathbb{Z}_3$, we have that $-2 = 1$.

Optional: We check our solutions by plugging them into our original system.

$$\begin{array}{rcrcrcrcl} & & 2 & + & 0 & & & = & 2 \\ 2(t+1) & + & 2 & & & + & t & = & 1 \\ 2(t+1) & + & 2 & + & 0 & + & t & = & 1 \\ (t+1) & & & + & 2 \cdot 0 & + & 2t & = & 1 \end{array}$$

By simplifying the left-hand-side, we see that all the equalities above are correct. Here, it is important to remember that we are working in $\mathbb{Z}_3$. For example, the left-hand-side of the second equality simplifies as follows:

$$2(t+1)+2+t \;=\; 2t+2+2+t \;=\; (\underbrace{2+1}_{=0})t+\underbrace{2+2}_{=1} \;=\; 1,$$

which is what we were supposed to get. □

Example 1.3.15. *Solve the linear systems* $(\star)$ *and* $(\star\star)$ *below (with coefficients understood to be in* $\mathbb{Z}_2$*), and specify how many solutions they have.*

$$\left.\begin{array}{rcrcrcrcrcl} x_1 & + & x_2 & + & x_3 & + & x_4 & + & x_5 & = & 1 \\ x_1 & + & x_2 & + & x_3 & & & & & = & 1 \\ x_1 & + & x_2 & & & & & + & x_5 & = & 1 \end{array}\right\} (\star)$$

$$\left.\begin{array}{rcrcrcl} x_1 & + & x_2 & + & x_3 & = & 1 \\ x_1 & & & + & x_3 & = & 0 \\ & & x_2 & & & = & 1 \\ x_1 & & & + & x_3 & = & 1 \end{array}\right\} (\star\star)$$

Solution. We begin by solving the linear system $(\star)$. Its augmented matrix is the matrix C_1 below.

$$C_1 \;=\; \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{array}\right]$$

This is precisely the matrix C_1 from Example 1.3.11. The reduced row echelon form of this matrix (computed in Example 1.3.11) is

$$\mathrm{RREF}(C_1) \;=\; \left[\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right].$$

We see from $\mathrm{RREF}(C_1)$ that the rightmost column (the one to the right of the vertical dotted line) of the augmented matrix C_1 of the linear system $(\star)$ is **not** a pivot column; therefore, the linear system $(\star)$ is consistent. We further see from $\mathrm{RREF}(C_1)$ that the pivot columns of C_1 are its first, third, and fourth column; so, the basic variables of the linear system $(\star)$ are x_1, x_3, x_4, whereas the remaining variables (namely, x_2, x_5) are free. Now, $\mathrm{RREF}(C_1)$ is the augmented matrix of the linear system below, which is equivalent to our original linear system $(\star)$.

$$\begin{array}{rcrcrcrcl} x_1 & + & x_2 & & & + & x_5 & = & 1 \\ & & & x_3 & & + & x_5 & = & 0 \\ & & & & x_4 & + & x_5 & = & 0 \end{array}$$

We read off the solutions as follows:

$$\begin{array}{rcll} x_1 &=& s+t+1 & \\ x_2 &=& s & \\ x_3 &=& t & \\ x_4 &=& t & \\ x_5 &=& t & \text{where } s,t \in \mathbb{Z}_2. \end{array}$$

The system $(\star)$ is consistent and has two free variables. Since the field $\mathbb{Z}_2$ has two elements, the number of solutions is $2^2 = 4$.

Remark: Remember, in $\mathbb{Z}_2$, we have that $-1 = 1$, and consequently, $-s = s$ and $-t = t$. We used this to solve for our basic variables (x_1, x_3, x_4).

Optional: We check our solutions by plugging them into our original system $(\star)$.

$$\begin{array}{rcrcrcrcrcl} (s+t+1) &+& s &+& t &+& t &+& t &=& 1 \\ (s+t+1) &+& s &+& t & & & & & =& 1 \\ (s+t+1) &+& s & & & & & + & t &=& 1 \end{array}$$

By simplifying the left-hand-side, we see that all the equalities above are correct. Here, it is important to keep in mind that we are working in $\mathbb{Z}_2$. For example, the left-hand-side of the first equality simplifies to

$$(s+t+1)+s+t+t+t \;=\; (\underbrace{1+1}_{=0})s + (\underbrace{1+1+1+1}_{=0})t + 1 \;=\; 1,$$

which is what we were supposed to get.

It remains to solve the linear system $(\star\star)$. Its augmented matrix is the matrix C_2 below.

$$C_2 \;=\; \left[\begin{array}{ccc:c} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{array}\right]$$

This is precisely the matrix C_2 from Example 1.3.11. The reduced row echelon form of this matrix (computed in Example 1.3.11) is

$$\mathrm{RREF}(C_2) \;=\; \left[\begin{array}{ccc:c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

We see from $\mathrm{RREF}(C_2)$ that the rightmost column (the one to the right of the vertical dotted line) of the augmented matrix C_2 of the linear system $(\star\star)$ is a pivot column; therefore, the linear system $(\star\star)$ is inconsistent. (The number of solutions is zero.)

Remark: We cannot check this answer (since there are no solutions to plug into the system). We can only hope that we did not make any mistakes in our calculation! When in doubt, redo the whole calculation from scratch. □

Example 1.3.16. *Solve the linear systems* $(\star)$ *and* $(\star\star)$ *below (with coefficients understood to be in* $\mathbb{Z}_5$*), and specify how many solutions they have.*

$$\left.\begin{array}{rcrcrcrcl} 2x_1 & + & x_2 & & & + & 2x_4 & = & 3 \\ 4x_1 & + & 2x_2 & + & 2x_3 & + & x_4 & = & 2 \\ 3x_1 & + & 4x_2 & + & x_3 & + & 2x_4 & = & 2 \end{array}\right\} (\star)$$

$$\left.\begin{array}{rcrcrcrcl} 4x_1 & + & 3x_2 & + & 2x_3 & = & 1 \\ & & x_2 & + & 2x_3 & = & 3 \\ x_1 & + & 2x_2 & + & x_3 & = & 3 \\ 2x_1 & + & x_2 & + & 3x_3 & = & 3 \end{array}\right\} (\star\star)$$

Solution. We begin by solving the linear system $(\star)$. Its augmented matrix is the matrix D_1 below.

$$D_1 := \left[\begin{array}{cccc:c} 2 & 1 & 0 & 2 & 3 \\ 4 & 2 & 2 & 1 & 2 \\ 3 & 4 & 1 & 2 & 2 \end{array}\right]$$

This is precisely the matrix D_1 from Example 1.3.12. The reduced row echelon form of this matrix (computed in Example 1.3.12) is

$$\operatorname{RREF}(D_1) = \left[\begin{array}{cccc:c} 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right].$$

We see from $\operatorname{RREF}(D_1)$ that the rightmost column (the one to the right of the vertical dotted line) of the augmented matrix D_1 of the linear system $(\star)$ is **not** a pivot column; therefore, the linear system $(\star)$ is consistent. We further see from $\operatorname{RREF}(D_1)$ that the pivot columns of D_1 are its first, third, and fourth column; so, the basic variables of the linear system $(\star)$ are x_1, x_3, x_4, whereas the remaining variable (namely, x_2) is free. Now, $\operatorname{RREF}(D_1)$ is the augmented matrix of the linear system below, which is equivalent to our original linear system $(\star)$.

$$\begin{array}{rcrcrcl} x_1 & + & 3x_2 & & & = & 0 \\ & & & x_3 & & = & 4 \\ & & & & x_4 & = & 4 \end{array}$$

We read off the solutions as follows:

$$\begin{array}{rcll} x_1 & = & 2t & \\ x_2 & = & t & \\ x_3 & = & 4 & \\ x_4 & = & 4 & \text{where } t \in \mathbb{Z}_5. \end{array}$$

The system $(\star)$ is consistent and has one free variable. Since the field $\mathbb{Z}_5$ has five elements, the number of solutions is $5^1 = 5$.

Optional: We check our solutions by plugging them into our original system $(\star)$.

$$\begin{array}{rcrcrcrcl} 2(2t) & + & t & & & + & 2\cdot 4 & = & 3 \\ 4(2t) & + & 2t & + & 2\cdot 4 & + & 4 & = & 2 \\ 3(2t) & + & 4t & + & 4 & + & 2\cdot 4 & = & 2 \end{array}$$

By simplifying the left-hand-side, we see that all the equalities above are correct. It is important to keep in mind that we are working in $\mathbb{Z}_5$. For example, the left-hand-side of the first equality simplifies to

$$2(2t) + t + 2\cdot 4 \;=\; (\underbrace{2\cdot 2+1}_{=0})t + (\underbrace{2\cdot 4}_{=3}) \;=\; 3,$$

which is what we were supposed to get.

It remains to solve the linear system $(\star\star)$. Its augmented matrix is the matrix D_2 below.

$$D_2 \;:=\; \left[\begin{array}{ccc|c} 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & 3 \\ 1 & 2 & 1 & 3 \\ 2 & 1 & 3 & 3 \end{array}\right]$$

This is precisely the matrix D_2 from Example 1.3.12. The reduced row echelon form of this matrix (computed in Example 1.3.12) is

$$\mathrm{RREF}(D_2) \;=\; \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

We see from $\mathrm{RREF}(D_2)$ that the rightmost column (the one to the right of the vertical dotted line) of the augmented matrix D_2 of the linear system $(\star\star)$ is **not** a pivot column; therefore, the linear system $(\star\star)$ is consistent. We further see from $\mathrm{RREF}(D_2)$ that all the columns of D_2 other than the rightmost one (i.e. all the columns to the left of the vertical dotted line) are pivot columns. So, all three variables of the linear system $(\star\star)$ are basic, and consequently, this linear system has a unique solution. $\mathrm{RREF}(D_2)$ is the augmented matrix of the linear system below.

$$\begin{array}{rcrcrcl} x_1 & & & & & = & 1 \\ & & x_2 & & & = & 2 \\ & & & & x_3 & = & 3 \\ & & & & 0 & = & 0 \end{array}$$

We now see that

$$\begin{array}{rcl} x_1 & = & 1 \\ x_2 & = & 2 \\ x_3 & = & 3 \end{array}$$

is the unique solution of the linear system ($\star\star$). In particular, the linear system ($\star\star$) has exactly one solution.

Optional: We check our solution by plugging it into our original system ($\star\star$).

$$\begin{array}{rcrcrcl} 4 \cdot 1 & + & 3 \cdot 2 & + & 2 \cdot 3 & = & 1 \\ & & 2 & + & 2 \cdot 3 & = & 3 \\ 1 & + & 2 \cdot 2 & + & 3 & = & 3 \\ 2 \cdot 1 & + & 2 & + & 3 \cdot 3 & = & 3 \end{array}$$

By simplifying the left-hand-side, we see that all the equalities above are correct. (Again, we must keep in mind that we are computing in $\mathbb{Z}_5$). □

Homogeneous linear systems. A *homogeneous linear system* is a linear system of the form

$$\begin{array}{rcrcccrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \dots & + & a_{1,m}x_m & = & 0 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \dots & + & a_{2,m}x_m & = & 0 \\ & & & & & & & \vdots & \\ a_{n,1}x_1 & + & a_{n,2}x_2 & + & \dots & + & a_{n,m}x_m & = & 0 \end{array}$$

where the coefficients $a_{i,j}$ are all from some field $\mathbb{F}$ (and 0 is also understood to be from that same field $\mathbb{F}$). Such a system is always consistent: $x_1 = x_2 = \dots = x_m = 0$ is a solution, called the *trivial solution*. A *non-trivial solution* of a homogeneous linear system is a solution that is not trivial. Some homogeneous linear systems have only the trivial solution, whereas others also have non-trivial solutions. (The former happens when there are no free variables, and the latter happens when there is at least one free variable.)

We note that when working with homogeneous linear systems, we typically row reduce only the coefficient matrix, and not the augmented matrix (see Example 1.3.17 below).

Example 1.3.17. *Solve the homogeneous linear system below, with coefficients understood to be in* $\mathbb{R}$.

$$\begin{array}{rcrcrcrcl} 2x_1 & - & 4x_2 & & & + & 6x_4 & = & 0 \\ 2x_1 & - & 4x_2 & + & 2x_3 & - & 2x_4 & = & 0 \end{array}$$

How many solutions does this homogeneous linear system have? Does it have any non-trivial solutions?

Proof. The **coefficient** matrix of our homogeneous linear system is

$$A := \begin{bmatrix} 2 & -4 & 0 & 6 \\ 2 & -4 & 2 & -2 \end{bmatrix}$$

We row reduce this matrix as follows:

$$\begin{aligned} A &= \begin{bmatrix} 2 & -4 & 0 & 6 \\ 2 & -4 & 2 & -2 \end{bmatrix} \\ &\overset{R_2 \to R_2 - R_1}{\sim} \begin{bmatrix} 2 & -4 & 0 & 6 \\ 0 & 0 & 2 & -8 \end{bmatrix} \\ &\overset{\substack{R_1 \to \frac{1}{2} R_1 \\ R_2 \to \frac{1}{2} R_2}}{\sim} \begin{bmatrix} 1 & -2 & 0 & 3 \\ 0 & 0 & 1 & -4 \end{bmatrix}. \end{aligned}$$

The last matrix from the calculation above is in reduced row echelon form, and so

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & -2 & 0 & 3 \\ 0 & 0 & 1 & -4 \end{bmatrix}.$$

Remark: We must keep in mind that A is the **coefficient** matrix of our linear system. The **augmented** matrix of our linear system would be $\left[\begin{array}{c|c} A & \mathbf{0} \end{array}\right]$. Since zero columns remain unchanged when we perform elementary row operations, the matrix $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & \mathbf{0} \end{array}\right]\right)$ is obtained by adding a zero column to the right of $\mathrm{RREF}(A)$. However, we do not normally write all this. We simply keep track of it mentally.

We now continue our computation. We see from the matrix $\mathrm{RREF}(A)$ that the pivot columns of the coefficient matrix A are its first and third column. So, x_1, x_3 are the basic variables, and x_2, x_4 are the free variables. Further, we see from $\mathrm{RREF}(A)$ that our original linear system is equivalent to the linear system below.

$$\begin{array}{rcrcrcl} x_1 & - & 2x_2 & & + & 3x_4 & = & 0 \\ & & & x_3 & - & 4x_4 & = & 0 \end{array}$$

We now read off the solutions as follows:

$$\begin{array}{rcll} x_1 &=& 2s - 3t & \\ x_2 &=& s & \\ x_3 &=& 4t & \\ x_4 &=& t & \text{where } s, t \in \mathbb{R}. \end{array}$$

Since our system has free variables (in fact, two of them), and since we are working over the infinite field $\mathbb{R}$, we see that our system has infinitely many solutions. In particular, our system has a non-trivial solution (in fact, it has infinitely many of them).

Optional: We can check our solutions by plugging them into the original linear system, as follows.

$$\begin{array}{rcrcrcrcl} 2(2s-3t) & - & 4s & & & + & 6t & = & 0 \\ 2(2s-3t) & - & 4s & + & 2(4t) & - & 2t & = & 0 \end{array}$$

By simplifying the left-hand-side, we see that both equalities above are correct. □

1.3.5 Solving systems of linear equations via back substitution

In this subsection, we present a way of solving linear systems by performing only the forward phase of the row reduction algorithm on the augmented matrix (and thus transforming it into a matrix in row echelon form, but not necessarily reduced row echelon form), and then solving the equivalent system via "back substitution." Rather than explaining the general principle, we give a couple of examples. We note that the systems in Examples 1.3.18 and 1.3.19 (below) are precisely those from Examples 1.3.13 and 1.3.14, respectively, and so, unsurprisingly, we get the same answer.

Example 1.3.18. *Find the solution set of the following system of linear equations (with coefficients in $\mathbb{R}$).*

$$\begin{array}{rcrcrcrcrcl} & & -3x_2 & - & 6x_3 & + & 3x_4 & + & 4x_5 & = & -1 \\ 2x_1 & + & x_2 & - & 4x_3 & + & 13x_4 & - & 4x_5 & = & 3 \\ 2x_1 & + & 3x_2 & & & + & 11x_4 & - & 6x_5 & = & 5 \end{array}$$

Solution. The augmented matrix of this linear system is the matrix A below.

$$A = \left[\begin{array}{ccccc:c} 0 & -3 & -6 & 3 & 4 & -1 \\ 2 & 1 & -4 & 13 & -4 & 3 \\ 2 & 3 & 0 & 11 & -6 & 5 \end{array}\right]$$

This is precisely the matrix from Example 1.3.9. By performing only the forward phase of the row reduction algorithm on the matrix A (see the solution of Example 1.3.9 for the details), we see that

$$A \sim \left[\begin{array}{ccccc:c} 2 & 3 & 0 & 11 & -6 & 5 \\ 0 & -2 & -4 & 2 & 2 & -2 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right],$$

where the pivot columns are in red for emphasis. Since the rightmost column (the one to the right of the vertical dotted line) is not a pivot column, our system is consistent. The pivot columns are the first, second, and fifth column. Therefore, x_1, x_2, x_5 are the basic variables, while the remaining variables (namely, x_3, x_4) are

free. Now, the matrix above is the augmented matrix of the linear system below (which is equivalent to our original system).

$$\begin{array}{rcrcrcrcrcr} 2x_1 & + & 3x_2 & & & + & 11x_4 & - & 6x_5 & = & 5 \\ & - & 2x_2 & - & 4x_3 & + & 2x_4 & + & 2x_5 & = & -2 \\ & & & & & & & & x_5 & = & 2 \end{array}$$

The free variables x_3, x_4 become arbitrary parameters, say $x_3 = s$ and $x_4 = t$ (where $s, t \in \mathbb{R}$). We plug this into the linear system above, and we obtain the following.

$$\begin{array}{rcrcrcrcrcr} 2x_1 & + & 3x_2 & & & + & 11t & - & 6x_5 & = & 5 \\ & - & 2x_2 & - & 4s & + & 2t & + & 2x_5 & = & -2 \\ & & & & & & & & x_5 & = & 2 \end{array}$$

We now solve for the basic variables, working our way from the bottom up. From the bottom equation, we get $x_5 = 2$. If we plug $x_5 = 2$ into the equation above (the second-from-bottom one), we get $-2x_2 - 4s + 2t + 2 \cdot 2 = -2$. By solving for x_2, we get $x_2 = -2s + t + 3$. We now plug in both $x_2 = -2s + t + 3$ and $x_5 = 2$ into the top equation, which yields $2x_1 + 3(-2s + t + 3) + 11t - 6 \cdot 2 = 5$. By solving for x_1, we obtain $x_1 = 3s - 7t + 4$. By putting all this together, we obtain the general solution of our linear system:

$$\begin{array}{rcll} x_1 & = & 3s - 7t + 4 & \\ x_2 & = & -2s + t + 3 & \\ x_3 & = & s & \\ x_4 & = & t & \\ x_5 & = & 2 & \text{where } s, t \in \mathbb{R}. \end{array}$$

We note that this is exactly the same as the solution that we obtained in Example 1.3.13. □

Example 1.3.19. *Find the solution set of the following system of linear equations (with coefficients in $\mathbb{Z}_3$).*

$$\begin{array}{rcrcrcrcr} & & x_2 & + & x_3 & & & = & 2 \\ 2x_1 & + & x_2 & & & + & x_4 & = & 1 \\ 2x_1 & + & x_2 & + & x_3 & + & x_4 & = & 1 \\ x_1 & & & + & 2x_3 & + & 2x_4 & = & 1 \end{array}$$

Solution. The augmented matrix of this linear system is the matrix B below.

$$B = \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 1 \end{array}\right]$$

This is precisely the matrix from Example 1.3.10. By only performing the forward phase of the row reduction algorithm on the matrix B (see the solution of Example 1.3.10 for the details), we obtain

$$B \sim \left[\begin{array}{cccc:c} 1 & 0 & 2 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

where the pivot columns are in red for emphasis. Since the rightmost column (the one to the right of the vertical dotted line) is not a pivot column, our system is consistent. The pivot columns are the first, second, and third column. Therefore, x_1, x_2, x_3 are the basic variables, whereas the variable x_4 is free. Now, the matrix above is the augmented matrix of the system below (which is equivalent to our original system).

$$\begin{array}{rcrcrcl} x_1 & & + & 2x_3 & + & 2x_4 & = & 1 \\ & x_2 & + & 2x_3 & & & = & 2 \\ & & & x_3 & & & = & 0 \\ & & & & & 0 & = & 0 \end{array}$$

The free variable x_4 becomes an arbitrary parameter, say $x_4 = t$ (where $t \in \mathbb{Z}_3$). We now plug this into the linear system above to obtain the following.

$$\begin{array}{rcrcrcl} x_1 & & + & 2x_3 & + & 2t & = & 1 \\ & x_2 & + & 2x_3 & & & = & 2 \\ & & & x_3 & & & = & 0 \\ & & & & & 0 & = & 0 \end{array}$$

Finally, we solve for the basic variables, working our way from the bottom up. The bottom equation ("$0 = 0$") gives us no information, so we ignore it. The equation above it yields $x_3 = 0$. We then plug that into the equation right above to obtain $x_2 + 2 \cdot 0 = 2$; solving for x_2, we get $x_2 = 2$. We now plug in both $x_2 = 2$ and $x_3 = 0$ into our top equation, and we obtain $x_1 + 2 \cdot 0 + 2t = 1$. Solving for x_1 (and keeping in mind that in $\mathbb{Z}_3$, we have $-2t = t$), we obtain $x_1 = t + 1$. By putting all this together, we obtain the general solution of our linear system:

$$\begin{array}{rcll} x_1 & = & t+1 & \\ x_2 & = & 2 & \\ x_3 & = & 0 & \\ x_4 & = & t & \text{where } t \in \mathbb{Z}_3. \end{array}$$

We note that this is exactly the same as the solution that we obtained in Example 1.3.14. □

Remark: Computers typically use back substitution to solve linear systems. However, when computing by hand (especially when the numbers are reasonably nice), it is

more convenient to find the **reduced** row echelon form of the augmented matrix and then read off the solutions, as described in subsection 1.3.4.

1.3.6 A few more remarks about the (reduced) row echelon form

Submatrices. A *submatrix* of a matrix A is any matrix obtained from A by possibly deleting some rows and some columns. For example, the matrix

$$B = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 0 & 2 & 3 & 5 \\ 3 & 5 & 5 & 3 \end{bmatrix}$$

is a submatrix of the matrix

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 8 & 4 & 2 & 1 & 2 & 4 \\ 3 & 4 & 5 & 5 & 4 & 3 \end{bmatrix}$$

because B can be obtained from A by deleting the third row and the second and fifth column, as shown below.

$$\left[\begin{array}{c|c|c|c|c|c} 1 & 2 & 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 8 & 4 & 2 & 1 & 2 & 4 \\ \hline 3 & 4 & 5 & 5 & 4 & 3 \end{array}\right]$$

Submatrices and row equivalence. Suppose that A and B are row equivalent matrices. If A' is a submatrix of A obtained by possibly deleting some **columns** of A (and no rows), and B' is the submatrix of B obtained by deleting the corresponding columns of B, then A' and B' are also row equivalent. Indeed, any sequence of elementary row operations that transforms A into B will transform A' into B'. For example, we have the following (matrix entries are assumed to be in $\mathbb{R}$):

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 2 \\ 1 & 2 & 1 & 2 & 1 & 1 \\ 3 & 8 & 3 & 8 & 5 & 5 \end{bmatrix} \overset{R_1 \leftrightarrow R_2}{\sim} \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 3 & 8 & 3 & 8 & 5 & 5 \end{bmatrix}$$

$$\overset{R_3 \to R_3 - 3R_1}{\sim} \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 2 & 0 & 2 & 2 & 2 \end{bmatrix}$$

$$\overset{R_3 \to \frac{1}{2}R_3}{\sim} \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

If we delete, say, the first, fourth, and fifth column throughout, we get the following:

$$\left[\begin{array}{cc|cc|cc} 0 & 1 & 2 & 0 & 1 & 2 \\ 1 & 2 & 1 & 2 & 1 & 1 \\ 3 & 8 & 3 & 8 & 5 & 5 \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{cc|cc|cc} 1 & 2 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 3 & 8 & 3 & 8 & 5 & 5 \end{array}\right]$$

$$\overset{R_3 \to R_3 - 3R_1}{\sim} \left[\begin{array}{cc|cc|cc} 1 & 2 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 2 & 0 & 2 & 2 & 2 \end{array}\right]$$

$$\overset{R_3 \to \frac{1}{2} R_3}{\sim} \left[\begin{array}{cc|cc|cc} 1 & 2 & 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{array}\right].$$

Warning: Do not delete rows! Deleting rows may destroy row equivalence.

Submatrices and the row echelon form. If A and B are matrices with the same number of rows, then we denote by $\left[\begin{array}{c|c} A & B \end{array}\right]$ the matrix obtained by placing A and B next to each other (A is to the left and B is to the right). For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 3 \\ 2 & 0 \\ 0 & 1 \end{bmatrix},$$

then

$$\left[\begin{array}{c|c} A & B \end{array}\right] = \left[\begin{array}{cccc|cc} 1 & 2 & 3 & 4 & 0 & 3 \\ 2 & 3 & 4 & 5 & 2 & 0 \\ 3 & 4 & 5 & 6 & 0 & 1 \end{array}\right].$$

Further, when we write $\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right] \sim \left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$ or $\mathrm{RREF}\left(\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right]\right) = \left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$, we implicitly assume that A_1 is of the same size as A_2 (i.e. A_1 and A_2 have the same number of rows and the same number of columns), and that B_1 and B_2 are of the same size.

It follows immediately from the appropriate definitions that if a matrix A is in row echelon form, then any submatrix of A that lies in the upper-left corner of A is also in row echelon form (see the diagram below for illustration; the submatrix in question is in red). However, other submatrices of a matrix in row echelon form need not be in row echelon form.

$$\begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Similarly, if a matrix A is in reduced row echelon form, then any submatrix of A that lies in the upper-left corner of A is also in reduced row echelon form (see the

diagram below for illustration; the submatrix in question is in red). However, other submatrices of a matrix in reduced row echelon form need not be in reduced row echelon form.

$$\begin{bmatrix} 0 & 1 & * & 0 & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In particular, this means that if a matrix $\left[\begin{array}{c|c} A & B \end{array}\right]$ is in row echelon form (resp. reduced row echelon form), then A is also in row echelon form (resp. reduced row echelon form).

Further, note that if A is a matrix in row echelon form (resp. reduced row echelon form), then adding an arbitrary number of zero columns to the right of the matrix A, or an arbitrary number of zero rows on the bottom of the matrix A, produces another matrix in row echelon form (resp. reduced row echelon form).[13]

Proposition 1.3.20. *Let $\mathbb{F}$ be a field, let $A_1, A_2 \in \mathbb{F}^{n\times m}$, and let $B_1, B_2 \in \mathbb{F}^{n\times p}$. Then both the following hold:*

(a) if $\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right] \sim \left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$, then $A_1 \sim A_2$ and $B_1 \sim B_2$;

(b) if $RREF\left(\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right]\right) = \left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$, then $RREF(A_1) = A_2$.

Warning: The converse of (a) does **not** hold, that is, it is possible that $A_1 \sim A_2$ and $B_1 \sim B_2$, but $\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right] \not\sim \left[\begin{array}{c|c} A_1 & B_1 \end{array}\right]$. Furthermore, in (b), it is possible that $\mathrm{RREF}(B_1) \neq B_2$ (because B_2 need not be in reduced row echelon form).

Proof. (a) Assume that $\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right] \sim \left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$. Then some sequence of elementary row operations transforms the matrix $\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right]$ into the matrix $\left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$. If we apply that same sequence of elementary row operations to the matrix A_1, we obtain the matrix A_2; similarly, if we apply that same sequence of elementary row operations to the matrix B_1, we obtain the matrix B_2. This proves that $A_1 \sim A_2$ and $B_1 \sim B_2$.

(b) Assume that $\mathrm{RREF}\left(\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right]\right) = \left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$. Then $\left[\begin{array}{c|c} A_1 & B_1 \end{array}\right] \sim \left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$, and so by (a), we have that $A_1 \sim A_2$. Moreover, since the matrix $\left[\begin{array}{c|c} A_2 & B_2 \end{array}\right]$ is in reduced row echelon form, so is the matrix A_2.[14] So, $\mathrm{RREF}(A_1) = A_2$. □

[13] As a matter of fact, we can insert zero **columns** into a matrix in row echelon form anywhere (left, right, middle), and we will obtain another matrix in row echelon form; similar remarks apply to matrices in **reduced** row echelon form. However, zero **rows** can only be added to the bottom, or at least below any non-zero rows (if we insert zero rows elsewhere, a matrix in row echelon form will no longer be in row echelon form).

[14] However, B_2 need not be in reduced row echelon form!

Recall that the $n \times m$ *zero matrix* in $\mathbb{F}^{n\times m}$ (where $\mathbb{F}$ is some field) is the $n \times m$ matrix, all of whose entries are 0;[15] this matrix is denoted by $O_{n\times m}$. As our next proposition shows, the converse of Proposition 1.3.20 holds in the special case when $B_1 = B_2 = O_{n\times p}$.

Proposition 1.3.21. *Let $\mathbb{F}$ be a field, let $A_1, A_2 \in \mathbb{F}^{n\times m}$, and let $O_{n\times p}$ be the zero matrix in $\mathbb{F}^{n\times p}$. Then both the following hold:*

(a) $A_1 \sim A_2$ if and only if $\left[\begin{array}{c:c} A_1 & O_{n\times p} \end{array}\right] \sim \left[\begin{array}{c:c} A_2 & O_{n\times p} \end{array}\right]$;

(b) $RREF(A_1) = A_2$ if and only if $RREF\left(\left[\begin{array}{c:c} A_1 & O_{n\times p} \end{array}\right]\right) = \left[\begin{array}{c:c} A_2 & O_{n\times p} \end{array}\right]$.

Proof. (a) Note that elementary row operations leave any zero columns unaffected. Therefore, any sequence of elementary row operations that transforms A_1 into A_2 also transforms $\left[\begin{array}{c:c} A_1 & O_{n\times p} \end{array}\right]$ into $\left[\begin{array}{c:c} A_2 & O_{n\times p} \end{array}\right]$, and vice versa. So, $A_1 \sim A_2$ if and only if $\left[\begin{array}{c:c} A_1 & O_{n\times p} \end{array}\right] \sim \left[\begin{array}{c:c} A_2 & O_{n\times p} \end{array}\right]$.

(b) By (a), we have that $A_1 \sim A_2$ if and only if $\left[\begin{array}{c:c} A_1 & O_{n\times p} \end{array}\right] \sim \left[\begin{array}{c:c} A_2 & O_{n\times p} \end{array}\right]$. Moreover, it follows immediately from the definition that A_2 is in reduced row echelon form if and only if the matrix $\left[\begin{array}{c:c} A_2 & O_{n\times p} \end{array}\right]$ is in reduced row echelon form. This proves that $\mathrm{RREF}(A_1) = A_2$ if and only if $\mathrm{RREF}\left(\left[\begin{array}{c:c} A_1 & O_{n\times p} \end{array}\right]\right) = \left[\begin{array}{c:c} A_2 & O_{n\times p} \end{array}\right]$. □

1.3.7 Proof of Theorem 1.3.6 and Corollaries 1.3.7 and 1.3.8

The existence part of Theorem 1.3.6 essentially follows from the row reduction algorithm. To prove uniqueness, we need the following lemma.

Lemma 1.3.22. *Let $\mathbb{F}$ be a field. If two reduced row echelon matrices in $\mathbb{F}^{n\times m}$ are row equivalent, then they are in fact equal.*

Proof. We keep the number of rows (n) fixed, and we proceed by induction on the number of columms (m). More precisely, we fix a positive integer n, and we prove (by induction on m) that for all positive integers m, if two reduced row echelon matrices in $\mathbb{F}^{n\times m}$ are row equivalent, then they are in fact equal.

Base case: $m = 1$. There are exactly two reduced row echelon matrices in $\mathbb{F}^{n\times 1}$, namely,

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

[15] The 0 is from our field $\mathbb{F}$.

(Here, the matrix/vector on the left has n many 0's, and the matrix/vector on the right has $n-1$ many 0's.) Since no elementary row operation will transform a zero column into a non-zero column, we see that these two matrices are not row equivalent.

Induction step: Fix a positive integer m, and assume inductively that the claim is true for m. We must prove that it is true for $m+1$. Fix two row equivalent reduced row echelon matrices A and B in $\mathbb{F}^{n\times(m+1)}$. We must show that $A = B$.

Let A' and B' be the $n \times m$ matrices obtained by deleting the rightmost column of A and B, respectively. Since A and B are in reduced row echelon form, so are A' and B'.[16] Next, since A and B are row equivalent, so are A' and B' (indeed, any sequence of elementary row operations that transforms A into B also transforms A' into B'). So, by the induction hypothesis, $A' = B'$. Thus, A and B are of the form

- $A = \left[\begin{array}{ccc|c} a_{1,1} & \dots & a_{1,m} & c_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & \dots & a_{n,m} & c_n \end{array}\right]$, - $B = \left[\begin{array}{ccc|c} a_{1,1} & \dots & a_{1,m} & d_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & \dots & a_{n,m} & d_n \end{array}\right]$,

where the $a_{i,j}$'s, the c_i's, and the d_i's are some elements of the field $\mathbb{F}$. (Here, the submatrix of both A and B to the left of the vertical dotted line is $A' = B' = \left[\begin{array}{c} a_{i,j} \end{array}\right]_{n\times m}$.) We must show that $c_1 = d_1, \dots, c_n = d_n$.

Assume that $A' = B'$ has exactly k pivot columns, and assume that those pivot columns are columns number $j_1, \dots, j_k$ (appearing from left to right in $A' = B'$, so that $j_1 < \dots j_k \leq m$). Thus, the matrix $A' = B'$ has precisely k non-zero rows, and for each $i \in \{1, \dots, m\}$, the leading 1 of the i-th row of $A' = B'$ is in the j_i-th column. So, schematically, the matrix $A' = B'$ looks as shown in the diagram below (the first k rows are non-zero, the pivot columns are in red, indices of the pivot columns are shown on top, row indices are shown on the left, and the horizontal dotted line separates the non-zero rows from the zero rows).

$$\begin{array}{c} \\ 1 \\ 2 \\ \\ k \\ \\ n \end{array} \begin{array}{c} \begin{array}{cccccccccc} & j_1 & & j_2 & & & & j_k & & \end{array} \\ \left[\begin{array}{cccccccccc} 0 & 1 & * & 0 & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

Now, consider the linear systems $(\star)$ and $(\star\star)$ below, whose augmented matrices are A and B, respectively. (Note that this means that they have the same coefficient matrix, namely $A' = B'$.)

[16] This follows straight from the definition of a matrix in reduced row echelon form.

$$\left.\begin{array}{rcccccc} a_{1,1}x_1 & + & \dots & + & a_{1,m}x_m & = & c_1 \\ & & & & & \vdots & \\ a_{n,1}x_1 & + & \dots & + & a_{n,m}x_m & = & c_n \end{array}\right\} (\star)$$

$$\left.\begin{array}{rcccccc} a_{1,1}x_1 & + & \dots & + & a_{1,m}x_m & = & d_1 \\ & & & & & \vdots & \\ a_{n,1}x_1 & + & \dots & + & a_{n,m}x_m & = & d_n \end{array}\right\} (\star\star)$$

Since A and B are row equivalent, the linear systems $(\star)$ and $(\star\star)$ are equivalent, i.e. they have exactly the same solutions. In particular, $(\star)$ and $(\star\star)$ are both either consistent or both inconsistent.

Suppose first that the linear systems $(\star)$ and $(\star\star)$ are both inconsistent. Since $(\star)$ is inconsistent, the rightmost column of A is a pivot column. Since A is in reduced row echelon form, and since it has exactly k pivot columns to the left of its rightmost column (i.e. to the left of the vertical dotted line), we see that the rightmost column of A has 1 in the $(k+1)$-th row and 0's everywhere else. Schematically, the matrix A is of the following form (where to the left of the vertical dotted line, we have the matrix $A' = B'$, and the horizontal dotted line is inherited from $A' = B'$):

$$\left[\begin{array}{cccccccccc:c} 0 & 1 & * & 0 & 0 & * & * & 0 & * & * & 0 \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

In other words, we have that $c_{k+1} = 1$ and $c_i = 0$ for all $i \in \{1, \dots, n\} \setminus \{k\}$. Since $(\star\star)$ is also inconsistent, a completely analogous argument shows that $d_{k+1} = 1$ and $d_i = 0$ for all $i \in \{1, \dots, n\} \setminus \{k\}$. It follows that $c_1 = d_1, \dots, c_n = d_n$, and we are done.

Suppose now that the linear systems $(\star)$ and $(\star\star)$ are both consistent. Since $(\star)$ is consistent, the rightmost column of its augmented matrix A is not a pivot column. Since A is in reduced row echelon form, and since it has exactly k pivot columns to the left of its rightmost column (i.e. to the left of the vertical dotted line), we see that the bottom $n-k$ many rows of A are all zero, and in particular, $c_{k+1} = \dots = c_n = 0$. Schematically, the matrix A is of the following form (where to the left of the vertical dotted line, we have the matrix $A' = B'$, and the horizontal dotted line is inherited from $A' = B'$):

$$\left[\begin{array}{cccccccccc|c} 0 & 1 & * & 0 & 0 & * & * & 0 & * & * & c_1 \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & * & * & c_2 \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & * & * & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * & c_k \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Since $(\star\star)$ is also consistent, a completely analogous argument shows establishes that $d_{k+1} = \cdots = d_n = 0$. Schematically, the matrix B is of the following form (where to the left of the vertical dotted line, we have the matrix $A' = B'$, and the horizontal dotted line is inherited from $A' = B'$):

$$\left[\begin{array}{cccccccccc|c} 0 & 1 & * & 0 & 0 & * & * & 0 & * & * & d_1 \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & * & * & d_2 \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & * & * & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * & d_k \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We have now shown that $c_{k+1}, \ldots, c_n, d_{k+1}, \ldots, d_n$ are all zero. It remains to show that $c_1 = d_1, \ldots, c_k = d_k$.

We consider the systems $(\star)$ and $(\star\star)$. In both of those systems, the basic variables are the variables $x_{j_1}, \ldots, x_{j_k}$, and the remaining variables are free. We first consider the linear system $(\star)$. If we set all the free variables to 0 and solve for the basic variables, we get the following solution of the system $(\star)$:

- $x_j = 0$ for all $j \in \{1, \ldots, m\} \setminus \{j_1, \ldots, j_k\}$;
- and $x_{j_i} = c_i$ for all $i \in \{1, \ldots, k\}$.

Now, since the linear systems $(\star)$ and $(\star\star)$ are equivalent, this solution of $(\star)$ is also a solution of $(\star\star)$. But if we plug it into $(\star\star)$, we obtain the following.

$$\left.\begin{array}{rcl} c_1 &=& d_1 \\ c_2 &=& d_2 \\ &\vdots& \\ c_k &=& d_k \\ 0 &=& 0 \\ 0 &=& 0 \\ &\vdots& \\ 0 &=& 0 \end{array}\right\} n - k$$

In particular, $c_1 = d_1, \ldots, c_k = d_k$. This completes the argument. □

We are now ready to prove Theorem 1.3.6, restated below.

Theorem 1.3.6. *Every matrix (with entries in some field) is row equivalent to a* ***unique*** *matrix in reduced row echelon form.*

Proof. The row reduction algorithm transforms any matrix into one in reduced row echelon form; these two matrices are row equivalent because the row reduction algorithm is simply a particular sequence of elementary row operations. This proves the existence part of the theorem: every matrix is row equivalent to at least one matrix in reduced row echelon form.

It remains to prove uniqueness. Fix any matrix A (with entries in some field), and suppose that it is row equivalent to matrices A_1 and A_2, both in reduced row echelon form. But then $A_1 \sim A \sim A_2$; consequently (by the transitivity of row equivalence, see Proposition 1.3.5(c)), we have that $A_1 \sim A_2$. But now Lemma 1.3.22 guarantees that $A_1 = A_2$. This proves uniqueness, and we are done. □

We complete this subsection by proving Corollaries 1.3.7 and 1.3.8.

Corollary 1.3.7. *If two row equivalent matrices (with entries in some field) are both in row echelon form, then they have exactly the same pivot positions and exactly the same pivot columns.*

Proof. Let A and B be row equivalent matrices in row echelon form, both with entries in some field $\mathbb{F}$. By performing the backward phase of the row reduction algorithm on the matrix A, we obtain the matrix $\mathrm{RREF}(A)$, which is in reduced row echelon form, is row equivalent to A, and (by the description of the backward phase of the row reduction algorithm) has exactly the same pivot columns (and consequently, exactly the same pivot positions) as A. Similarly, $\mathrm{RREF}(B)$ is in reduced row echelon form, is row equivalent to B, and has exactly the same pivot columns (and consequently, exactly the same pivot positions) as B. Now, we have that

$$\mathrm{RREF}(A) \;\sim\; A \;\overset{(*)}{\sim}\; B \;\sim\; \mathrm{RREF}(B),$$

where (*) is true by hypothesis. So, by the transitivity of row equivalence, it follows that $\mathrm{RREF}(A) \sim \mathrm{RREF}(B)$. Since both $\mathrm{RREF}(A)$ and $\mathrm{RREF}(B)$ are in row echelon form, Theorem 1.3.6 implies that they are in fact equal. But now both A and B have exactly the same pivot columns (and consequently exactly the same pivot positions) as the matrix $\mathrm{RREF}(A) = \mathrm{RREF}(B)$, and the result follows. □

Corollary 1.3.8. *Two matrices (with entries in some field) are row equivalent if and only if they have the same reduced row echelon form.*

Proof. Fix two matrices A and B (with entries in some field). By Theorem 1.3.6, the matrix A is row equivalent to a unique matrix in reduced row echelon form, denoted by $\mathrm{RREF}(A)$. Similarly, the matrix B is row equivalent to a unique matrix in reduced row echelon form, denoted by $\mathrm{RREF}(B)$.

Suppose first that $A \sim B$. Then $\mathrm{RREF}(A) \sim A \sim B \sim \mathrm{RREF}(B)$, and so by the transitivity of row equivalence, we have that $\mathrm{RREF}(A) \sim \mathrm{RREF}(B)$. But now Lemma 1.3.22 guarantees that $\mathrm{RREF}(A) = \mathrm{RREF}(B)$.

Conversely, suppose that $\mathrm{RREF}(A) = \mathrm{RREF}(B)$. Then $A \sim \mathrm{RREF}(A) = \mathrm{RREF}(B) \sim B$. By the transitivity of row equivalence, it follows that $A \sim B$, and we are done. □

1.4 Algebraic operations on vectors and matrices

1.4.1 Vector addition, vector subtraction, and scalar-vector multiplication

We can add and subtract vectors, and we can also multiply them by scalars, as follows. Let $\mathbb{F}$ be some field.

- Given two vectors in $\mathbb{F}^n$, say

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

  we define the *sum* of $\mathbf{x}$ and $\mathbf{y}$ by

$$\mathbf{x} + \mathbf{y} := \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix},$$

  and the *difference* of $\mathbf{x}$ and $\mathbf{y}$ by

$$\mathbf{x} - \mathbf{y} := \begin{bmatrix} x_1 - y_1 \\ \vdots \\ x_n - y_n \end{bmatrix},$$

  where the sums $x_i + y_i$ and differences $x_i - y_i$ (for $i = 1, \ldots, n$) are computed in the field $\mathbb{F}$.

- Given a vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

  in $\mathbb{F}^n$ and a scalar $\alpha \in \mathbb{F}$, we define the *scalar-vector product* of α and $\mathbf{x}$ by

$$\alpha\mathbf{x} := \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{bmatrix},$$

where the products $\alpha x_1, \ldots, \alpha x_n$ are computed in the field $\mathbb{F}$.

Terminology: A *scalar multiple* of a vector $\mathbf{x} \in \mathbb{F}^n$ (where $\mathbb{F}$ is some field) is any vector of the form $\alpha\mathbf{x}$, where $\alpha \in \mathbb{F}$.

Notation: By convention, for a vector $\mathbf{x}$ and a scalar α, we write $\alpha\mathbf{x}$, but we do **not** write $\mathbf{x}\alpha$. In other words, by convention, we have "scalar times vector," but not "vector times scalar."

Example 1.4.1. *Consider the vectors*

$$\boldsymbol{x} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 2 \end{bmatrix} \quad \textit{and} \quad \boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

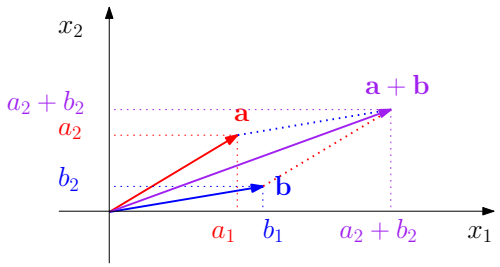
in $\mathbb{Z}_3^4$. *Then*

$$\mathbf{x}+\mathbf{y} = \begin{bmatrix} 0+1 \\ 1+0 \\ 2+2 \\ 2+1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \quad \textit{and} \quad 2\mathbf{x} = \begin{bmatrix} 2\cdot 0 \\ 2\cdot 1 \\ 2\cdot 2 \\ 2\cdot 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \end{bmatrix}.$$

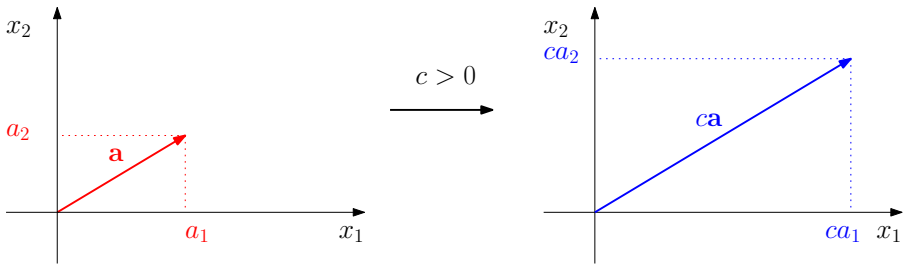
*(**Reminder:** In* $\mathbb{Z}_3$, *we have that* $2+2=1$, $2+1=0$, *and* $2\cdot 2 = 1$.*)*

1.4.2 Vector addition and scalar multiplication in $\mathbb{R}^2$

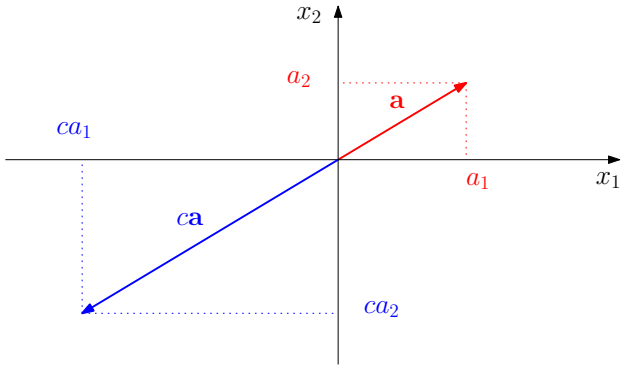
To add two vectors in $\mathbb{R}^2$, say $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, we apply the "parallelogram rule," as shown below.



Scalar multiplication can be interpreted as follows. Suppose we are given a vector $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and a scalar $c \in \mathbb{R}$. If $c > 0$, then $c\mathbf{a}$ is the vector that points in the same direction as $\mathbf{a}$, but whose length is scaled by c.

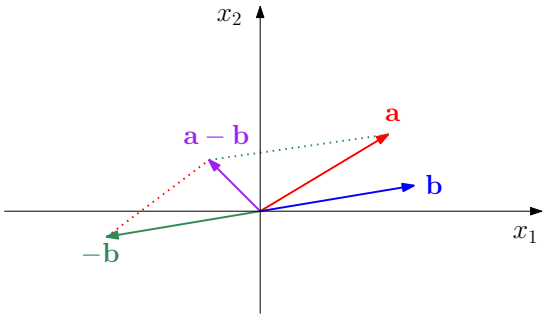


On the other hand, if $c < 0$, then $c\mathbf{a}$ is the vector that points in the opposite direction to $\mathbf{a}$, but whose length is scaled by $|c| = -c$.



If $c = 0$, then $c\mathbf{a} = \mathbf{0}$, which is simply the origin.

For vectors $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ in $\mathbb{R}^2$, we note that $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-1)\mathbf{b}$, which yields the geometric interpretation below.



For vectors in $\mathbb{R}^3$, we have a similar geometric interpretation, only in the three-dimensional Euclidean space.

1.4.3 Linear combinations of vectors

Suppose $\mathbb{F}$ is some field. A *linear combination* of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{F}^n$ is any sum of the form

$$\sum_{i=1}^{k} \alpha_i \mathbf{v}_i \quad = \quad \alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k,$$

where $\alpha_1, \dots, \alpha_k$ are scalars from the field $\mathbb{F}$.

For example, in $\mathbb{R}^3$, vectors $\begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$, and $\begin{bmatrix} -3 \\ -9 \\ -3 \end{bmatrix}$ are linear combinations of the vectors $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ because

- $\begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$;
- $\begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + (-1)\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$;
- $\begin{bmatrix} -3 \\ -9 \\ -3 \end{bmatrix} = -3\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = (-3)\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

Similarly, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is a linear combination of the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ in $\mathbb{Z}_3^2$ because

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

We note that in $\mathbb{F}^n$ (where $\mathbb{F}$ is a field), the zero vector $\mathbf{0}$ is a linear combination of **any** vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ because

$$\mathbf{0} = 0\mathbf{v}_1 + \dots + 0\mathbf{v}_k.$$

Moreover, we define the "empty sum" of vectors in $\mathbb{F}^n$ (or the sum of an "empty list" of vectors in $\mathbb{F}^n$) to be $\mathbf{0}$, where $\mathbf{0}$ is the zero vector in $\mathbb{F}^n$.

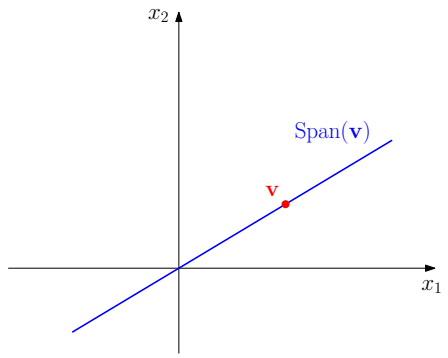
Linear span. The *linear span* (or simply *span*) of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ in $\mathbb{F}^n$ (where $\mathbb{F}$ is a field), denoted by $\mathrm{Span}(\{\mathbf{v}_1, \dots, \mathbf{v}_k\})$ or simply $\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$, is the set of all linear combinations of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. In other words,

$$\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) = \Big\{ \sum_{i=1}^{k} \alpha_i \mathbf{v}_i \mid \alpha_1, \dots, \alpha_k \in \mathbb{F} \Big\}.$$

So, by definition, a vector $\mathbf{v}$ belongs to $\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ if and only if it can be written as a linear combination the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. As a special case, the empty sum of vectors is equal to the zero vector, and so $\mathrm{Span}(\emptyset) = \{\mathbf{0}\}$. Obviously, $\mathrm{Span}(\mathbf{0}) = \{\mathbf{0}\}$.

We will study the linear span in more generality in chapter 3 (see subsection 3.1.2). Here, let us try to give a geometric intuition for the special case of $\mathbb{R}^n$. As we discussed

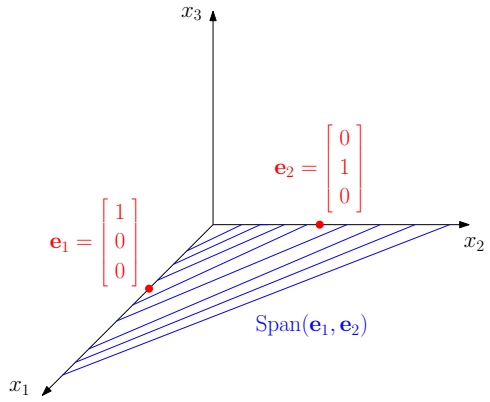
above, $\text{Span}(\emptyset) = \{\mathbf{0}\}$ and $\text{Span}(\mathbf{0}) = \{\mathbf{0}\}$. If $\mathbf{v} \neq \mathbf{0}$, then $\text{Span}(\mathbf{v}) = \{\alpha\mathbf{v} \mid \alpha \in \mathbb{R}\}$ is the line through the origin containing $\mathbf{v}$: indeed, $\text{Span}(\mathbf{v})$ is the set of all scalar multiples of $\mathbf{v}$, which is precisely the line through $\mathbf{0}$ and $\mathbf{v}$ (this is illustrated below for the special case of $\mathbb{R}^2$).



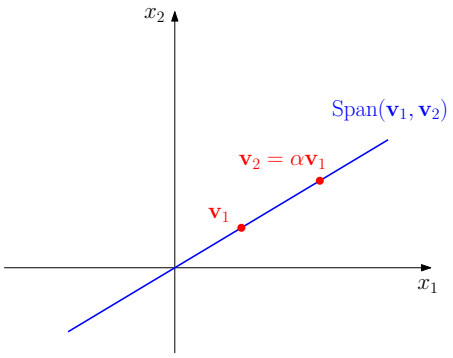
What if we have two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$? If neither of those vectors is a scalar multiple of the other (and in particular, neither of the two vectors is $\mathbf{0}$), then $\text{Span}(\mathbf{v}_1, \mathbf{v}_2)$ is the plane through $\mathbf{0}, \mathbf{v}_1, \mathbf{v}_2$. The case that is particularly easy to visualize is that of the vectors $\mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ in $\mathbb{R}^3$:

$$\text{Span}(\mathbf{e}_1, \mathbf{e}_2) \quad = \quad \Big\{a_1\mathbf{e}_1 + a_2\mathbf{e}_2 \mid a_1, a_2 \in \mathbb{R}\Big\} \quad = \quad \Big\{ \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix} \mid a_1, a_2 \in \mathbb{R}\Big\},$$

which is simply the x_1x_2-plane in $\mathbb{R}^3$, shown below.



But what if we have two vectors, one of which is a scalar multiple of the other? If $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, with $\mathbf{v}_2 = \alpha\mathbf{v}_1$ for some scalar $\alpha \in \mathbb{R}$ and $\mathbf{v}_1 \neq \mathbf{0}$, then $\text{Span}(\mathbf{v}_1, \mathbf{v}_2)$ is the line through the origin, $\mathbf{v}_1$, and $\mathbf{v}_2$. In the case of $\mathbb{R}^2$, this is illustrated below.



In general, for vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$, the set $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ is the smallest "flat" (point, line, plane, or higher dimensional generalization) containing the **origin** and all the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

1.4.4 Matrix-vector multiplication

Suppose that $\mathbb{F}$ is some field. Given a matrix $A \in \mathbb{F}^{n\times m}$ and a vector $\mathbf{x} \in \mathbb{F}^m$, say

$$A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix},$$

we define the *matrix-vector product* $A\mathbf{x}$ as follows:

$$A\mathbf{x} := \sum_{i=1}^{m} x_i\mathbf{a}_i = x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m.$$

Thus, $A\mathbf{x}$ is a linear combination of the columns of A, and the weights/scalars in front of the columns are determined by the entries of the vector $\mathbf{x}$.

Note that, for the matrix-vector product $A\mathbf{x}$ to be defined, two conditions must be satisfied:

- entries of the matrix A and entries of the vector $\mathbf{x}$ must belong to the same field;
- the number of **columns** of A must be the same as the number of **entries** of $\mathbf{x}$.

Schematically, we have the following:

$$\underbrace{A}_{\in\mathbb{F}^{n\times m}} \quad \underbrace{\mathbf{x}}_{\in\mathbb{F}^m} = \underbrace{A\mathbf{x}}_{\in\mathbb{F}^n}.$$

Example 1.4.2. *Consider the matrix $A \in \mathbb{R}^{3\times 2}$ and vector $\mathbf{x} \in \mathbb{R}^2$, given below:*

$$A = \begin{bmatrix} -1 & 2 \\ 2 & 0 \\ 3 & -2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Then

$$A\mathbf{x} = \begin{bmatrix} -1 & 2 \\ 2 & 0 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2\begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix} + 3\begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix}.$$

Example 1.4.3. *Consider the matrix $A \in \mathbb{Z}_2^{2\times 3}$ and vector $\mathbf{x} \in \mathbb{Z}_2^3$, given below:*

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Then

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Remark: Suppose that $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ is a matrix in $\mathbb{F}^{n\times m}$ (where $\mathbb{F}$ is some field). Then

$$\begin{aligned} \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m) &= \Big\{ x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m \mid x_1, \dots, x_m \in \mathbb{F} \Big\} \\ &= \left\{ \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \mid x_1, \dots, x_m \in \mathbb{F} \right\} \\ &= \Big\{ A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m \Big\}. \end{aligned}$$

So, $\text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$, which we defined as the set of all linear combinations of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$, is in fact the set of all possible matrix-vector products $A\mathbf{x}$ (where our matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ is fixed, and the vector $\mathbf{x} \in \mathbb{F}^m$ is allowed to vary). We note that $\text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$, the span of the columns of A, has a special name: it is called the "column space" of the matrix A, and it is denoted by $\text{Col}(A)$. We will study the column space of a matrix in more detail in chapter 3 (see section 3.3).

The standard basis vectors. Let $\mathbb{F}$ be a field. For each positive integer n and index $i \in \{1, \dots, n\}$, the vector $\mathbf{e}_i^n$ is the vector in $\mathbb{F}^n$ whose i-th entry is 1, and all of whose other entries are 0's (here, both 0 and 1 are understood to belong to the field $\mathbb{F}$). Schematically, for each index $i \in \{1, \dots, n\}$, the vector $\mathbf{e}_i^n$ is given by

$$\mathbf{e}_i^n \quad = \quad \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow i\text{-th entry}$$

(here, we have exactly one 1, and we have $n-1$ many 0's). When n is clear from context, we drop the superscript n, and we write $\mathbf{e}_1, \ldots, \mathbf{e}_n$ instead of $\mathbf{e}_1^n, \ldots, \mathbf{e}_n^n$, respectively. Vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are called the *standard basis vectors* of $\mathbb{F}^n$, and the set $\mathcal{E}_n := \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is called the *standard basis* of $\mathbb{F}^n$. We note that any vector $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{F}^n$ can be expressed as a linear combination of the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in a unique way, namely

$$\mathbf{v} \quad = \quad v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n.$$

As our next proposition shows, multiplying a matrix by the i-th standard basis vector yields the i-th column of the matrix that we started with.

Proposition 1.4.4. *Let $\mathbb{F}$ be a field, and let $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix}$ be a matrix in $\mathbb{F}^{n\times m}$. Then for all indices $i \in \{1, \ldots, m\}$, we have that $A\mathbf{e}_i^m = \mathbf{a}_i$.*

Proof. Fix $i \in \{1, \ldots, m\}$. Then

$$A\mathbf{e}_i^m \quad = \quad \begin{bmatrix} \mathbf{a}_1 & \ldots \mathbf{a}_{i-1} & \mathbf{a}_i & \mathbf{a}_{i+1} & \ldots & \mathbf{a}_m \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow i\text{-th entry}$$

$$= \quad 0\mathbf{a}_1 + \cdots + 0\mathbf{a}_{i-1} + 1\mathbf{a}_i + 0\mathbf{a}_{i+1} + \cdots + 0\mathbf{a}_m \quad = \quad \mathbf{a}_i,$$

which is what we needed to show. □

The identity matrix. For a field $\mathbb{F}$, the *identity matrix* in $\mathbb{F}^{n\times n}$ is the $n \times n$ matrix

$$I_n \quad := \quad \begin{bmatrix} \mathbf{e}_1^n & \ldots & \mathbf{e}_n^n \end{bmatrix}.$$

In other words, the identity matrix I_n is the $n \times n$ matrix with 1's on the main diagonal and 0's elsewhere (where the 1's and the 0's are from the field $\mathbb{F}$). Schematically, we have that

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{n \times n}$$

for all positive integers n. For small values of n, we have:

$$I_1 = \begin{bmatrix} 1 \end{bmatrix}, \qquad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

As our next proposition shows, if we multiply the identity matrix by a vector, we obtain that same vector.

Proposition 1.4.5. *Let $\mathbb{F}$ be a field. Then for all vectors $\mathbf{v} \in \mathbb{F}^n$, we have that $I_n \mathbf{v} = \mathbf{v}$.*

Proof. For any vector $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{F}^n$, we have that

$$\begin{aligned} I_n \mathbf{v} &= \begin{bmatrix} \mathbf{e}_1^n & \mathbf{e}_2^n & \dots & \mathbf{e}_n^n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= v_1 \mathbf{e}_1^n + v_2 \mathbf{e}_2^n + \cdots + v_n \mathbf{e}_n^n \\ &= v_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \mathbf{v}, \end{aligned}$$

which is what we needed. □

Multiplying by zero. Recall that, for a field $\mathbb{F}$, the *zero matrix* in $\mathbb{F}^{n\times m}$, denoted by $O_{n\times m}$, is the $n \times m$ matrix, all of whose entries are 0 (where the 0 is from the field $\mathbb{F}$.

Proposition 1.4.6. *Let $\mathbb{F}$ be a field. Then both the following hold:*

(a) for all $\mathbf{v} \in \mathbb{F}^m$, we have that $O_{n\times m}\mathbf{v} = \mathbf{0}$;[17]

(b) for all matrices $A \in \mathbb{F}^{n\times m}$, we have that $A\mathbf{0} = \mathbf{0}$.[18]

Proof. This readily follows from the definition of matrix-vector multiplication. □

1.5 Matrix-vector equations

A *matrix-vector equation* is an equation of the form $A\mathbf{x} = \mathbf{b}$, where the matrix A and vector $\mathbf{b}$ are known, and the vector $\mathbf{x}$ is unknown. Once again, the entries of A and $\mathbf{b}$ must come from the same field $\mathbb{F}$. Moreover, the number of **rows** of A must be the same as the number of **entries** of $\mathbf{b}$. Any solution $\mathbf{x}$ will then be a vector in $\mathbb{F}^m$, where m is the number of **columns** of A.

As we shall see, a matrix-vector equation is equivalent to a system of linear equations. Before considering the general case, let us first take a look at an example. Consider the following matrix A and vector $\mathbf{b}$ (with entries understood to be in $\mathbb{R}$):

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}.$$

We now transform the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ into a system of linear equations, as follows.

$$A\mathbf{x} = \mathbf{b} \iff \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{=\mathbf{x}} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

$$\iff x_1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

$$\iff \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

[17] Here, the zero vector $\mathbf{0}$ belongs to $\mathbb{F}^n$.

[18] Here, the first $\mathbf{0}$ belongs to $\mathbb{F}^m$, whereas the second $\mathbf{0}$ belongs to $\mathbb{F}^n$. Or, if we color code for convenience, in the expression $A\mathbf{0} = \mathbf{0}$, we have that $\mathbf{0} \in \mathbb{F}^m$ and $\mathbf{0} \in \mathbb{F}^n$.

$$\iff \quad \left\{ \begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & 3x_3 & = & 7 \\ 4x_1 & + & 5x_2 & + & 6x_3 & = & 8 \end{array} \right.$$

Note that the augmented matrix of the linear system that we obtained is

$$\left[\begin{array}{c:c} A & \mathbf{b} \end{array} \right] = \left[\begin{array}{ccc:c} 1 & 2 & 3 & 7 \\ 4 & 5 & 6 & 8 \end{array} \right].$$

Let us now consider the general case. Suppose that $\mathbb{F}$ is a field, $A \in \mathbb{F}^{n \times m}$ is a matrix, and $\mathbf{b} \in \mathbb{F}^n$ is a vector. Set $A = \left[\ a_{i,j} \ \right]_{n \times m}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$. We transform the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ into a system of linear equations, as follows.

$$A\mathbf{x} = \mathbf{b} \iff \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{bmatrix} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}}_{=\mathbf{x}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\iff \quad x_1 \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \end{bmatrix} + x_2 \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \end{bmatrix} + \dots + x_m \begin{bmatrix} a_{1,m} \\ a_{2,m} \\ \vdots \\ a_{n,m} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\iff \quad \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,m}x_m \\ a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,m}x_m \\ \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \dots + a_{n,m}x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\iff \quad \left\{ \begin{array}{rcrcrcl} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \dots + a_{1,m}x_m & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \dots + a_{2,m}x_m & = & b_2 \\ & & & & & \vdots & \\ a_{n,1}x_1 & + & a_{n,2}x_2 & + & \dots + a_{n,m}x_m & = & b_n \end{array} \right.$$

So, solving the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ boils down to solving a system of linear equations. The augmented matrix of this linear system is the matrix

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,m} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} & b_n \end{array}\right].$$

The matrix $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$ will also be referred to as the *augmented matrix* of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. As in the case of linear systems, a matrix-vector equation $A\mathbf{x} = \mathbf{b}$ may have no solutions, may have exactly one solution, or may have more than one solution. A matrix-vector equation that has at least one solution is called *consistent*; a matrix-vector equation that has no solutions is said to be *inconsistent*.

Example 1.5.1. *Solve the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, where*

$$A = \left[\begin{array}{cc} 1 & 2 \\ 3 & 6 \end{array}\right] \quad \textit{and} \quad \mathbf{b} = \left[\begin{array}{c} 2 \\ 6 \end{array}\right],$$

with entries understood to be in $\mathbb{R}$. How many solutions does the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ have?

Solution. The augmented matrix of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 3 & 6 & 6 \end{array}\right].$$

We now row reduce in order to find $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right)$, as follows:

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 3 & 6 & 6 \end{array}\right] \overset{R_2 \to R_2 - 3R_1}{\sim} \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 0 & 0 \end{array}\right].$$

The last matrix from the computation above is in reduced row echelon form, and we deduce that

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right) = \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 0 & 0 \end{array}\right].$$

The matrix $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right)$ is the augmented matrix of the linear system below.

$$\begin{array}{rcrcl} x_1 & + & 2x_2 & = & 2 \\ & & 0 & = & 0 \end{array}$$

The system is consistent, with one free variable (namely, x_2). We read off the solutions as follows.

$$\begin{array}{rcll} x_1 & = & -2s + 2 & \\ x_2 & = & s, & \text{where } s \in \mathbb{R}. \end{array}$$

So, the general solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x} = \begin{bmatrix} -2s+2 \\ s \end{bmatrix}, \qquad \text{where } s \in \mathbb{R}.$$

Here is another way to write the general solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$:

$$\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \qquad \text{where } s \in \mathbb{R}.$$

Remark: We obtained this second form of the solution by separating the constant part of $\mathbf{x}$ form the part with the parameter, and then factoring out the parameter, as follows:

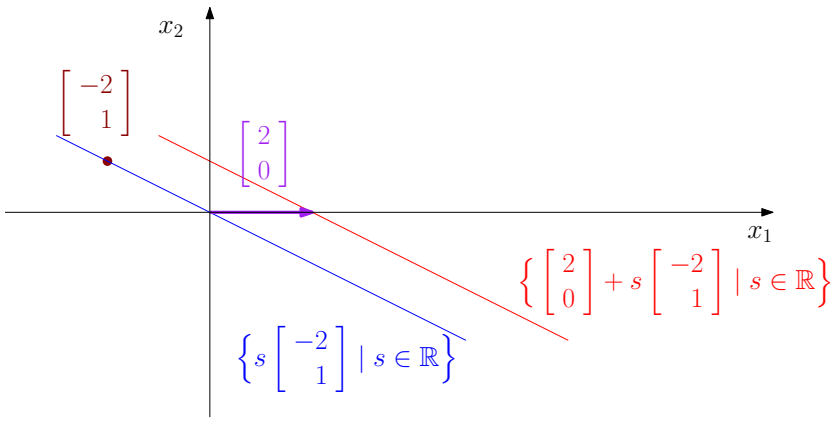
$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} -2s+2 \\ s \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} -2s \\ s \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \qquad \text{where } s \in \mathbb{R}. \end{aligned}$$

The set of solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\left\{ \begin{bmatrix} -2s+2 \\ s \end{bmatrix} \mid s \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \end{bmatrix} \mid s \in \mathbb{R} \right\}.$$

Since the parameter s can take infinitely many values (because $\mathbb{R}$ is infinite), the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions. □

Remark: The solution set from Example 1.5.1 has a geometric interpretation. Indeed, $\left\{ s \begin{bmatrix} -2 \\ 1 \end{bmatrix} \mid s \in \mathbb{R} \right\} = \text{Span}(\begin{bmatrix} -2 \\ 1 \end{bmatrix})$ is the line that passes through the origin and the point $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ (this is the blue line in the picture below). The solution set $\left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \end{bmatrix} \mid s \in \mathbb{R} \right\}$ is obtained by shifting this line by the vector $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$, i.e. by adding the vector $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$ to each point on the line (this vector is shown in purple in the picture below). The solution set $\left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 1 \end{bmatrix} \mid s \in \mathbb{R} \right\}$ is the red line in the picture below.



Example 1.5.2. *Solve the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$*, where*

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 2 & 2 & 1 & 1 \end{bmatrix} \quad \textit{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_3$*. How many solutions does the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *have?*

Solution. The augmented matrix of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 1 & 0 & 1 & 0 & 2 \\ 2 & 2 & 1 & 1 & 0 \end{array}\right].$$

We now row reduce in order to find $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right)$, as follows:

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 1 & 0 & 1 & 0 & 2 \\ 2 & 2 & 1 & 1 & 0 \end{array}\right]$$

$$\overset{\substack{R_2 \to R_2 + 2R_1 \\ R_3 \to R_3 + R_1}}{\sim} \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 2 & 2 \end{array}\right]$$

$$\overset{R_3 \to R_3 + 2R_2}{\sim} \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{array}\right]$$

$$\overset{R_3 \to 2R_3}{\sim} \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\overset{R_1 \to R_1 + R_3}{\sim} \quad \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\overset{R_1 \to R_1 + R_2}{\sim} \quad \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

The last matrix from the computation above is in reduced row echelon form, and we deduce that

$$\mathrm{RREF}\Big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\Big) \quad = \quad \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

We see from $\mathrm{RREF}\big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\big)$ that the rightmost column of $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$ is a pivot column; consequently, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is inconsistent, i.e. the solution set of the equation $A\mathbf{x} = \mathbf{b}$ is $\emptyset$. (The number of solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is zero.) □

Remark: In the solution of Example 1.5.2, we could in fact have stopped as soon as we got the red matrix (despite the fact that this matrix is not in reduced row echelon form). This is because the bottom row of the red matrix encodes the equation $0 = 2$, which is has no solutions. Indeed, as soon as we obtain a row of the form $\left[\begin{array}{c|c} 0 \quad \dots \quad 0 & \blacksquare \end{array}\right]$, where $\blacksquare$ is a non-zero number, we can stop row reducing, and we can deduce that the system has no solutions (because this row encodes the equation $0 = \blacksquare$, and $\blacksquare$ is non-zero).

Example 1.5.3. *Solve the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, where*

$$A \quad = \quad \left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array}\right] \quad \textit{and} \quad \mathbf{b} \quad = \quad \left[\begin{array}{c} 1 \\ 1 \end{array}\right],$$

with entries understood to be in $\mathbb{Z}_2$. How many solutions does the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ have?

Solution. The augmented matrix of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] \quad = \quad \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 0 & 1 \end{array}\right].$$

We now row reduce in order to find $\mathrm{RREF}\big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\big)$, as follows:

$$\begin{array}{rcl} \left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] & = & \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 0 & 1 \end{array}\right] \\ & \overset{R_2 \to R_2 + R_1}{\sim} & \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 0 \end{array}\right] \\ & \overset{R_1 \to R_1 + R_2}{\sim} & \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right]. \end{array}$$

The last matrix from the computation above is in reduced row echelon form, and we deduce that

$$\mathrm{RREF}\Big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\Big) \;=\; \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right].$$

We now see that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution, namely

$$\mathbf{x} \;=\; \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The solution set of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$. (The number of solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is one.) □

Example 1.5.4. *Solve the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, where*

$$A \;=\; \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad \textit{and} \quad \mathbf{b} \;=\; \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_2$. How many solutions does the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ have?

Solution. The augmented matrix of the the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] \;=\; \left[\begin{array}{cccc|c} 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{array}\right].$$

We now row reduce in order to find $\mathrm{RREF}\big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\big)$, as follows:

$$\begin{array}{rcl} \left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] & = & \left[\begin{array}{cccc|c} 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{array}\right] \end{array}$$

$$\begin{array}{l} \\ R_2 \to R_2 + R_1 \\ \sim \\ \\ R_3 \to R_3 + R_2 \\ \sim \\ \\ R_1 \to R_1 + R_2 \\ \sim \end{array} \quad \left[\begin{array}{cccc|c} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{array}\right] \quad \left[\begin{array}{cccc|c} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

The last matrix from the computation above is in reduced row echelon form, and we deduce that

$$\mathrm{RREF}\Big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\Big) \quad = \quad \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We see from $\mathrm{RREF}\big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\big)$ that the rightmost column of $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$ is not a pivot column, and so the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent. The matrix $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$ has two non-pivot columns to the left of the vertical dotted line, namely, the third and fourth column. So, the third and fourth entry of the solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ become arbitrary parameters. The general solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is the following:[19]

[19]Normally, we do indeed read off the solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ straight from the matrix $\mathrm{RREF}\big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\big)$. However, let us give a slightly more detailed explanation of how exactly we got our answer. The matrix $\mathrm{RREF}\big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\big)$ is the augmented matrix of the linear system below.

$$\begin{array}{rcrcrcrcl} x_1 & & & + & x_3 & & & = & 0 \\ & & x_2 & + & x_3 & + & x_4 & = & 1 \\ & & & & & & 0 & = & 0 \end{array}$$

The system is consistent, with two free variables (namely, x_3 and x_4). We read off its solution as follows.

$$\begin{array}{lcll} x_1 & = & s & \\ x_2 & = & s + t + 1 & \\ x_3 & = & s & \\ x_4 & = & t & \text{where } s, t \in \mathbb{Z}_2. \end{array}$$

So, the general solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x} \quad = \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad = \quad \begin{bmatrix} s \\ s + t + 1 \\ s \\ t \end{bmatrix}, \qquad \text{where } s, t \in \mathbb{Z}_2.$$

$$\mathbf{x} = \begin{bmatrix} s \\ s+t+1 \\ s \\ t \end{bmatrix}, \qquad \text{where } s,t \in \mathbb{Z}_2.$$

We can also write the general solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ as follows:

$$\mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \text{where } s,t \in \mathbb{Z}_2.$$

Remark: We obtained this second form of the general solution by separating the constant part from the parts associated with each parameter:

$$\mathbf{x} = \begin{bmatrix} s \\ s+t+1 \\ s \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} s \\ s \\ s \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ t \\ 0 \\ t \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \text{where } s,t \in \mathbb{Z}_2.$$

The solution set of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\left\{ \begin{bmatrix} s \\ s+t+1 \\ s \\ t \end{bmatrix} \mid s,t \in \mathbb{Z}_2 \right\} = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \mid s,t \in \mathbb{Z}_2 \right\}.$$

There are two parameters (namely, s and t), and each of them can take two values (because $|\mathbb{Z}_2|$). So, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has $2^2 = 4$ solutions.[20] □

[20] Since there are only four solutions, we could easily list all of them:

- for $s = 0$ and $t = 0$, we have the solution $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$;
- for $s = 0$ and $t = 1$, we have the solution $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$;

1.5.1 Matrix-vector equations and linear span

Let $\mathbb{F}$ be a field, and consider vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ ($m \geq 1$) and $\mathbf{b}$ in $\mathbb{F}^n$. How do we determine if $\mathbf{b}$ is a linear combination of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$, that is, if $\mathbf{b} \in \mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$? Set $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$, and recall from subsection 1.4.3 that

$$\mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m) \quad = \quad \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m\}.$$

So, $\mathbf{b} \in \mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$ if and only if there exists some $\mathbf{x} \in \mathbb{F}^m$ such that $A\mathbf{x} = \mathbf{b}$. In other words, $\mathbf{b} \in \mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$ if and only if the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent. Thus, determining whether $\mathbf{b}$ is a linear combination of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ boils down to solving the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, which we know how to do.

We give two examples. The solution to the first one (Example 1.5.5) aims to carefully justify all the steps. The solution to the second one (Example 1.5.6) is more concise, but follows the same process. When solving examples by yourself, you should aim to give the amount of detail given in the solution to Example 1.5.6.

Example 1.5.5. *Consider the vectors*

$$\mathbf{a}_1 = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \qquad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{a}_4 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 2 \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{d} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_3$. For each of the vectors $\mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}$, determine if it can be expressed as a linear combination of the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ (that is, if it

---

- for $s = 1$ and $t = 0$, we have the solution $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$;
- for $s = 1$ and $t = 1$, we have the solution $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

So, the solution set is $\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$. However, we usually leave the final answer in parametric form, rather than listing all possible solutions one by one.

belongs to Span$(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$*), and if so, express it as such a linear combination, and explain whether your answer is unique.*

Solution. Set

$$A := \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{array}\right] = \left[\begin{array}{cccc} 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

We have that $\mathbf{b} \in \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$ if and only if the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent (and similar for the vectors $\mathbf{c}$, $\mathbf{d}$, and $\mathbf{e}$). So, the obvious way to proceed would be to solve four matrix-vector equations, namely, $A\mathbf{x} = \mathbf{b}$, $A\mathbf{x} = \mathbf{c}$, $A\mathbf{x} = \mathbf{d}$, and $A\mathbf{x} = \mathbf{e}$. However, this would require row reducing four times! We can solve the problem more efficiently by forming the matrix (color coded for emphasis)

$$\left[\begin{array}{c:cccc} A & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \end{array}\right] = \left[\begin{array}{cccc:cccc} 2 & 1 & 1 & 1 & 2 & 0 & 2 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 2 & 1 & 1 & 0 & 1 \end{array}\right],$$

and then row reducing to obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c:cccc} A & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \end{array}\right]\right) = \left[\begin{array}{cccc:cccc} 1 & 2 & 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

$$=: \left[\begin{array}{c:cccc} A' & \mathbf{b}' & \mathbf{c}' & \mathbf{d}' & \mathbf{e}' \end{array}\right].$$

We now have that $\left[\begin{array}{c:cccc} A & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \end{array}\right] \sim \left[\begin{array}{c:cccc} A' & \mathbf{b}' & \mathbf{c}' & \mathbf{d}' & \mathbf{e}' \end{array}\right]$, and consequently,

- $\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right] \sim \left[\begin{array}{c:c} A' & \mathbf{b}' \end{array}\right]$;
- $\left[\begin{array}{c:c} A & \mathbf{c} \end{array}\right] \sim \left[\begin{array}{c:c} A' & \mathbf{c}' \end{array}\right]$;
- $\left[\begin{array}{c:c} A & \mathbf{d} \end{array}\right] \sim \left[\begin{array}{c:c} A' & \mathbf{d}' \end{array}\right]$;
- $\left[\begin{array}{c:c} A & \mathbf{e} \end{array}\right] \sim \left[\begin{array}{c:c} A' & \mathbf{e}' \end{array}\right]$.

(Indeed, the same sequence of elementary row operations that transforms the matrix $\left[\begin{array}{c:cccc} A & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \end{array}\right]$ into $\left[\begin{array}{c:cccc} A' & \mathbf{b}' & \mathbf{c}' & \mathbf{d}' & \mathbf{e}' \end{array}\right]$ will transform matrices $\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]$, $\left[\begin{array}{c:c} A & \mathbf{c} \end{array}\right]$, $\left[\begin{array}{c:c} A & \mathbf{d} \end{array}\right]$, and $\left[\begin{array}{c:c} A & \mathbf{e} \end{array}\right]$ into matrices $\left[\begin{array}{c:c} A' & \mathbf{b}' \end{array}\right]$, $\left[\begin{array}{c:c} A' & \mathbf{c}' \end{array}\right]$, $\left[\begin{array}{c:c} A' & \mathbf{d}' \end{array}\right]$, and $\left[\begin{array}{c:c} A' & \mathbf{e}' \end{array}\right]$, respectively.)

From this point on, we deal with the vectors $\mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}$ separately.

Vector $\mathbf{b}$. We need to check whether there exist scalars $x_1, x_2, x_3, x_4 \in \mathbb{Z}_3$ such that

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4,$$

and if so, to find such x_i's. The above is equivalent to solving the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. Since $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] \sim \left[\begin{array}{c|c} A' & \mathbf{b}' \end{array}\right]$, we can "read off" the solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ (if they exist) from the matrix

$$\left[\begin{array}{c|c} A' & \mathbf{b}' \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We see that our equation $A\mathbf{x} = \mathbf{b}$ is consistent, and that the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} s + 2t + 2 \\ s \\ t + 1 \\ t \end{bmatrix}, \qquad \text{where } s, t \in \mathbb{Z}_3.$$

However, we were asked to find just one particular solution, and not the general solution. In principle, we could choose any values from $\mathbb{Z}_3$ for the parameters s and t, but it is easiest to choose $s = t = 0$, which yields

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

So, $\mathbf{b}$ is indeed a linear combination of the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$, that is, $\mathbf{b} \in \text{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$, and we have that

$$\mathbf{b} = 2\mathbf{a}_1 + 0\mathbf{a}_2 + 1\mathbf{a}_3 + 0\mathbf{a}_4 = 2\mathbf{a}_1 + \mathbf{a}_3.$$

(Both $\mathbf{b} = 2\mathbf{a}_1 + 0\mathbf{a}_2 + 1\mathbf{a}_3 + 0\mathbf{a}_4$ and $\mathbf{b} = 2\mathbf{a}_1 + \mathbf{a}_3$ are acceptable as a final answer, though the second form is more common.) We note that our solution is **not** unique, because we had parameters for which we chose particular values.

Remark: Here is a slightly different way to proceed. By looking at the matrix $\left[\begin{array}{c|c} A & \mathbf{b}' \end{array}\right]$, we see that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent. We do not need the general solution, but only one possible solution. So, we cross out (or simply ignore) the non-pivot columns of $\left[\begin{array}{c|c} A' & \mathbf{b}' \end{array}\right]$ to the left of the vertical dotted line,[21] and we read off the (unique) solution that remains after we eliminated those columns. Here, it is important to remember which x_i corresponds to which column.

[21] This has the effect of assigning the value 0 to free variables that correspond to the crossed out non-pivot columns.

$$\left[\begin{array}{c|c} A' & \mathbf{b}' \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{array}{cc} & \\ & \\ & \\ & \\ \uparrow & \uparrow \\ x_1 & x_3 \end{array}$$

We now read off $x_1 = 2$ and $x_3 = 1$, and we get $\mathbf{b} = 2\mathbf{a}_1 + \mathbf{a}_3$. Because we crossed out some non-pivot columns (which correspond to arbitrary parameters), our solution is not unique.

Vector c. We need to check whether there exist scalars $x_1, x_2, x_3, x_4 \in \mathbb{Z}_3$ such that

$$\mathbf{c} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4,$$

and if so, to find such x_i's. The above is equivalent to solving the matrix-vector equation $A\mathbf{x} = \mathbf{c}$. Since $\left[\begin{array}{c|c} A & \mathbf{c} \end{array}\right] \sim \left[\begin{array}{c|c} A' & \mathbf{c}' \end{array}\right]$, we can "read off" the solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{c}$ (if they exist) from the matrix

$$\left[\begin{array}{c|c} A' & \mathbf{c}' \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

The rightmost column (the one to the right of the vertical dotted line) is a pivot column, and it follows that the matrix-vector equation $A\mathbf{x} = \mathbf{c}$ is inconsistent. Consequently, the vector $\mathbf{c}$ is **not** a linear combination of the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$, i.e. $\mathbf{c} \notin \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$.

Vector d. We need to check whether there exist scalars $x_1, x_2, x_3, x_4 \in \mathbb{Z}_3$ such that

$$\mathbf{d} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4,$$

and if so, to find such x_i's. The above is equivalent to solving the matrix-vector equation $A\mathbf{x} = \mathbf{d}$. Since $\left[\begin{array}{c|c} A & \mathbf{d} \end{array}\right] \sim \left[\begin{array}{c|c} A' & \mathbf{d}' \end{array}\right]$, we can "read off" the solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{d}$ (if they exist) from the matrix

$$\left[\begin{array}{c|c} A' & \mathbf{d}' \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

The rightmost column (the one to the right of the vertical dotted line) is a pivot column, and it follows that the matrix-vector equation $A\mathbf{x} = \mathbf{d}$ is inconsistent. Consequently, the vector $\mathbf{d}$ is **not** a linear combination of the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$, i.e. $\mathbf{d} \notin \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$.

Remark: The matrix $\left[\begin{array}{c|c} A' & \mathbf{d}' \end{array}\right]$ is **not** in reduced row echelon form, but this does not matter. The important point is that we have a row of the form $\left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & \blacksquare \end{array}\right]$, where $\blacksquare$ is non-zero. Since the equation $0 = \blacksquare$ is inconsistent (whenever $\blacksquare$ is non-zero), we see that our matrix-vector equation $A\mathbf{x} = \mathbf{d}$ is inconsistent, and consequently, $\mathbf{d} \notin \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$.

Vector e. We need to check whether there exist scalars $x_1, x_2, x_3, x_4 \in \mathbb{Z}_3$ such that

$$\mathbf{e} \;=\; x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4,$$

and if so, to find such x_i's. The above is equivalent to solving the matrix-vector equation $A\mathbf{x} = \mathbf{e}$. Since $\left[\begin{array}{c|c} A & \mathbf{e} \end{array}\right] \sim \left[\begin{array}{c|c} A' & \mathbf{e}' \end{array}\right]$, we can "read off" the solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{e}$ (if they exist) from the matrix

$$\left[\begin{array}{c|c} A' & \mathbf{e}' \end{array}\right] \;=\; \left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

We see that our equation $A\mathbf{x} = \mathbf{e}$ is consistent, and that the general solution is

$$\mathbf{x} \;=\; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \;=\; \begin{bmatrix} s + 2t + 1 \\ s \\ t + 1 \\ t \end{bmatrix}, \qquad \text{where } s, t \in \mathbb{Z}_3.$$

We only need one solution, and so we set $s = t = 0$,[22] which yields

$$\mathbf{x} \;=\; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \;=\; \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

So, $\mathbf{e}$ is indeed a linear combination of the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$, that is, $\mathbf{e} \in \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$, and we have that

$$\mathbf{e} \;=\; 1\mathbf{a}_1 + 0\mathbf{a}_2 + 1\mathbf{a}_3 + 0\mathbf{a}_4 \;=\; \mathbf{a}_1 + \mathbf{a}_3.$$

Remark: As in the case of the vector **b**, we could also simply cross out the non-pivot columns of $\left[\begin{array}{c|c} A' & \mathbf{e}' \end{array}\right]$ to the left of the vertical dotted line, and read off the (unique) solution that remains after we eliminated those columns. Again, we must keep track of which x_i corresponds to which column.

[22] We could choose other values for s and t (for example, $s = 2$ and $t = 1$), but it is simplest to choose $s = t = 0$.

$$\left[\begin{array}{c|c} A' & \mathbf{e}' \end{array}\right] = \left[\begin{array}{cc|cc|c} 1 & 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$
$$\begin{array}{cc} \uparrow & \uparrow \\ x_1 & x_3 \end{array}$$

We now read off $x_1 = 1$ and $x_3 = 1$, and we get $\mathbf{e} = \mathbf{a}_1 + \mathbf{a}_3$. Because we crossed out some non-pivot columns (which correspond to arbitrary parameters), our solution is not unique. □

Example 1.5.6. *Consider the vectors*

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \\ 10 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 7 \\ 12 \\ 10 \\ -1 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 3 \\ 8 \\ 11 \\ 13 \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 6 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. For each of the vectors $\mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}$, determine if it can be expressed as a linear combination of the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ (that is, if it belongs to Span$(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$), and if so, express it as such a linear combination, and explain whether your answer is unique.

Solution. Set

$$A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{array}\right] = \begin{bmatrix} 1 & 0 & -2 & 0 \\ 2 & 1 & -1 & 0 \\ 2 & 2 & 2 & 1 \\ 1 & 4 & 10 & 2 \end{bmatrix}.$$

We now form the following matrix (color coded for emphasis):

$$\left[\begin{array}{c|cccc} A & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \end{array}\right] = \left[\begin{array}{cccc|cccc} 1 & 0 & -2 & 0 & 2 & 7 & 3 & 1 \\ 2 & 1 & -1 & 0 & 2 & 12 & 8 & 2 \\ 2 & 2 & 2 & 1 & 3 & 10 & 11 & 4 \\ 1 & 4 & 10 & 2 & 1 & -1 & 13 & 6 \end{array}\right].$$

We find the reduced row echelon form of the matrix $\left[\begin{array}{c|cccc} A & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \end{array}\right]$, and we cross out any non-pivot columns to the left of the vertical dotted line.

$$\mathrm{RREF}\Big(\big[\ A\ \big|\ \mathbf{b}\ \ \mathbf{c}\ \ \mathbf{d}\ \ \mathbf{e}\ \big]\Big) \;=\; \left[\begin{array}{cccc|cccc} 1 & 0 & -2 & 0 & 0 & 7 & 3 & -1 \\ 0 & 1 & 3 & 0 & 0 & -2 & 2 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

We see from the matrix above that $\mathbf{b}, \mathbf{e} \notin \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$, whereas $\mathbf{c}, \mathbf{d} \in \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$. Moreover, vectors $\mathbf{c}$ and $\mathbf{d}$ can be expressed as linear combinations of the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$, as follows:

- $\mathbf{c} = 7\mathbf{a}_1 - 2\mathbf{a}_2$;
- $\mathbf{d} = 3\mathbf{a}_1 + 2\mathbf{a}_2 + \mathbf{a}_4$.

Since we crossed out a non-pivot column (which corresponds to an arbitrary parameter), the two expressions above are **not** unique. □

1.6 The rank of a matrix

The *rank* of a matrix A (with entries in some field $\mathbb{F}$), denoted by $\mathrm{rank}(A)$, is the number of pivot columns of A. Equivalently, $\mathrm{rank}(A)$ is the number of pivot positions of A, or the number of non-zero rows of any row echelon form of A. To find the rank of a matrix, we first find some row echelon form of that matrix (e.g. by performing the forward phase of row reduction; the backward phase is optional), and we count the number of pivot columns (or alternatively, the number of pivot positions, or the number of non-zero rows) of that row echelon matrix.

Example 1.6.1. *Find the rank of each of the following matrices.*

(a) $A = \begin{bmatrix} 0 & -3 & -6 & 3 & 4 & -1 \\ 2 & 1 & -4 & 13 & -4 & 3 \\ 2 & 3 & 0 & 11 & -6 & 5 \end{bmatrix}$, *with entries understood to be in* $\mathbb{R}$*;*

(b) $B = \begin{bmatrix} 0 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 1 \end{bmatrix}$, *with entries understood to be in* $\mathbb{Z}_3$.

Solution#1. (a) In Example 1.3.9, we computed

$$\mathrm{RREF}(A) \;=\; \begin{bmatrix} 1 & 0 & -3 & 7 & 0 & 4 \\ 0 & 1 & 2 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

The matrix $\mathrm{RREF}(A)$ has three pivot columns (equivalently: three pivot positions or three non-zero rows), and so $\mathrm{rank}(A) = 3$.

(b) In Example 1.3.10, we computed

$$\text{RREF}(B) = \begin{bmatrix} 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The matrix $\text{RREF}(B)$ has three pivot columns (equivalently: three pivot positions or three non-zero rows), and so $\text{rank}(B) = 3$. □

Solution#2. (a) In Example 1.3.9, we saw that the matrix A is row equivalent to the following matrix in row echelon form:

$$\begin{bmatrix} 2 & 3 & 0 & 11 & -6 & 5 \\ 0 & -2 & -4 & 2 & 2 & -2 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

This row echelon matrix has three pivot columns (equivalently: three pivot positions or three non-zero rows), and so $\text{rank}(A) = 3$.

(b) In Example 1.3.10, we saw that the matrix B is row equivalent to the following matrix in row echelon form:

$$\begin{bmatrix} 1 & 0 & 2 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This row echelon matrix has three pivot columns (equivalently: three pivot positions or three non-zero rows), and so $\text{rank}(B) = 3$. □

Proposition 1.6.2. *Row equivalent matrices (with entries in some field) have the same rank.*

Proof. Fix row equivalent matrices A and B (with entries in some field). By the definition of rank, $\text{rank}(A)$ is equal to the number of pivot columns of A, which is precisely the number of pivot columns of $\text{RREF}(A)$. Similarly, $\text{rank}(B)$ is equal to the number of pivot columns of $\text{RREF}(B)$. But since A and B are row equivalent, Corollary 1.3.8 guarantees that $\text{RREF}(A) = \text{RREF}(B)$. So, $\text{rank}(A) = \text{rank}(B)$. □

Proposition 1.6.3. *Let A be an $n \times m$ matrix (with entries in some field $\mathbb{F}$). Then rank$(A) \leq \min\{n, m\}$.*[23]

[23] This means that $\text{rank}(A) \leq n$ (i.e. $\text{rank}(A)$ is at most the number of rows of A) and $\text{rank}(A) \leq m$ (i.e. $\text{rank}(A)$ is at most the number of columns of A).

Proof. By definition, $\text{rank}(A)$ is equal to the number of pivot columns of A, and consequently, $\text{rank}(A)$ is at most the number of columns of A, which is m. So, $\text{rank}(A) \leq m$.

On the other hand, $\text{rank}(A)$ is equal to the number of non-zero rows of $\text{RREF}(A)$, and consequently, $\text{rank}(A)$ is at most the number of rows of $\text{RREF}(A)$; since A and $\text{RREF}(A)$ have the same number of rows, we deduce that $\text{rank}(A)$ is at most the number of rows of A, which is n. So, $\text{rank}(A) \leq n$. □

Terminology: For a field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n\times m}$ (so, A has n rows and m columns):

- if $\text{rank}(A) = n$, then A is said to have *full row rank*;[24]
- if $\text{rank}(A) = m$, then A is said to have *full column rank*;[25]
- if $\text{rank}(A) = \min\{n, m\}$, then A is said to have *full rank*;[26]
- if $\text{rank}(A) < \min\{n, m\}$, then A is said to be *rank-deficient*.

1.6.1 Rank and the number of solutions of a matrix-vector equation

As our next theorem shows, the number of solutions of a matrix-vector equation $A\mathbf{x} = \mathbf{b}$ can easily be determined if we know the size of the matrix A (i.e. the number of rows and columns of A) and we also know $\text{rank}(A)$ and $\text{rank}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right)$.

Theorem 1.6.4. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$ and $\mathbf{b} \in \mathbb{F}^n$. Then*

$$\text{rank}(A) \quad \leq \quad \text{rank}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right) \quad \leq \quad \text{rank}(A) + 1.$$

Moreover, all the following hold:

(a) if $\text{rank}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right) \neq \text{rank}(A)$ (and consequently, $\text{rank}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right) = \text{rank}(A)+1$), then the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is inconsistent.

(b) if $\text{rank}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right) = \text{rank}(A) = m$, then the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution.

(c) if $\text{rank}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right) = \text{rank}(A) < m$, then the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has more than one solution, and more precisely,

(c.1) if the field $\mathbb{F}$ is finite, then the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has exactly $|\mathbb{F}|^{m-\text{rank}(A)}$ many solutions,[27]

[24] In this case, Proposition 1.6.3 guarantees that $n \leq m$, i.e. the number of rows of A is no greater than the number of columns.

[25] In this case, Proposition 1.6.3 guarantees that $m \leq n$, i.e. the number of columns of A is no greater than the number of rows.

[26] So, A has full rank if and only if it has full row rank **or** full column rank.

[27] As usual, $|\mathbb{F}|$ is the cardinality of $\mathbb{F}$, i.e. the number of elements of $\mathbb{F}$.

(c.2) if the field $\mathbb{F}$ is infinite, then the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions.

Proof. First, set $\left[\begin{array}{c:c} U & \mathbf{c} \end{array}\right] = \mathrm{RREF}\left(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]\right)$, so that $\mathrm{RREF}(A) = U$.[28] Next, let $(\star)$ be the linear system whose augmented matrix is $\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]$. Obviously, the linear system $(\star)$ and the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ have the same number of solutions, and we can move back and forth between the linear system $(\star)$ and the matrix-vector equations $A\mathbf{x} = \mathbf{b}$ in a natural way.

Let us first prove that $\mathrm{rank}(A) \leq \mathrm{rank}(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]) \leq \mathrm{rank}(A) + 1$. The pivot columns of $\left[\begin{array}{c:c} U & \mathbf{c} \end{array}\right]$ are precisely the pivot columns of U, plus possibly the rightmost column (namely, the column $\mathbf{c}$ to the right of the vertical dotted line). If the rightmost column of $\left[\begin{array}{c:c} U & \mathbf{c} \end{array}\right]$ is a pivot column, then $\mathrm{rank}(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]) = \mathrm{rank}(A) + 1$, and otherwise, $\mathrm{rank}(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]) = \mathrm{rank}(A)$. This proves that

$$\mathrm{rank}(A) \;\leq\; \mathrm{rank}(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]) \;\leq\; \mathrm{rank}(A) + 1.$$

We now prove (a). Suppose that $\mathrm{rank}(A) \neq \mathrm{rank}(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right])$. By what we just showed, this implies that $\mathrm{rank}(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]) = \mathrm{rank}(A) + 1$ and that the rightmost column of $\left[\begin{array}{c:c} U & \mathbf{c} \end{array}\right] = \mathrm{RREF}\left(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]\right)$ is a pivot column. It follows that the linear system $(\star)$ is inconsistent, and consequently, that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is inconsistent. This proves (a).

It remains to prove (b) and (c), which we prove simultaneously. Assume that $\mathrm{rank}(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]) = \mathrm{rank}(A)$. Then the rightmost column of $\left[\begin{array}{c:c} U & \mathbf{c} \end{array}\right] = \mathrm{RREF}\left(\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]\right)$ is not a pivot column,[29] and it follows that the system $(\star)$ is consistent. Further, exactly $\mathrm{rank}(A)$ many columns of U are pivot columns, and consequently, the system $(\star)$ has precisely $\mathrm{rank}(A)$ many basic variables. The remaining $m - \mathrm{rank}(A)$ many variables of $(\star)$ are free variables. If $\mathrm{rank}(A) = m$, it follows that $(\star)$ has no free variables, and we deduce that the linear system $(\star)$ has a unique solution, and consequently, that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution. This proves (b). We may now assume that $\mathrm{rank}(A) < m$. So, the linear system $(\star)$ has at least one free variable, and therefore, it has more than one solution. Each free variable can take any value from the field $\mathbb{F}$, and the values of the basic variables are fully determined by the values of the free variables. So, if $\mathbb{F}$ is infinite, then the number of solutions of $(\star)$ is infinite,[30] and if $\mathbb{F}$ is finite, then $(\star)$ has precisely $|\mathbb{F}|^{m-\mathrm{rank}(A)}$ many solutions.[31] Since the number of solutions of the linear system $(\star)$ is the same as the number of solutions of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, this proves (c). □

[28] This is "obvious," but it also follows from Proposition 1.3.20(b).

[29] This follows from the discussion above (the second paragraph of the proof).

[30] This is because each free variable can take infinitely many values, and there is at least one free variable.

[31] This is because each free variable can take $|\mathbb{F}|$ many different values, and there are precisely $m - \mathrm{rank}(A)$ many free variables.

1.6.2 Matrices of full rank

In this subsection, we prove a couple of corollaries of Theorem 1.6.4 for matrices of full rank (see Corollaries 1.6.5 and 1.6.6 below). By definition, a matrix of full rank has full column rank or full row rank (possibly both). We deal with these two cases separately. Finally, at the end of the subsection, we prove Theorem 1.6.8, which deals with **square** matrices of full rank (note that such matrices have both full column rank and full row rank).

Matrices of full column rank. In a matrix of full column rank, all columns are pivot columns. So, the reduced row echelon form of such a matrix is of the form

$$\left[\begin{array}{cccccc} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ \hdashline 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \end{array}\right],$$

where the zero rows on the bottom are optional. More precisely, if we have an $n \times m$ matrix of full column rank,[32] then the reduced row echelon form of that matrix is obtained from the identity matrix I_m by adding $n - m$ many zero rows to the bottom.

A *homogeneous matrix-vector equation* is a matrix-vector equation of the form $A\mathbf{x} = \mathbf{0}$. Note that such an equation is always consistent: indeed, $\mathbf{x} = \mathbf{0}$ is a solution, called the trivial solution.

Corollary 1.6.5. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$. Then the following are equivalent:*

(a) $\text{rank}(A) = m$ (i.e. A has full column rank);

(b) the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$);

(c) there exists some vector $\mathbf{b} \in \mathbb{F}^n$ such that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;

[32] Note that this means that $\text{rank}(A) = m \leq n$. Indeed, since the $n \times m$ matrix A has full column rank, we have that $\text{rank}(A) = m$. On the other hand, by Proposition 1.6.3, we have that $\text{rank}(A) \leq n$. So, $\text{rank}(A) = m \leq n$.

(d) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *has at most one solution.*

Proof. It is enough to prove the implications shown in the diagram below.

$$\begin{array}{ccc} \text{(a)} & \Longrightarrow & \text{(d)} \\ \Uparrow & & \Downarrow \\ \text{(c)} & \Longleftarrow & \text{(b)} \end{array}$$

In fact, the implications "(d) $\Longrightarrow$ (b)" and "(b) $\Longrightarrow$ (c)" are obvious. It remains to prove the implications "(c) $\Longrightarrow$ (a)" and "(a) $\Longrightarrow$ (d)."

We first prove the implication "(c) $\Longrightarrow$ (a)." Assume that (c) is true, and fix a vector $\mathbf{b} \in \mathbb{F}^n$ such that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution. In particular, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent, and so Theorem 1.6.4(a) guarantees that $\operatorname{rank}(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]) = \operatorname{rank}(A)$. Moreover, by Proposition 1.6.3 and Theorem 1.6.4(c), we have that $\operatorname{rank}(A) = m$.[33] Thus, (a) holds.

It remains to prove the implication "(a) $\Longrightarrow$ (d)." Assume that (a) is true, i.e. that $\operatorname{rank}(A) = m$, and fix a vector $\mathbf{b} \in \mathbb{F}^n$. We must show that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has at most one solution. If $\operatorname{rank}(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]) \neq \operatorname{rank}(A)$, then Theorem 1.6.4(a) guarantees that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has no solutions. On the other hand, if $\operatorname{rank}(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]) = \operatorname{rank}(A)$, then since $\operatorname{rank}(A) = m$, Theorem 1.6.4(b) guarantees that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution. In either case, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has at most one solution, i.e. (d) holds. □

Matrices of full row rank. Note that matrices of full row rank are precisely those matrices whose reduced row echelon form has no zero rows.

Corollary 1.6.6. *Let* $\mathbb{F}$ *be a field, and let* $A \in \mathbb{F}^{n \times m}$*. Then the following are equivalent:*

(a) $\operatorname{rank}(A) = n$ *(i.e. A has full row rank);*

(b) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *is consistent.*

Proof. Suppose first that (a) holds. We must prove (b). Fix any $\mathbf{b} \in \mathbb{F}^n$. Then

$$\begin{array}{rcll} n & = & \operatorname{rank}(A) & \text{by (a)} \\ & \leq & \operatorname{rank}(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]) & \text{by Theorem 1.6.4} \\ & \leq & n & \begin{array}{l}\text{by Proposition 1.6.3,} \\ \text{since } \left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] \text{ is an} \\ n \times (m+1) \text{ matrix,}\end{array} \end{array}$$

[33] Indeed, by Proposition 1.6.3, we have that $\operatorname{rank}(A) \leq m$. If $\operatorname{rank}(A) < m$, then Theorem 1.6.4(c) would imply that $A\mathbf{x} = \mathbf{b}$ has more than one solution, a contradiction. So, $\operatorname{rank}(A) = m$.

and it follows that $\operatorname{rank}(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]) = \operatorname{rank}(A) = n$. But now Theorem 1.6.4 guarantees that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent. Thus, (b) holds.

Suppose now that (a) is false; we must show that (b) is false, i.e. that there exists some $\mathbf{b} \in \mathbb{F}^n$ such that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is inconsistent. Since A is an $n \times m$ matrix and $\operatorname{rank}(A) \neq n$, Proposition 1.6.3 guarantees that $\operatorname{rank}(A) \leq n-1$. Now, set $U := \operatorname{RREF}(A)$, and let $R_1, \ldots, R_k$ be some sequence of elementary row operations that transforms A into U, and for each $i \in \{1, \ldots, k\}$, let R'_i be the elementary row operation that reverses (undoes) the elementary row operation R_i.[34] Since U has n rows and $r := \operatorname{rank}(A) \leq n-1$, we see that the $(r+1)$-th row of U is a zero row. Then the rightmost column of the matrix $\left[\begin{array}{c|c} U & \mathbf{e}_{r+1} \end{array}\right]$ is a pivot column,[35] and consequently, the matrix-vector equation $U\mathbf{x} = \mathbf{e}_{r+1}$ is inconsistent. Now, we perform the elementary row operations $R'_k, \ldots, R'_1$ on the matrix $\left[\begin{array}{c|c} U & \mathbf{e}_{r+1} \end{array}\right]$, and we obtain the matrix $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$ for some vector $\mathbf{b} \in \mathbb{F}^n$. Since matrices $\left[\begin{array}{c|c} U & \mathbf{e}_{r+1} \end{array}\right]$ and $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$ are row equivalent, the matrix-vector equations $U\mathbf{x} = \mathbf{e}_{r+1}$ and $A\mathbf{x} = \mathbf{b}$ are equivalent. Since the matrix-vector equation $U\mathbf{x} = \mathbf{e}_{r+1}$ is inconsistent, it follows that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is also inconsistent. Thus, (b) is false. □

Square matrices of full rank. We now consider the special case of square matrices of full rank. We begin with a simple proposition.

Proposition 1.6.7. *Let $\mathbb{F}$ be a field. Then for all square matrices $A \in \mathbb{F}^{n \times n}$, we have that rank$(A) = n$ if and only if RREF$(A) = I_n$. In particular, rank$(I_n) = n$.*

Proof. I_n is a matrix in reduced row echelon form, and it has n pivot columns; so, $\operatorname{rank}(I_n) = n$. Moreover, it is clear that I_n is the **only** reduced row echelon form matrix in $\mathbb{F}^{n \times n}$ of rank n.

Now, fix any matrix $A \in \mathbb{F}^{n \times n}$. By Proposition 1.6.2, we have that $\operatorname{rank}(A) = \operatorname{rank}\big(\operatorname{RREF}(A)\big)$. Since I_n is the only reduced row echelon form matrix in $\mathbb{F}^{n \times n}$ of rank n, it follows that $\operatorname{rank}(A) = n$ if and only if $\operatorname{RREF}(A) = I_n$. □

Proposition 1.6.7 and Corollaries 1.6.5 and 1.6.6 readily yield the following theorem.

Theorem 1.6.8. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$ be a **square** matrix. Then the following are equivalent:*

(a) rank$(A) = n$ (i.e. the square matrix A has full rank);

(b) RREF$(A) = I_n$;

[34] See subsection 1.3.2.

[35] Here, $\mathbf{e}_{r+1}$ is the $(r+1)$-th standard basis vector of $\mathbb{F}^n$, i.e. the vector whose $(r+1)$-th entry is 1, and all of whose other entries are 0.

(c) the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$);

(d) there exists some vector $\mathbf{b} \in \mathbb{F}^n$ such that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;

(e) for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;

(f) for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has at most one solution;

(g) for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent.

Proof. By Proposition 1.6.7, (a) and (b) are equivalent, and by Corollary 1.6.6, (a) and (g) are equivalent. Further, Corollary 1.6.5 guarantees that (a), (c), (d), and (f) are equivalent. Obviously, (e) implies (f). We complete the proof by showing that (a) implies (e). Assume that (a) holds, and fix a vector $\mathbf{b} \in \mathbb{F}^n$. Since A is a square matrix, (a) guarantees that A has both full column rank and full row rank. Since A has full column rank, Corollary 1.6.5 guarantees that $A\mathbf{x} = \mathbf{b}$ has at most one solution. On the other hand, since A has full row rank, Corollary 1.6.6 guarantees that $A\mathbf{x} = \mathbf{b}$ is consistent, i.e. has at least one solution. It now follows that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution, i.e. (e) holds. □

1.7 Matrix operations

1.7.1 Matrix addition, matrix subtraction, and scalar-matrix multiplication

Suppose that $\mathbb{F}$ is a field. Given matrices $A = \left[\ a_{i,j}\ \right]_{n\times m}$ and $B = \left[\ b_{i,j}\ \right]_{n\times m}$ in $\mathbb{F}^{n\times m}$, and given a scalar c, we define

- $A + B := \left[\ a_{i,j} + b_{i,j}\ \right]_{n\times m}$;
- $A - B := \left[\ a_{i,j} - b_{i,j}\ \right]_{n\times m}$;
- $cA := \left[\ ca_{i,j}\ \right]$.

Thus, we add (resp. subtract) matrices by adding (resp. subtracting) corresponding entries, i.e.

- $\left[\ a_{i,j}\ \right]_{n\times m} + \left[\ b_{i,j}\ \right]_{n\times m} = \left[\ a_{i,j} + b_{i,j}\ \right]_{n\times m}$;
- $\left[\ a_{i,j}\ \right]_{n\times m} - \left[\ b_{i,j}\ \right]_{n\times m} = \left[\ a_{i,j} - b_{i,j}\ \right]_{n\times m}$.

Similarly, we multiply a matrix by a scalar (on the left) by multiplying each entry of the matrix by that scalar, i.e.

- $c \left[\ a_{i,j} \ \right]_{n\times m} = \left[\ ca_{i,j} \ \right]_{n\times m}$.

Notation: By convention, for a matrix A and scalar c, we write cA, but we do **not** write Ac. In other words, by convention, we have "scalar times matrix," but not "matrix times scalar."

1.7.2 Matrix multiplication

Let $\mathbb{F}$ be a field, and suppose that we are given two matrices, $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times p}$, where $B = \left[\ \mathbf{b}_1 \ \ \ldots \ \ \mathbf{b}_p \ \right]$. We define

$$AB \ := \ \left[\ A\mathbf{b}_1 \ \ \ldots \ \ A\mathbf{b}_p \ \right]$$

Note that $AB \in \mathbb{F}^{n\times p}$.

Note that, for the product AB to be defined, the number of **columns** of A must be the same as the number of **rows** of B. The matrix AB has the same number of **rows** as A, and the same number of **columns** as B. Schematically, we get:

$$(n \times m) \ \cdot \ (m \times p) \ = \ (n \times p).$$

Example 1.7.1. *Let*

$$A \ = \ \begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 1 \end{bmatrix} \quad \textit{and} \quad B \ = \ \begin{bmatrix} 1 & 1 \\ -2 & 1 \\ 0 & -1 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. *Compute* AB.

Solution. We set

$$\mathbf{b}_1 \ = \ \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_2 \ = \ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix},$$

so that $B = \left[\ \mathbf{b}_1 \ \ \mathbf{b}_2 \ \right]$. Then $AB = \left[\ A\mathbf{b}_1 \ \ A\mathbf{b}_2 \ \right]$.

We compute

$$A\mathbf{b}_1 \ = \ \begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} \ = \ 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (-2) \begin{bmatrix} 2 \\ -3 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \ \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -4 \\ 6 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ = \ \begin{bmatrix} -3 \\ 6 \end{bmatrix}$$

and

$$A\mathbf{b}_2 = \begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ -3 \end{bmatrix} + (-1) \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ -3 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \end{bmatrix},$$

which yields

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} -3 & 4 \\ 6 & -4 \end{bmatrix}.$$

□

Multiplication by the identity or zero matrix. As our next proposition shows, by multiplying a matrix by an appropriately sized identity matrix (on the left or on the right), we obtain the original matrix. On the other hand, multiplying a matrix by an appropriately sized zero matrix (on the left or on the right) yields an appropriately sized zero matrix.

Proposition 1.7.2. *Let $\mathbb{F}$ be a field, let m, n, p be positive integers, and let $A \in \mathbb{F}^{n \times m}$ be a matrix. Then all the following hold:*

(a) $I_n A = A I_m = A$;

(b) $A O_{m \times p} = O_{n \times p}$;

(c) $O_{p \times n} A = O_{p \times m}$.

Proof. Parts (b) and (c) readily follow from the appropriate definitions (the details are left as an easy exercise). Let us prove (a). Set $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$. To show that $I_n A = A$, we compute:

$$\begin{aligned} I_n A &= I_n \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix} \\ &= \begin{bmatrix} I_n \mathbf{a}_1 & \dots & I_n \mathbf{a}_m \end{bmatrix} && \text{by the definition of matrix multiplication} \\ &= \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix} && \text{by Proposition 1.4.5} \\ &= A. \end{aligned}$$

On the other hand, to show that $A I_m = A$, we compute:

$$A I_m = A \begin{bmatrix} \mathbf{e}_1^m & \dots & \mathbf{e}_m^m \end{bmatrix}$$

$$
\begin{aligned}
&= \begin{bmatrix} A\mathbf{e}_1^m & \dots & A\mathbf{e}_m^m \end{bmatrix} && \text{by the definition of matrix multiplication} \\
&= \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix} && \text{by Proposition 1.4.4} \\
&= A.
\end{aligned}
$$

This proves (a). □

Another way to compute the product of two matrices. Suppose we are given a matrix $A \in \mathbb{F}^{n\times m}$ and a vector $\mathbf{v} \in \mathbb{F}^m$. Set $A = \begin{bmatrix} a_{i,j} \end{bmatrix}_{n\times m}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$. Then by the definition of a matrix-vector product, we have that $A\mathbf{v} \in \mathbb{F}^n$, and moreover, we have the following:

$$
\begin{aligned}
A\mathbf{v} &= \begin{bmatrix} a_{1,1} & \dots & a_{1,k} & \dots & a_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1} & \dots & a_{i,k} & \dots & a_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,k} & \dots & a_{n,m} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_k \\ \vdots \\ v_m \end{bmatrix} \\
&= v_1 \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{i,1} \\ \vdots \\ a_{n,1} \end{bmatrix} + \dots + v_k \begin{bmatrix} a_{1,k} \\ \vdots \\ a_{i,k} \\ \vdots \\ a_{n,k} \end{bmatrix} + \dots + v_m \begin{bmatrix} a_{1,m} \\ \vdots \\ a_{i,m} \\ \vdots \\ a_{n,m} \end{bmatrix} \\
&= \begin{bmatrix} a_{1,1}v_1 + \dots + a_{1,k}v_k + \dots + a_{1,m}v_m \\ \vdots \\ a_{i,1}v_1 + \dots + a_{i,k}v_k + \dots + a_{i,m}v_m \\ \vdots \\ a_{n,1}v_1 + \dots + a_{n,k}v_k + \dots + a_{n,m}v_m \end{bmatrix}
\end{aligned}
$$

$$= \begin{bmatrix} \sum_{k=1}^{m} a_{1,k}v_k \\ \vdots \\ \sum_{k=1}^{m} a_{i,k}v_k \\ \vdots \\ \sum_{k=1}^{m} a_{n,k}v_k \end{bmatrix}.$$

So, the i-th entry of the vector $A\mathbf{v}$ is $\sum_{k=1}^{m} a_{i,k}v_k$.

Let us now consider the product of two matrices. Suppose we are given matrices $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times p}$, and set $A = \begin{bmatrix} a_{i,j} \end{bmatrix}_{n\times m}$ and $B = \begin{bmatrix} b_{i,j} \end{bmatrix}_{m\times p}$. The matrix AB belongs to $\mathbb{F}^{n\times p}$. We would like to compute the i,j-th entry of the matrix AB in terms of the entries of A and B. The i,j-th entry of AB is precisely the i-th entry of the j-th column of AB, and by the definition of matrix product, the j-th column of AB is the vector $A\mathbf{b}_j$, where $\mathbf{b}_j = \begin{bmatrix} b_{1,j} \\ \vdots \\ b_{m,j} \end{bmatrix}$ is the j-th column of B. Using the formula for the matrix-vector product that we obtained above, we see that the i-th entry of the vector $A\mathbf{b}_j$ is $\sum_{k=1}^{m} a_{i,k}b_{k,j}$. So, the i,j-th entry of the $n \times p$ matrix AB is $\sum_{k=1}^{m} a_{i,k}b_{k,j}$.

Here is a way to visualize the product of two matrices. To obtain the i,j-th entry of the matrix AB, we focus on the i-th row of A and j-th column of B. We then take the sum of the products of the corresponding entries of this row and column, and we obtain the i,j-th entry of AB. Schematically, this is represented below. The matrix A is on the bottom-left, the matrix B is on the top-right, and the matrix AB is on the bottom-right (squeezed between A and B). The i-th row of A is in red, the j-th column of B is in blue, and the i,j-th entry of AB is $\sum_{k=1}^{m} a_{i,k}b_{k,j}$.

$$\begin{bmatrix} a_{1,1} & \dots & a_{1,k} & \dots & a_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1} & \dots & a_{i,k} & \dots & a_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,k} & \dots & a_{n,m} \end{bmatrix} \begin{bmatrix} b_{1,1} & \dots & b_{1,j} & \dots & b_{1,p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{k,1} & \dots & b_{k,j} & \dots & b_{k,p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{m,1} & \dots & b_{m,j} & \dots & b_{m,p} \end{bmatrix} = \begin{bmatrix} & & \\ & \sum_{k=1}^{m} a_{i,k}b_{k,j} & \\ & & \end{bmatrix}$$

Another way to write this is as follows:

$$\left[\ a_{i,j}\ \right]_{n\times m} \left[\ b_{i,j}\ \right]_{m\times p} = \left[\ \sum_{k=1}^{m} a_{i,k}b_{k,j}\ \right]_{n\times p},$$

where in each of the three matrices, the expression between the square brackets is the general form of the i,j-th entry (i.e. the entry in the i-th row and j-th column) of the matrix in question.

Example 1.7.3. *Let*

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_2$. Compute the matrix AB.

Solution. We compute as shown below (the rows of A are color coded, as are the columns of B).

$$\begin{bmatrix} & 1 & & 0 & & 1 & \\ & 1 & & 1 & & 0 & \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1\cdot 1+0\cdot 1 & 1\cdot 0+0\cdot 1 & 1\cdot 1+0\cdot 0 \\ 1\cdot 1+1\cdot 1 & 1\cdot 0+1\cdot 1 & 1\cdot 1+1\cdot 0 \end{bmatrix}$$

By performing arithmetic (in $\mathbb{Z}_2$) on the entries of the matrix that we obtained, we get:

$$AB = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

□

Example 1.7.4. *Let*

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 0 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$*. Compute the matrix* AB*.*

Solution. We compute as shown below.

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 2 & 4 \end{bmatrix} \quad \begin{bmatrix} (1\cdot 1+2\cdot 2) & (1\cdot 0+2\cdot 3) & (1\cdot(-1)+2\cdot 0) \\ ((-1)\cdot 1+0\cdot 2) & ((-1)\cdot 0+0\cdot 3) & ((-1)\cdot(-1)+0\cdot 0) \\ (2\cdot 1+4\cdot 2) & (2\cdot 0+4\cdot 3) & (2\cdot(-1)+4\cdot 0) \end{bmatrix}$$

By performing arithmetic (in $\mathbb{R}$) on the entries of the matrix that we obtained, we get

$$AB = \begin{bmatrix} 5 & 6 & -1 \\ -1 & 0 & 1 \\ 10 & 12 & -2 \end{bmatrix}.$$

□

1.7.3 Properties of matrix addition and multiplication

Theorem 1.7.5. *For any matrices* A*,* B*, and* C*, and any scalars* α *and* β*, the following hold (provided the matrices are of compatible size for the operation in question, and the entries of our matrices and our scalars all belong to the same field* $\mathbb{F}$*):*

(a) $(\alpha+\beta)A = \alpha A + \beta A$*;*

(b) $(\alpha\beta)A = \alpha(\beta A)$

(c) $A + B = B + A$*;*

(d) $(A+B)+C = A+(B+C)$*;*

(e) $(A+B)C = AC + BC$*;*

(f) $A(B+C) = AB + AC$*;*

(g) $(AB)C = A(BC)$*;*

(h) $(\alpha A)B = \alpha(AB)$*;*

(i) $A(\alpha B) = \alpha(AB)$*.*

Proof. Parts (a)-(d) readily follow from the definition of the relevant operations and from the properties of addition and multiplication in the field $\mathbb{F}$ (listed in section 1.1). We prove (a) to illustrate the principle; the proof of (b)-(d) is left as an exercise. Fix

a matrix $A = \left[\ a_{i,j}\ \right]_{n\times m}$ in $\mathbb{F}^{n\times m}$, and fix scalars $\alpha, \beta \in \mathbb{F}$. Then

$$\begin{aligned}
(\alpha+\beta)A &= (\alpha+\beta)\left[\ a_{i,j}\ \right]_{n\times m} \\
&= \left[\ (\alpha+\beta)a_{i,j}\ \right]_{n\times m} && \text{by the definition of the scalar-matrix product} \\
&= \left[\ \alpha a_{i,j} + \beta a_{i,j}\ \right]_{n\times m} && \text{by the distributive property of multiplication over addition in } \mathbb{F} \\
&= \left[\ \alpha a_{i,j}\ \right]_{n\times m} + \left[\ \beta a_{i,j}\ \right]_{n\times m} && \text{by the definition of matrix addition} \\
&= \alpha\left[\ a_{i,j}\ \right]_{n\times m} + \beta\left[\ a_{i,j}\ \right]_{n\times m} && \text{by the definition of the scalar-matrix product} \\
&= \alpha A + \beta A.
\end{aligned}$$

Thus, (a) holds.

Next, we prove (e). Fix matrices $A = \left[\ a_{i,j}\ \right]_{n\times m}$ and $B = \left[\ b_{i,j}\ \right]_{n\times m}$ in $\mathbb{F}^{n\times m}$, and fix a matrix $C = \left[\ c_{i,j}\ \right]_{m\times p}$ in $\mathbb{F}^{m\times p}$. We compute:

$$\begin{aligned}
(A+B)C &= \left(\left[\ a_{i,j}\ \right]_{n\times m} + \left[\ b_{i,j}\ \right]_{n\times m}\right)\left[\ c_{i,j}\ \right]_{m\times p} \\
&= \left[\ a_{i,j}+b_{i,j}\ \right]_{n\times m}\ \left[\ c_{i,j}\ \right]_{m\times p} \\
&= \left[\ \sum_{k=1}^{m}(a_{i,k}+b_{i,k})c_{k,j}\ \right]_{n\times p} \\
&\overset{(*)}{=} \left[\ \left(\sum_{k=1}^{m} a_{i,k}c_{k,j}\right) + \left(\sum_{k=1}^{m} b_{i,k}c_{k,j}\right)\ \right]_{n\times p} \\
&= \left[\ \sum_{k=1}^{m} a_{i,k}c_{k,j}\ \right]_{n\times p} + \left[\ \sum_{k=1}^{m} b_{i,k}c_{k,j}\ \right]_{n\times p} \\
&\overset{(**)}{=} AC + BC,
\end{aligned}$$

where (*) follows from the fact that addition distributes over multiplication in the field $\mathbb{F}$, (**) follows from the formula for matrix multiplication that we obtained in

subsection 1.7.2, and the rest follows from the appropriate definitions. This proves (e). The proof of (f) is similar.

We now prove (g). Fix matrices $A = \left[\ a_{i,j}\ \right]_{n_1 \times n_2}$ in $\mathbb{F}^{n_1 \times n_2}$, $B = \left[\ b_{i,j}\ \right]_{n_2 \times n_3}$ in $\mathbb{F}^{n_2 \times n_3}$, and $C = \left[\ c_{i,j}\ \right]_{n_3 \times n_4}$ in $\mathbb{F}^{n_3 \times n_4}$. Clearly, both $(AB)C$ and $A(BC)$ are matrices in $\mathbb{F}^{n_1 \times n_4}$. To prove that these two matrices are equal, it suffices to prove that their corresponding entries are equal. So, fix indices $i \in \{1, \ldots, n_1\}$ and $j \in \{1, \ldots, n_4\}$. We must show that the i,j-th entry of $(AB)C$ is equal to the i,j-th entry of $A(BC)$.

We first compute the i,j-th entry of $(AB)C$. The i-th row of the $n_1 \times n_3$ matrix AB is $\left[\ \sum_{k=1}^{n_2} a_{i,k}b_{k,1} \quad \sum_{k=1}^{n_2} a_{i,k}b_{k,2} \quad \ldots \quad \sum_{k=1}^{n_2} a_{i,k}b_{k,n_3}\ \right]$. The j-th column of the $n_3 \times n_4$ matrix C is $\begin{bmatrix} c_{1,j} \\ c_{2,j} \\ \vdots \\ c_{n_3,j} \end{bmatrix}$. So, the i,j-th entry of the $n_1 \times n_4$ matrix $(AB)C$ is $\sum_{\ell=1}^{n_3} \Big((\sum_{k=1}^{n_2} a_{i,k}b_{k,\ell})c_{\ell,j} \Big)$.

We now compute the i,j-th entry of $A(BC)$. The i-th row of the $n_1 \times n_2$ matrix A is $\left[\ a_{i,1} \quad a_{i,2} \quad \ldots \quad a_{i,n_2}\ \right]$. The j-th column of the $n_2 \times n_4$ matrix BC is $\begin{bmatrix} \sum_{k=1}^{n_3} b_{1,k}c_{k,j} \\ \sum_{k=1}^{n_3} b_{2,k}c_{k,j} \\ \vdots \\ \sum_{k=1}^{n_3} b_{n_2,k}c_{k,j} \end{bmatrix}$. So, the i,j-th entry of the $n_1 \times n_4$ matrix $(AB)C$ is $\sum_{\ell=1}^{n_2} \Big(a_{i,\ell} (\sum_{k=1}^{n_3} b_{\ell,k}c_{k,j}) \Big)$.

It now remains to show that $\sum_{\ell=1}^{n_3} \Big((\sum_{k=1}^{n_2} a_{i,k}b_{k,\ell})c_{\ell,j} \Big) = \sum_{\ell=1}^{n_2} \Big(a_{i,\ell} (\sum_{k=1}^{n_3} b_{\ell,k}c_{k,j}) \Big)$. For this, we compute:

$$\begin{aligned}
\sum_{\ell=1}^{n_3} \Big((\sum_{k=1}^{n_2} a_{i,k}b_{k,\ell})c_{\ell,j} \Big) &= \sum_{\ell=1}^{n_3} \Big(\sum_{k=1}^{n_2} a_{i,k}b_{k,\ell}c_{\ell,j} \Big) && \text{by the distributive property of multiplication over addition in } \mathbb{F} \\
&= \sum_{k=1}^{n_2} \Big(\sum_{\ell=1}^{n_3} a_{i,k}b_{k,\ell}c_{\ell,j} \Big) && \text{by swapping the two } \textstyle\sum\text{'s}
\end{aligned}$$

$$= \sum_{k=1}^{n_2} \Big(a_{i,k} (\sum_{\ell=1}^{n_3} b_{k,\ell} c_{\ell,j}) \Big) \qquad \text{by the distributive property of multiplication over addition in } \mathbb{F}$$

$$= \sum_{\ell=1}^{n_2} \Big(a_{i,\ell} (\sum_{k=1}^{n_3} b_{\ell,k} c_{k,j}) \Big) \qquad \text{by swapping the names of the dummy variables } k \text{ and } \ell,$$

and we obtain the equality that we needed. Thus, $(AB)C = A(BC)$. This proves (g).

The proof of (h) and (i) is left as an exercise. □

Warning: Matrix multiplication is **not** commutative, that is, for matrices A and B,

$$AB \not= BA.$$

In fact, it is possible that one of AB and BA is defined, while the other one is not. (For instance, if $A \in \mathbb{F}^{2\times 3}$ and $B \in \mathbb{F}^{3\times 4}$, where $\mathbb{F}$ is some field, then AB is defined, but BA is not.) Moreover, it is possible that both AB and BA are defined, but are not of the same size. (For instance, if $A \in \mathbb{F}^{2\times 3}$ and $B \in \mathbb{F}^{3\times 2}$, where $\mathbb{F}$ is some field, then $AB \in \mathbb{F}^{2\times 2}$ and $BA \in \mathbb{F}^{3\times 3}$.) Finally, it is possible that AB and BA are both defined, and are of the same size, but $AB \neq BA$. Consider, for example, matrices $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, with the 1's and 0's understood to be in some field $\mathbb{F}$. Then $AB = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, but $BA = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, and we see that $AB \neq BA$.

Recall that a vector is simply a matrix with one column. So, the following corollary is a special case of Theorem 1.7.5.

Corollary 1.7.6. *For any matrices A, B, vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, and scalars α and β, the following hold (provided the matrices and vectors are of compatible size for the operation in question, and the entries of our matrices, the entries of our vectors, and our scalars all belong to the same field $\mathbb{F}$):*

(a) $(\alpha+\beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$;

(b) $(\alpha\beta)\mathbf{u} = \alpha(\beta\mathbf{u})$;

(c) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$;

(d) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$;

(e) $(A + B)\mathbf{u} = A\mathbf{u} + B\mathbf{u}$;

(f) $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$;

(g) $(AB)\mathbf{u} = A(B\mathbf{u})$;

(h) $(\alpha A)\mathbf{u} = \alpha(A\mathbf{u})$;

(i) $A(\alpha\mathbf{u}) = \alpha(A\mathbf{u})$.

1.7.4 Matrix powers

We can define powers of **square** matrices in a natural way, as follows. For a field $\mathbb{F}$ and a square matrix $A \in \mathbb{F}^{n\times n}$, we define

- $A^0 := I_n$;
- $A^{m+1} := A^m A$ for all non-negative integers m.

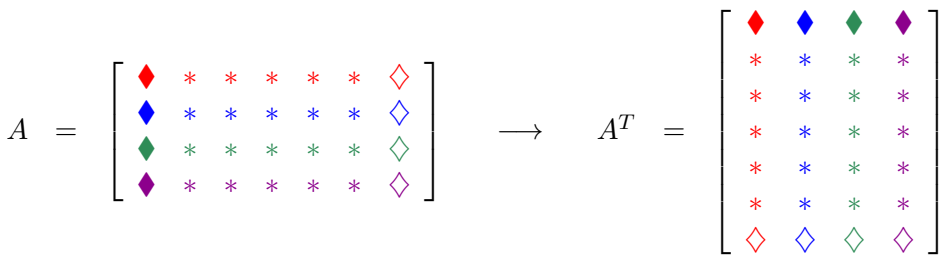
So, by convention, we set $A^0 := I_n$, and for any positive integer m, we have that

$$A^m = \underbrace{A \dots A}_{m},$$

where we did not have to indicate parentheses since, by Theorem 1.7.5(g), matrix multiplication is associative.

1.8 The transpose of a matrix

Given a matrix $A \in \mathbb{F}^{n\times m}$ (where $\mathbb{F}$ is a field), the *transpose* of A, denoted by A^T, is the matrix in $\mathbb{F}^{m\times n}$ such that the i,j-th entry of A^T is the j,i-th entry of A, for all indices $i \in \{1,\dots,m\}$ and $j \in \{1,\dots,n\}$. In other words, to form A^T from A, the columns of A (from left to right) become the rows of A^T (from top to bottom), and likewise, the rows of A (from top to bottom) become the columns of A^T (from left to right). Schematically, we have the picture below.



For example, if $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$.

Column vectors as transposes of row vectors. In order to save space, we often specify column vectors in terms of transposes of row vectors. For instance, we often write something like $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}^T$ instead of $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$.

Proposition 1.8.1. *For any matrices A and B, and any scalar α, the following hold (provided the matrices are of compatible size for the operation in question, and the entries of our matrices and our scalar belong to the same field $\mathbb{F}$):*

(a) $(A^T)^T = A$;

(b) $(A+B)^T = A^T + B^T$;

(c) $(\alpha A)^T = \alpha A^T$

(d) $(AB)^T = B^T A^T$.

Proof. Parts (a), (b), and (c) are obvious. Let us prove (d). Fix matrices $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times p}$, and set $A = \begin{bmatrix} a_{i,j} \end{bmatrix}_{n\times m}$ and $B = \begin{bmatrix} b_{i,j} \end{bmatrix}_{m\times p}$. Clearly, $AB \in \mathbb{F}^{n\times p}$, and so $(AB)^T \in \mathbb{F}^{p\times n}$. On the other hand, we have that $B^T \in \mathbb{F}^{p\times m}$ and $A^T \in \mathbb{F}^{m\times n}$, and so $B^T A^T \in \mathbb{F}^{p\times n}$. So, both $(AB)^T$ and $B^T A^T$ are $p \times n$ matrices with entries in $\mathbb{F}$. It remains to show that the corresponding entries of $(AB)^T$ and $B^T A^T$ are the same. Fix indices $i \in \{1,\dots,p\}$ and $j \in \{1,\dots,n\}$; we will show that the i,j-th entry of $(AB)^T$ is equal to the i,j-th entry of $B^T A^T$.

By the definition of matrix transpose, the i,j-th entry of $(AB)^T$ is equal to the j,i-th entry of AB, which is equal to $\sum_{k=1}^{m} a_{j,k} b_{k,i}$.

We now compute the i,j-th entry of $B^T A^T$. We observe that i-th row of the matrix B^T is $\begin{bmatrix} b_{1,i} & b_{2,i} & \dots & b_{m,i} \end{bmatrix}$,[36] whereas the j-th column of the matrix A^T is $\begin{bmatrix} a_{j,1} & a_{j,2} & \dots & a_{j,m} \end{bmatrix}^T$.[37] So, the i,j-th entry of the matrix $B^T A^T$ is $b_{1,i}a_{j,1} + b_{2,i}a_{j,2} + \dots + b_{m,i}a_{j,m} = \sum_{k=1}^{m} b_{k,i}a_{j,k} = \sum_{k=1}^{m} a_{j,k}b_{k,i}$.

We have now shown that the corresponding entries of the $p \times n$ matrices $(AB)^T$ and $B^T A^T$ are the same, and we deduce that $(AB)^T = B^T A^T$. This proves (d). □

Remark: Proposition 1.8.1(d) and an easy induction on k readily imply that if $A_1,\dots,A_k$ are matrices with entries in some field $\mathbb{F}$, and of sizes that are compatible for the product $A_1 \dots A_k$ to be defined, then $(A_1 \dots A_k)^T = A_k^T \dots A_1^T$. The details are left as an exercise.

Recall that, by Proposition 1.4.4, if we multiply a matrix A by the i-th standard basis vector (on the right), we obtain the i-th column of the matrix. To obtain the i-th row of a matrix, we should multiply it on the left by the transpose of the i-th standard basis vector (see Proposition 1.8.2 below). For example,

$$\underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}}_{=\mathbf{e}_3^T} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 5 & 6 \end{bmatrix},$$

[36] Indeed, this is the transpose of the i-th column of B.

[37] Indeed, this is the transpose of the j-th row of A.

as we can easily verify by routine matrix multiplication.

Proposition 1.8.2. *Let $\mathbb{F}$ be a field, and let*

$$A = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix}$$

be a matrix in $\mathbb{F}^{n\times m}$.[38] *Then for all $i \in \{1, \dots, n\}$, we have that*

$$\mathbf{e}_i^T A = \mathbf{r}_i,$$

where $\mathbf{e}_i$ is the i-th standard basis vector of $\mathbb{F}^n$.

Proof. First of all, we note that $A^T = \begin{bmatrix} \mathbf{r}_1^T & \dots & \mathbf{r}_n^T \end{bmatrix}$, i.e. vectors $\mathbf{r}_1^T, \dots, \mathbf{r}_n^T$ are the columns of A^T, appearing from left to right in A^T. We will apply Proposition 1.4.4 to A^T, as follows. For any index $i \in \{1, \dots, n\}$, we have that

$$\begin{aligned} \mathbf{e}_i^T A &= (A^T \mathbf{e}_i)^T && \text{by Proposition 1.8.1} \\ &= (\mathbf{r}_i^T)^T && \text{by Proposition 1.4.4, since } \mathbf{r}_i^T \text{ is the } i\text{-th column of } A^T \\ &= \mathbf{r}_i && \text{by Proposition 1.8.1,} \end{aligned}$$

which is what we needed to show. □

1.9 Solving matrix equations of the form $AX = B$ and $XA = B$

1.9.1 Solving matrix equations of the form $AX = B$

Example 1.9.1. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ -1 & 3 & 1 & -2 \\ 0 & 1 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 3 \\ -3 & 1 \\ 3 & 0 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Solve the matrix equation $AX = B$.[39] *How many solutions does the equation $AX = B$ have?*

[38] So, $\mathbf{r}_1, \dots, \mathbf{r}_n$ are the rows of A, appearing in that order from top to bottom in A.

[39] Note that solutions of the matrix equation $AX = B$ are 4×2 real matrices.

Solution #1. Set $X = \left[\begin{array}{cc} \mathbf{x}_1 & \mathbf{x}_2 \end{array}\right]$ and $B = \left[\begin{array}{cc} \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right]$. Then $AX = \left[\begin{array}{cc} A\mathbf{x}_1 & A\mathbf{x}_2 \end{array}\right]$, and so the equation $AX = B$ is equivalent to

$$\left[\begin{array}{cc} A\mathbf{x}_1 & A\mathbf{x}_2 \end{array}\right] = \left[\begin{array}{cc} \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right].$$

So, we need to solve two matrix-vector equations, namely $A\mathbf{x}_1 = \mathbf{b}_1$ and $A\mathbf{x}_2 = \mathbf{b}_2$. We solve these two equations one by one.

First, we solve the matrix-vector equation $A\mathbf{x}_1 = \mathbf{b}_1$. We form the augmented matrix $\left[\begin{array}{c|c} A & \mathbf{b}_1 \end{array}\right]$ and we row reduce to obtain its reduced row echelon form:

$$\left[\begin{array}{c|c} A & \mathbf{b}_1 \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 5 \\ -1 & 3 & 1 & -2 & -3 \\ 0 & 1 & 0 & 3 & 3 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & \frac{31}{4} & \frac{35}{4} \\ 0 & 1 & 0 & 3 & 3 \\ 0 & 0 & 1 & -\frac{13}{4} & -\frac{13}{4} \end{array}\right].$$

We now read off the solutions for $\mathbf{x}_1$:

$$\mathbf{x}_1 = \left[\begin{array}{c} -\frac{31}{4}s + \frac{35}{4} \\ -3s + 3 \\ \frac{13}{4}s - \frac{13}{4} \\ s \end{array}\right], \quad \text{where } s \in \mathbb{R}.$$

We now solve the matrix-vector equation $A\mathbf{x}_2 = \mathbf{b}_2$. We form the augmented matrix $\left[\begin{array}{c|c} A & \mathbf{b}_2 \end{array}\right]$ and we row reduce to obtain its reduced row echelon form:

$$\left[\begin{array}{c|c} A & \mathbf{b}_2 \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 3 \\ -1 & 3 & 1 & -2 & 1 \\ 0 & 1 & 0 & 3 & 0 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & \frac{31}{4} & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & -\frac{13}{4} & 1 \end{array}\right].$$

We now read off the solutions for $\mathbf{x}_2$:

$$\mathbf{x}_2 = \left[\begin{array}{c} -\frac{31}{4}t \\ -3t \\ \frac{13}{4}t + 1 \\ t \end{array}\right], \quad \text{where } t \in \mathbb{R}.$$

We now read off the general solution for $X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}$:

$$X = \begin{bmatrix} -\frac{31}{4}s + \frac{35}{4} & -\frac{31}{4}t \\ -3s + 3 & -3t \\ \frac{13}{4}s - \frac{13}{4} & \frac{13}{4}t + 1 \\ s & t \end{bmatrix}, \quad \text{where } s, t \in \mathbb{R}.$$

There are two parameters (namely, s and t), and they can each take infinitely many values (because $\mathbb{R}$ is infinite). So, the equation $AX = B$ has infinitely many solutions.

Remark: Note that the parameters (namely, s and t) from the solution above are different for different columns! This is because the equations $A\mathbf{x}_1 = \mathbf{b}_1$ and $A\mathbf{x}_2 = \mathbf{b}_2$ are solved independently, and so the parameter that appears in $\mathbf{x}_1$ is independent of the one that appears in $\mathbf{x}_2$. □

Remark: Solution #1 is correct, but rather inefficient. We had to solve a separate matrix-vector equation for each column of B,[40] and each of these matrix-vector equations involved forming an augmented matrix and finding its reduced row echelon form. Luckily, we can do better by essentially solving these two matrix-vector equations simultaneously.

Solution #2. We first form the matrix $\left[\begin{array}{c|c} A & B \end{array}\right]$ and row reduce to find its reduced row echelon form.

$$\left[\begin{array}{c|c} A & B \end{array}\right] = \left[\begin{array}{cccc|cc} 1 & 2 & 3 & 4 & 5 & 3 \\ -1 & 3 & 1 & -2 & -3 & 1 \\ 0 & 1 & 0 & 3 & 3 & 0 \end{array}\right]$$

After row reducing, we obtain the following matrix (the columns to the right of the vertical dotted line are color coded for easier reference):

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A & B \end{array}\right]\right) = \left[\begin{array}{ccc|cc} 1 & 0 & 0 & \frac{31}{4} & \frac{35}{4} & 0 \\ 0 & 1 & 0 & 3 & 3 & 0 \\ 0 & 0 & 1 & -\frac{13}{4} & -\frac{13}{4} & 1 \end{array}\right].$$

We now read off the columns of X one by one. We read off the first column of X by reading off the solutions of the matrix-vector equation encoded by the matrix obtained by taking the submatrix to the left of the vertical dotted line, plus the first column

[40] Since B has two columns, this translated into two matrix-vector equations. In general, if B has m columns, we get m matrix-vector equations.

to the right of the vertical dotted line (i.e. the red column) of $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & B \end{array}\right]\right)$.[41] We read off the second column of X by reading off the solutions of the matrix-vector equation encoded by the matrix obtained by taking the submatrix to the left of the vertical dotted line, plus the second column to the right of the vertical dotted line (i.e. the blue column) of $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & B \end{array}\right]\right)$.[42] The solutions are as follows:[43]

$$X = \begin{bmatrix} -\frac{31}{4}s + \frac{35}{4} & -\frac{31}{4}t \\ -3s + 3 & -3t \\ \frac{13}{4}s - \frac{13}{4} & \frac{13}{4}t + 1 \\ s & t \end{bmatrix}, \quad \text{where } s, t \in \mathbb{R}.$$

There are two parameters (namely, s and t), and they can each take infinitely many values (because $\mathbb{R}$ is infinite). So, the equation $AX = B$ has infinitely many solutions. □

Recipe for solving matrix equations of the form $AX = B$. Suppose that A is an $n \times m$ matrix and B is an $n \times p$ matrix (both with entries in some field $\mathbb{F}$), and we wish to solve the matrix equation $AX = B$.[44] We proceed as follows:

1. We form the $n \times (m+p)$ matrix $\left[\begin{array}{c|c} A & B \end{array}\right]$ and find its reduced row echelon form.
2. We check if $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & B \end{array}\right]\right)$ has a row of the form

$$\left[\begin{array}{ccc|ccc} 0 & \dots & 0 & * & \dots & * \end{array}\right],$$

where at least one of the $*$'s (to the right of the vertical dotted line) is non-zero.

[41] This is the matrix in question:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{31}{4} & \frac{35}{4} \\ 0 & 1 & 0 & 3 & 3 \\ 0 & 0 & 1 & -\frac{13}{4} & -\frac{13}{4} \end{array}\right].$$

[42] This is the matrix in question:

$$\left[\begin{array}{ccc|cc} 1 & 0 & 0 & \frac{31}{4} & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & -\frac{13}{4} & 1 \end{array}\right].$$

[43] Remember to use different parameters for different columns!

[44] Note that solutions of the matrix equation $AX = B$ are $m \times p$ matrices.

(a) If such a row exists, then the matrix equation $AX = B$ is inconsistent (i.e. has no solutions).[45]

(b) If no such row exits, then the matrix equation $AX = B$ is consistent (i.e. has at least one solution). For each $k \in \{1, \dots, p\}$,[46] we read off the k-th column of X by focusing on the part of $\mathrm{RREF}\big(\left[\begin{array}{c:c} A & B \end{array}\right]\big)$ to the left of the vertical dotted line, plus the k-th column of $\mathrm{RREF}\big(\left[\begin{array}{c:c} A & B \end{array}\right]\big)$ to the right of the vertical dotted line.

- If there are any free variables, remember to use different letters for the parameters in different columns, as in the solution of Example 1.9.1.

Example 1.9.2. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & 2 & -1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 3 & 1 & 3 \\ 4 & 3 & 1 & 3 \\ 2 & 1 & 1 & 3 \\ 2 & 1 & 2 & 3 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Solve the matrix equation $AX = B$.[47] *How many solutions does the equation $AX = B$ have?*

Solution. We first form the matrix

$$\left[\begin{array}{c:c} A & B \end{array}\right] = \left[\begin{array}{ccc:cccc} 1 & 1 & 1 & 4 & 3 & 1 & 3 \\ 1 & 1 & 1 & 4 & 3 & 1 & 3 \\ 1 & 2 & -1 & 2 & 1 & 1 & 3 \\ 1 & 2 & -1 & 2 & 1 & 2 & 3 \end{array}\right].$$

After row reducing, we obtain

$$\mathrm{RREF}\Big(\left[\begin{array}{c:c} A & B \end{array}\right]\Big) = \left[\begin{array}{ccc:cccc} 1 & 0 & 3 & 6 & 5 & 0 & 3 \\ 0 & 1 & -2 & -2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

By considering the third row of $\mathrm{RREF}\big(\left[\begin{array}{c:c} A & B \end{array}\right]\big)$, we see that the matrix equation $AX = B$ is inconsistent, i.e. it has no solutions. □

Example 1.9.3. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

[45] Indeed, suppose we got a row of the form $\left[\begin{array}{c:ccc} 0 & \dots & 0 & * & \dots & * \end{array}\right]$, where the k-th $*$ to the right of the vertical dotted line is non-zero. Then there are no solutions for the k-th column of the matrix X, and therefore, there are no solutions for the matrix X, either.

[46] Remember: p is the number of columns of B, and therefore, the number of columns of X.

[47] Note that solutions of the matrix equation $AX = B$ are 3×4 real matrices.

with entries understood to be in $\mathbb{Z}_2$. Solve the matrix equation $AX = B$. How many solutions does the equation $AX = B$ have?

Solution. We first form the matrix

$$\left[\begin{array}{c|c} A & B \end{array}\right] = \left[\begin{array}{cccc|ccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{array}\right].$$

After row reducing, we obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A & B \end{array}\right]\right) = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{array}\right].$$

We now read off the solutions:

$$X = \left[\begin{array}{ccc} t_1+1 & t_2+1 & t_3 \\ t_1+1 & t_2 & t_3+1 \\ t_1+1 & t_2 & t_3 \\ t_1 & t_2 & t_3 \end{array}\right], \quad \text{where } t_1, t_2, t_3 \in \mathbb{Z}_2.$$

There are three parameters (namely, t_1, t_2, t_3), and each of them can take two values (because $|\mathbb{Z}_2| = 2$). So, the total number of solutions of the equation $AX = B$ is $2^3 = 8$. □

Example 1.9.4. *Consider the matrices*

$$A = \left[\begin{array}{ccc} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 2 \end{array}\right] \quad \textit{and} \quad B = \left[\begin{array}{cccc} 0 & 1 & 1 & 2 \\ 2 & 1 & 1 & 0 \\ 1 & 2 & 2 & 2 \end{array}\right],$$

with entries understood to be in $\mathbb{Z}_3$. Solve the matrix equation $AX = B$. How many solutions does the equation $AX = B$ have?

Solution. We first form the matrix

$$\left[\begin{array}{c|c} A & B \end{array}\right] = \left[\begin{array}{ccc|cccc} 1 & 1 & 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 2 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 2 & 2 & 2 \end{array}\right].$$

After row reducing, we obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A & B \end{array}\right]\right) = \left[\begin{array}{ccc|cccc} 1 & 0 & 0 & 2 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{array}\right].$$

We now see that the equation $AX = B$ has a unique solution, namely,

$$X = \left[\begin{array}{cccc} 2 & 0 & 0 & 2 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right].$$

(The number of solutions of the matrix equation $AX = B$ is one.) □

1.9.2 Solving matrix equations of the form $XA = B$

Suppose we are asked to solve a matrix equation of the form $XA = B$. This equation is equivalent to the equation $(XA)^T = B^T$, which is, in turn, equivalent to $A^T X^T = B^T$.[48] Using the methods from the previous section, we solve the equation $A^T X^T = B^T$ for X^T, and then we take the transpose of the solution(s) to obtain X.

Example 1.9.5. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 3 & 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 5 & 1 & -2 \\ 4 & 3 & 1 & -1 \\ 2 & 4 & 0 & -2 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Solve the matrix equation $XA = B$.[49] *How many solutions does the equation $XA = B$ have?*

Solution. Note that $XA = B$ if and only if $A^T X^T = B^T$. We first find all the matrices X^T that satisfy $A^T X^T = B^T$, and then we take the transpose to obtain all the matrices X that satisfy $XA = B$. First, we have

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad B^T = \begin{bmatrix} 5 & 4 & 2 \\ 5 & 3 & 4 \\ 1 & 1 & 0 \\ -2 & -1 & -2 \end{bmatrix}.$$

We now form the matrix

$$\left[\begin{array}{c|c} A^T & B^T \end{array}\right] = \left[\begin{array}{cc|ccc} 1 & 3 & 5 & 4 & 2 \\ 2 & 1 & 5 & 3 & 4 \\ 0 & 1 & 1 & 1 & 0 \\ -1 & 0 & -2 & -1 & -2 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A^T & B^T \end{array}\right]\right) = \left[\begin{array}{cc|ccc} 1 & 0 & 2 & 1 & 2 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Using the matrix above, we can solve for X^T. There is only one solution, namely:

$$X^T = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 0 \end{bmatrix}.$$

[48] We are using Proposition 1.8.1(d).

[49] Note that solutions of the matrix equation $XA = B$ are 3×2 real matrices.

Thus, the equation $XA = B$ has a unique solution, namely:

$$X = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}.$$

(The number of solutions of the matrix equation $XA = B$ is one.) □

Example 1.9.6. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_2$. Solve the matrix equation $XA = B$.[50] *How many solutions does the equation $XA = B$ have?*

Solution. Note that $XA = B$ if and only if $A^T X^T = B^T$. We first find all the matrices X^T that satisfy $A^T X^T = B^T$, and then we take the transpose to obtain all the matrices X that satisfy $XA = B$. First, we have

$$A^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad B^T = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

We now form the matrix

$$\left[\begin{array}{c|c} A^T & B^T \end{array}\right] = \left[\begin{array}{ccc|cc} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{array}\right],$$

and by row reducing, we obtain:

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A^T & B^T \end{array}\right]\right) = \left[\begin{array}{ccc|cc} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

By considering the third row of $\mathrm{RREF}\left(\left[\begin{array}{c|c} A^T & B^T \end{array}\right]\right)$, we see that the equation $A^T X^T = B^T$ has no solutions. Consequently, the original matrix equation $XA = B$ has no solutions either (i.e. the number of solutions of $XA = B$ is zero). □

[50] Note that solutions of the matrix equation $XA = B$ are 2×3 matrices with entries in $\mathbb{Z}_2$.

Example 1.9.7. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 \\ 3 & 0 & 3 & 0 & 3 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 2 & 3 & 2 & 3 & 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_5$. Solve the matrix equation $XA = B$.[51] *How many solutions does the equation $XA = B$ have?*

Proof. Note that $XA = B$ if and only if $A^T X^T = B^T$. We first find all the matrices X^T that satisfy $A^T X^T = B^T$, and then we take the transpose to obtain all the matrices X that satisfy $XA = B$. First, we have

$$A^T = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 2 & 0 \\ 1 & 1 & 3 \\ 1 & 2 & 0 \\ 1 & 1 & 3 \end{bmatrix} \quad \text{and} \quad B^T = \begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 1 & 2 \\ 0 & 3 \\ 1 & 2 \end{bmatrix}.$$

We now form the matrix

$$\left[\begin{array}{c|c} A^T & B^T \end{array}\right] = \left[\begin{array}{ccc|cc} 1 & 1 & 3 & 1 & 2 \\ 1 & 2 & 0 & 0 & 3 \\ 1 & 1 & 3 & 1 & 2 \\ 1 & 2 & 0 & 0 & 3 \\ 1 & 1 & 3 & 1 & 2 \end{array}\right],$$

and by row reducing, we obtain:

$$\operatorname{RREF}\left(\left[\begin{array}{c|c} A^T & B^T \end{array}\right]\right) = \left[\begin{array}{ccc|cc} 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 4 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Using the matrix above, we can solve for X^T:

$$X^T = \begin{bmatrix} 4s+2 & 4t+1 \\ 3s+4 & 3t+1 \\ s & t \end{bmatrix}, \quad \text{where } s, t \in \mathbb{Z}_5.$$

Thus, the general solution of the matrix equation $XA = B$ is

$$X = \begin{bmatrix} 4s+2 & 3s+4 & s \\ 4t+1 & 3t+1 & t \end{bmatrix}, \quad \text{where } s, t \in \mathbb{Z}_5.$$

Since we have two parameters (namely, s and t), each of which can take five values (because $|\mathbb{Z}_5| = 5$), we see that the matrix equation $XA = B$ has $5^2 = 25$ solutions. □

[51] Note that solutions of the matrix equation $XA = B$ are 2×3 matrices with entries in $\mathbb{Z}_5$.

1.10 A first look at linear functions and their matrices

In this section, we introduce "linear functions" from $\mathbb{F}^m$ to $\mathbb{F}^n$ (where $\mathbb{F}$ is some field). In chapter 4, we will study linear functions in a more general setting.

1.10.1 Linear functions: definition and examples

For a field $\mathbb{F}$, a function $f : \mathbb{F}^m \to \mathbb{F}^n$ is said to be a *linear function* (or a *linear transformation*) if it satisfies the following two conditions (axioms):

1. for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{F}^m$, we have that $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$;
2. for all vectors $\mathbf{u} \in \mathbb{F}^m$ and scalars $\alpha \in \mathbb{F}$, we have that $f(\alpha\mathbf{u}) = \alpha f(\mathbf{u})$.

Proposition 1.10.1. *Let $\mathbb{F}$ be a field, and let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function. Then for all vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}^m$ and all scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$, we have that*

$$f\Big(\sum_{i=1}^{k} \alpha_i \mathbf{v}_i\Big) \;=\; \sum_{i=1}^{k} \alpha_i f(\mathbf{v}_i),$$

or, written in another way, that

$$f\Big(\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k\Big) \;=\; \alpha_1 f(\mathbf{v}_1) + \cdots + \alpha_k f(\mathbf{v}_k).$$

Proof. This follows from the definition of a linear function via an easy induction on k. The details are left as an exercise. □

Example 1.10.2. *Determine whether the following functions are linear (and prove your answer):*

(a) the function $f : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$f\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\Big) \;=\; \begin{bmatrix} x_1 - x_2 + x_3 \\ x_1 + x_2 \end{bmatrix}$$

for all $x_1, x_2, x_3 \in \mathbb{R}$.

(b) the function $g : \mathbb{Z}_2^2 \to \mathbb{Z}_2^4$ given by

$$g\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) \;=\; \begin{bmatrix} x_1 \\ x_1 + x_2 \\ x_2 \\ 1 \end{bmatrix}$$

for all $x_1, x_2 \in \mathbb{Z}_2$.

(c) The function $h : \mathbb{Z}_3^3 \to \mathbb{Z}_3^2$ *given by*

$$h\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\Big) = \begin{bmatrix} x_1 + x_2 \\ x_1 x_2 \end{bmatrix}$$

for all $x_1, x_2, x_3 \in \mathbb{Z}_3$.

Remark: To show that a function is linear, we must show that it satisfies both axioms from the definition of a linear function; in particular, axiom 1 must hold for **all** vectors **u** and **v**, and axiom 2 must hold for **all** vectors **u** and scalars α. On the other hand, to show that a function is **not** linear, it is enough to show that it fails to satisfy at least one of the axioms 1 and 2 from the definition of a linear function. To show that a function does **not** satisfy axiom 1, it is enough to exhibit **one particular pair of vectors u** and **v** for which that axiom does not hold. Similarly, to show that a function does **not** satisfy axiom 2, it is enough to exhibit **one particular vector u** and **one particular scalar** α for which axiom 2 fails.

Solution of Example 1.10.2. (a) The function f is linear. We prove this by verifying the axioms of a linear function for the function f, as follows.

1. Fix vectors $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ in $\mathbb{R}^3$. We must show that $f(\mathbf{u}+\mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$. For this, we compute:

$$\begin{aligned}
f(\mathbf{u}+\mathbf{v}) &= f\Big(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\Big) = f\Big(\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix}\Big) \\
&\overset{(*)}{=} \begin{bmatrix} (u_1 + v_1) - (u_2 + v_2) + (u_3 + v_3) \\ (u_1 + v_1) + (u_2 + v_2) \end{bmatrix} \\
&= \begin{bmatrix} (u_1 - u_2 + u_3) + (v_1 - v_2 + v_3) \\ (u_1 + u_2) + (v_1 + v_2) \end{bmatrix} \\
&= \begin{bmatrix} u_1 - u_2 + u_3 \\ u_1 + u_2 \end{bmatrix} + \begin{bmatrix} v_1 - v_2 + v_3 \\ v_1 + v_2 \end{bmatrix} \\
&\overset{(**)}{=} f\Big(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\Big) + f\Big(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\Big) \\
&= f(\mathbf{u}) + f(\mathbf{v}),
\end{aligned}$$

where both (*) and (**) follow from the definition of f.

2. Fix a vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ in $\mathbb{R}^3$ and a scalar $\alpha \in \mathbb{R}$. We must show that $f(\alpha\mathbf{u}) = \alpha f(\mathbf{u})$. For this, we compute

$$\begin{aligned} f(\alpha\mathbf{u}) &= f\Big(\alpha \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\Big) = f\Big(\begin{bmatrix} \alpha u_1 \\ \alpha u_2 \\ \alpha u_3 \end{bmatrix}\Big) \\ &\overset{(*)}{=} \begin{bmatrix} \alpha u_1 - \alpha u_2 + \alpha u_3 \\ \alpha u_1 + \alpha u_2 \end{bmatrix} \\ &= \begin{bmatrix} \alpha(u_1 - u_2 + u_3) \\ \alpha(u_1 + u_2) \end{bmatrix} \\ &= \alpha \begin{bmatrix} u_1 - u_2 + u_3 \\ u_1 + u_2 \end{bmatrix} \\ &\overset{(**)}{=} \alpha f\Big(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\Big) \\ &= \alpha f(\mathbf{u}), \end{aligned}$$

where both (*) and (**) follow from the definition of f.

We have now shown that f satisfies both axioms from the definition of a linear function. So, f is linear, as we had claimed.

(b) The function g is **not** linear because it does not satisfy axiom 1 of the definition of a linear function.[52] To see this, we consider, for example, the vectors $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in $\mathbb{Z}_2^2$, and we observe that

$$g(\mathbf{u}+\mathbf{v}) = g\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big) = g\Big(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\Big) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

whereas

[52] In fact, g also fails to satisfy axiom 2 (details?). However, to show that g is not linear, it is enough to show that it fails to satisfy **at least one** of the two axioms.

$$g(\mathbf{u})+g(\mathbf{v}) \;=\; g\Big(\begin{bmatrix}1\\1\end{bmatrix}\Big)+g\Big(\begin{bmatrix}1\\1\end{bmatrix}\Big) \;=\; \begin{bmatrix}1\\0\\1\\1\end{bmatrix}+\begin{bmatrix}1\\0\\1\\1\end{bmatrix} \;=\; \begin{bmatrix}0\\0\\0\\0\end{bmatrix}.$$

As we can see, $g(\mathbf{u}+\mathbf{v}) \neq g(\mathbf{u})+g(\mathbf{v})$, and we deduce that g is not linear.

(c) The function h is **not** linear because it does not satisfy axiom 2 of the definition of a linear function. To see this, we consider, for example, the vector $\mathbf{u} = \begin{bmatrix}1\\2\\0\end{bmatrix}$ in $\mathbb{Z}_3^3$ and the scalar $\alpha = 2$ in $\mathbb{Z}_3$, and we observe that

- $h(\alpha\mathbf{u}) = h\Big(2\begin{bmatrix}1\\2\\0\end{bmatrix}\Big) = h\Big(\begin{bmatrix}2\\1\\0\end{bmatrix}\Big) = \begin{bmatrix}2+1\\2\cdot 1\end{bmatrix} = \begin{bmatrix}0\\2\end{bmatrix}$;
- $\alpha h(\mathbf{u}) = 2h\Big(\begin{bmatrix}1\\2\\0\end{bmatrix}\Big) = 2\begin{bmatrix}1+2\\1\cdot 2\end{bmatrix} = 2\begin{bmatrix}0\\2\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}$.

As we can see, $h(\alpha\mathbf{u}) \neq \alpha h(\mathbf{u})$, and we deduce that h is not linear. □

Proposition 1.10.3. *Let $\mathbb{F}$ be a field, and let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function. Then $f(\mathbf{0}) = \mathbf{0}$.*[53]

Proof. We observe that

$$f(\mathbf{0}) \;=\; f(0\cdot\mathbf{0}) \;\overset{(*)}{=}\; 0f(\mathbf{0}) \;=\; \mathbf{0},$$

where (*) follows from the fact that f is linear.[54] □

Remark: Proposition 1.10.3 can sometimes be used to show that a function is not linear. For example, for the function g from Example 1.10.2(b), we have that $g(\mathbf{0}) \neq \mathbf{0}$, and so g is not linear. However, note that the converse of Proposition 1.10.3 fails: it is possible that a function $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is some field) satisfies $f(\mathbf{0}) = \mathbf{0}$, but that the function f is still not linear. For instance, the function h from Example 1.10.2(c) satisfies $h(\mathbf{0}) = \mathbf{0}$, but h is nevertheless not linear.

[53] Note that in $f(\mathbf{0}) = \mathbf{0}$, we have that $\mathbf{0} \in \mathbb{F}^m$, whereas $\mathbf{0} \in \mathbb{F}^n$. So, the two zero vectors aren't actually the same (unless $m = n$). Furthermore, 0 (from the proof of Proposition 1.10.3) is the zero element of the field $\mathbb{F}$.

[54] In particular, we are using axiom 2 of the definition of a linear function.

1.10.2 The images of lines under linear functions $f : \mathbb{R}^m \to \mathbb{R}^n$

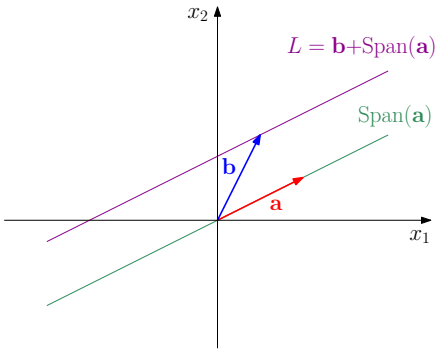
Suppose that $f : \mathbb{R}^m \to \mathbb{R}^n$ is a linear function. It turns out that the image of any line in $\mathbb{R}^m$ under f is either a line in $\mathbb{R}^n$ or a point in $\mathbb{R}^n$ (technically, a set that contains only one point/vector of $\mathbb{R}^n$; we can think of such one-point sets as "degenerate lines"). This is one of the reasons why linear functions are called linear. Let us give a formal proof. Lines through the origin in $\mathbb{R}^m$ are simply sets of the form

$$\text{Span}(\mathbf{a}) \quad = \quad \{\alpha\mathbf{a} \mid \alpha \in \mathbb{R}\},$$

where $\mathbf{a}$ is a non-zero vector in $\mathbb{R}^m$. Any line in $\mathbb{R}^m$ is obtained by shifting a line through the origin by some vector $\mathbf{b} \in \mathbb{R}^m$ (if $\mathbf{b} = \mathbf{0}$, then our line still passes though the origin). So, consider some line

$$L \quad := \quad \mathbf{b} + \text{Span}(\mathbf{a}) \quad = \quad \{\mathbf{b} + \alpha\mathbf{a} \mid \alpha \in \mathbb{R}\},$$

where $\mathbf{a} \neq \mathbf{0}$ and $\mathbf{b}$ are fixed vectors in $\mathbb{R}^m$ (this is illustrated below for the special case of $\mathbb{R}^2$).



For any point $\mathbf{b} + \alpha\mathbf{a}$ ($\alpha \in \mathbb{R}$) on the line L, we have that

$$f(\mathbf{b} + \alpha\mathbf{a}) \quad \overset{(*)}{=} \quad f(\mathbf{b}) + f(\alpha\mathbf{a}) \quad \overset{(**)}{=} \quad f(\mathbf{b}) + \alpha f(\mathbf{a})$$

where both (*) and (**) follow from the linearity of f, but in (*) we used axiom 1 from the definition of a linear function, and in (**) we used axiom 2. So, the image of our line L under f, denoted by $f[L]$, is

$$f[L] \quad = \quad \{f(\mathbf{b}) + \alpha f(\mathbf{a}) \mid \alpha \in \mathbb{R}\} \quad = \quad f(\mathbf{b}) + \text{Span}\Big(f(\mathbf{a})\Big).$$

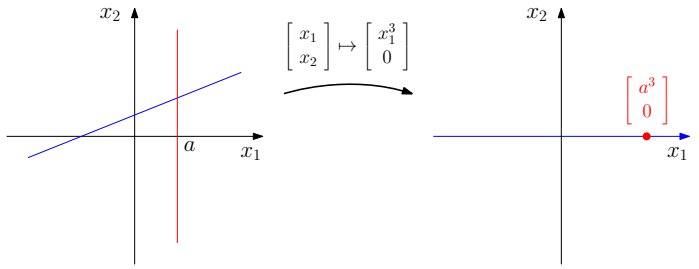
If $f(\mathbf{a}) \neq \mathbf{0}$, then $f[L]$ is a line in $\mathbb{R}^n$. On the other hand, if $f(\mathbf{a}) = \mathbf{0}$, then $f[L] = \{f(\mathbf{b})\}$, which is a one-point subset ("degenerate line") of $\mathbb{R}^n$.

We also remark that linear functions $f : \mathbb{R}^m \to \mathbb{R}^n$ map line segments onto line segments (possibly degenerate ones, i.e. those that contain only one point). The proof is similar to the above and is left as an exercise.

We note, however, that not all functions $f : \mathbb{R}^m \to \mathbb{R}^n$ that map lines to lines (or points) are linear. An obvious example might be a function $f : \mathbb{R}^n \to \mathbb{R}^n$ given by $f(\mathbf{x}) = \mathbf{x} + \mathbf{b}$ for all $\mathbf{x} \in \mathbb{R}^n$, where $\mathbf{b}$ is a fixed non-zero vector in $\mathbb{R}^n$. This function is not linear because $f(\mathbf{0}) \neq \mathbf{0}$, and we know (by Proposition 1.10.3) that all linear functions map $\mathbf{0}$ to $\mathbf{0}$. However, even if a function $f : \mathbb{R}^m \to \mathbb{R}^n$ maps lines to lines (or points) and maps $\mathbf{0}$ to $\mathbf{0}$, it might still fail to be linear. For example, consider the function $g : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$g\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) \;=\; \begin{bmatrix} x_1^3 \\ 0 \end{bmatrix} \qquad \text{for all } x_1, x_2 \in \mathbb{R}.$$

This function is not linear,[55] although it does map all lines onto either lines or points, and it does map $\mathbf{0}$ to $\mathbf{0}$. In particular, g maps any non-vertical line in $\mathbb{R}^2$ onto the x_1-axis, and it maps any vertical line onto a one-point set, as illustrated in the picture below.



1.10.3 Matrix transformations. The standard matrix of a linear function

Proposition 1.10.4. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times m}$ be a matrix, and define $f : \mathbb{F}^m \to \mathbb{F}^n$ by setting $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$. Then f is a linear function.*

Proof. By Corollary 1.7.6, the following hold:

(i) for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{F}^m$, we have $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$;

(ii) for all vectors $\mathbf{u} \in \mathbb{F}^m$ and scalars $\alpha \in \mathbb{F}$, we have that $A(\alpha\mathbf{u}) = \alpha(A\mathbf{u})$.

But now we have the following:

1. for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{F}^m$, we have that

$$f(\mathbf{u}+\mathbf{v}) \;=\; A(\mathbf{u}+\mathbf{v}) \;\overset{(i)}{=}\; A\mathbf{u} + A\mathbf{v} \;=\; f(\mathbf{u}) + f(\mathbf{v});$$

[55] This is "obvious," but here is a formal proof:

$$g\Big(2\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big) \;=\; g\Big(\begin{bmatrix} 2 \\ 2 \end{bmatrix}\Big) \;=\; \begin{bmatrix} 8 \\ 0 \end{bmatrix} \;\neq\; \begin{bmatrix} 2 \\ 0 \end{bmatrix} \;=\; 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} \;=\; 2g\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big),$$

and so g does not satisfy axiom 2 from the definition of a linear function, and it follows that g is not linear.

2. for all vectors $\mathbf{u} \in \mathbb{F}^m$ and scalars $\alpha \in \mathbb{F}$, we have that

$$f(\alpha\mathbf{u}) \;=\; A(\alpha\mathbf{u}) \;\overset{(ii)}{=}\; \alpha(A\mathbf{u}) \;=\; \alpha f(\mathbf{u}).$$

So, f is linear. □

Mappings of the form $\mathbf{x} \mapsto A\mathbf{x}$, where A is some matrix, are sometimes called *matrix transformations*. By Proposition 1.10.4, all matrix transformations are linear. Let us try to describe matrix transformations in a bit more detail. Suppose we are given a matrix $A = \begin{bmatrix} a_{i,j} \end{bmatrix}_{n\times m}$ in $\mathbb{F}^{n\times m}$ (where $\mathbb{F}$ is some field), and define the function $f : \mathbb{F}^m \to \mathbb{F}^n$ by setting $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$. But now for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_m \end{bmatrix}^T$ in $\mathbb{F}^m$, we have the following:

$$f(\mathbf{x}) \;=\; A\mathbf{x} \;=\; \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$= \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,m}x_m \\ a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,m}x_m \\ \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \dots + a_{n,m}x_m \end{bmatrix}.$$

So, our matrix transformation maps each vector $\mathbf{x} \in \mathbb{F}^m$ to a vector in $\mathbb{F}^n$, each of whose entries is a linear combination of the entries of $\mathbf{x}$, and the scalars/weights are determined by the corresponding row of the matrix A. Note that the function f from Example 1.10.2 has this form, whereas the functions g and h from the same example do not.

By Proposition 1.10.4, every matrix transformation is a linear function. Interestingly, a converse of sorts also holds (see Theorem 1.10.6 below). We begin with another important theorem, which readily implies Theorem 1.10.6.

Theorem 1.10.5. *Let $\mathbb{F}$ be a field, and let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be any vectors in $\mathbb{F}^n$. Then there exists a **unique** linear function $f : \mathbb{F}^m \to \mathbb{F}^n$ that satisfies $f(\mathbf{e}_1) = \mathbf{a}_1, \dots, f(\mathbf{e}_m) = \mathbf{a}_m$, where $\mathbf{e}_1, \dots, \mathbf{e}_m$ are the standard basis vectors of $\mathbb{F}^m$. Moreover, this linear function f is given by $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$, where $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$.*

Proof. **Existence.** Define $f : \mathbb{F}^m \to \mathbb{F}^n$ by setting $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$. Then f is a matrix transformation, and so by Proposition 1.10.4, it is linear. Moreover, for all indices $i \in \{1, \dots, m\}$, we have that

$$f(\mathbf{e}_i) \;=\; A\mathbf{e}_i \;=\; \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix} \mathbf{e}_i \;\overset{(*)}{=}\; \mathbf{a}_i,$$

where (*) follows from Proposition 1.4.4.

Uniqueness. Suppose that $f : \mathbb{F}^m \to \mathbb{F}^n$ is **any** linear function that satisfies $f(\mathbf{e}_1) = \mathbf{a}_1, \ldots, f(\mathbf{e}_m) = \mathbf{a}_m$. We must show that $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$. Fix any vector $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_m \end{bmatrix}^T$ in $\mathbb{F}^m$. Then we have that $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_m\mathbf{e}_m$, and we compute:

$$\begin{aligned} f(\mathbf{x}) &= f\Big(x_1\mathbf{e}_1 + \cdots + x_m\mathbf{e}_m\Big) \\ &\overset{(*)}{=} x_1 f(\mathbf{e}_1) + \cdots + x_m f(\mathbf{e}_m) \\ &\overset{(**)}{=} x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m \\ &\overset{(***)}{=} \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \\ &= A\mathbf{x}, \end{aligned}$$

where f follows from the linearity of f (and more precisely, from Proposition 1.10.1), (**) follows from the fact that $f(\mathbf{e}_1) = \mathbf{a}_1, \ldots, f(\mathbf{e}_m) = \mathbf{a}_m$, and (***) follows from the definition of matrix-vector multiplication. □

Remark: Theorem 1.10.5 essentially states that we can fully determine a linear function $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field) by simply specifying what the standard basis vectors of $\mathbb{F}^m$ get mapped to. Moreover, we can choose what the standard basis vectors get mapped to arbitrarily (i.e. we can map them to any vectors of $\mathbb{F}^n$ that we like).

As a corollary of Theorem 1.10.5, we obtain the following theorem, which essentially states that all linear functions $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field) are in fact matrix transformations.

Theorem 1.10.6. *Let $\mathbb{F}$ be a field, and let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function. Then there exists a unique matrix A (called the* standard matrix of f*) such that for all $\mathbf{x} \in \mathbb{F}^m$, we have that $f(\mathbf{x}) = A\mathbf{x}$. Moreover, the standard matrix A of f is given by*

$$A = \begin{bmatrix} f(\mathbf{e}_1) & \ldots & f(\mathbf{e}_m) \end{bmatrix},$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_m$ are the standard basis vectors of $\mathbb{F}^m$.

Proof. **Existence.** Set $\mathbf{a}_1 := f(\mathbf{e}_1), \ldots, \mathbf{a}_m := f(\mathbf{e}_m)$ and

$$A := \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} f(\mathbf{e}_1) & \ldots & f(\mathbf{e}_m) \end{bmatrix}.$$

Then by Theorem 1.10.5, we have that $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$.[56] This proves existence.

Uniqueness. Let $B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_m \end{bmatrix}$ be **any** matrix in $\mathbb{F}^{n\times m}$ such that $f(\mathbf{x}) = B\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$. Then for all $i \in \{1, \dots, m\}$, we have that

$$f(\mathbf{e}_i) \;=\; B\mathbf{e}_i \;\overset{(*)}{=}\; \mathbf{b}_i$$

where (*) follows from Proposition 1.4.4. Consequently,

$$B \;=\; \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_m \end{bmatrix} \;=\; \begin{bmatrix} f(\mathbf{e}_1) & \dots & f(\mathbf{e}_m) \end{bmatrix}.$$

This proves uniqueness. □

Example 1.10.7. *Find the standard matrix of the linear function $f : \mathbb{R}^3 \to \mathbb{R}^2$ given by*

$$f\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\Big) \;=\; \begin{bmatrix} x_1 - x_2 + x_3 \\ x_1 + x_2 \end{bmatrix}$$

for all $x_1, x_2, x_3 \in \mathbb{R}$. (The fact that f is linear was proven in the solution of Example 1.10.2(a).)

Solution. The standard matrix of f is

$$A \;:=\; \begin{bmatrix} f(\mathbf{e}_1) & f(\mathbf{e}_2) & f(\mathbf{e}_3) \end{bmatrix} \;=\; \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Remark: Note that for all vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in $\mathbb{R}^3$, we really do have

$$A\mathbf{x} \;=\; \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \;\overset{(*)}{=}\; \begin{bmatrix} x_1 - x_2 + x_3 \\ x_1 + x_2 \end{bmatrix} \;=\; f(\mathbf{x}),$$

where (*) was obtained by matrix-vector multiplication. □

For any set X, the *identity function* on X is the function $\mathrm{Id}_X : X \to X$ given by $\mathrm{Id}_X(x) = x$ for all $x \in X$. The following proposition is obvious, but useful to keep in mind.

Proposition 1.10.8. *Let $\mathbb{F}$ be a field. Then the identity function $Id_{\mathbb{F}^n} : \mathbb{F}^n \to \mathbb{F}^n$ is linear, and its standard matrix is the identity matrix I_n.*

[56] Indeed, $f : \mathbb{F}^m \to \mathbb{F}^n$ is a linear function that satisfies $f(\mathbf{e}_1) = \mathbf{a}_1, \dots, f(\mathbf{e}_m) = \mathbf{a}_m$. So, by Theorem 1.10.5, we have that $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$.

Proof. Obviously, the identity function $\mathrm{Id}_{\mathbb{F}^n}$ satisfies the two axioms from the definition of a linear function, and by Theorem 1.10.6, its standard matrix is

$$\begin{bmatrix} \mathrm{Id}_{\mathbb{F}^n}(\mathbf{e}_1) & \dots & \mathrm{Id}_{\mathbb{F}^n}(\mathbf{e}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \dots & \mathbf{e}_n \end{bmatrix} = I_n.$$

Alternatively, we observe that for any vector $\mathbf{x} \in \mathbb{F}^n$, we have that

$$\mathrm{Id}_{\mathbb{F}^n}(\mathbf{x}) \overset{(*)}{=} \mathbf{x} \overset{(**)}{=} I_n\mathbf{x},$$

where (*) follows from the definition of the identity function, and (**) follows from Proposition 1.4.5. So, $\mathrm{Id}_{\mathbb{F}^n}$ is a matrix transformation and is therefore linear (by Proposition 1.10.4), and its standard matrix is I_n. □

1.10.4 Checking the existence and uniqueness of linear functions with certain specifications

Suppose that $\mathbb{F}$ is a field, and $\mathbf{b}_1, \dots, \mathbf{b}_k \in \mathbb{F}^m$ and $\mathbf{c}_1, \dots, \mathbf{c}_k \in \mathbb{F}^n$ are vectors. How would we determine if there exists a linear function $f : \mathbb{F}^m \to \mathbb{F}^n$ that satisfies $f(\mathbf{b}_i) = f(\mathbf{c}_i)$ for all $i \in \{1, \dots, k\}$? If it is exists, how do we tell if it is unique? Since linear functions $f : \mathbb{F}^m \to \mathbb{F}^n$ are precisely the matrix transformations, this boils down to solving matrix equations of the form $XA = B$, which we studied in section 1.9. We will see this in the examples below, but first, let us make some general remarks.

A linear function with specifications of the sort described above does not always exist. For example, no linear function $f : \mathbb{R}^2 \to \mathbb{R}^2$ satisfies $f\Big(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\Big) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $f\Big(\begin{bmatrix} 2 \\ 0 \end{bmatrix}\Big) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, since any such function would violate axiom 2 of the definition of a linear function.[57] Moreover, if a function of this type does exist, it need not be unique. For example, there is more than one linear function $g : \mathbb{R}^2 \to \mathbb{R}^2$ that satisfies $g(\mathbf{e}_1) = \mathbf{e}_2$. In fact, there are infinitely many such functions, since we can arbitrarily choose the value of $g(\mathbf{e}_2)$, as per Theorem 1.10.5.

Example 1.10.9. *Prove that there exists a unique linear function $f : \mathbb{Z}_2^3 \to \mathbb{Z}_2^5$ that satisfies*

- $f\Big(\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T\Big) = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \end{bmatrix}^T$,

[57] Indeed, if $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear function that satisfies $f\Big(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\Big) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, then it also satisfies

$$f\Big(\begin{bmatrix} 2 \\ 0 \end{bmatrix}\Big) = f\Big(2\begin{bmatrix} 1 \\ 0 \end{bmatrix}\Big) \overset{(*)}{=} 2f\Big(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\Big) = 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix},$$

where (*) follows from the linearity of f.

- $f\Big(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T\Big) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$,
- $f\Big(\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T\Big) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$,

and find its standard matrix.

Solution. To simplify notation, we set

$$\mathbf{b}_1 := \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{b}_2 := \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{b}_3 := \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},$$

$$\mathbf{c}_1 := \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{c}_2 := \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{c}_3 := \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

We need to prove that there exists a unique linear function $f : \mathbb{Z}_2^3 \to \mathbb{Z}_2^5$ that satisfies $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, 2, 3\}$. This is equivalent to proving that there exists a unique matrix $A \in \mathbb{Z}_2^{5\times 3}$ (the standard matrix of f) such that $A\mathbf{b}_i = \mathbf{c}_i$ for all $i \in \{1, 2, 3\}$. So, we are looking for the matrix $A \in \mathbb{Z}_2^{5\times 3}$ that satisfies

$$A\mathbf{b}_1 = \mathbf{c}_1, \qquad A\mathbf{b}_2 = \mathbf{c}_2, \qquad A\mathbf{b}_3 = \mathbf{c}_3.$$

This is equivalent to

$$A\underbrace{\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix}}_{=:B} = \underbrace{\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \end{bmatrix}}_{=:C},$$

in which the matrix A is the unknown (and matrices B and C are known). So, we proceed as in subsection 1.9.2. We take the transpose of both sides of the equation above to obtain $B^TA^T = C^T$, we form the matrix

$$\begin{bmatrix} B^T \,\vdots\, C^T \end{bmatrix} = \left[\begin{array}{c:c} \mathbf{b}_1^T & \mathbf{c}_1^T \\ \mathbf{b}_2^T & \mathbf{c}_2^T \\ \mathbf{b}_3^T & \mathbf{c}_3^T \end{array}\right] = \left[\begin{array}{ccc:ccc} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right],$$

and we row reduce to obtain

$$\operatorname{RREF}\Big(\begin{bmatrix} B^T \,\vdots\, C^T \end{bmatrix}\Big) = \left[\begin{array}{ccc:ccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right].$$

We now see that the equation $B^TA^T = C^T$ has a unique solution for A^T, namely,

$$A^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

So, the equation $AB = C$ has a unique solution for A, namely,

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We have now shown that there exists a unique linear function $f : \mathbb{Z}_2^3 \to \mathbb{Z}_3^5$ that satisfies $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, 2, 3\}$, and that its standard matrix is the matrix A above. (The existence and uniqueness of f follow from the existence and uniqueness of A.)

Remark: Now that we have computed the standard matrix A of f, we can easily compute a formula for f, as follows. For all vectors $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T$ in $\mathbb{Z}_2^3$, we have:

$$f(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \overset{(*)}{=} \begin{bmatrix} x_3 \\ x_3 \\ x_2 \\ x_2 \\ x_3 \end{bmatrix},$$

where (*) was obtained via standard matrix-vector multiplication. □

Example 1.10.10. *Determine if there exists a linear function $f : \mathbb{Z}_3^3 \to \mathbb{Z}_3^2$ that satisfies all the following:*

- $f\Big(\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\Big) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$;
- $f\Big(\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}\Big) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$;
- $f\Big(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\Big) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$;
- $f\Big(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\Big) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

If such a linear function f exists, determine if it is unique.

Solution. To simplify notation, we set

$$\mathbf{b}_1 := \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \qquad \mathbf{b}_2 := \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{b}_3 := \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \qquad \mathbf{b}_4 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

$$\mathbf{c}_1 := \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{c}_2 := \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \mathbf{c}_3 := \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \qquad \mathbf{c}_4 := \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

We need to check whether there exists a linear function $f : \mathbb{Z}_3^3 \to \mathbb{Z}_3^2$ that satisfies $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, 2, 3, 4\}$, and if so, whether it is unique. This is equivalent to determining whether there exists a matrix $A \in \mathbb{Z}_3^{2\times 3}$ (the standard matrix of f) such that $A\mathbf{b}_i = \mathbf{c}_i$ for all $i \in \{1, 2, 3, 4\}$. So, we have a system of four equations (in which the unknown is the matrix A):

$$A\mathbf{b}_1 = \mathbf{c}_1, \qquad A\mathbf{b}_2 = \mathbf{c}_2, \qquad A\mathbf{b}_3 = \mathbf{c}_3, \qquad A\mathbf{b}_4 = \mathbf{c}_4.$$

This is equivalent to the equation

$$A \underbrace{\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix}}_{=:B} = \underbrace{\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4 \end{bmatrix}}_{=:C},$$

in which the matrix A is the unknown (and matrices B and C are known). We proceed as in subsection 1.9.2. We take the transpose of both sides of the equation above to obtain $B^T A^T = C^T$, we form the matrix

$$\left[\begin{array}{c|c} B^T & C^T \end{array}\right] = \left[\begin{array}{c|c} \mathbf{b}_1^T & \mathbf{c}_1^T \\ \mathbf{b}_2^T & \mathbf{c}_2^T \\ \mathbf{b}_3^T & \mathbf{c}_3^T \\ \mathbf{b}_4^T & \mathbf{c}_4^T \end{array}\right] = \left[\begin{array}{ccc|cc} 1 & 2 & 0 & 1 & 1 \\ 2 & 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 2 & 0 \\ 0 & 0 & 1 & 1 & 2 \end{array}\right],$$

and we row reduce to obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} B^T & C^T \end{array}\right]\right) = \left[\begin{array}{ccc|cc} 1 & 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We now see that the equation $B^T A^T = C^T$ is consistent, and that the following is the general solution for A^T:

$$A^T = \begin{bmatrix} t_1 + 1 & t_2 + 1 \\ t_1 & t_2 \\ 1 & 2 \end{bmatrix}, \qquad \text{where } t_1, t_2 \in \mathbb{Z}_3.$$

By taking the transpose of the matrix above, we obtain the general solution of the equation $AB = C$:

$$A = \begin{bmatrix} t_1+1 & t_1 & 1 \\ t_2+1 & t_2 & 2 \end{bmatrix}, \qquad \text{where } t_1, t_2 \in \mathbb{Z}_3.$$

So, the equation $AB = C$ (with the unknown A) has a solution, but because of the two parameters (namely, t_1 and t_2), it is not unique. It follows that there exists a linear function $f : \mathbb{Z}_3^3 \to \mathbb{Z}_3^2$ that satisfies $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, 2, 3, 4\}$, but such a linear function f is **not** unique. □

Example 1.10.11. *Determine if there exists a linear function $f : \mathbb{R}^3 \to \mathbb{R}^2$ that satisfies all the following:*

- $f(\begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}) = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$;
- $f(\begin{bmatrix} 2 \\ -4 \\ 4 \end{bmatrix}) = \begin{bmatrix} 4 \\ -6 \end{bmatrix}$;
- $f(\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}) = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$;
- $f(\begin{bmatrix} -2 \\ 7 \\ -4 \end{bmatrix}) = \begin{bmatrix} -2 \\ 7 \end{bmatrix}$.

If such a linear function f exists, determine if it is unique.

Solution. To simplify notation, we set

$$\mathbf{b}_1 := \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \quad \mathbf{b}_2 := \begin{bmatrix} 2 \\ -4 \\ 4 \end{bmatrix}, \quad \mathbf{b}_3 := \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{b}_4 := \begin{bmatrix} -2 \\ 7 \\ -4 \end{bmatrix}$$

$$\mathbf{c}_1 := \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \quad \mathbf{c}_2 := \begin{bmatrix} 4 \\ -6 \end{bmatrix}, \quad \mathbf{c}_3 := \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \quad \mathbf{c}_4 := \begin{bmatrix} -2 \\ 7 \end{bmatrix}.$$

We need to check whether there exists a linear function $f : \mathbb{R}^3 \to \mathbb{R}^2$ that satisfies $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, 2, 3, 4\}$, and if so, whether it is unique. This is equivalent to determining whether there exists a matrix $A \in \mathbb{R}^{2\times 3}$ (the standard matrix of f) such that $A\mathbf{b}_i = \mathbf{c}_i$ for all $i \in \{1, 2, 3, 4\}$. So, we have a system of four equations (in which the unknown is the matrix A):

$$A\mathbf{b}_1 = \mathbf{c}_1, \qquad A\mathbf{b}_2 = \mathbf{c}_2, \qquad A\mathbf{b}_3 = \mathbf{c}_3, \qquad A\mathbf{b}_4 = \mathbf{c}_4.$$

This is equivalent to the equation

$$A\underbrace{\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix}}_{=:B} = \underbrace{\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4 \end{bmatrix}}_{=:C},$$

in which the matrix A is the unknown (and matrices B and C are known). We proceed as in subsection 1.9.2. We take the transpose of both sides of the equation above to obtain $B^TA^T = C^T$, we form the matrix

$$\left[\begin{array}{c|c} B^T & C^T \end{array}\right] = \left[\begin{array}{c|c} \mathbf{b}_1^T & \mathbf{c}_1^T \\ \mathbf{b}_2^T & \mathbf{c}_2^T \\ \mathbf{b}_3^T & \mathbf{c}_3^T \\ \mathbf{b}_4^T & \mathbf{c}_4^T \end{array}\right] = \left[\begin{array}{ccc|cc} 1 & -2 & 2 & 2 & -3 \\ 2 & -4 & 4 & 4 & -6 \\ 0 & -1 & 0 & -2 & -1 \\ -2 & 7 & -4 & -2 & 7 \end{array}\right],$$

and we row reduce to obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} B^T & C^T \end{array}\right]\right) = \left[\begin{array}{ccc|cc} 1 & 0 & 2 & 0 & -4 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We see from the third row of the matrix above that the equation $AB = C$ (where A is the unknown) is inconsistent. Therefore, there does not exist a linear function $f : \mathbb{R}^3 \to \mathbb{R}^2$ that satisfies the property that $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, 2, 3, 4\}$. □

Example 1.10.12. *Determine if there exists a linear function $f : \mathbb{Z}_3^4 \to \mathbb{Z}_3^3$ that satisfies all the following:*

- $f\Big(\begin{bmatrix} 1 & 2 & 1 & 2 \end{bmatrix}^T\Big) = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$;
- $f\Big(\begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix}^T\Big) = \begin{bmatrix} 2 & 0 & 1 \end{bmatrix}^T$;
- $f\Big(\begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}^T\Big) = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}^T$;
- $f\Big(\begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix}^T\Big) = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T$;
- $f\Big(\begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix}^T\Big) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$;
- $f\Big(\begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}^T\Big) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$.

If such a linear function f exists, determine if it is unique.

Solution. To simply notation, we set

$$\mathbf{b}_1 := \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \qquad \mathbf{b}_2 := \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}, \qquad \mathbf{b}_3 := \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

$$\mathbf{b}_4 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{b}_5 := \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{b}_6 := \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix},$$

and we further set

$$\mathbf{c}_1 := \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{c}_2 := \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{c}_3 := \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix},$$

$$\mathbf{c}_4 := \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{c}_5 := \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{c}_6 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

We need to check whether there exists a linear function $f : \mathbb{Z}_3^4 \to \mathbb{Z}_3^3$ that satisfies $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, \dots, 6\}$, and if so, whether it is unique. This is equivalent to determining whether there exists a matrix $A \in \mathbb{Z}_3^{3\times 4}$ (the standard matrix of f) such that $A\mathbf{b}_i = \mathbf{c}_i$ for all $i \in \{1, \dots, 6\}$. So, we have a system of six equations (in which the unknown is the matrix A):

$$A\mathbf{b}_1 = \mathbf{c}_1, \quad A\mathbf{b}_2 = \mathbf{c}_2, \quad A\mathbf{b}_3 = \mathbf{c}_3, \quad A\mathbf{b}_4 = \mathbf{c}_4, \quad A\mathbf{b}_5 = \mathbf{c}_5, \quad A\mathbf{b}_6 = \mathbf{c}_6.$$

This is equivalent to the equation

$$A \underbrace{\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 & \mathbf{b}_5 & \mathbf{b}_6 \end{bmatrix}}_{=:B} = \underbrace{\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4 & \mathbf{c}_5 & \mathbf{c}_6 \end{bmatrix}}_{=:C},$$

in which the matrix A is the unknown (and matrices B and C are known). We proceed as in subsection 1.9.2. We take the transpose of both sides of the equation above to obtain $B^T A^T = C^T$, we form the matrix

$$\begin{bmatrix} B^T & C^T \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{b}_1^T & \mathbf{c}_1^T \\ \mathbf{b}_2^T & \mathbf{c}_2^T \\ \mathbf{b}_3^T & \mathbf{c}_3^T \\ \mathbf{b}_4^T & \mathbf{c}_4^T \\ \mathbf{b}_5^T & \mathbf{c}_5^T \\ \mathbf{b}_6^T & \mathbf{c}_6^T \end{array}\right] = \left[\begin{array}{cccc|ccc} 1 & 2 & 1 & 2 & 1 & 1 & 0 \\ 2 & 2 & 2 & 2 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} B^T & C^T \end{array}\right]\right) = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 0 & 2 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

So, the equation $B^T A^T = C^T$ has a unique solution for A^T, namely,

$$A^T = \left[\begin{array}{ccc} 0 & 2 & 2 \\ 1 & 0 & 0 \\ 1 & 0 & 2 \\ 2 & 1 & 1 \end{array}\right],$$

and we deduce that the equation $AB = C$ has a unique solution for A, namely,

$$A = \left[\begin{array}{cccc} 0 & 1 & 1 & 2 \\ 2 & 0 & 0 & 1 \\ 2 & 0 & 2 & 1 \end{array}\right].$$

It now follows that there exists a unique linear function $f : \mathbb{Z}_3^4 \to \mathbb{Z}_3^3$ such that $f(\mathbf{b}_i) = \mathbf{c}_i$ for all $i \in \{1, \ldots, 6\}$, and moreover, the standard matrix of f is the matrix A above.

Remark: Now that we have computed the standard matrix A of f, we can easy compute a formula for f:

$$f(\mathbf{x}) = A\mathbf{x} = \left[\begin{array}{cccc} 0 & 1 & 1 & 2 \\ 2 & 0 & 0 & 1 \\ 2 & 0 & 2 & 1 \end{array}\right]\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array}\right] = \left[\begin{array}{c} x_2 + x_3 + 2x_4 \\ 2x_1 + x_4 \\ 2x_1 + 2x_3 + x_4 \end{array}\right].$$

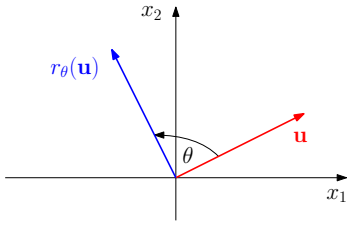
for all vectors $\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array}\right]^T$ in $\mathbb{Z}_3^4$. □

1.10.5 Some geometric examples

In this subsection, we consider some linear functions $f : \mathbb{R}^2 \to \mathbb{R}^2$ that have a nice geometric interpretation, and we find their standard matrices. We will not formally prove that these functions are all linear. To convince yourself that they are linear, think about what happens geometrically to sums and scalar multiples of vectors under these functions. We also note that the functions that we consider in this subsection have higher-dimensional analogs, which you can try to think about. However, we will not discuss higher dimensions in this subsection. In what follows, $\mathbf{e}_1$ and $\mathbf{e}_2$ are the standard basis vectors of $\mathbb{R}^2$.

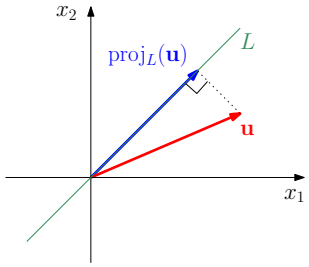
Rotation. The function $r_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates each vector about the origin counterclockwise by the angle θ (see the picture below) is linear, and its standard matrix is

$$\begin{bmatrix} r_\theta(\mathbf{e}_1) & r_\theta(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

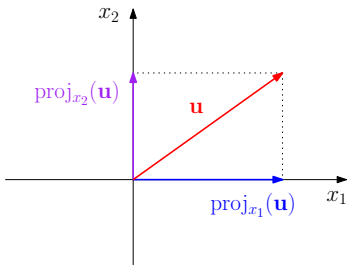


Note that rotating by the angle θ clockwise is the same as rotating by the angle $-\theta$ counterclockwise (which is why it is enough to consider only counterclockwise rotation, as long as we allow negative angles as well).

Orthogonal projection. Given a line L in $\mathbb{R}^2$ that passes through the origin, the orthogonal projection $\text{proj}_L : \mathbb{R}^2 \to \mathbb{R}^2$ onto L (see the picture below) is linear.



We cannot yet compute the standard matrix of orthogonal projection onto an arbitrary line through the origin; we will be able to do so only after we have developed a lot more theory (see Corollary 6.6.4). However, we can already compute this matrix in some special cases. Consider the projection $\text{proj}_{x_1} : \mathbb{R}^2 \to \mathbb{R}^2$ onto the x_1-axis and the projection $\text{proj}_{x_2} : \mathbb{R}^2 \to \mathbb{R}^2$ onto the x_2-axis (illustrated below). Note that for a vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ in $\mathbb{R}^2$, we have $\text{proj}_{x_1}(\mathbf{u}) = \begin{bmatrix} u_1 \\ 0 \end{bmatrix}$ and $\text{proj}_{x_2}(\mathbf{u}) = \begin{bmatrix} 0 \\ u_2 \end{bmatrix}$.



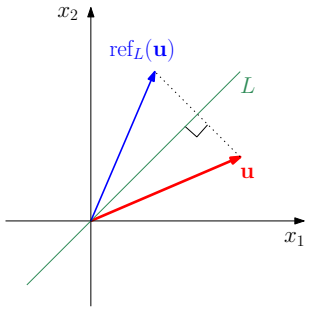
The standard matrix of proj_{x_1} is

$$\begin{bmatrix} \text{proj}_{x_1}(\mathbf{e}_1) & \text{proj}_{x_1}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{0} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

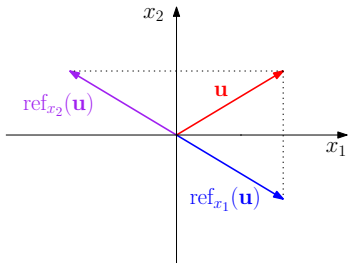
and the standard matrix of proj_{x_2} is

$$\begin{bmatrix} \text{proj}_{x_2}(\mathbf{e}_1) & \text{proj}_{x_2}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Reflection. Given a line L in $\mathbb{R}^2$ that passes through the origin, the reflection $\text{ref}_L : \mathbb{R}^2 \to \mathbb{R}^2$ about the line L (see the picture below) is linear.



As in the case of orthogonal projections, we cannot yet compute the standard matrix of the reflection about an arbitrary line through the origin; we will only be able to do so once we have developed a lot more theory (see subsection 6.8.3). However, we can already compute this matrix in some special cases. Consider the reflection $\text{ref}_{x_1} : \mathbb{R}^2 \to \mathbb{R}^2$ about the x_1-axis and the reflection $\text{ref}_{x_2} : \mathbb{R}^2 \to \mathbb{R}^2$ about the x_2-axis (illustrated below). Note that for a vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ in $\mathbb{R}^2$, we have $\text{ref}_{x_1}(\mathbf{u}) = \begin{bmatrix} u_1 \\ -u_2 \end{bmatrix}$ and $\text{ref}_{x_2}(\mathbf{u}) = \begin{bmatrix} -u_1 \\ u_2 \end{bmatrix}$.



The standard matrix of ref_{x_1} is

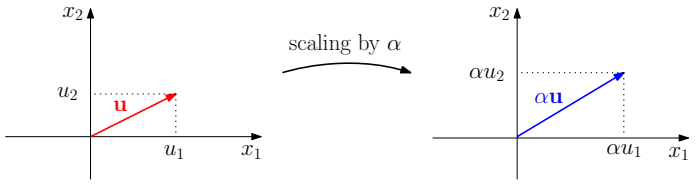
$$\begin{bmatrix} \text{ref}_{x_1}(\mathbf{e}_1) & \text{ref}_{x_1}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & -\mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and the standard matrix of ref_{x_2} is

$$\begin{bmatrix} \text{ref}_{x_2}(\mathbf{e}_1) & \text{ref}_{x_2}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -\mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

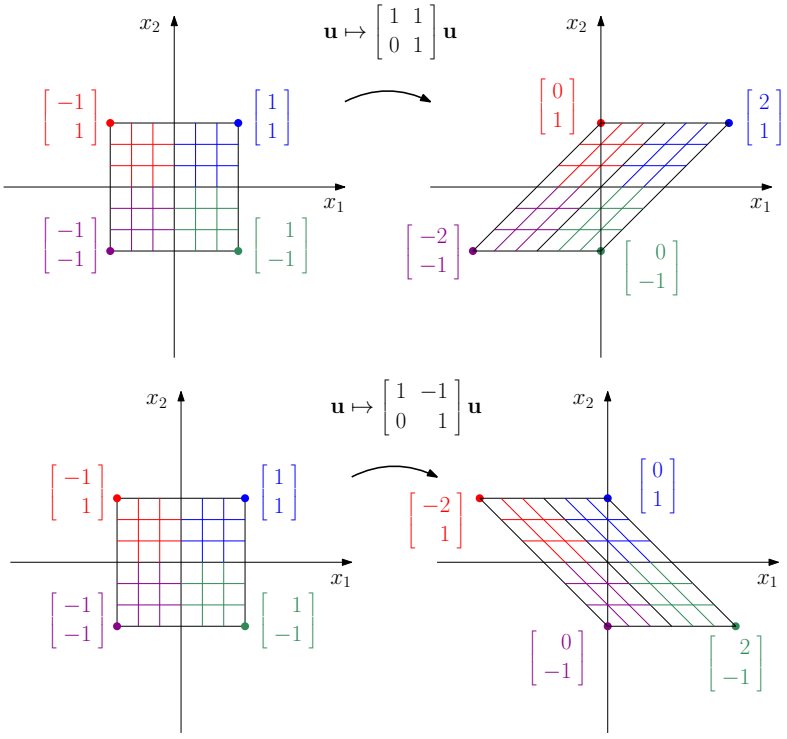
Scaling. Given a scalar $\alpha \in \mathbb{R}$, the function that scales each vector in $\mathbb{R}^2$ by α (see the picture below) is linear. The standard matrix of this linear function is

$$\begin{bmatrix} \alpha\mathbf{e}_1 & \alpha\mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}.$$

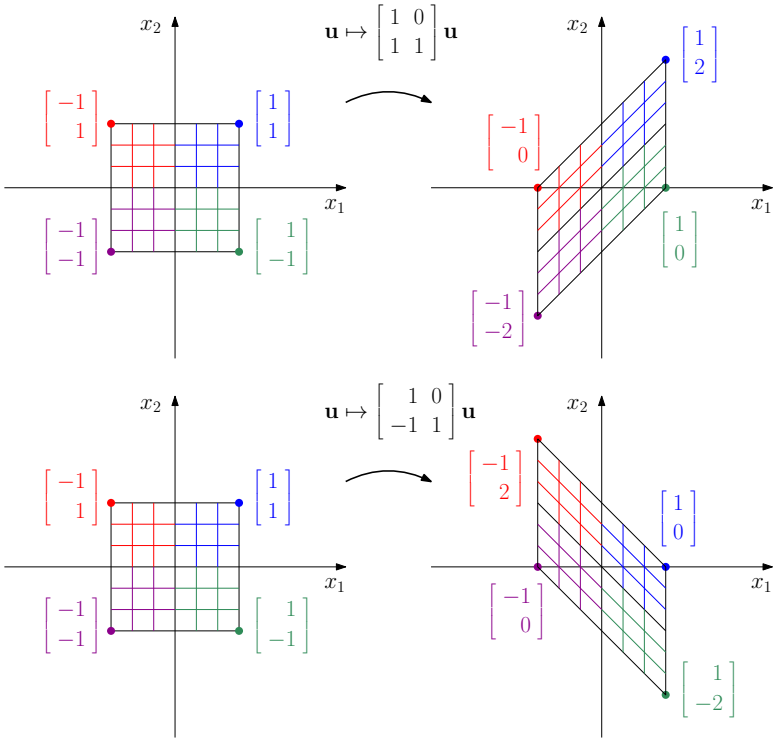


(If $\alpha = 0$, then scaling by α is the same as mapping each vector to the origin.)

Horizontal Shear. A *horizontal shear* in $\mathbb{R}^2$ is a mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by the formula $\mathbf{u} \mapsto \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \mathbf{u}$, i.e. by the formula $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \mapsto \begin{bmatrix} u_1 + ku_2 \\ u_2 \end{bmatrix}$, where k is a fixed real constant. This mapping has the effect of horizontally tilting objects in the coordinate plane (while keeping the vertical component unchanged). This is illustrated below for the cases when $k = 1$ and $k = -1$.

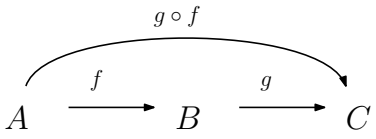


Vertical Shear. A *vertical shear* in $\mathbb{R}^2$ is a mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by the formula $\mathbf{u} \mapsto \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \mathbf{u}$, i.e. by the formula $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \mapsto \begin{bmatrix} u_1 \\ ku_1 + u_2 \end{bmatrix}$, where k is a fixed real constant. This mapping has the effect of vertically tilting objects in the coordinate plane (while keeping the horizontal component unchanged). This is illustrated below for the cases when $k = 1$ and $k = -1$.



1.10.6 Making new linear functions out of old ones

Given functions $f : A \to B$ and $g : B \to C$ (where A, B, and C are sets), we define the *composition of functions* g and f to be the function $g \circ f : A \to C$ given by $(g \circ f)(a) = g\big(f(a)\big)$ for all $a \in A$ (see the diagram below).



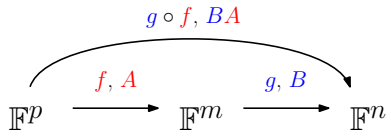
Proposition 1.10.13. *Let $\mathbb{F}$ be a field. Then all the following hold:*

(a) for all linear functions $f, g : \mathbb{F}^m \to \mathbb{F}^n$, the function $f + g$ is linear,[58] *and moreover, if A and B (both in $\mathbb{F}^{n \times m}$) are the standard matrices of f and g, respectively, then $A + B$ is the standard matrix of $f + g$;*

[58] As usual, the function $f + g : \mathbb{F}^m \to \mathbb{F}^n$ is defined by $(f + g)(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{F}^m$.

(b) for all linear functions $f : \mathbb{F}^m \to \mathbb{F}^n$ and scalars $\alpha \in \mathbb{F}$, the function αf is linear,[59] *and moreover, if $A \in \mathbb{F}^{n\times m}$ is the standard matrix of f, then αA is the standard matrix of αf;*

(c) for all linear functions $f : \mathbb{F}^p \to \mathbb{F}^m$ and $g : \mathbb{F}^m \to \mathbb{F}^n$, the function $g \circ f$ is linear,[60] *and moreover, if $A \in \mathbb{F}^{m\times p}$ and $B \in \mathbb{F}^{n\times m}$ are the standard matrices of f and g, respectively, then BA is the standard matrix of $g \circ f$.*[61]



Proof. We prove (c). Parts (a) and (b) are left as an exercise. Fix linear functions $f : \mathbb{F}^p \to \mathbb{F}^m$ and $g : \mathbb{F}^m \to \mathbb{F}^n$. Let $A \in \mathbb{F}^{m\times p}$ be the standard matrix of f, and let $B \in \mathbb{F}^{n\times m}$ be the standard matrix of g. Then for any $\mathbf{u} \in \mathbb{F}^p$, we have that

$$(g \circ f)(\mathbf{u}) \;=\; g(f(\mathbf{u})) \overset{(*)}{=} g(A\mathbf{u}) \overset{(**)}{=} B(A\mathbf{u}) \overset{(***)}{=} (BA)\mathbf{u},$$

where (*) follows from the fact that A is the standard matrix of f, (**) follows from the fact that B is the standard matrix of g, and (***) follows from Corollary 1.7.6(g). We have now shown that $g\circ f$ is a matrix transformation, and so (by Proposition 1.10.4) it is linear. Moreover, since (by the calculation above) we have that $(g \circ f)(\mathbf{u}) = (BA)\mathbf{u}$ for all vectors $\mathbf{u} \in \mathbb{F}^p$, we see that BA is the standard matrix of $g \circ f$. □

Example 1.10.14.

(a) Find the standard matrix of the linear function $f_1 : \mathbb{R}^2 \to \mathbb{R}^2$ that first reflects about the x_1-axis and then rotates about the origin counterclockwise by $90°$.

(b) Find the standard matrix of the linear function $f_2 : \mathbb{R}^2 \to \mathbb{R}^2$ that first rotates about the origin counterclockwise by $90°$ and then reflects about the x_1-axis.

You may assume that f_1 and f_2 are indeed linear.

Solution. We solve the problem in two ways: first, by checking what the linear functions f_1 and f_2 map the standard basis vectors to, and second, by multiplying matrices as in Proposition 1.10.13(c).

Computing directly. (a) We observe that $f_1(\mathbf{e}_1) = \mathbf{e}_2$ and $f_1(\mathbf{e}_2) = \mathbf{e}_1$. Consequently, the standard matrix of f_1 is

$$\begin{bmatrix} f_1(\mathbf{e}_1) & f_1(\mathbf{e}_2) \end{bmatrix} \;=\; \begin{bmatrix} \mathbf{e}_2 & \mathbf{e}_1 \end{bmatrix} \;=\; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

[59] As usual, the function $\alpha f : \mathbb{F}^m \to \mathbb{F}^n$ is defined by $(\alpha f)(\mathbf{u}) = \alpha(f(\mathbf{u}))$ for all $\mathbf{u} \in \mathbb{F}^m$.
[60] As usual, the function $g \circ f : \mathbb{F}^p \to \mathbb{F}^n$ is defined by $(g \circ f)(\mathbf{u}) = g(f(\mathbf{u}))$ for all $\mathbf{u} \in \mathbb{F}^p$.
[61] Note that $BA \in \mathbb{F}^{n\times p}$.

(b) We observe that $f_2(\mathbf{e}_1) = -\mathbf{e}_2$ and $f_2(\mathbf{e}_2) = -\mathbf{e}_1$. Consequently, the standard matrix of f_1 is

$$\begin{bmatrix} f_2(\mathbf{e}_1) & f_2(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -\mathbf{e}_2 & -\mathbf{e}_1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Computing via matrix multiplication. We use Proposition 1.10.13(c). First, we note that the standard matrix of $\mathrm{ref}_{x_1} : \mathbb{R}^2 \to \mathbb{R}^2$, the reflection about the x_1-axis, is

$$A = \begin{bmatrix} \mathrm{ref}_{x_1}(\mathbf{e}_1) & \mathrm{ref}_{x_1}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & -\mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

whereas the standard matrix of $r_{90^\circ} : \mathbb{R}^2 \to \mathbb{R}^2$, the counterclockwise rotation by 90° about the origin, is

$$B = \begin{bmatrix} r_{90^\circ}(\mathbf{e}_1) & r_{90^\circ}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_2 & -\mathbf{e}_1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Note that

$$f_1 = r_{90^\circ} \circ \mathrm{ref}_{x_1} \qquad \text{and} \qquad f_2 = \mathrm{ref}_{x_1} \circ r_{90^\circ}.$$

So, by Proposition 1.10.13(c), the standard matrix of f_1 is

$$BA = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

whereas by the standard matrix of f_2 is

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Remark: Our two methods produced the same final answer, as expected. □

1.10.7 One-to-one and onto functions. Bijections and their inverses

For a set X, we define the function $\mathrm{Id}_X : X \to X$ by setting $\mathrm{Id}_X(x) = x$ for all $x \in X$. Id_X is called the *identity function* on X.

A function $f : A \to B$ is said to be

- *one-to-one* (or *injective*, or an *injection*) if for all $a_1, a_2 \in A$ such that $a_1 \neq a_2$, we have $f(a_1) \neq f(a_2)$;[62]

[62] Equivalently, $f : A \to B$ is *one-to-one* if for all $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$, we have that $a_1 = a_2$.

- *onto* (or *surjective*, or a *surjection*) if for all $b \in B$, there exists some $a \in A$ such that $f(a) = b$;
- *bijective* or a *bijection* if it is both one-to-one and onto.

Proposition 1.10.15. *Let $f : A \to B$ be a function. Then the following are equivalent:*

(a) f is a bijection;

(b) there exists some function $g : B \to A$ such that $g \circ f = Id_A$ and $f \circ g = Id_B$.

Proof. Suppose first that (a) holds. Then for all $b \in B$, there exists a unique $a \in A$ such that $f(a) = b$.[63] We now define $g : B \to A$ by, for each $b \in B$, letting $g(b)$ be the unique $a \in A$ such that $f(a) = b$. Then clearly, $g \circ f = \mathrm{Id}_A$ and $f \circ g = \mathrm{Id}_B$.[64]

Suppose now that (b) holds, and fix a function $g : B \to A$ such that $g \circ f = \mathrm{Id}_A$ and $f \circ g = \mathrm{Id}_B$. We first show that f is one-to-one. Fix $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$. Then

$$\begin{aligned} a_1 &= \mathrm{Id}_A(a_1) \\ &= (g \circ f)(a_1) && \text{because } g \circ f = \mathrm{Id}_A \\ &= g\big(f(a_1)\big) \\ &= g\big(f(a_2)\big) && \text{because } f(a_1) = f(a_2) \\ &= (g \circ f)(a_2) \\ &= \mathrm{Id}_A(a_2) && \text{because } g \circ f = \mathrm{Id}_A \\ &= a_2. \end{aligned}$$

So, f is one-to-one. We now show that f is onto. Fix $b \in B$, and set $a := g(b)$. Then

$$f(a) \;=\; f(g(b)) \;=\; (f \circ g)(b) \;=\; \mathrm{Id}_B(b) \;=\; b.$$

So, f is onto. We have now shown that f is both one-to-one and onto, and so f is a bijection, i.e. (a) holds. □

Proposition 1.10.16. *Let $f : A \to B$ be a bijection. Then there exists a **unique** function $g : B \to A$ such that $g \circ f = Id_A$ and $f \circ g = Id_B$.*

[63] The existence of such an a follows from the fact that f is onto, and the uniqueness of a follows from the fact that f is one-to-one.

[64] Indeed, for all $a \in A$, we have that $(g \circ f)(a) = g(f(a)) = a$. On the other hand, fix $b \in B$, and let a be the unique element of A such that $f(a) = b$; then $(f \circ g)(b) = f(g(b)) = f(a) = b$.

Proof. The existence of g follows immediately from Proposition 1.10.15. It remains to prove uniqueness. So, suppose that functions $g_1, g_2 : B \to A$ satisfy

- $g_1 \circ f = \mathrm{Id}_A$ and $f \circ g_1 = \mathrm{Id}_B$;
- $g_2 \circ f = \mathrm{Id}_A$ and $f \circ g_2 = \mathrm{Id}_B$.

We must show that $g_1 = g_2$. Fix $b \in B$. Since f is onto, there exists some $a \in A$ such that $f(a) = b$. We now have that

$$\begin{aligned} g_1(b) &= g_1(f(a)) && \text{because } f(a) = b \\ &= (g_1 \circ f)(a) \\ &= \mathrm{Id}_A(a) && \text{because } g_1 \circ f = \mathrm{Id}_A \\ &= (g_2 \circ f)(a) && \text{because } g_2 \circ f = \mathrm{Id}_A \\ &= g_2(f(a)) \\ &= g_2(b) && \text{because } f(a) = b. \end{aligned}$$

So, $g_1 = g_2$. □

Terminology/Notation: If $f : A \to B$ is a bijection, then the unique function $g : B \to A$ that satisfies $g \circ f = \mathrm{Id}_A$ and $f \circ g = \mathrm{Id}_B$ (i.e. the function g from Proposition 1.10.16) is called the *inverse* of f and is denoted by f^{-1}. Note that this means that:

- $f^{-1} \circ f = \mathrm{Id}_A$;
- $f \circ f^{-1} = \mathrm{Id}_B$;
- for all $a \in A$ and $b \in B$, we have that $b = f(a)$ if and only if $a = f^{-1}(b)$.

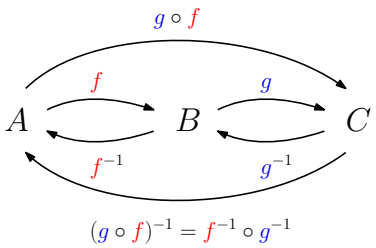
Note that the inverse of a bijection is also a bijection (by Proposition 1.10.15), and moreover, $(f^{-1})^{-1} = f$.

Proposition 1.10.17. *Let $f : A \to B$ and $g : B \to C$ be functions. Then all the following hold:*

(a) if f and g are one-to-one, then $g \circ f$ is also one-to-one;

(b) if f and g are onto, then $g \circ f$ is also onto;

(c) if f and g are bijections, then $g \circ f$ is also a bijection, and moreover, $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$ (see the diagram below).



Proof. (a) Assume that f and g are one-to-one; we must show that $g \circ f$ is one-to-one. Fix $a_1, a_2 \in A$ such that $(g \circ f)(a_1) = (g \circ f)(a_2)$, i.e. $g\big(f(a_1)\big) = g\big(f(a_2)\big)$. Since g is one-to-one, we have that $f(a_1) = f(a_2)$. Since f is one-to-one, we have that $a_1 = a_2$. This proves that $g \circ f$ is one-to-one.

(b) Assume that f and g are onto; we must show that $g \circ f$ is onto. Fix $c \in C$. Since g is onto, there exists some $b \in B$ such that $g(b) = c$. Since f is onto, there exists some $a \in A$ such that $f(a) = b$. But now

$$(g \circ f)(a) \quad = \quad g\big(f(a)\big) \quad = \quad g(b) \quad = \quad c.$$

This proves that $g \circ f$ is onto.

(c) Assume that f and g are bijections. By definition, this means that they are one-to-one and onto, and so by (a) and (b), $g \circ f$ is one-to-one and onto, i.e. $g \circ f$ is a bijection. It remains to show that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. Fix any $c \in C$. Set $b := g^{-1}(c)$ and $a := f^{-1}(b)$, so that $g(b) = c$ and $f(a) = b$. Then

$$(g \circ f)(a) \quad = \quad g\big(f(a)\big) \quad = \quad g(b) \quad = \quad c,$$

and consequently, $(g \circ f)^{-1}(c) = a$. On the other hand,

$$(f^{-1} \circ g^{-1})(c) \quad = \quad f^{-1}\big(g^{-1}(c)\big) \quad = \quad f^{-1}(b) \quad = \quad a.$$

Thus, $(g \circ f)^{-1}(c) = (f^{-1} \circ g^{-1})(c)$, and we deduce that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. □

1.10.8 A first look at isomorphisms

As the following theorem shows, we can easily check whether a linear function is one-to-one or onto by computing the rank of its standard matrix.

Theorem 1.10.18. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function, and let $A \in \mathbb{F}^{n \times m}$ be the standard matrix of f. Then both the following hold:*

(a) f is one-to-one if and only if $\text{rank}(A) = m$ (i.e. A has full column rank);

(b) f is onto if and only if $\text{rank}(A) = n$ (i.e. A has full row rank).

Proof. (a) We have the following sequence of equivalent statements:

$$\begin{aligned} f \text{ is one-to-one} &\overset{(*)}{\iff} \text{for all } \mathbf{b} \in \mathbb{F}^n,\ f(\mathbf{x}) = \mathbf{b} \\ &\qquad\qquad \text{has at most one solution} \\ &\overset{(**)}{\iff} \text{for all } \mathbf{b} \in \mathbb{F}^n,\ A\mathbf{x} = \mathbf{b} \\ &\qquad\qquad \text{has at most one solution,} \\ &\overset{(***)}{\iff} \operatorname{rank}(A) = m, \end{aligned}$$

where (*) follows from the definition of a one-to-one function, (**) follows from the fact that A is the standard matrix of f, and (***) follows from Corollary 1.6.5.

(b) We have the following sequence of equivalent statements:

$$\begin{aligned} f \text{ is onto} &\overset{(*)}{\iff} \text{for all } \mathbf{b} \in \mathbb{F}^n,\ f(\mathbf{x}) = \mathbf{b} \\ &\qquad\qquad \text{has at least one solution} \\ &\overset{(**)}{\iff} \text{for all } \mathbf{b} \in \mathbb{F}^n,\ A\mathbf{x} = \mathbf{b} \\ &\qquad\qquad \text{has at least one solution} \\ &\qquad\qquad \text{(i.e. } A\mathbf{x} = \mathbf{b} \text{ is consistent)} \\ &\overset{(***)}{\iff} \operatorname{rank}(A) = n, \end{aligned}$$

where (*) follows from the definition of an onto function, (**) follows from the fact that A is the standard matrix of f, and (***) follows from Corollary 1.6.6. □

Let $\mathbb{F}$ be a field. A function $f : \mathbb{F}^m \to \mathbb{F}^n$ is an *isomorphism* if it is both linear and a bijection. As Theorem 1.10.19 (below) shows, if we know the standard matrix of a linear function, then we can easily determine whether that linear function is an isomorphism. Moreover, Theorem 1.10.19 implies, in particular, that for a field $\mathbb{F}$, there can be no isomorphism from $\mathbb{F}^m$ to $\mathbb{F}^n$ for $m \neq n$.

Theorem 1.10.19. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function, and let $A \in \mathbb{F}^{n\times m}$ be the standard matrix of f. Then the following are equivalent:*

(a) f is an isomorphism;

(b) $\operatorname{rank}(A) = m = n$ (i.e. A is a square matrix of full rank).

Proof. Suppose first that (a) holds. Since f is a one-to-one linear function, Theorem 1.10.18(a) guarantees that $\operatorname{rank}(A) = m$. On the other hand, since f is an onto linear function, Theorem 1.10.18(b) guarantees that $\operatorname{rank}(A) = n$. But now $m = \operatorname{rank}(A) = n$, and (b) follows.

Suppose now that (b) holds. Then by Theorem 1.10.18(a), f is one-to-one, and by Theorem 1.10.18(b), f is onto. So, f is a bijection. Since f is also linear (by hypothesis), we deduce that f is an isomorphism, i.e. (a) holds. □

Proposition 1.10.20. *Let $\mathbb{F}$ be a field, and let $f : \mathbb{F}^n \to \mathbb{F}^n$ be an isomorphism. Then $f^{-1} : \mathbb{F}^n \to \mathbb{F}^n$ is also an isomorphism.*

Proof. Since $f : \mathbb{F}^n \to \mathbb{F}^n$ is an isomorphism, it is, in particular, a bijection; consequently, f has an inverse $f^{-1} : \mathbb{F}^n \to \mathbb{F}^n$, which is also a bijection. So, to show that f^{-1} is an isomorphism, it suffices to show that f^{-1} is linear.

First, fix $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{F}^n$. We must show that $f^{-1}(\mathbf{v}_1 + \mathbf{v}_2) = f^{-1}(\mathbf{v}_1) + f^{-1}(\mathbf{v}_2)$. Set $\mathbf{u}_1 := f^{-1}(\mathbf{v}_1)$ and $\mathbf{u}_2 := f^{-1}(\mathbf{v}_2)$, so that $f(\mathbf{u}_1) = \mathbf{v}_1$ and $f(\mathbf{u}_2) = \mathbf{v}_2$. Then

$$\begin{aligned}
f^{-1}(\mathbf{v}_1 + \mathbf{v}_2) &= f^{-1}\big(f(\mathbf{u}_1) + f(\mathbf{u}_2)\big) \\
&= f^{-1}\big(f(\mathbf{u}_1 + \mathbf{u}_2)\big) && \text{because } f \text{ is linear} \\
&= (f^{-1} \circ f)(\mathbf{u}_1 + \mathbf{u}_2) \\
&= \mathrm{Id}_{\mathbb{F}^n}(\mathbf{u}_1 + \mathbf{u}_2) \\
&= \mathbf{u}_1 + \mathbf{u}_2 \\
&= f^{-1}(\mathbf{v}_1) + f^{-1}(\mathbf{v}_2).
\end{aligned}$$

Next, fix $\mathbf{v} \in \mathbb{F}^n$ and $\alpha \in \mathbb{F}$. We must show that $f^{-1}(\alpha\mathbf{v}) = \alpha f^{-1}(\mathbf{v})$. Set $\mathbf{u} := f^{-1}(\mathbf{v})$, so that $f(\mathbf{u}) = \mathbf{v}$. Then

$$\begin{aligned}
f^{-1}(\alpha\mathbf{v}) &= f^{-1}\big(\alpha f(\mathbf{u})\big) \\
&= f^{-1}\big(f(\alpha\mathbf{u})\big) && \text{because } f \text{ is linear} \\
&= (f^{-1} \circ f)(\alpha\mathbf{u}) \\
&= \mathrm{Id}_{\mathbb{F}^n}(\alpha\mathbf{u}) \\
&= \alpha\mathbf{u} \\
&= \alpha f^{-1}(\mathbf{v}).
\end{aligned}$$

We have now proven that f^{-1} linear. This completes the argument. □

1.11 Invertible matrices

1.11.1 Invertible matrices: definition and uniqueness of inverses

A **square** matrix $A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) is *invertible* if there exists a matrix $B \in \mathbb{F}^{n\times n}$, called an *inverse* of A, such that $AB = BA = I_n$. A square matrix that is not invertible is called *non-invertible*. As we shall see, the inverse of an invertible matrix is in fact unique (see Proposition 1.11.1 below). In subsection 1.11.2, we will describe a simple procedure for determining whether a square matrix is invertible, and if so, for finding its inverse.

Terminology: Inverible matrices are also called *non-singular* or *non-degenerate*, whereas non-invertible matrices are also called *singular* or *degenerate*. The Czech term for an invertible matrix is "regulární matice," and for this reason, Czech mathematicians sometimes use the term "regular matrix" instead of "invertible matrix"; however, this usage ("regular matrix") is quite rare in the English speaking world. In these notes, we will consistently use the term "invertible matrix."

Proposition 1.11.1. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be an invertible matrix. Then A has a unique inverse.*

Notation: The unique inverse of A is denoted by A^{-1}.

Proof. Since A is invertible, it has an inverse, and we just need to show that it is unique. So, suppose that $B, C \in \mathbb{F}^{n\times n}$ are both inverses of A, so that $AB = BA = I_n$ and $AC = CA = I_n$. Then

$$\begin{aligned}
B &= BI_n && \text{by Proposition 1.7.2}\\
&= B(AC) && \text{because } AC = I_n\\
&= (BA)C && \text{by the associativity of matrix multiplication (see Theorem 1.7.5(g))}\\
&= I_nC && \text{because } BA = I_n\\
&= C && \text{by Proposition 1.7.2.}
\end{aligned}$$

This completes the argument. □

Example 1.11.2. *The matrix $A := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ in $\mathbb{R}^{2\times 2}$ is invertible, and its inverse is $A^{-1} := \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, which we can easily verify by checking that*

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = I_2 \quad \text{and} \quad \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = I_2.$$

We complete this subsection with a technical proposition whose proof is very similar to that of Proposition 1.11.1 (above).

Proposition 1.11.3. *Let $\mathbb{F}$ be a field, and let $A, B \in \mathbb{F}^{n\times n}$. Assume that A is invertible and that $AB = I_n$ or $BA = I_n$. Then $A^{-1} = B$.*

Proof. We prove the proposition for the case when $BA = I_n$. The proof of the other case (i.e. the case when $AB = I_n$) is similar and is left as an easy exercise. Since A is invertible, it has an inverse A^{-1}, and we have that $AA^{-1} = I_n$. We now compute:

$$\begin{aligned} B &= BI_n && \text{by Proposition 1.7.2} \\ &= B(AA^{-1}) && \text{because } AA^{-1} = I_n \\ &= (BA)A^{-1} && \text{by the associativity of matrix multiplication (see Theorem 1.7.5(g))} \\ &= I_nA^{-1} && \text{because } BA = I_n \\ &= A^{-1} && \text{by Proposition 1.7.2.} \end{aligned}$$

This completes the argument. □

Remark: Note that Proposition 1.11.3 can only be applied if we already know that A is invertible. Once we have developed a lot more theory, we will be able to eliminate this hypothesis and show that if $A, B \in \mathbb{F}^{n\times n}$ are **square** matrices that satisfy $AB = I_n$, then both A and B are invertible and are each other's inverses (see Corollary 3.3.18). However, we cannot prove this stronger statement yet, and therefore, we cannot use it yet.

1.11.2 Computing the inverse of an invertible matrix

The following theorem (whose proof we postpone to subsection 1.11.6) gives us a recipe for determining whether a square matrix is invertible, and if so, for finding the inverse of that matrix. We state the theorem and give a few examples. However, as we develop our theory in the remainder of this section, we will **not** rely on Theorem 1.11.4 (in fact, we will need to develop sufficient theory in order to actually prove this theorem).

Theorem 1.11.4. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ be a square matrix, and set $\left[\begin{array}{c|c} U & B \end{array}\right] = RREF\left(\left[\begin{array}{c|c} A & I_n \end{array}\right]\right)$, where each of U and B has n columns. Then*

(a) if $U = I_n$, then A is invertible and $B = A^{-1}$;

(b) if $U \neq I_n$, then A is not invertible.

Example 1.11.5. *Consider the following matrices.*

(a) $A = \left[\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}\right]$, *with entries understood to be in $\mathbb{R}$;*

(b) $B = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{array}\right]$, *with entries understood to be in $\mathbb{Z}_2$;*

(c) $C = \left[\begin{array}{ccc} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 2 & 0 & 1 \end{array}\right]$, *with entries understood to be in $\mathbb{Z}_3$.*

For each of these three matrices, determine if the matrix is invertible, and if so, find its inverse.

Solution. (a) We form the matrix

$$\left[\begin{array}{c|c} A & I_2 \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A & I_2 \end{array}\right]\right) = \left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right].$$

The submatrix of $\mathrm{RREF}\left(\left[\begin{array}{c|c} A & I_2 \end{array}\right]\right)$ to the left of the vertical dotted line is I_2. So, A is invertible, and its inverse is

$$A^{-1} = \left[\begin{array}{cc} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{array}\right].$$

(b) We form the matrix

$$\left[\begin{array}{c|c} B & I_3 \end{array}\right] = \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}\Big(\left[\begin{array}{c:c} B & I_3 \end{array}\right]\Big) = \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right].$$

The submatrix of $\mathrm{RREF}\Big(\left[\begin{array}{c:c} B & I_3 \end{array}\right]\Big)$ to the left of the vertical dotted line is I_3. So, B is invertible, and its inverse is

$$B^{-1} = \left[\begin{array}{ccc} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{array}\right].$$

(c) We form the matrix

$$\left[\begin{array}{c:c} C & I_3 \end{array}\right] = \left[\begin{array}{ccc:ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 & 1 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}\Big(\left[\begin{array}{c:c} C & I_3 \end{array}\right]\Big) = \left[\begin{array}{ccc:ccc} 1 & 0 & 2 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 2 \end{array}\right].$$

The submatrix of $\mathrm{RREF}\Big(\left[\begin{array}{c:c} C & I_3 \end{array}\right]\Big)$ to the left of the vertical dotted line is not I_3. So, C is **not** invertible. □

Remark: Because it is easy to miscompute when row reducing, it is not a bad idea to check our answers. Suppose that we have computed the inverse A^{-1} of an invertible $n \times n$ matrix A (with entries in some field $\mathbb{F}$). We can check if our answer is correct by computing the matrix products AA^{-1} and $A^{-1}A$, and verifying that we get I_n in both cases. Actually, in practice, it is more or less enough to check that one of $AA^{-1} = I_n$ and $A^{-1}A = I_n$ holds. The theoretical justification for this is given by Corollary 3.3.18 (which we cannot prove yet), but for now, the point is that this answer checking is not a formal part of our solution/calculation: we only do it in order to increase our own confidence that we have not miscomputed.

1.11.3 Basic properties of invertible matrices

Matrix invertibility and matrix-vector equations. Theorem 1.11.6 (below) is one of the main reasons we care about invertible matrices. Note that it implies that if the **coefficient** matrix of a linear system is invertible, then that linear system has a unique solution.

Theorem 1.11.6. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be an invertible matrix. Then for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution, and that solution is $A^{-1}\mathbf{b}$.*

Proof. Fix any vector $\mathbf{b} \in \mathbb{F}^n$. To show that $A^{-1}\mathbf{b}$ is indeed a solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, we compute

$$A(A^{-1}\mathbf{b}) \overset{(*)}{=} (\underbrace{AA^{-1}}_{=I_n})\mathbf{b} = I_n\mathbf{b} \overset{(**)}{=} \mathbf{b},$$

where (*) follows from Corollary 1.7.6(g), and (**) follows from Proposition 1.4.5.

So far, we have proven that $A^{-1}\mathbf{b}$ is a solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. It remains to prove uniqueness. Fix any solution $\mathbf{x}_0 \in \mathbb{F}^n$ of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. Then $A\mathbf{x}_0 = \mathbf{b}$, and consequently, $A^{-1}(A\mathbf{x}_0) = A^{-1}\mathbf{x}_0$. We now compute:

$$A^{-1}\mathbf{b} = A^{-1}(A\mathbf{x}_0) \overset{(*)}{=} (\underbrace{A^{-1}A}_{=I_n})\mathbf{x}_0 = I_n\mathbf{x}_0 \overset{(**)}{=} \mathbf{x}_0.$$

where once again, (*) follows from Corollary 1.7.6(g), and (**) follows from Proposition 1.4.5. This proves that $A^{-1}\mathbf{b}$ is in fact the unique solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. □

Example 1.11.7. *Set*

$$A := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} := \begin{bmatrix} 2 \\ -3 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Solve the matrix-vector equation $A\mathbf{x} = \mathbf{b}$.

Solution. As we saw in Example 1.11.2, the matrix A is invertible, and its inverse is

$$A^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

So, by Theorem 1.11.6, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution, namely

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

□

Remark: We saw in subsection 1.11.2 how one can check if a square matrix (with entries in some field) is invertible, and if so, how one can compute its inverse. However, if we do not already know whether A is invertible (or we know that A is invertible, but have not yet computed its inverse), then the most efficient way to

solve our matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is by row reducing the augmented matrix $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$. Using the formula $\mathbf{x} = A^{-1}\mathbf{b}$ is only efficient if we already happen to know that A is invertible and have already computed its inverse A^{-1} for some reason other than solving the equation $A\mathbf{x} = \mathbf{b}$.

Making new invertible matrices out of old ones. We complete this subsection by proving an easy, but important, proposition about invertible matrices and their inverses.

Proposition 1.11.8. *Let $\mathbb{F}$ be a field. Then all the following hold:*

(a) the identity matrix I_n is invertible and is its own inverse (i.e. $I_n^{-1} = I_n$);

(b) if a matrix $A \in \mathbb{F}^{n\times n}$ is invertible, then its inverse A^{-1} is also invertible, and moreover, $(A^{-1})^{-1} = A$;

(c) if a matrix $A \in \mathbb{F}^{n\times n}$ is invertible, then its transpose A^T is also invertible, and moreover, $(A^T)^{-1} = (A^{-1})^T$;

(d) if matrices $A, B \in \mathbb{F}^{n\times n}$ are invertible matrices, then AB is also invertible, and moreover, $(AB)^{-1} = B^{-1}A^{-1}$;

(e) if matrices $A_1, \ldots, A_k \in \mathbb{F}^{n\times n}$ are invertible, then the matrix $A_1 \ldots A_k$ is also invertible, and moreover, $(A_1 \ldots A_k)^{-1} = A_k^{-1} \ldots A_1^{-1}$;

(f) if a matrix $A \in \mathbb{F}^{n\times n}$ is invertible, then for all non-negative integers m, the matrix A^m is also invertible, and moreover, $(A^m)^{-1} = (A^{-1})^m$.

Proof. Part (a) follows immediately from the fact that $I_n I_n = I_n$.

Let us prove (b). Fix an invertible matrix $A \in \mathbb{F}^{n\times n}$. Since $AA^{-1} = A^{-1}A = I_n$, we see that A^{-1} is invertible that that its inverse is A. This proves (b).

Next, we prove (c). Fix an invertible matrix $A \in \mathbb{F}^{n\times n}$. Then

$$A^T(A^{-1})^T \overset{(*)}{=} (A^{-1}A)^T = I_n^T = I_n,$$

where (*) follows from Proposition 1.8.1(d). An analogous argument shows that $(A^{-1})^T A^T = I_n$. So, A^T is invertible and its inverse is $(A^{-1})^T$. This proves (c).

We now prove (d). Fix invertible matrices $A, B \in \mathbb{F}^{n\times n}$. It suffices to show that $(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I_n$. For this, we compute (using the associativity of matrix multiplication):

- $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$;
- $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n$.

This proves (d).

Part (e) follows from (d) via an easy induction on k (the details are left as an exercise). Part (f) follows from (a) when $m = 0$ (this is because $A^0 = I_n$ for all matrices $A \in \mathbb{F}^{n\times n}$), and is a special case of (e) when $m \geq 1$. □

Notation: For a field $\mathbb{F}$, an invertible matrix $A \in \mathbb{F}^{n\times n}$, and a positive integer m, we define $A^{-m} := (A^{-1})^m$. By Proposition 1.11.8(f), we also have that $A^{-m} = (A^m)^{-1}$, as we would expect. Note that this is only defined if A is invertible (and is undefined otherwise).

1.11.4 Invertible matrices, isomorphisms, and rank

As our next theorem shows, invertible matrices are precisely the standard matrices of isomorphisms, or equivalently, the square matrices of full rank. Recall from Theorem 1.10.19 that if $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is some field) is an isomorphism, then $m = n$.

Theorem 1.11.9. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ be a square matrix, and let $f : \mathbb{F}^n \to \mathbb{F}^n$ be given by $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$. Then f is linear and its standard matrix is A. Furthermore, the following are equivalent:*

(a) f is an isomorphism;

(b) A is invertible;

(c) $RREF(A) = I_n$;

(d) $rank(A) = n$.

Moreover, in this case,[65] *f^{-1} is an isomorphism and its standard matrix is A^{-1}.*

Proof. The function f is a matrix transformation, and so by Proposition 1.10.4, it is linear. The fact that A is its standard matrix follows from the definition of a standard matrix.

By Theorem 1.10.19, (a) and (d) are equivalent, and by Proposition 1.6.7, (c) and (d) are equivalent. So, (a), (c), and (d) are equivalent. Moreover, Proposition 1.10.20 guarantees that if f is an isomorphism, then so is f^{-1}. It now suffices to prove the following:

(1) if f is an isomorphism, then A is invertible, and moreover, the standard matrix of f^{-1} is A^{-1};[66]

[65] "In this case" means "if (a), (b), (c), and (d) hold," or equivalently (since (a), (b), (c), and (d) are equivalent): "if one of (a), (b), (c), and (d) holds."

[66] Note that (1) states that (a) implies (b), and moreover, that if (a) holds, then the standard matrix of f^{-1} is A^{-1}.

(2) if A is invertible, then f is an isomorphism.[67]

We first prove (1). Assume that f is an isomorphism. Then by Proposition 1.10.20, $f^{-1} : \mathbb{F}^n \to \mathbb{F}^n$ is an isomorphism; let $B \in \mathbb{F}^{n\times n}$ be the standard matrix of the isomorphism f^{-1}. We must show that A is invertible and that $B = A^{-1}$. Since f and f^{-1} are linear, Proposition 1.10.13(c) guarantees that $f \circ f^{-1}$ and $f^{-1} \circ f$ are also linear, and moreover, that their standard matrices are AB and BA, respectively. On the other hand, we have that $f^{-1} \circ f = f \circ f^{-1} = \mathrm{Id}_{\mathbb{F}^n}$, and clearly, the standard matrix of $\mathrm{Id}_{\mathbb{F}^n}$ is I_n.[68] So, $AB = BA = I_n$. But now A is invertible and B is its inverse, i.e. $B = A^{-1}$.

It remains to prove (2). Assume that A is invertible. We must show that f is an isomorphism. By hypothesis, f is linear; it remains to show that f is a bijection. Define $g : \mathbb{F}^n \to \mathbb{F}^n$ by setting $g(\mathbf{u}) = A^{-1}\mathbf{u}$ for all $\mathbf{u} \in \mathbb{F}^n$. (So, $g : \mathbb{F}^n \to \mathbb{F}^n$ is the linear function whose standard matrix is A^{-1}.) Our goal is to show that $f \circ g = g \circ f = \mathrm{Id}_{\mathbb{F}^n}$. In view of Proposition 1.10.15, this will imply that f is a bijection, which is what we need. But indeed, for any $\mathbf{u} \in \mathbb{F}^n$, we have that

- $(f \circ g)(\mathbf{u}) = f\big(g(\mathbf{u})\big) = A(A^{-1}\mathbf{u}) = (AA^{-1})\mathbf{u} = I_n\mathbf{u} = \mathbf{u}$;
- $(g \circ f)(\mathbf{u}) = g\big(f(\mathbf{u})\big) = A^{-1}(A\mathbf{u}) = (A^{-1}A)\mathbf{u} = I_n\mathbf{u} = \mathbf{u}$.

This proves that $f \circ g = g \circ f = \mathrm{Id}_{\mathbb{F}^n}$, and it follows that f is indeed a bijection. □

As an easy corollary of Proposition 1.11.8 and Theorem 1.11.9, we obtain the following.

Corollary 1.11.10. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then the following are equivalent:*

(a) A is invertible;

(b) A^T is invertible;

(c) rank$(A) = n$;

(d) rank$(A^T) = n$.

Proof. By Theorem 1.11.9 applied to the matrix A, we have that (a) and (c) are equivalent. Similarly, by Theorem 1.11.9 applied to the matrix A^T, we have that (b) and (d) are equivalent. By Proposition 1.11.8(c) applied to the matrix A, we have that (a) implies (b). On the other hand, Proposition 1.11.8(c) applied to A^T guarantees that if A^T is invertible, then so is $(A^T)^T = A$, and so (b) implies (a). This completes the argument. □

Remark: By Corollary 1.11.10, a square matrix (with entries in some field) has full rank if and only if its transpose has full rank. In fact, the rank of any matrix is equal to the rank of its transpose (see Corollary 3.3.11), but we cannot prove this yet.

[67] Note that (2) states that (b) implies (a).

[68] This is obvious, but it also follows from Proposition 1.10.8.

1.11.5 Elementary matrices and row reduction

An *elementary matrix* is any matrix obtained by performing one elementary row operation on an identity matrix I_n. For an elementary row operation performed on a matrix with n rows, the elementary matrix that *corresponds* to this elementary row operation is the matrix obtained by performing that same elementary row operation on the identity matrix I_n. Let us consider some examples.

1. The elementary matrix that corresponds to swapping rows 2 and 4 ("$R_2 \leftrightarrow R_4$") of a matrix with 5 rows is
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

2. The elementary matrix that corresponds to multiplying the second row of a matrix with three rows by a scalar $\alpha \neq 0$ ("$R_2 \to \alpha R_2$") is
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

3. The elementary matrix that corresponds to adding α times the third row to the second row ("$R_2 \to R_2 + \alpha R_3$") of a matrix with three rows is
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{bmatrix}.$$

Proposition 1.11.11. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times m}$ be a matrix. Then both the following hold:*

(a) if R is any elementary row operation (performed on a matrix with n rows and with entries in $\mathbb{F}$) and E is the corresponding elementary matrix, then the matrix obtained from A by performing R on it is precisely the matrix EA;

(b) if $R_1, \ldots, R_k$ are elementary row operations (performed on a matrix with n rows and with entries in $\mathbb{F}$) and $E_1, \ldots, E_k \in \mathbb{F}^{n \times n}$ are, respectively, the corresponding elementary matrices, then the matrix obtained from A by performing $R_1, \ldots, R_k$ (in that order) on it is precisely the matrix $E_k \ldots E_1 A$.

Remark: In part (b), note the swapping of order between $R_1, \ldots, R_k$ on the one hand and $E_k \ldots E_1$ on the other.

Proof. We first prove (a). Consider any elementary row operation R performed on a matrix with n rows (and with entries in the field $\mathbb{F}$). Define $f_R : \mathbb{F}^n \to \mathbb{F}^n$ by, for each $\mathbf{u} \in \mathbb{F}^n$, letting $f(\mathbf{u})$ be the vector obtained by performing the elementary row operation R on $\mathbf{u}$. It is easy to see that f_R is linear.[69] So, f_R has a standard matrix. But clearly, the standard matrix of f_R is precisely the matrix E.[70]

Now, fix any matrix $A \in \mathbb{F}^{n\times m}$, and set $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$. Then

$$EA = \begin{bmatrix} E\mathbf{a}_1 & \dots & E\mathbf{a}_m \end{bmatrix} \overset{(*)}{=} \begin{bmatrix} f_R(\mathbf{a}_1) & \dots & f_R(\mathbf{a}_m) \end{bmatrix} =: M,$$

where (*) follows from the fact that E is the standard matrix of f_R. But obviously, the matrix M is precisely the matrix obtained by performing the elementary row operation R on A. This proves (a).

Part (b) follows from part (a) via an easy induction on k (the details are left as an exercise). □

Remark: Schematically (but somewhat informally), Proposition 1.11.11(b) yields the following:

$$A \overset{R_1}{\sim} E_1A \overset{R_2}{\sim} E_2E_1A \overset{R_3}{\sim} E_3E_2E_1A \overset{R_4}{\sim} \dots \overset{R_k}{\sim} E_k \dots E_3E_2E_1A.$$

Proposition 1.11.12. *Let $\mathbb{F}$ be a field. Then all the following hold:*

(a) elementary matrices in $\mathbb{F}^{n\times n}$ are invertible;

(b) the inverse of an elementary matrix in $\mathbb{F}^{n\times n}$ is an elementary matrix in $\mathbb{F}^{n\times n}$;

(c) a matrix $A \in \mathbb{F}^{n\times n}$ is invertible if and only if there exist elementary matrices $E_1, \dots, E_k$ such that $A = E_1 \dots E_k$ (that is, a matrix is invertible if and only if it can be written as a product of elementary matrices).

Proof. We prove (a) and (b) simultaneously. Let R be an elementary row operation performed on a matrix with n rows (and with entries in the field $\mathbb{F}$), and let E be the elementary matrix that corresponds to R. Let R' be the elementary row operation that "undoes" R,[71] and let E' be the elementary matrix that corresponds to R'. But now Proposition 1.11.11 guarantees that $EE' = E'E = I_n$.[72] This proves that E is invertible, and that its inverse is the elementary matrix E'. This proves (a) and (b).

[69] Check this!

[70] Indeed, the standard matrix of R is $\begin{bmatrix} f_R(\mathbf{e}_1) & \dots & f_R(\mathbf{e}_n) \end{bmatrix}$, which is precisely the matrix obtained from I_n by applying the elementary row operation R to it, and this matrix is precisely the elementary matrix E.

[71] See subsection 1.3.2.

[72] Let us explain this in detail. By Proposition 1.11.11, applying the elementary row operation R (resp. R') to a matrix in $\mathbb{F}^{n\times n}$ is the same as multiplying that matrix on the left by the elementary matrix E (resp. E'). If we apply R to the matrix I_n, and then apply R' to the resulting matrix, we obtain I_n back. So, if we multiply I_n by E on the left, and then multiply the resulting matrix by E' on the left, we obtain I_n; so, $E'EI_n = I_n$, and consequently, $E'E = I_n$. Analogously, $EE' = I_n$.

Let us now prove (c). The fact that products of elementary matrices are invertible follows immediately from part (a) and from the fact that (by Proposition 1.11.8(e)) products of invertible matrices are invertible. For the reverse direction, we fix an arbitrary invertible matrix $A \in \mathbb{F}^{n\times n}$, and we show that A can be written as a product of elementary matrices. Since A is invertible, Proposition 1.11.9 guarantees that $\mathrm{RREF}(A) = I_n$. In particular, A and I_n are row equivalent, and it follows that we can transform I_n into A via some sequence $R_1, \dots, R_k$ of elementary row operations. For each index $i \in \{1, \dots, k\}$, let $E_i \in \mathbb{F}^{n\times n}$ be the elementary matrix that corresponds to the elementary row operation R_i. But then by Proposition 1.11.11(b), we have that $A = E_k \dots E_1 I_n = E_k \dots E_1$. This proves (c). □

As a corollary of Propositions 1.11.11(b) and 1.11.12(c), we obtain the following theorem.

Theorem 1.11.13. *Let $\mathbb{F}$ be a field, and let $A, B \in \mathbb{F}^{n\times m}$. Then the following are equivalent:*

(a) $A \sim B$;

(b) there exist elementary matrices $E_1, \dots, E_k \in \mathbb{F}^{n\times n}$ such that $B = E_1 \dots E_k A$;

(c) there exists an invertible matrix $C \in \mathbb{F}^{n\times n}$ such that $B = CA$.

Proof. By definition, (a) is equivalent to:

(a') B can be obtained from A via some sequence of elementary row operations.

But Proposition 1.11.11(b) guarantees that (a') and (b) are equivalent, and Proposition 1.11.12(c) guarantees that (b) and (c) are equivalent. This completes the argument. □

1.11.6 Proof of Theorem 1.11.4

We are now ready to prove Theorem 1.11.4, restated below for the reader's convenience.

Theorem 1.11.4. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ be a square matrix, and set $\left[\begin{array}{c|c} U & B \end{array}\right] = \mathrm{RREF}\left(\left[\begin{array}{c|c} A & I_n \end{array}\right]\right)$, where each of U and B has n columns. Then*

(a) if $U = I_n$, then A is invertible and $B = A^{-1}$;

(b) if $U \neq I_n$, then A is not invertible.

Proof. By Theorem 1.11.9, we have that A is invertible if and only if $\mathrm{RREF}(A) = I_n$, and since $\left[\begin{array}{c|c} U & B \end{array}\right] = \mathrm{RREF}\left(\left[\begin{array}{c|c} A & I_n \end{array}\right]\right)$, we have that $\mathrm{RREF}(A) = U$.[73] So, if $U \neq I_n$, then A is not invertible; this proves (b) holds. Assume now that $U = I_n$,

[73] This is "obvious," but it also follows from Proposition 1.3.20(b).

so that A is invertible. To prove (a), it now remains to show that $B = A^{-1}$. Since $\left[\begin{array}{c|c} A & I_n \end{array}\right]$ and $\left[\begin{array}{c|c} I_n & B \end{array}\right]$ are row equivalent,[74] Theorem 1.11.13 guarantees that there exists an invertible matrix $C \in \mathbb{F}^{n\times n}$ such that $C\left[\begin{array}{c|c} A & I_n \end{array}\right] = \left[\begin{array}{c|c} I_n & B \end{array}\right]$. But note that $C\left[\begin{array}{c|c} A & I_n \end{array}\right] = \left[\begin{array}{c|c} CA & C \end{array}\right]$.[75] So, $\left[\begin{array}{c|c} CA & C \end{array}\right] = C\left[\begin{array}{c|c} A & I_n \end{array}\right] = \left[\begin{array}{c|c} I_n & B \end{array}\right]$, which in turn implies that $CA = I_n$ and $C = B$, and consequently, $BA = I_n$. But we already saw that A is invertible, and so Proposition 1.11.3 guarantees that $A^{-1} = B$. □

1.11.7 The Invertible matrix theorem (version 1)

The following theorem, which essentially summarizes the results of this section (plus Theorems 1.6.8 and 1.10.18), gives several equivalent characterizations of invertible matrices. Later in these lecture notes (see subsections 3.3.6, 7.4.1, and 8.2.6), we will add several more equivalent characterizations of invertible matrices.

Warning: The Invertible Matrix Theorem only works for **square** matrices. Do **not** attempt to apply it to matrices that are not square (or to linear functions whose standard matrices are not square)!

The Invertible Matrix Theorem (version 1). *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a **square** matrix. Further, let $f : \mathbb{F}^n \to \mathbb{F}^n$ be given by $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$.*[76] *Then the following are equivalent:*

(a) A is invertible (i.e. A has an inverse);

(b) A^T is invertible;

(c) $RREF(A) = I_n$;

(d) $RREF\left(\left[\begin{array}{c|c} A & I_n \end{array}\right]\right) = \left[\begin{array}{c|c} I_n & B \end{array}\right]$ for some matrix $B \in \mathbb{F}^{n\times n}$;

(e) $rank(A) = n$;

[74] This is because $\left[\begin{array}{c|c} I_n & B \end{array}\right] = \left[\begin{array}{c|c} U & B \end{array}\right] = \text{RREF}\left(\left[\begin{array}{c|c} A & I_n \end{array}\right]\right)$.

[75] This is "obvious," but here are the details. Set $A = \left[\begin{array}{ccc} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{array}\right]$ and set $C = \left[\begin{array}{ccc} \mathbf{c}_1 & \dots & \mathbf{c}_n \end{array}\right]$. Then

$$\begin{aligned} C\left[\begin{array}{c|c} A & I_n \end{array}\right] &= C\left[\begin{array}{ccc|ccc} \mathbf{a}_1 & \dots & \mathbf{a}_n & \mathbf{e}_1 & \dots & \mathbf{e}_n \end{array}\right] \\ &= \left[\begin{array}{ccc|ccc} C\mathbf{a}_1 & \dots & C\mathbf{a}_n & C\mathbf{e}_1 & \dots & C\mathbf{e}_n \end{array}\right] \\ &= \left[\begin{array}{ccc|ccc} C\mathbf{a}_1 & \dots & C\mathbf{a}_n & \mathbf{c}_1 & \dots & \mathbf{c}_n \end{array}\right] \\ &= \left[\begin{array}{c|c} CA & C \end{array}\right]. \end{aligned}$$

[76] Since f is a matrix equation, Proposition 1.10.4 guarantees that f is linear. Moreover, A is the standard matrix of f.

(f) $rank(A^T) = n$;

(g) A *is a product of elementary matrices;*

(h) the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ *has only the trivial solution (i.e. the solution* $\mathbf{x} = \mathbf{0}$*);*

(i) there exists some vector $\mathbf{b} \in \mathbb{F}^n$ *such that the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *has a unique solution;*

(j) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *has a unique solution;*

(k) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *has at most one solution;*

(l) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *is consistent;*

(m) f *is one-to-one;*

(n) f *is onto;*

(o) f *is an isomorphism.*

Proof. By Theorem 1.11.9, (a), (c), (e), and (o) are equivalent, and by Corollary 1.11.10, (a), (b), and (f) are equivalent. By Theorem 1.11.4, (a) and (d) are equivalent. By Proposition 1.11.12(c), we have that (a) and (g) are equivalent. So far, we have shown that (a), (b), (c), (d), (e), (f), (g), and (o) are equivalent.

Next, by Theorem 1.6.8, (e), (h), (i), (j), (k), and (l) are equivalent.

Finally, by Theorem 1.10.18(a), we have that (m) and (e) are equivalent, and by Theorem 1.10.18(b), we have that (n) and (e) are equivalent. This completes the argument. □

Chapter 2

Groups and permutations. Fields

2.1 Monoids

A *monoid* is an ordered pair $(S, \circ)$, where S is a set and $\circ$ is a binary operation on S (i.e. $\circ : S \times S \to S$), satisfying the following two axioms:

1. the operation $\circ$ is associative, i.e. for all $a, b, c \in S$, we have that
$$a \circ (b \circ c) \quad = \quad (a \circ b) \circ c;$$
2. there exists some $e \in S$, called the *identity element* of $(S, \circ)$, such that for all $a \in S$, we have that
$$e \circ a \quad = \quad a \qquad \text{and} \qquad a \circ e \quad = \quad a.$$

Proposition 2.1.1. *Every monoid has a unique identity element.*

Proof. Let $(S, \circ)$ be a monoid. By definition (in particular, by axiom 2), the monoid $(S, \circ)$ has an identity element; we must show that this identity element is unique. Suppose that e_1, e_2 are identity elements of $(S, \circ)$.[1] Then

$$e_1 \quad \overset{(*)}{=} \quad e_1 \circ e_2 \quad \overset{(**)}{=} \quad e_2$$

where (*) follows from the fact that e_2 is the identity element of the monoid $(S, \circ)$, and (**) follows from the fact that e_1 is the identity element of the monoid $(S, \circ)$. So, the identity element of the monoid $(S, \circ)$ is unique. □

[1] This means that the following hold:

- for all $a \in S$, we have that $e_1 \circ a = a$ and $a \circ e_1 = a$;
- for all $a \in S$, we have that $e_2 \circ a = a$ and $a \circ e_2 = a$.

Example 2.1.2. *All the following are monoids:*

1. $(\mathbb{N}_0, +)$*;* *2.* $(\mathbb{Z}, +)$*;* *3.* $(\mathbb{Q}, +)$*;* *4.* $(\mathbb{R}, +)$*;* *5.* $(\mathbb{C}, +)$.

In each of the above, 0 *is the identity element.*

Remark: $(\mathbb{N}, +)$ is **not** a monoid, since it does not have an identity element.

Example 2.1.3. *All the following are monoids ("$\cdot$" denotes multiplication):*

1. $(\mathbb{N}_0, \cdot)$*;* *3.* $(\mathbb{Z}, \cdot)$*;* *5.* $(\mathbb{R}, \cdot)$*;*

2. $(\mathbb{N}, \cdot)$*;* *4.* $(\mathbb{Q}, \cdot)$*;* *6.* $(\mathbb{C}, \cdot)$.

In each of the above, 1 *is the identity element.*

Example 2.1.4. *All the following are monoids ("$\cdot$" denotes multiplication):*

1. $(\mathbb{N}, \cdot)$*;* *3.* $(\mathbb{Q} \setminus \{0\}, \cdot)$*;* *5.* $(\mathbb{C} \setminus \{0\}, \cdot)$.

2. $(\mathbb{Z} \setminus \{0\}, \cdot)$*;* *4.* $(\mathbb{R} \setminus \{0\}, \cdot)$*;*

In each of the above, 1 *is the identity element.*

2.2 Groups

2.2.1 Groups: definition and basic properties

A *group* is an ordered pair $(G, \circ)$, where G is a set and $\circ$ is a binary operation on G (i.e. $\circ : G \times G \to G$) that satisfy the following three axioms:

1. the operation $\circ$ is associative, i.e. for all $a, b, c \in G$, we have that
$$a \circ (b \circ c) \;=\; (a \circ b) \circ c;$$
2. there exists some $e \in G$, called the *identity element* of $(G, \circ)$, such that for all $a \in G$, we have that
$$e \circ a \;=\; a \qquad \text{and} \qquad a \circ e \;=\; a;$$
3. for all $a \in G$, there exists some $a' \in G$, called the *inverse* of a, such that
$$a \circ a' \;=\; e \qquad \text{and} \qquad a' \circ a \;=\; e.$$

An *abelian group* is a group $(G, \circ)$ that satisfies the following additional axiom:

4. the operation $\circ$ is commutative, i.e. for all $a, b \in G$, we have that

$$a \circ b = b \circ a.$$

A *non-abelian* group is a group that is not abelian.

Remark: Note that the first two axioms (axioms 1 and 2) from the definition of a group are precisely the monoid axioms. So, every group is a monoid. By Proposition 2.1.1, it follows that the identity element e of a group is unique. In particular, the third axiom (axiom 3) makes sense.

Terminology/Notation: If the operation $\circ$ of the group $(G, \circ)$ is clear from context, then we may say that G is a group, rather than that $(G, \circ)$ is a group. However, this is only done if there is no chance of confusion, and so when in doubt, you should specify the operation. Sometimes, we say "G is a group under the operation $\circ$," which means exactly the same thing as "$(G, \circ)$ is a group."

Proposition 2.2.1. *Each element of a group has a* ***unique*** *inverse.*

Proof. Let $(G, \circ)$ be a group, and let e be its identity element. Fix some $g \in G$. By the definition of a group (and in particular, by axiom 3), g has an inverse in the group $(G, \circ)$; we must show that this inverse is unique. Let g_1 and g_2 be inverses of g in the group $(G, \circ)$.[2] Then

$$\begin{aligned}
g_1 &= g_1 \circ e && \text{because } e \text{ is the identity element of } (G, \circ) \\
&= g_1 \circ (g \circ g_2) && \text{because } g_2 \text{ is an inverse of } g \\
&= (g_1 \circ g) \circ g_2 && \text{because } \circ \text{ is associative} \\
&= e \circ g_2 && \text{because } g_1 \text{ is an inverse of } g \\
&= g_2 && \text{because } e \text{ is the identity element of } (G, \circ).
\end{aligned}$$

We have now shown that $g_1 = g_2$. So, the inverse of g is unique. □

Notation: Typically, the (unique) inverse of an element g of a group $(G, \circ)$ is denoted by g^{-1}. However, when the group operation is denoted by $+$ (note: this is typically done only if the group is abelian), then the inverse of an element g is denoted by $-g$.

[2] This means that both the following hold:

- $g \circ g_1 = e$ and $g_1 \circ g = e$;
- $g \circ g_2 = e$ and $g_2 \circ g = e$.

Example 2.2.2. *All the following are abelian groups:*

1. $(\mathbb{Z}, +)$; *2.* $(\mathbb{Q}, +)$; *3.* $(\mathbb{R}, +)$; *4.* $(\mathbb{C}, +)$.

In each of the above cases, the identity element is 0*, and the inverse of a group element* g *is* $-g$.[3]

Note that the monoid $(\mathbb{N}_0, +)$ is **not** a group because elements other than 0 do not have inverses, and so axiom 3 from the definition of a group is not satisfied.

Example 2.2.3. *All the following are abelian groups:*

1. $(\mathbb{Q} \setminus \{0\}, \cdot)$; *2.* $(\mathbb{R} \setminus \{0\}, \cdot)$; *3.* $(\mathbb{C} \setminus \{0\}, \cdot)$.

In each of the above cases, the identity element is 1*, and the inverse of a group element* g *is* $g^{-1} = \frac{1}{g}$.[4]

Remark: Monoids $(\mathbb{Q}, \cdot)$, $(\mathbb{R}, \cdot)$, and $(\mathbb{C}, \cdot)$ are **not** groups because, in each of those cases, 0 does not have an inverse element. Note also that $(\mathbb{Z} \setminus \{0\}, \cdot)$ is **not** a group because elements other than 1 and -1 do not have inverses.

Remark: It might now seem that all groups are abelian. However, this is not the case: we will see examples of non-abelian groups in subsection 2.2.2 and in section 2.3.

Proposition 2.2.4. *Let* $(G, \circ)$ *be a group with identity element* e*. Then all the following hold (here, the inverse of a group element* g *is denoted by* g^{-1}*):*

(a) for all $a, b, c \in G$*, if* $a \circ b = a \circ c$*, then* $b = c$*;*

(b) for all $a, b, c \in G$*, if* $b \circ a = c \circ a$*, then* $b = c$*;*

(c) for all $a, b \in G$*, there exists a unique* $x \in G$ *such that* $a \circ x = b$*;*

(d) for all $a, b \in G$*, there exists a unique* $x \in G$ *such that* $x \circ a = b$*;*

(e) for all $a \in G$*,* $(a^{-1})^{-1} = a$;[5]

(f) for all $a, b \in G$*,* $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$.

Proof. We first prove (a). Fix $a, b, c \in G$, and assume that $a \circ b = a \circ c$. Then

[3] For example, in the group $(\mathbb{R}, +)$, the inverse of $\sqrt{13}$ is $-\sqrt{13}$.

[4] For example, in the group $(\mathbb{R} \setminus \{0\}, \cdot)$, the inverse of $\sqrt{13}$ is $\frac{1}{\sqrt{13}}$.

[5] So, the inverse of the inverse of a is equal to a.

$$\begin{aligned}
b &= e \circ b && \text{because } e \text{ is the identity element of } (G, \circ) \\
&= (a^{-1} \circ a) \circ b && \text{because } a^{-1} \circ a = e \\
&= a^{-1} \circ (a \circ b) && \text{because } \circ \text{ is associative} \\
&= a^{-1} \circ (a \circ c) && \text{because } a \circ b = a \circ c \\
&= (a^{-1} \circ a) \circ c && \text{because } \circ \text{ is associative} \\
&= e \circ c && \text{because } a^{-1} \circ a = e \\
&= c && \text{because } e \text{ is the identity element of } (G, \circ).
\end{aligned}$$

This proves (a). The proof of (b) is similar.

Next, we prove (c). Fix $a, b \in G$. We must show that there exists a unique $x \in G$ such that $a \circ x = b$. For existence, we set $x := a^{-1} \circ b$, and we observe that

$$\begin{aligned}
a \circ x &= a \circ (a^{-1} \circ b) && \text{because } x = a^{-1} \circ b \\
&= (a \circ a^{-1}) \circ b && \text{because } \circ \text{ is associative} \\
&= e \circ b && \text{because } a \circ a^{-1} = e \\
&= b && \text{becuase } e \text{ is the identity element of } (G, \circ).
\end{aligned}$$

Uniqueness follows from (a). This proves (c). The proof of (d) is similar.

We now prove (e). Fix $a \in G$. It suffices to show that $a^{-1} \circ (a^{-1})^{-1} = a^{-1} \circ a$, for then (a) will guarantee that $(a^{-1})^{-1} = a$, which is what we need. Since $(a^{-1})^{-1}$ is the inverse of a^{-1}, we know that $a^{-1} \circ (a^{-1})^{-1} = e$. On the other hand, since a^{-1} is the inverse of a, we have that $a^{-1} \circ a = e$. Thus, $a^{-1} \circ (a^{-1})^{-1} = a^{-1} \circ a$. As explained above, this implies that $(a^{-1})^{-1} = a$. This proves (e).

It remains to prove (f). Fix $a, b \in G$. We observe that

$$\begin{aligned}
(a \circ b) \circ (b^{-1} \circ a^{-1}) &= a \circ (b \circ b^{-1}) \circ a && \text{because } \circ \text{ is associative} \\
&= a \circ e \circ a^{-1} && \text{because } b \circ b^{-1} = e
\end{aligned}$$

$$\begin{aligned} &= a \circ a^{-1} && \text{because } e \text{ is the identity element of } (G, \circ) \\ &= e && \text{because } a \circ a^{-1} = e, \end{aligned}$$

and similarly,

$$\begin{aligned} (b^{-1} \circ a^{-1}) \circ (a \circ b) &= b^{-1} \circ (a^{-1} \circ a) \circ b && \text{because } \circ \text{ is associative} \\ &= b^{-1} \circ e \circ b && \text{because } a^{-1} \circ a = e \\ &= b^{-1} \circ b && \text{because } e \text{ is the identity element of } (G, \circ) \\ &= e && \text{because } b^{-1} \circ b = e. \end{aligned}$$

We have now shown that $(a \circ b) \circ (b^{-1} \circ a^{-1}) = e$ and $(b^{-1} \circ a^{-1}) \circ (a \circ b) = e$. It follows that $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$. This proves (f). □

The case of $\mathbb{Z}_n$ and $\mathbb{Z}_p$. For $\mathbb{Z}_n$ (where n is a positive integer) and $\mathbb{Z}_p$ (where p is a prime number), we have Proposition 2.2.5 (below). We note that part (b) crucially relies on Fermat's Little Theorem, stated and proven in subsection 0.2.2, and restated below for the reader's convenience.

Fermat's Little Theorem. *If $p \in \mathbb{N}$ is a prime number and $a \in \mathbb{Z}_p \setminus \{0\}$, then $a^{p-1} = 1$.*

Proposition 2.2.5.

(a) For all positive integers n, $(\mathbb{Z}_n, +)$ is an abelian group whose identity element is $0 := [0]_n$.

*(b) For all **prime** numbers p, $(\mathbb{Z}_p \setminus \{0\}, \cdot)$ is an abelian group whose identity element is $1 := [1]_p$.*

Proof. (a) Fix a positive integer n. The fact that $+$ ("addition") is an associative and commutative binary operation on $\mathbb{Z}_n$ follows from Proposition 0.2.11. The identity element of $\mathbb{Z}_n$ is $0 := [0]_n$. For each element $[a]_n$ in $\mathbb{Z}_n$ (where $a \in \mathbb{Z}$), the additive inverse of $[a]_n$ is $[-a]_n = [n - a]_n$. So, $(\mathbb{Z}_n, +)$ is an abelian group with identity element $[0]_n$.

(b) Fix a prime number p. By Proposition 0.2.11, we know that $\cdot$ ("multiplication") is an associative and commutative binary operation on $\mathbb{Z}_p$. However, the question

is whether multiplication remains a binary operation on $\mathbb{Z}_p \setminus \{0\}$, that is, whether $\mathbb{Z}_p \setminus \{0\}$ is "closed under multiplication," that is, whether the product of two numbers in $\mathbb{Z}_p \setminus \{0\}$ is always another number in $\mathbb{Z}_p \setminus \{0\}$.[6] So, fix $a, b \in \mathbb{Z}$ such that $[a]_p$ and $[b]_p$ are both non-zero (in $\mathbb{Z}_p$), i.e. p divides neither a nor b. Since p is **prime**, p does not divide the product ab,[7] and consequently, $[a]_p[b]_p = [ab]_p \neq 0$. So, multiplication is indeed a binary operation on $\mathbb{Z}_p \setminus \{0\}$. The identity element of $(\mathbb{Z}_p \setminus \{0\}, \cdot)$ is $1 := [1]_p$. Moreover, by Fermat's Little Theorem, each number $a \in \mathbb{Z}_p \setminus \{0\}$ has a multiplicative inverse, namely, a^{p-2}. This proves that $(\mathbb{Z}_p \setminus \{0\}, \cdot)$ is indeed an abelian group. □

Remark: If n is a positive integer that is not prime, then $(\mathbb{Z}_n \setminus \{0\}, \cdot)$ is **not** a group. Indeed, if $n = 1$, then $\mathbb{Z}_n \setminus \{0\}$ is empty and therefore not a group under any operation (no group is empty, since it must, at a minimum, contain an identity element). On the other hand, if $n \geq 2$ is a composite number, say $n = pq$ for some integers $p, q \geq 2$, then we have that $[p]_n, [q]_n \in \mathbb{Z}_n \setminus \{0\}$, but $[p]_n[q]_n = [pq]_n = [n]_n = 0$, and it follows that $\mathbb{Z}_n \setminus \{0\}$ is not closed under multiplication, i.e. multiplication is not a binary operation on $\mathbb{Z}_n \setminus \{0\}$.

2.2.2 Groups of matrices and vectors. The general linear group

Let $\mathbb{F}$ is a field. Since we have not formally studied fields yet, you may assume for now that $\mathbb{F}$ is one of the following: $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p$ (where p is a prime number). However, the examples given in this subsection work for all fields, not just the four listed above. First of all, it is obvious that $(\mathbb{F}^{n\times m}, +)$ is an abelian group whose identity element is the zero matrix $O_{n\times m}$; the (additive) inverse of a matrix $\left[\ a_{i,j}\ \right]_{n\times m}$ in the group $(\mathbb{F}^{n\times m}, +)$ is the matrix $\left[\ -a_{i,j}\ \right]_{n\times m}$ (i.e. the $n \times m$ matrix whose i,j-th entry is $-a_{i,j}$ for all indices $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$). In particular, $(\mathbb{F}^n, +)$ is an abelian group (with identity element $\mathbf{0}$).[8] More interestingly, consider the set $\mathrm{GL}_n(\mathbb{F})$ of all **invertible** matrices in $\mathbb{F}^{n\times n}$. $\mathrm{GL}_n(\mathbb{F})$ is a group under matrix multiplication, called the *general linear group of degree n over the field $\mathbb{F}$*. The identity element of $\mathrm{GL}_n(\mathbb{F})$ is the identity matrix I_n, and the inverse of a matrix A in $\mathrm{GL}_n(\mathbb{F})$ is the matrix A^{-1} (the usual matrix inverse that we studied in section 1.11). The group $\mathrm{GL}_1(\mathbb{F})$ is abelian (because multiplication is commutative in the field $\mathbb{F}$). However, for $n \geq 2$, the group $\mathrm{GL}_n(\mathbb{F})$ is **not** abelian. Let us first check this for $n = 2$, and then we will generalize. Consider the following two matrices in $\mathbb{F}^{2\times 2}$:

$$A_2 := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad B_2 := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

[6] In other words, the question is whether the product of two **non-zero** numbers in $\mathbb{Z}_p$ is always a **non-zero** number in $\mathbb{Z}_p$.

[7] **Remark:** This is why we care about p being prime! If p were not prime, then this implication would be invalid.

[8] We are using the fact that, by definition, $\mathbb{F}^n = \mathbb{F}^{n\times 1}$.

Both of these matrices have rank 2, and so by the Invertible Matrix Theorem (see subsection 1.11.7), they are both invertible and therefore belong to $\mathrm{GL}_2(\mathbb{F})$. However, we have that

- $A_2B_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1+1 & 1 \\ 1 & 1 \end{bmatrix}$,
- $B_2A_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1+1 \end{bmatrix}$.

Since $1+1 \neq 1$,[9] we see that $A_2B_2 \neq B_2A_2$, and so $\mathrm{GL}_2(\mathbb{F})$ is not abelian. Let us now generalize this. Fix an integer $n \geq 2$, and consider the following two matrices in $\mathbb{F}^{n\times n}$:

$$A_n := \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}, \qquad B_n := \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}.$$

It is then easy to see that $\mathrm{rank}(A_n) = \mathrm{rank}(B_n) = n$, and so by the Invertible Matrix Theorem (see subsection 1.11.7), A_n and B_n are both invertible and consequently belong to $\mathrm{GL}_n(\mathbb{F})$. Moreover, the 1,1-th entry (the one in the upper left corner) of A_nB_n is $1+1$, whereas the 1,1-th entry of B_nA_n is 1. Since $1+1 \neq 1$, we see that $A_nB_n \neq B_nA_n$, and it follows that the group $\mathrm{GL}_n(\mathbb{F})$ is not abelian.

Remark: The fact that $1+1 \neq 1$ is obviously true for the fields that we are familiar with. But in fact, it is true in **any** field $\mathbb{F}$, not just those that we have seen so far, and it essentially follows from the fact that $1 \neq 0$ (which is true for any field; see axiom 3 from the definition of a field in section 2.4). On the other hand, $1+1+1 = 1$ is true in some fields (for example, it is true for the field $\mathbb{Z}_2$).

2.2.3 Subgroups

A *subgroup* of a group $(G, \circ)$ is a group $(H, \diamond)$ such that $H \subseteq G$ and for all $a, b \in H$, we have that $a \diamond b = a \circ b$. If $(H, \diamond)$ is a subgroup of $(G, \circ)$, then we write $(H, \diamond) \leq (G, \circ)$. Here, $\diamond$ is the restriction of $\circ$ to H, and it is important that $a \diamond b = a \circ b \in H$ for all $a, b \in H$ (otherwise, H is not "closed under" $\diamond$, which means that $\diamond$ is not a binary operation on H, and in particular, $(H, \diamond)$ is not a group). Normally, we do not notationally distinguish between $\diamond$ and $\circ$, and we speak about $(H, \circ)$ being a subgroup of $(G, \circ)$, where it is understood from context that the operation $\circ$ from $(H, \circ)$ is the restriction of the the binary operation $\circ$ on G to H.

[9] See the Remark at the end of this subsection.

Example 2.2.6. *Every group $(G, \circ)$ has at least two subgroups: $(G, \circ)$ and $(\{e\}, \circ)$, where e is the identity element of G.*

Example 2.2.7. $(\mathbb{Z}, +) \leq (\mathbb{Q}, +) \leq (\mathbb{R}, +) \leq (\mathbb{C}, +)$.

Example 2.2.8. $(\mathbb{Q} \setminus \{0\}, \cdot) \leq (\mathbb{R} \setminus \{0\}, \cdot) \leq (\mathbb{C} \setminus \{0\}, \cdot)$.

Theorem 2.2.9. *Let $(G, \circ)$ be a group with identity element e, and with the inverse of an element $a \in G$ denoted by a^{-1}. Then for all $H \subseteq G$, we have that $(H, \circ)$ is a subgroup of $(G, \circ)$ if and only if all the following hold:*

(i) $e \in H$;

(ii) H is closed under $\circ$, that is, for all $a, b \in H$, we have that $a \circ b \in H$;

(iii) H is closed under inverses, that is, for all $a \in H$, we have that $a^{-1} \in H$.

Proof. Fix $H \subseteq G$. Suppose first that (i), (ii), and (iii) hold. By (ii), the binary operation $\circ$ on G can be restricted to H (so that it becomes a binary operation on H). The fact that $\circ$ is associative in $(H, \circ)$ follows simply from the fact that $\circ$ is inherited from the group $(G, \circ)$, where it is associative. By (i), H contains an identity element, and by (iii), every element of H has an inverse in $(H, \circ)$.

Suppose, conversely, that $(H, \circ)$ is a subgroup of $(G, \circ)$. Then (ii) holds, because $\circ$ (properly restricted) is a binary operation on H. It remains to prove that (i) and (iii) hold. Since H is a group, it must have an identity element e_H, and each element of H must have inverse in $(H, \circ)$. The question is whether the identity element of $(H, \circ)$ is the same as in $(G, \circ)$, and similar for inverses.[10] We first deal with the identity element. If we compute in $(H, \circ)$, we have that $e_H \circ e_H = e_H$ (because e_H is the identity element of $(H, \circ)$), and if we compute in $(G, \circ)$, then we have that $e_H \circ e = e_H$ (because e is the identity element of $(G, \circ)$). But now $e_H \circ e_H = e_H \circ e$, and so by Proposition 2.2.4(a) applied to $(G, \circ)$, we have that $e_H = e$. So, $e \in H$, and it follows that (i) holds. Finally, fix $a \in H$. Since $(H, \circ)$ is a group, a has an inverse a' in $(H, \circ)$, so that $a \circ a' = e_H = e$. On the other hand, if we compute in $(G, \circ)$, we get that $a \circ a^{-1} = e$. It follows that $a \circ a' = a \circ a^{-1}$, and so by Proposition 2.2.4(a) applied to $(G, \circ)$, we have that $a' = a^{-1}$, and consequently, $a^{-1} \in H$. This proves (ii). □

[10] Could it be that $e_H \neq e$, i.e. that $(H, \circ)$ has an identity element, but one that is different from the identity element of $(G, \circ)$? Could something similar happen with inverses? Actually, this cannot happen, but we need to prove that!

2.3 Permutations and the symmetric group

A *permutation* of a set X is any bijection from X to itself. The set of all permutations of X is denoted by $\mathrm{Sym}(X)$. As usual, Id_X is the identity function on X, i.e. $\mathrm{Id}_X : X \to X$ is given by $\mathrm{Id}_X(x) = x$ for all $x \in X$.

Note that for any set X, $(\mathrm{Sym}(X), \circ)$ is a group, called the *symmetric group on* X (here, $\circ$ is the composition of functions). Let us justify this. First of all, by Proposition 1.10.17(c),[11] the composition of two permutations of X is a permutation of X, and so $\circ$ is indeed a binary operation on $\mathrm{Sym}(X)$. Moreover, it is clear that $\circ$ is associative; indeed, for any $\pi, \sigma, \tau \in \mathrm{Sym}(X)$, we have that $\pi \circ (\sigma \circ \tau) = (\pi \circ \sigma) \circ \tau$, because for all $x \in X$, we have the following:

$$\begin{aligned} \big(\pi \circ (\sigma \circ \tau)\big)(x) &= \pi\big((\sigma \circ \tau)(x)\big) \\ &= \pi\big(\sigma(\tau(x))\big) \\ &= (\pi \circ \sigma)\big(\tau(x)\big) \\ &= \big((\pi \circ \sigma) \circ \tau\big)(x). \end{aligned}$$

The identity element of this group is the identity function Id_X on X. The inverse element of any permutation $\pi \in \mathrm{Sym}(X)$ is the inverse permutation π^{-1}. (Since permutations are bijections, they have inverse functions, and moreover, those inverses are also bijections; see the comment following the proof of Proposition 1.10.15. We deduce that the inverse of a permutation of X is another permutation of X.)

If a set X has at most two elements, then it is easy to see that the group $\mathrm{Sym}(X)$ is abelian. However, if X has at least three elements, then X is **not** abelian, as we now show. Suppose that $|X| \geq 3$, and let a, b, c be pairwise distinct elements of X. Let $\sigma, \tau : X \to X$ be defined as follows:[12]

- $\sigma(a) = b$, $\sigma(b) = a$, and $\sigma(x) = x$ for all $x \in X \setminus \{a, b\}$;
- $\tau(a) = c$, $\tau(c) = a$, and $\tau(x) = x$ for all $x \in X \setminus \{a, c\}$.

Clearly, $\sigma, \tau \in \mathrm{Sym}(X)$. But now

- $(\sigma \circ \tau)(a) = \sigma\big(\tau(a)\big) = \sigma(c) = c$;
- $(\tau \circ \sigma)(a) = \tau\big(\sigma(a)\big) = \tau(b) = b$.

Since $b \neq c$, we have that $(\sigma \circ \tau)(a) \neq (\tau \circ \sigma)(a)$. So, $\sigma \circ \tau \neq \tau \circ \sigma$, and it follows that $\mathrm{Sym}(X)$ is not abelian.

[11] We apply Proposition 1.10.17(c) for $A = B = C = X$.

[12] The permutation σ swaps ("transposes") a and b, while leaving all other elements of X fixed. Similarly, the permutation τ swaps ("transposes") a and c, while leaving all other elements of X fixed. For more on transpositions, see subsection 2.3.3.

We particularly often consider $\mathrm{Sym}(X)$ for the case when $X = \{1, \ldots, n\}$ for some positive integer n. The set $\mathrm{Sym}(\{1, \ldots, n\})$ is also denoted by $\mathrm{Sym}(n)$, Sym_n, or S_n. In these lecture notes, we will consistently use the notation S_n. The group $(S_n, \circ)$ is called the *symmetric group of degree n*. Note that $|S_n| = n!$.

A permutation $\pi \in S_n$ can be represented in the following way:

$$\pi = \begin{pmatrix} 1 & 2 & \ldots & n \\ \pi(1) & \pi(2) & \ldots & \pi(n) \end{pmatrix}.$$

So, in the top row, we have numbers $1, 2, \ldots, n$, and in the bottom row, we have those same numbers in some order (determined by the permutation π). For example, the permutation $\pi \in S_4$ given by $\pi(1) = 3$, $\pi(2) = 2$, $\pi(3) = 4$, and $\pi(4) = 1$ can be represented as follows:

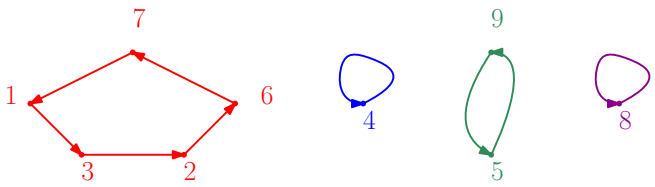
$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}.$$

2.3.1 Cycle notation

Suppose we are given the following permutation in S_9:

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 6 & 2 & 4 & 9 & 7 & 1 & 8 & 5 \end{pmatrix}.$$

We can represent this permutation geometrically, as shown below (the cycles are color coded for easier reference).



We can "encode" the picture that we obtained as a "product of disjoint cycles":

$$\pi = (13267)(4)(59)(8).$$

The above is also referred to as a "disjoint cycle decomposition" of the permutation π. The disjoint cycle decomposition of a permutation is unique up to cyclic permutation of the elements within each cycle, and up to a reordering of the cycles. For example, the permutation π above can also be expressed as follows:

$$\pi = (95)(26713)(8)(4).$$

However, the first disjoint cycle decomposition[13] is canonical/standard because it satisfies the following two properties:

[13] That is, the disjoint cycle decomposition $\pi = (13267)(4)(59)(8)$.

- within each cycle, the smallest number appears first;
- the first elements of the cycles from the disjoint cycle decomposition form an increasing sequence.[14]

Usually, the canonical representation is preferred, but occasionally, it may be more practical to use a non-canonical one. When the n from S_n is clear from context, one-element cycles may be omitted. So, if we know that we are working in S_9, then we may omit the one-element cycles (4) and (8) from the representation above, and write simply

$$\pi \quad = \quad (13267)(59).$$

In this case, the cycles (4) and (8) are understood from context. However, we can only do this when n has been specified beforehand! Otherwise, cycles of length one must be included.

Notation: When there is danger of confusion, we put commas between elements within cycles. For instance, if we are working in S_{12}, then (123) is ambiguous. To avoid ambiguity, we write $(1, 2, 3)$ or $(12, 3)$, as appropriate. However, if we are working in S_n, where n is a single-digit number, then there is no danger of confusion, and so we normally omit commas.

Let us consider some more examples.

Example 2.3.1. *Find the disjoint cycle decompositions of the following permutations.*

(a) $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 3 & 1 \end{pmatrix}$

(b) $\pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 4 & 1 & 6 & 5 \end{pmatrix}$

(c) $\pi_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 1 & 3 & 4 \end{pmatrix}$

Solution. We have:

(a) $\pi_1 = (125)(34)$;

(b) $\pi_2 = (134)(2)(56)$;

(c) $\pi_3 = (12543)$.

Note that in (b), we could also have written $\pi \in S_6$, $\pi = (134)(56)$. □

[14] Indeed, $1 < 4 < 5 < 8$.

It is also easy to go the other way around: from the disjoint cycle decomposition to the table representation, i.e. representation of the form

$$\pi = \begin{pmatrix} 1 & 2 & \dots & n \\ \pi(1) & \pi(2) & \dots & \pi(n) \end{pmatrix}.$$

For instance, we see that

$$(143)(26)(5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 6 & 1 & 3 & 5 & 2 \end{pmatrix}$$

and

$$(154362) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 1 & 6 & 3 & 4 & 2 \end{pmatrix}.$$

Compositions of permutations. By Proposition 1.10.17(c), the composition of two permutations in S_n is another permutation in S_n. For instance, in S_5, we have the following (with permutations color coded for easier reference):

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 5 & 4 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \end{pmatrix}.$$

As usual with function composition, we apply permutations from right to left with respect to $\circ$. So, in the case above, we first apply the blue permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{pmatrix}$, and then we apply the red permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 5 & 4 \end{pmatrix}$. For instance, 1 first gets mapped to 2 via the blue permutation, and then 2 gets mapped to 3 via the red permutation. So, the composition above maps 1 to 3.

We can similarly compose permutations specified in terms of their disjoint cycle decompositions. For instance,

$$(1)(23)(45) \circ (124)(35) = (134)(25).$$

Again we apply permutations from right to left with respect to $\circ$. So, in the case above, we first apply the blue permutation $(124)(35)$, and then we apply the red permutation $(1)(23)(45)$. However, within each permutation (separated by $\circ$'s from the other permutations), we read from left to right. For instance, in the blue permutation $(124)(35)$, 1 gets mapped to 2, 2 gets mapped to 4, and 4 gets mapped to 1.

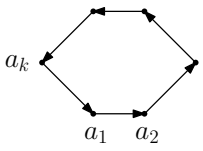
Again, when the n from S_n is clear from context, we may omit one-element cycles. For instance, in S_5, we have

$$(154) \circ (245)(13) \circ (25) = (135).$$

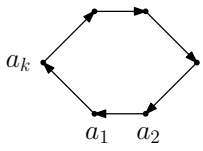
Here, certain one-element cycles are understood from context. In particular, $(154) = (154)(2)(3)$, $(25) = (1)(25)(3)(4)$, and $(135) = (135)(2)(4)$. So, the above expression can be rewritten as

$$(154)(2)(3) \circ (245)(13) \circ (1)(25)(3)(4) \quad = \quad (135)(2)(4).$$

Inverses of permutations. The inverse of a permutation π in S_n can be obtained by starting with a disjoint cycle decomposition of π, and then reversing the order of elements in all cycles, i.e. turning each cycle of the form $(a_1 a_2 \dots a_k)$ into $(a_k \dots a_2 a_1)$. Pictorially, we get the same cycles, only with arrows reversed (see the picture below).



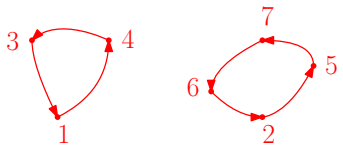
$(a_1 a_2 \dots a_k)$



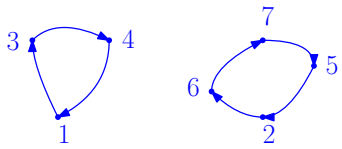
$(a_k \dots a_2 a_1)$

For example, in S_7:

- if $\pi_1 = (143)(2576)$, then $\pi_1^{-1} = (341)(6752) = (134)(2675)$;

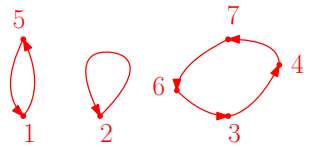


π_1

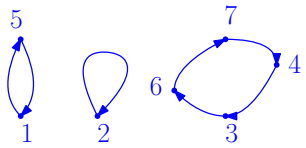


π_1^{-1}

- if $\pi_2 = (15)(2)(3476)$, then $\pi_2^{-1} = (51)(2)(6743) = (15)(2)(3674)$.



π_2



π_2^{-1}

Notation: The identity permutation in S_n is often denoted simply by 1. So, in this context, we have that

$$1 \quad = \quad (1)(2)\dots(n).$$

If we wish to emphasize n (or if we need to avoid confusion with other kinds of 1 that may appear in our proof/computation), then we can denote the identity permutation in S_n by 1_n.

2.3.2 The sign of a permutation. Even and odd permutations

Given a positive integer n and a permutation $\pi \in S_n$, the *sign* of π, denoted by $\mathrm{sgn}(\pi)$, is given by

$$\mathrm{sgn}(\pi) \quad = \quad (-1)^{n-k},$$

where k is the number of cycles in the disjoint cycle decomposition of π **including the one-element cycles**. For instance, for $\pi_1 = (1367)(2)(45)$ in S_7, we have

$$\mathrm{sgn}(\pi_1) \quad = \quad (-1)^{7-3} \quad = \quad 1,$$

whereas for $\pi_2 = (12)(345)(6)(7)$ in S_7, we have

$$\mathrm{sgn}(\pi_2) \quad = \quad (-1)^{7-4} \quad = \quad -1.$$

Equivalently, for $\pi \in S_n$, we have that

$$\mathrm{sgn}(\pi) \quad = \quad (-1)^{n'-k'},$$

where k' is the number of cycles in some disjoint cycles in some disjoint cycle decomposition of π (possibly with some one-element cycles omitted), and n' is the number of elements in those k' cycles. The two definitions are equivalent because if d is the number of omitted one-element cycles in some disjoint cycle decomposition of π , then $n = n' + d$, and if we write the complete disjoint cycle decomposition of π including all one-element cycles, then we get $k = k' + d$ many cycles. So, $n - k = n' - k'$, and consequently, $(-1)^{n-k} = (-1)^{n'-k'}$. For instance, for $\pi_3 = (123)(45)$ in S_7, we have

$$\mathrm{sgn}(\pi_3) \quad = \quad (-1)^{5-2} = -1.$$

Note that the one-element cycles (6) and (7) are implicitly understood for π_3, that is, $\pi_3 = (123)(45)(6)(7)$. And indeed, we have

$$\mathrm{sgn}(\pi_3) \quad = \quad (-1)^{7-4} = -1,$$

as before.

Remark: Note that for all positive integers n, the identity permutation in S_n has sign 1. This is because the identity permutation in S_n has disjoint cycle decomposition $(1)(2)\dots(n)$, and so its sign is $(-1)^{n-n} = (-1)^0 = 1$.

Terminology: Permutations whose sign is $+1$ are called *even*, and permutations whose sign is -1 are called *odd*. Since the sign of the identity permutation is $+1$, the identity permutation is even.

Proposition 2.3.2. *Let $n \geq 2$ be an integer, and let π be a permutation in S_n. Then $sgn(\pi^{-1}) = sgn(\pi)$.*

Proof. This follows from the fact that π and π^{-1} have the same number of cycles in their disjoint cycle decompositions (when the one-element cycles are included). □

2.3.3 Transpositions

Slightly informally, a transposition is a permutation that swaps two elements and fixes all the remaining ones. More formally, given an integer $n \geq 2$, a *transposition* in S_n is a permutation $\pi \in S_n$ for which there exist distinct $i, j \in \{1, \ldots, n\}$ such that $\pi(i) = j$, $\pi(j) = i$, and $\pi(\ell) = \ell$ for all $\ell \in \{1, \ldots, n\} \setminus \{i, j\}$. Such a transposition is typically denoted by (ij), and the $n-2$ many one-element cycles are implicitly understood. For instance, the following permutation in S_5 is a transposition:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 4 & 2 \end{pmatrix} = (25).$$

Note that this transposition could also have been written in the form $(1)(25)(3)(4)$. More commonly, one-element cycles are omitted.

Remark: Every transposition is its own inverse, that is, for any transposition $\tau = (ij)$ in S_n $(n \geq 2)$, we have that $\tau^{-1} = \tau$.

Remark: The sign of any transposition is -1,[15] and so transpositions are odd.

As we shall see, for $n \geq 2$, any permutation can be written as a composition of transpositions. For instance, in S_7, we have

$$(134)(2657) = (13) \circ (34) \circ (26) \circ (65) \circ (57).$$

The correctness of the above can easily be verified by checking that the image of each element of $\{1, \ldots, 7\}$ under the permutations $(134)(2657)$ and $(13) \circ (34) \circ (26) \circ (65) \circ (57)$ is the same. Moreover, this works in general, as the following proposition shows.

Proposition 2.3.3. *Let $n \geq 2$ be an integer. Then any permutation in S_n can be written as a composition of transpositions.*

Proof. The identity permutation in S_n can be written in the form $(12) \circ (12)$.[16] Let us now suppose that π is some permutation in S_n other than the identity. Then π can be written as the product of one or more disjoint cycles of length at least two (one-element cycles are omitted in our expression, but are understood from context). Let us say we have k cycles of length at least two, as follows (to help the reader, the cycles are color coded):

$$\pi = (a_1^1 a_2^1 \ldots a_{\ell_1}^1) \ldots (a_1^k a_2^k \ldots a_{\ell_k}^k),$$

[15] This follows straight from the definition of the sign of a permutation. Indeed, if τ is a transposition in S_n $(n \geq 2)$, then the disjoint cycle decomposition of τ consists of one cycle of length two and $n-2$ many cycles of length one, and consequently, it consists of $n-1$ cycles total (when cycles of length one are included). So, $\operatorname{sgn}(\tau) = (-1)^{n-(n-1)} = -1$.

[16] Actually, the identity permutation can also be written as an "empty" composition of transpositions.

where the a_i^j's are pairwise distinct, and $\ell_1, \ldots, \ell_k \geq 2$. But then we have

$$\pi = (a_1^1 a_2^1) \circ (a_2^1 a_3^1) \circ \cdots \circ (a_{\ell_1-1}^1 a_{\ell_1}^1) \circ \cdots \circ (a_1^k a_2^k) \circ (a_2^k a_3^k) \circ \cdots \circ (a_{\ell_k-1}^k a_{\ell_k}^k),$$

and so π is the composition of transpositions. □

Example 2.3.4. *Express each of the following permutations in S_6 as the composition of transpositions.*

(a) $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 5 & 2 & 3 & 4 & 6 \end{pmatrix}$;

(b) $\pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix}$;

(c) $\pi_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 6 & 5 & 2 & 1 & 4 \end{pmatrix}$.

Solution. To help the reader, we color code the cycles that we obtain, as well as the transpositions that correspond to them.

(a) $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 5 & 2 & 3 & 4 & 6 \end{pmatrix} = (2543) = (25) \circ (54) \circ (43)$;

(b) $\pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix} = (12)(456) = (12) \circ (45) \circ (56)$;

(c) $\pi_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 6 & 5 & 2 & 1 & 4 \end{pmatrix} = (135)(264) = (13) \circ (35) \circ (26) \circ (64)$.

□

We note that the same permutation can be expressed as the composition of transpositions in more than one way. For instance, in S_5, we have:

- $(12345) = (12) \circ (23) \circ (34) \circ (45)$;
- $(12345) = (12) \circ (23) \circ (34) \circ (45) \circ (35) \circ (35)$;
- $(12345) = (15) \circ (14) \circ (13) \circ (12)$;
- $(12345) = (35) \circ (35) \circ (23) \circ (23) \circ (15) \circ (14) \circ (13) \circ (12) \circ (35) \circ (35)$.

However, as we shall see, for any given permutation π in S_n, where $n \geq 2$, in all representations of π as a composition of transpositions, the number of transpositions is of the same parity (i.e. it is either always even or always odd). We prove this in Theorem 2.3.6. However, to prove Theorem 2.3.6, we need the following technical proposition, which essentially states that composing a permutation with a transposition results in a sign change.

Proposition 2.3.5. *Let $n \geq 2$ be an integer. Then for all $\pi, \tau \in S_n$ such that τ is a* ***transposition****, we have that $sgn(\tau \circ \pi) = sgn(\pi \circ \tau) = -sgn(\pi)$.*

Warning: In general, $\tau \circ \pi \not= \pi \circ \tau$.

Proof. The Claim below proves one part of the proposition ("$\operatorname{sgn}(\tau \circ \pi) = -\operatorname{sgn}(\pi)$"). The other part ("$\operatorname{sgn}(\pi \circ \tau) = -\operatorname{sgn}(\pi)$") can be proven using the Claim and certain basic properties of permutations (as we shall see below).

> **Claim.** For all $\pi, \tau \in S_n$ such that τ is a transposition, we have that $\operatorname{sgn}(\tau \circ \pi) = -\operatorname{sgn}(\pi)$.

Proof of the Claim. Fix $\pi, \tau \in S_n$, and assume that $\tau = (ij)$ is a transposition (here, i and j are some two distinct elements of $\{1, \ldots, n\}$). There are two cases to consider: when i and j are in the same cycle of the disjoint cycle decomposition of π, and when they are in different cycles.

Case 1: i and j are in the same cycle of the disjoint cycle decomposition of π. After possibly swapping the order of our disjoint cycles, and cyclically permuting the elements of the cycle that contains i and j, we may assume that our disjoint cycle decomposition of π is given by

$$\pi = (i\ a_1 \ldots a_p\ j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r).$$

Note: Here, i and j are both in the red cycle. The remaining cycles (the ones that do not contain i and j) are colored blue.[17]

In the permutation $\tau \circ \pi$, the red cycle essentially gets "split up" into two, while the blue cycles remain unaffected, as follows:

$$\begin{aligned} \tau \circ \pi &= (ij) \circ (i\ a_1 \ldots a_p\ j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r) \\ &= \underbrace{(i\ a_1 \ldots a_p)(j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r)}_{=:\pi'}. \end{aligned}$$

The correctness of the above can easily be verified by checking that the permutations $\tau \circ \pi$ and π' do indeed map each element of

$$\{1, \ldots, n\} = \{i, a_1, \ldots, a_p, j, b_1, \ldots, b_q, c_1^1, \ldots, c_{\ell_1}^1, \ldots, c_1^r, \ldots, c_{\ell_r}^r\}$$

to the same element.[18]

[17] It is possible that $r = 0$, so that π consists only of the red cycle. It is also possible that $p = 0$ (in this case, the red cycle is $(i\ j\ b_1 \ldots b_q)$), or that $q = 0$ (in this case, the red cycle is $(i\ a_1 \ldots a_p\ j)$). If $p = q = 0$, then the red cycle is simply (ij).

[18] Here, a picture may help. The diagram below represents the permutation π' (obviously). But by considering what each element of $\{1, \ldots, n\} = \{i, a_1, \ldots, a_p, j, b_1, \ldots, b_q, c_1^1, \ldots, c_{\ell_1}^1, \ldots, c_1^r, \ldots, c_{\ell_r}^r\}$

We now see that the disjoint cycle decomposition of $\tau \circ \pi$ has one cycle more than the disjoint cycle decomposition of π, and it follows that $\operatorname{sgn}\big(\tau \circ \pi\big) = -\operatorname{sgn}(\pi)$,[19] which is what we needed to show.

Case 2: i and j are in different cycles of the disjoint cycle decomposition of π. After possibly swapping the order of our disjoint cycles, and cyclically permuting the elements of the cycles that contain i and j, we may assume that our disjoint cycle decomposition of π is given by

$$\pi \;=\; (i\ a_1 \ldots a_p)(j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r)$$

Note: Here, i and j are in the two red cycles. The remaining cycles (the ones that do not contain i and j are colored blue.[20]

We then have that

$$\begin{aligned} \pi \;&=\; (i\ a_1 \ldots a_p)(j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r) \\ &\overset{(*)}{=}\; (ij) \circ (i\ a_1 \ldots a_p\ j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r), \end{aligned}$$

where (*) follows from the argument given in Case 1. We now compose both sides with $\tau = (ij)$ on the left, and we obtain

$$(ij) \circ \pi \;=\; (ij) \circ (ij) \circ (i\ a_1 \ldots a_p\ j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r).$$

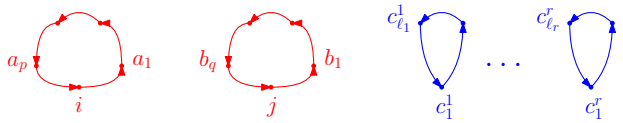
Since $(ij) = \tau$ and $(ij) \circ (ij) = 1_n$,[21] we deduce that

$$\tau \circ \pi \;=\; (i\ a_1 \ldots a_p\ j\ b_1 \ldots b_q)(c_1^1 \ldots c_{\ell_1}^1) \ldots (c_1^r \ldots c_{\ell_r}^r).$$

As we can see, in the permutation $\tau \circ \pi$, the two red cycles of π essentially get "merged" into one, while the blue cycles remain unaffected. But now the disjoint cycle decomposition of $\tau \circ \pi$ has one cycle less than the disjoint cycle decomposition of π, and it follows that $\operatorname{sgn}\big(\tau \circ \pi\big) = -\operatorname{sgn}(\pi)$,[22] which is what we needed to show. This completes the proof of the Claim. ♦

---

gets mapped to by the permutation $\tau \circ \pi$, we see that the diagram below also represents the permutation $\tau \circ \pi$.



[19] Indeed, the disjoint cycle decomposition of π has $r+1$ cycles, whereas the disjoint cycle decomposition of $\tau \circ \pi$ has $r+2$ cycles. Therefore, $\operatorname{sgn}\big(\tau \circ \pi\big) = (-1)^{n-(r+2)} = (-1)^{n-(r+1)-1} = -(-1)^{n-(r+1)} = -\operatorname{sgn}(\pi)$.

[20] It is possible that $p = 0$, $q = 0$, or $r = 0$. Similar remarks apply as in Case 1.

[21] Here, 1_n is the identity permutation in S_n.

[22] Indeed, the disjoint cycle decomposition of π has $r+2$ cycles, whereas the disjoint cycle decomposition of $\tau \circ \pi$ has $r+1$ cycles. Therefore, $\operatorname{sgn}\big(\tau \circ \pi\big) = (-1)^{n-(r+1)} = (-1)^{n-(r+2)+1} = -(-1)^{n-(r+2)} = -\operatorname{sgn}(\pi)$.

Now, fix $\pi, \tau \in S_n$ such that τ is a transposition. By the Claim, we have that $\operatorname{sgn}(\tau \circ \pi) = -\pi$. On the other hand,

$$\begin{aligned}
\operatorname{sgn}(\pi \circ \tau) &= \operatorname{sgn}\Big((\pi \circ \tau)^{-1}\Big) && \text{by Proposition 2.3.2} \\
&= \operatorname{sgn}\big(\tau^{-1} \circ \pi^{-1}\big) && \text{by Proposition 1.10.17(c)} \\
& && \text{(or by Proposition 2.2.4(f))} \\
&= \operatorname{sgn}\big(\tau \circ \pi^{-1}\big) && \text{because } \tau \text{ is a transposition,} \\
& && \text{and so } \tau^{-1} = \tau \\
&= -\operatorname{sgn}(\pi^{-1}) && \text{by the Claim applied to} \\
& && \pi^{-1} \text{ and } \tau \\
&= -\operatorname{sgn}(\pi) && \text{by Proposition 2.3.2.}
\end{aligned}$$

This completes the argument. □

Theorem 2.3.6. *Let $n \geq 2$. Then for any permutation $\pi \in S_n$, if π can be expressed as a composition of r transpositions, then*

(a) $sgn(\pi) = (-1)^r$;

(b) π is an even permutation if and only if r is even;

(c) π is an odd permutation if and only if r is odd.

Proof. Clearly, (b) and (c) follow from (a). Part (a) follows from Proposition 2.3.5 by an easy induction on r. Let us give the details. We prove the following statement: "for every positive integer r and permutation $\pi \in S_n$, if π is the composition of r transpositions, then $\operatorname{sgn}(\pi) = (-1)^r$."

Base case: $r = 1$. Note that if π is the composition of one transposition, i.e. π is itself a transposition, then π is odd, and we have that $\operatorname{sgn}(\pi) = -1 = (-1)^r$.

Induction step: Fix a positive integer r, and assume that for any permutation $\pi \in S_n$, if π is the composition of r transpositions, then $\operatorname{sgn}(\pi) = (-1)^r$. Now, fix a permutation $\pi \in S_n$ in S_n such that π can be expressed as the composition of $r+1$ transpositions, say $\pi = (a_0 a_0') \circ (a_1 a_1') \circ \cdots \circ (a_r a_r')$. Then by the induction hypothesis, $\pi' := (a_1 a_1') \circ \cdots \circ (a_r a_r')$ satisfies $\operatorname{sgn}(\pi') = (-1)^r$. But since $\pi = (a_0 a_0') \circ \pi'$, Proposition 2.3.5 guarantees that $\operatorname{sgn}(\pi) = -\operatorname{sgn}(\pi')$. So, $\operatorname{sgn}(\pi) = -\operatorname{sgn}(\pi') = -(-1)^r = (-1)^{r+1}$. This completes the induction. □

Theorem 2.3.7. *Let $n \geq 2$ be an integer, and let $\sigma, \pi \in S_n$. Then $sgn(\sigma \circ \pi) = sgn(\sigma) sgn(\pi)$.*

Proof. This easily follows from Proposition 2.3.3 and Theorem 2.3.6. Let us give the details. By Proposition 2.3.3, we can express σ and π as compositions of transpositions, say

- $\sigma = (s_1 s_1') \circ (s_2 s_2') \circ \cdots \circ (s_k s_k')$;
- $\pi = (t_1 t_1') \circ (t_2 t_2') \circ \cdots \circ (t_\ell t_\ell')$.

By Theorem 2.3.6(a), we have that $\operatorname{sgn}(\sigma) = (-1)^k$ and $\operatorname{sgn}(\pi) = (-1)^\ell$. On the other hand, $\sigma \circ \pi = (s_1 s_1') \circ (s_2 s_2') \circ \cdots \circ (s_k s_k') \circ (t_1 t_1') \circ (t_2 t_2') \circ \cdots \circ (t_\ell t_\ell')$, and so again by Theorem 2.3.6(a), we have that $\operatorname{sgn}(\sigma \circ \pi) = (-1)^{k+\ell}$. So, $\operatorname{sgn}(\sigma \circ \pi) = (-1)^{k+\ell} = (-1)^k (-1)^\ell = \operatorname{sgn}(\sigma)\operatorname{sgn}(\pi)$. □

2.3.4 The alternating group A_n

For an integer $n \geq 2$, let A_n be the set of all even permutations in S_n. Let us show that $(A_n, \circ)$ is a subgroup of $(S_n, \circ)$, where $\circ$ is the composition of functions. We apply Theorem 2.2.9. The identity element of S_n is the identity permutation 1_n, which is obviously even, and therefore belongs to A_n. Next, by Theorem 2.3.7, a composition of two even permutations is even, and consequently, A_n is closed under $\circ$. Finally, by Proposition 2.3.2, the sign of a permutation in S_n is equal to the sign of its inverse, and in particular, the inverse of an even permutation is even; so, A_n is closed under inverses. Theorem 2.2.9 now guarantees that A_n is indeed a subgroup of S_n.

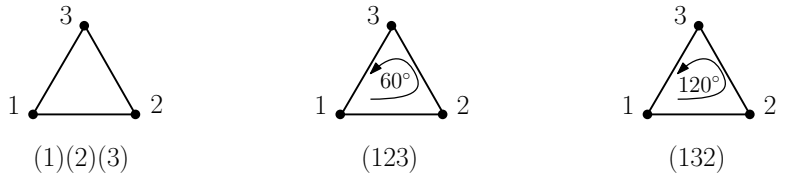
Terminology: For an integer $n \geq 2$, the group $(A_n, \circ)$ is called the *alternating group of degree n*. Typically, we just say that A_n is the alternating group of degree n, and the operation $\circ$ (composition of functions) is understood from context.

We remark that the set of odd permutations in S_n ($n \geq 2$), call it O_n,[23] does **not** form a subgroup of S_n. Indeed, the identity permutation 1_n is even and therefore does not belong to O_n; so, by Theorem 2.2.9, O_n is not a subgroup of S_n.

2.3.5 Symmetries of polygons

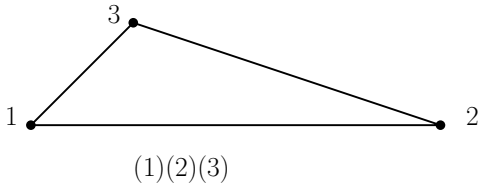
If we start with a (not necessarily symmetric, and not necessarily convex) polygon on n vertices, and we label its vertices $1, \ldots, n$, then the collection of symmetries of the polygon can be interpreted as a subgroup of the group S_n. For example, the group of symmetries of an equilateral triangle with vertices labeled $1, 2, 3$ is the entire group S_3. This corresponds to the identity function (which we can also think about as rotation by $0°$), two rotations, and three reflections, as indicated in the picture below.

[23] O_n is **not** standard notation for the set of odd permutations in S_n; in fact, no standard notation exists for this set. However, A_n is indeed the standard notation for the set of even permutations in S_n.



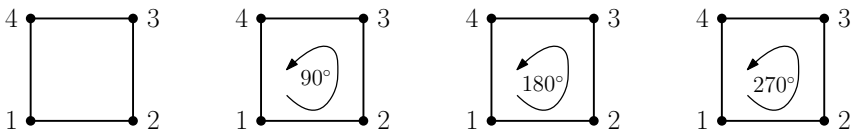
However, for an isosceles (but non-equilateral) triangle with vertices labeled $1, 2, 3$, where 12 is the base, the group of symmetries is $\{(1)(2)(3), (12)(3)\}$, where $(1)(2)(3)$ is simply the identity, and $(12)(3)$ is the reflection about the axis passing through the vertex 3 and the midpoint of the base 12.

Finally, if we have a triangle with three sides of different length, and with vertices labeled $1, 2, 3$, then its group of symmetries is just $\{(1)(2)(3)\}$, i.e. its only element is the identity permutation.

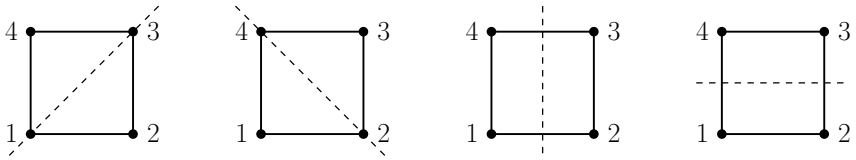


In the case of a square, we have eight symmetries: the identity (which we can also think of as rotation by $0°$), three rotations, and four reflections, shown below. The resulting subgroup of S_4 has 8 elements, whereas the group S_4 itself has $4! = 24$ elements. Of course, if we relabeled the vertices, we would get a different group. However, it would be the same as the group that we obtained with this labeling, up to a relabeling of the elements (vertices). The technical term is "isomorphic": the new group would be isomorphic to the old one.[24]

[24]We will study vector space isomorphism in chapter 4, but we will not go into group isomorphism in any detail.

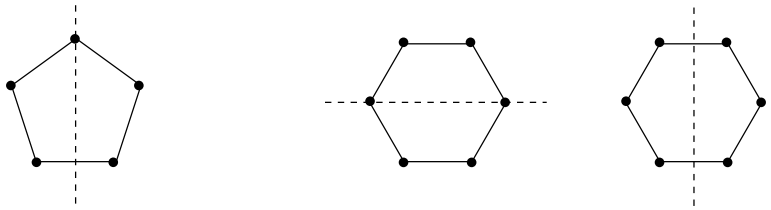


(1)(2)(3)(4) (1234) (13)(24) (1432)



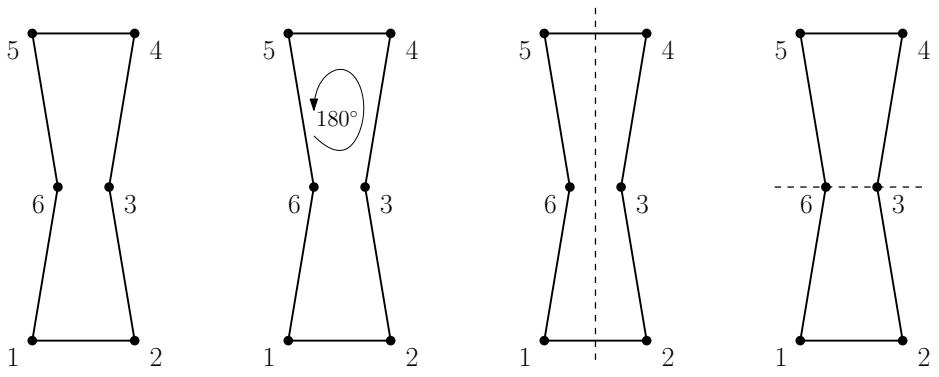
(1)(24)(3) (13)(2)(4) (12)(34) (14)(23)

We can, of course, also consider symmetries of polygons with more than three vertices. The regular n-gon has $2n$ symmetries: it has n rotations (by $\frac{k}{n}360°$ counterclockwise for $k = 0, 1, \ldots, n-1$, where for $k = 0$, we simply get the identity) and n reflections. If n is odd, then the reflections are about axes passing through one vertex and the midpoint of opposite side, and if n is even, then $\frac{n}{2}$ many reflections are through two opposite vertices, and the remaining $\frac{n}{2}$ reflections are through the midpoints of the opposite sides. (This is illustrated below for the cases when $n = 5$ and $n = 6$). Note that the symmetric group S_n has $n!$ many elements, whereas the group of symmetries of a regular n-gon has only $2n$ symmetries.



Terminology and notation: The group of symmetries of a regular polygon is called the *dihedral group* of that polygon. Unfortunately, notation is not entirely consistent in the literature. Some texts denote the dihedral group of the regular n-gon ($n \geq 3$) by D_n, whereas others denote it by D_{2n}.

It is also possible to consider non-regular, and even non-convex polygons. For example, for the polygon shown in the picture below, the group of symmetries has four elements, as in the picture.



(1)(2)(3)(4)(5)(6) (14)(25)(36) (12)(36)(45) (15)(24)(3)(6)

2.3.6 Inversions

In this subsection, we give another way of computing the sign of a permutation. Let n be a positive integer. An *inversion* of a permutation $\pi \in S_n$ is an ordered pair (i, j) of numbers in $\{1, \ldots, n\}$ such that $i < j$ and $\pi(i) > \pi(j)$.

Example 2.3.8. *The permutation*

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 3 & 6 & 5 & 4 & 7 \end{pmatrix}$$

in S_7 has the following four inversions: $(1, 2), (4, 5), (4, 6), (5, 6)$.

By Theorem 2.3.9 (below), the parity of a permutation is equal to the parity of the number of its inversions. Since the permutation π from Example 2.3.8 has an even number of inversions, Theorem 2.3.9 implies that π is even, i.e. $\operatorname{sgn}(\pi) = 1$. Note that the permutation π from Example 2.3.8 can be also written in the form $\pi = (12)(3)(46)(5)(7)$, and so by definition, its sign is $\operatorname{sgn}(\pi) = (-1)^{7-5} = (-1)^2 = 1$, which is the same as what we got using inversions. Let us now formally state and prove the theorem.

Theorem 2.3.9. *Let n be a positive integer. Then all permutations $\pi \in S_n$ satisfy $sgn(\pi) = (-1)^r$, where r is the number of inversions of π.*

Proof. We proceed by induction on the number r of inversions.

Base case: $r = 0$. The only permutation with no inversions is the identity permutation,[25] and its sign is 1. Since $(-1)^0 = 1$, this is what we needed.

[25]This is "obvious," but let us give the details. Fix a permutation $\pi \in S_n$ that has no inversions. Then, in particular, we have that $\pi(1) < \pi(2) < \cdots < \pi(n)$. Since $\pi(1), \ldots, \pi(n)$ all belong to the n-element set $\{1, \ldots, n\}$, it follows that $\pi(1) = 1, \ldots, \pi(n) = n$, i.e. π is the identity permutation.

Induction step: Fix a non-negative integer r, and assume inductively that any permutation in S_n that has exactly r inversions has sign $(-1)^r$. We must show that any permutation in S_n that has exactly $r+1$ inversions has sign $(-1)^{r+1}$.

Fix a permutation $\pi \in S_n$, and assume that it has exactly $r+1$ inversions. (Note that this implies that $n \geq 2$.) In particular, π has at least one inversion, and it follows that there exists some $p \in \{1, \ldots, n-1\}$ such that $(p, p+1)$ is an inversion of π (otherwise, we would have that $\pi(1) < \pi(2) < \cdots < \pi(n)$, and then π would be the identity permutation, contrary to the fact that it has at least one inversion). Now, consider the transposition $\tau := \Big(\pi(p)\pi(p+1)\Big)$ in S_n, and set $\pi' := \tau \circ \pi$, so that

$$\pi' = \begin{pmatrix} 1 & \ldots & p-1 & p & p+1 & p+2 & \ldots & n \\ \pi(1) & \ldots & \pi(p-1) & \pi(p+1) & \pi(p) & \pi(p+2) & \ldots & \pi(n) \end{pmatrix}$$

Then π' has exactly r inversions, i.e. exactly one inversion less than π has. To see this, we note the following:

- inversions (i, j) of π such that $i, j \notin \{p, p+1\}$ are still inversions of π';
- inversions of the form (i, p) of π correspond to inversions $(i, p+1)$ of π';
- inversions of the form $(i, p+1)$ of π, where $i < p$, correspond to inversions (i, p) of π';
- inversions of the form (p, j) of π, where $p+1 < j$, correspond to inversions $(p+1, j)$ of π';
- inversions of the form $(p+1, j)$ of π correspond to inversions (p, j) of π';
- π' has no other inversions, and in particular $(p, p+1)$ is **not** an inversion of π'.

But now

$$\begin{aligned} (-1)^r &= \operatorname{sgn}(\pi') && \text{by the induction hypothesis, since } \pi' \text{ has exactly } r \text{ inversions} \\ &= \operatorname{sgn}(\tau \circ \pi) && \text{because } \pi' = \tau \circ \pi \\ &= -\operatorname{sgn}(\pi) && \text{by Proposition 2.3.5, since } \tau \text{ is a transposition,} \end{aligned}$$

and it follows that $\operatorname{sgn}(\pi) = (-1)^{r+1}$. This completes the induction. □

Remark: In the induction step of the proof of Theorem 2.3.9, it was important that we chose an inversion of the form $(p, p+1)$, and not just any inversion of our

permutation π. To explain why, let us take a look at an example. Consider the permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 3 & 6 & 5 & 4 & 7 \end{pmatrix}$$

from Example 2.3.8. We could choose the inversion $(4,5)$, and consider the transposition $\tau := \Big(\pi(4)\pi(5)\Big) = (65) = (56)$ and the permutation

$$\pi' := \tau \circ \pi = (56) \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 3 & 6 & 5 & 4 & 7 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 3 & 5 & 6 & 4 & 7 \end{pmatrix}$$

Note that π' has three inversions,[26] whereas π has four.[27] If we had, instead, chosen an arbitrary inversion of π, then the number of inversions would not necessarily decrease by one, and we could not apply the induction hypothesis. Indeed, suppose we chose the inversion $(4,6)$ of our permutation π from Example 2.3.8 and then considered the transposition $\tau' := \Big(\pi(4)\pi(6)\Big) = (64) = (46)$ and the permutation

$$\pi'' := \tau' \circ \pi = (46) \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 3 & 6 & 5 & 4 & 7 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 3 & 4 & 5 & 6 & 7 \end{pmatrix}.$$

Note that π'' has only one inversion (namely, $(1,2)$), whereas π has four.

2.3.7 Permutation matrices

A *permutation matrix* is a square matrix that has exactly one 1 in each row and each column, and has 0's everywhere else. Below are all the possible 3×3 permutation matrices.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

[26] Those inversions are $(1,2), (4,6), (5,6)$.
[27] Those inversions are $(1,2), (4,5), (4,6), (5,6)$, as we saw in Example 2.3.8.

Remark: The 0's and 1's in permutation matrices may belong to any field $\mathbb{F}$ of our choice. Of course, we have not formally studied fields yet, and so for now, you may assume that the entries of our permutation matrices belong to one of the fields $\mathbb{F}$ that we already know, namely, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p$ for some prime number p. As usual, in what follows, we will denote by $\mathbf{e}_1, \ldots, \mathbf{e}_n$ the standard basis vectors of $\mathbb{F}^n$, where $\mathbb{F}$ is the field to which the 0's and 1's of our permutation matrices belong. The important point, though, is that in the remainder of this section, we will never need to add two non-zero numbers, and whenever we multiply two numbers, at least one of the two numbers will be 0 or 1. So, it does not matter which particular field we are working in, and therefore, for the remainder of the section, we will not emphasize this.

Obviously, identity matrices are permutation matrices. Moreover, $n \times n$ permutation matrices are precisely the matrices that can be obtained from the identity matrix I_n by reordering (i.e. permuting) rows, or alternatively, by reordering (i.e. permuting) columns. So, the columns of an $n \times n$ permutation matrix are the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ (appearing in some order in that matrix), whereas the rows are $\mathbf{e}_1^T, \ldots, \mathbf{e}_n^T$ (again, appearing in some order in that matrix).

For a positive integer n and a permutation $\pi \in S_n$, we define the *matrix of the permutation* π, denoted by P_π, to be the $n \times n$ matrix that has 1 in the $\Big(i, \pi(i)\Big)$-th entry for each each index $i \in \{1, \ldots, n\}$, and has 0 in all other entries. In other words, for each index $i \in \{1, \ldots, n\}$, the i-th row of the matrix P_π is $\mathbf{e}_{\pi(i)}^T$. For example, for the permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 1 & 6 & 5 & 3 \end{pmatrix},$$

in S_6, we obtain the 6×6 permutation matrix

$$P_\pi = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Obviously, for a positive integer n, the matrix of the identity permutation 1_n in S_n is precisely the identity matrix I_n, i.e. $P_{1_n} = I_n$.

Our next proposition states that, as we would expect, matrices of permutations are indeed permutation matrices (i.e. they have exactly one 1 in each row and each column, and they have 0's elsewhere).

Proposition 2.3.10. *Let n be a positive integer, and let $\pi \in S_n$. Then P_π is a permutation matrix.*

Proof. Obviously, P_π is an $n \times n$ matrix, all of whose entries are 0's and 1's. Moreover, by the definition of P_π, we have that for each index $i \in \{1, \ldots, n\}$, the i-th row of P_π is the row vector $\mathbf{e}_{\pi(i)}^T$. So, P_π has exactly one 1 in each row. Note that this means that the matrix P_π has exactly n entries that are 1, whereas all the other entries are 0's.

It remains to show that the matrix P_π has exactly one 1 in each column. Since P_π has exactly n many 1's, it is enough to show that no column has more than one 1. Since the rows of P_π (from top to bottom) are $\mathbf{e}_{\pi(1)}^T, \ldots, \mathbf{e}_{\pi(n)}^T$, and since all those row vectors are pairwise distinct (because π is a permutation), we see that no two rows of P_π have a 1 in the same position. So, no column of P_π has more than one 1, and we are done. □

Remark: By Proposition 2.3.10, the matrix of a permutation is a permutation matrix. What about the converse: is every permutation matrix the matrix of some permutation? The answer to this question is "yes," and it follows from a simple counting argument. Let n be a positive integer. The $n \times n$ permutation matrices are precisely those $n \times n$ matrices whose columns are the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, appearing in some order. There are $n!$ many ways to order the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, and consequently, there are $n!$ many $n \times n$ permutation matrices. On the other hand, $|S_n| = n!$, and consequently, there are $n!$ many matrices of permutations in S_n (we are using the fact that different permutations have different matrices). So, the number of $n \times n$ permutation matrices is the same as the number of matrices of permutations in S_n. It now follows from Proposition 2.3.10 that $n \times n$ permutation matrices are precisely the matrices of permutations in S_n.

Proposition 2.3.11. *Let n be a positive integer, and let $\pi \in S_n$ be a permutation. Then both the following hold:*

(a) for all indices $i \in \{1, \ldots, n\}$, we have that $\mathbf{e}_i^T P_\pi = \mathbf{e}_{\pi(i)}$, i.e. the i-th row of P_π is $\mathbf{e}_{\pi(i)}^T$;

(b) for all indices $j \in \{1, \ldots, n\}$, we have that $P_\pi \mathbf{e}_j = \mathbf{e}_{\pi^{-1}(j)}$, i.e. the j-th column of P_π is $\mathbf{e}_{\pi^{-1}(j)}$.

Consequently, in terms of its rows and columns, P_π can be written as follows:

$$P_\pi = \begin{bmatrix} \mathbf{e}_{\pi(1)}^T \\ \vdots \\ \mathbf{e}_{\pi(n)}^T \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{\pi^{-1}(1)} & \ldots & \mathbf{e}_{\pi^{-1}(n)} \end{bmatrix}.$$

Proof. The last statement of the proposition follows immediately from (a) and (b).[28] So, it is enough to prove (a) and (b).

(a) Fix an index $i \in \{1, \ldots, n\}$. By Proposition 1.8.2, $\mathbf{e}_i^T P_\pi$ is precisely the i-th row of the matrix P_π, and by the definition of the matrix P_π, its i-th row is precisely $\mathbf{e}_{\pi(i)}$.

(b) Fix an index $j \in \{1, \ldots, n\}$. By Proposition 1.4.4, $P_\pi \mathbf{e}_j$ is precisely the j-th column of the matrix P_π. Set $i := \pi^{-1}(j)$, so that $j = \pi(i)$. By (a), the i-th row of P_π is the row vector $\mathbf{e}_{\pi(i)}^T = \mathbf{e}_j^T$. So, P_π has 1 in its (i,j)-th entry. Since P_π is a permutation matrix (by Proposition 2.3.10), and therefore has exactly one 1 in each column, it follows that the j-th column of P_π is $\mathbf{e}_i = \mathbf{e}_{\pi^{-1}(j)}$. □

Propositions 2.3.12 and 2.3.13 (below) readily follow from Proposition 2.3.11.

Proposition 2.3.12. *Let n be a positive integer, and let $\pi \in S_n$. Then*

$$P_{\pi^{-1}} = P_\pi^T.$$

Proof. We have that

$$P_\pi^T \overset{(*)}{=} \left(\begin{bmatrix} \mathbf{e}_{\pi^{-1}(1)} & \ldots & \mathbf{e}_{\pi^{-1}(n)} \end{bmatrix} \right)^T = \begin{bmatrix} \mathbf{e}_{\pi^{-1}(1)}^T \\ \vdots \\ \mathbf{e}_{\pi^{-1}(n)}^T \end{bmatrix} \overset{(*)}{=} P_{\pi^{-1}},$$

where both instances of (*) follow from Proposition 2.3.11. □

Proposition 2.3.13. *Let n be a positive integer, and let σ and π be permutations in S_n. Then*

$$P_{\sigma \circ \pi} = P_\pi P_\sigma.$$

Remark: Note that swapping of order of σ and π: $P_{\sigma\circ\pi} = P_\pi P_\sigma$.

Proof. It suffices to show that matrices $P_{\sigma\circ\pi}$ and $P_\pi P_\sigma$ have the same corresponding rows. Fix an index $i \in \{1, \ldots, n\}$. By Proposition 1.8.2, the i-th row of the matrix $P_{\sigma\circ\pi}$ is $\mathbf{e}_i^T P_{\sigma\circ\pi}$, and the i-th row of the matrix $P_\pi P_\sigma$ is $\mathbf{e}_i^T (P_\pi P_\sigma)$. So, we just need to show that $\mathbf{e}_i^T P_{\sigma\circ\pi} = \mathbf{e}_i^T (P_\pi P_\sigma)$. But follows easily via repeated application of Proposition 2.3.11(a). Indeed, we have that

$$\mathbf{e}_i^T (P_\pi P_\sigma) = (\mathbf{e}_i^T P_\pi) P_\sigma \overset{(*)}{=} \mathbf{e}_{\pi(i)}^T P_\sigma \overset{(*)}{=} \mathbf{e}_{\sigma(\pi(i))} = \mathbf{e}_{(\sigma\circ\pi)(i)}^T \overset{(*)}{=} \mathbf{e}_i^T P_{\sigma\circ\pi},$$

where all three instances of (*) follow from Proposition 2.3.11(a). □

[28] The "last statement of the proposition" is the statement that

$$P_\pi = \begin{bmatrix} \mathbf{e}_{\pi(1)}^T \\ \vdots \\ \mathbf{e}_{\pi(n)}^T \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{\pi^{-1}(1)} & \ldots & \mathbf{e}_{\pi^{-1}(n)} \end{bmatrix}.$$

Theorem 2.3.14 (below) easily follows from Propositions 2.3.12 and 2.3.13, and it states that permutation matrices are invertible, and moreover, that the inverse of a permutation matrix is equal to the transpose of that permutation matrix.

Theorem 2.3.14. *Let n be a positive integer, and let $\pi \in S_n$. Then P_π is invertible, and moreover,*

$$P_\pi^{-1} \quad = \quad P_{\pi^{-1}} \quad = \quad P_\pi^T.$$

Proof. The fact that $P_{\pi^{-1}} = P_\pi^T$ follows immediately from Proposition 2.3.12. It remains to show that P_π is invertible, and that its inverse is $P_{\pi^{-1}}$. We will denote the identity permutation in S_n by 1_n, so that $\pi \circ \pi^{-1} = 1_n$ and $\pi^{-1} \circ \pi = 1_n$. We now compute:

$$P_\pi P_{\pi^{-1}} \quad \stackrel{(*)}{=} \quad P_{\pi^{-1} \circ \pi} \quad = \quad P_{1_n} \quad = \quad I_n,$$

where (*) follows immediately from Proposition 2.3.13. Analogously, $P_{\pi^{-1}} P_\pi = I_n$. So, P_π and $P_{\pi^{-1}}$ are invertible and are each other's inverses. This completes the argument. □

Remark: A matrix $Q \in \mathbb{R}^{n \times n}$ is *orthogonal* if it satisfies $Q^T Q = I_n$. We will study orthogonal matrices in chapter 6 (see section 6.8). For now, we note that Theorem 2.3.14 guarantees that permutation matrices are orthogonal (as long as we consider the 0's and 1's in those matrices as belonging to $\mathbb{R}$, rather than to some other field).

Permuting the rows and columns of a matrix. As our next theorem shows, multiplying a matrix by a permutation matrix on the left permutes the rows of the original matrix. On the other hand, multiplying a matrix by a permutation matrix on the right permutes the columns of the original matrix. More precisely, we have the following.

Theorem 2.3.15. *Let* $A = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ *be an $n \times m$ matrix with entries in some field $\mathbb{F}$.*[29] *Then all the following hold:*

(a) for all $\pi \in S_n$, we have that

$$P_\pi A \quad = \quad \begin{bmatrix} \mathbf{r}_{\pi(1)} \\ \vdots \\ \mathbf{r}_{\pi(n)} \end{bmatrix};$$

[29] Since we have not formally studied fields yet, you may assume for now that $\mathbb{F}$ is one of the fields that you are already familiar with, namely, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p$ for some prime number p. However, Theorem 2.3.15 is true for all fields $\mathbb{F}$, not just those particular ones.

(b) for all $\pi \in S_m$, we have that

$$AP_\pi = \left[\begin{array}{ccc} \mathbf{a}_{\pi^{-1}(1)} & \dots & \mathbf{a}_{\pi^{-1}(m)} \end{array} \right];$$

(c) for all $\pi \in S_m$, we have that

$$AP_\pi^T = \left[\begin{array}{ccc} \mathbf{a}_{\pi(1)} & \dots & \mathbf{a}_{\pi(m)} \end{array} \right].$$

Proof. We first prove (b). Fix any permutation $\pi \in S_m$. In what follows, $\mathbf{e}_1, \dots, \mathbf{e}_m$ are the standard basis vectors of $\mathbb{F}^m$. We compute:

$$\begin{array}{rcll} AP_\pi & = & A \left[\begin{array}{ccc} \mathbf{e}_{\pi^{-1}(1)} & \dots & \mathbf{e}_{\pi^{-1}(m)} \end{array} \right] & \text{by Proposition 2.3.11} \\ & = & \left[\begin{array}{ccc} A\mathbf{e}_{\pi^{-1}(1)} & \dots & A\mathbf{e}_{\pi^{-1}(m)} \end{array} \right] & \text{by the definition of matrix multiplication} \\ & = & \left[\begin{array}{ccc} \mathbf{a}_{\pi^{-1}(1)} & \dots & \mathbf{a}_{\pi^{-1}(m)} \end{array} \right] & \text{by Proposition 1.4.4.} \end{array}$$

This proves (b).

For (c), we note that for any permutation $\pi \in S_m$, we have that

$$\begin{array}{rcl} AP_\pi^T & \overset{(*)}{=} & AP_{\pi^{-1}} \\ & \overset{(**)}{=} & \left[\begin{array}{ccc} \mathbf{a}_{(\pi^{-1})^{-1}(1)} & \dots & \mathbf{a}_{(\pi^{-1})^{-1}(m)} \end{array} \right] \\ & \overset{(***)}{=} & \left[\begin{array}{ccc} \mathbf{a}_{\pi(1)} & \dots & \mathbf{a}_{\pi(m)} \end{array} \right], \end{array}$$

where (*) follows from Proposition 2.3.12, (**) follows from (b) applied to the matrix A and the permutation π^{-1}, and (***) follows from the fact that $(\pi^{-1})^{-1} = \pi$.

It remains to prove (a). Fix any permutation $\pi \in S_n$. We first consider the matrix $(P_\pi A)^T$, and we compute:

$$\begin{array}{rcll} (P_\pi A)^T & = & A^T P_\pi^T & \text{by Proposition 1.8.1(d)} \\ & = & \left[\begin{array}{ccc} \mathbf{r}_1^T & \dots & \mathbf{r}_n^T \end{array} \right] P_\pi^T & \\ & = & \left[\begin{array}{ccc} \mathbf{r}_{\pi(1)}^T & \dots & \mathbf{r}_{\pi(n)}^T \end{array} \right] & \text{by (c), applied to the matrix } A^T = \left[\begin{array}{ccc} \mathbf{r}_1^T & \dots & \mathbf{r}_n^T \end{array} \right] \text{ and the permutation } \pi \\ & = & \left(\begin{bmatrix} \mathbf{r}_{\pi(1)} \\ \vdots \\ \mathbf{r}_{\pi(n)} \end{bmatrix} \right)^T. & \end{array}$$

By taking the transpose of both sides, we obtain

$$P_\pi A = \begin{bmatrix} \mathbf{r}_{\pi(1)} \\ \vdots \\ \mathbf{r}_{\pi(n)} \end{bmatrix},$$

which is what we needed. This proves (a). □

2.4 Fields

2.4.1 Fields: definition, examples, and basic properties

A *field* is an ordered triple $(\mathbb{F}, +, \cdot)$, where $\mathbb{F}$ is a set, and $+$ and $\cdot$ are binary operations on $\mathbb{F}$ (i.e. functions from $\mathbb{F} \times \mathbb{F}$ to $\mathbb{F}$), called *addition* and *multiplication*, respectively, satisfying the following axioms:

1. addition and multiplication are associative, that is, for all $a, b, c \in \mathbb{F}$, we have that $a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
2. addition and multiplication are commutative, that is, for all $a, b \in \mathbb{F}$, we have that $a + b = b + a$ and $a \cdot b = b \cdot a$;
3. there exist distinct elements $0_\mathbb{F}, 1_\mathbb{F} \in \mathbb{F}$ such that for all $a \in \mathbb{F}$, $a + 0_\mathbb{F} = a$ and $a \cdot 1_\mathbb{F} = a$; $0_\mathbb{F}$ is called the *additive identity* of $\mathbb{F}$, and $1_\mathbb{F}$ is called the *multiplicative identity* of $\mathbb{F}$;
4. for every $a \in \mathbb{F}$, there exists an element in $\mathbb{F}$, denoted by $-a$ and called the *additive inverse* of a, such that $a + (-a) = 0_\mathbb{F}$;
5. for all $a \in \mathbb{F} \setminus \{0_\mathbb{F}\}$, there exists an element in $\mathbb{F}$, denoted by a^{-1} and called the *multiplicative inverse* of a, such that $a \cdot a^{-1} = 1_\mathbb{F}$;
6. multiplication is distributive over addition, that is, for all $a, b, c \in \mathbb{F}$, we have that $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

Example 2.4.1. *All the following are fields:*

1. $(\mathbb{Q}, +, \cdot)$; *2.* $(\mathbb{R}, +, \cdot)$; *3.* $(\mathbb{C}, +, \cdot)$.

Remark: Note that $(\mathbb{Z}, +, \cdot)$ is **not** a field. This is because elements of $\mathbb{Z} \setminus \{-1, 0, 1\}$ do not have multiplicative inverses. As we shall see, $(\mathbb{Z}_p, +, \cdot)$ is a field for every prime number p (see Theorem 2.4.3).

Notation:

- If operations $+$ and $\cdot$ are understood from context, then we typically just say "field $\mathbb{F}$" instead of "field $(\mathbb{F}, +, \cdot)$."
- For $a, b \in \mathbb{F}$, we typically write ab instead of $a \cdot b$, and we typically write $a - b$ instead of $a + (-b)$.
- As usual, unless parentheses indicate otherwise, we perform multiplication before performing addition. So, for $a, b, c \in \mathbb{F}$, we write $ab + c$ instead of $(a \cdot b) + c$, and similarly, we write $a + bc$ instead of $a + (b \cdot c)$.

Remarks:

- Axioms 1, 2, and 3 above imply that $(\mathbb{F}, +)$ and $(\mathbb{F}, \cdot)$ are monoids with identity elements $0_\mathbb{F}$ and $1_\mathbb{F}$, respectively. Proposition 2.1.1 guarantees that $0_\mathbb{F}$ and $1_\mathbb{F}$ are unique.
  - When there is no danger of confusion, we write 0 and 1 instead of $0_\mathbb{F}$ and $1_\mathbb{F}$, respectively.
- Axioms 1, 2, 3, and 4 imply that $(\mathbb{F}, +)$ is an abelian group with identity element $0_\mathbb{F}$. By Proposition 2.2.1, this implies that each element $a \in \mathbb{F}$ has a **unique** additive inverse $-a$.
- By Proposition 2.4.2 (below), for any $a, b \in \mathbb{F} \setminus \{0_\mathbb{F}\}$, we have $ab \neq 0_\mathbb{F}$, i.e. $ab \in \mathbb{F} \setminus \{0_\mathbb{F}\}$. This, together with axioms 1 and 3, implies that $(\mathbb{F} \setminus \{0_\mathbb{F}\}, \cdot)$ is a monoid with identity element $1_\mathbb{F}$. Next, by Proposition 2.4.2, and by axioms 2 (commutativity of addition) and 3 ($0_\mathbb{F} \neq 1_\mathbb{F}$), we have that we have that $a0_\mathbb{F} = 0_\mathbb{F}a = 0_\mathbb{F} \neq 1_\mathbb{F}$. This, together with axiom 5 implies that the multiplicative inverse of any element $a \in \mathbb{F} \setminus \{0_\mathbb{F}\}$ also belongs to $\mathbb{F} \setminus \{0_\mathbb{F}\}$. So, $(\mathbb{F} \setminus \{0_\mathbb{F}\}, \cdot)$ is an abelian group with identity element $1_\mathbb{F}$. By Proposition 2.2.1, it follows that every element $a \in \mathbb{F} \setminus \{0_\mathbb{F}\}$ has a **unique** multiplicative inverse a^{-1}.
- By axioms 2 and 6, for all $a, b, c \in \mathbb{F}$, we have that $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$, or written in a simplified manner, $(b + c)a = ba + ca$.[30]

Proposition 2.4.2. *Let $(\mathbb{F}, +, \cdot)$ be a field. Then all the following hold:*

(a) for all $a \in \mathbb{F}$, $0a = a0 = 0$;

(b) for all $a, b \in \mathbb{F}$, if $ab = 0$, then $a = 0$ or $b = 0$;

[30] Indeed, for $a, b, c \in \mathbb{F}$, we have that $(b + c)a \overset{\text{ax. 2.}}{=} a(b + c) \overset{\text{ax. 6.}}{=} ab + ac \overset{\text{ax. 2.}}{=} ba + ca$.

(c) for all $a \in \mathbb{F}$, $(-1)a = -a$.[31]

Proof. We first prove (a). Fix $a \in \mathbb{F}$. Since multiplication in the field $\mathbb{F}$ is commutative, we know that $0a = a0$. So, it suffices to show that $a0 = 0$. First, note that

$$a0 \overset{(*)}{=} a(0+0) \overset{(**)}{=} a0 + a0,$$

where (*) follows from the fact that $0 + 0 = 0$ (because 0 is the additive identity of the field), and (**) follows from axiom 6 of the definition of a field. We have now established that $a0 = a0 + a0$, and it follows that

$$\begin{aligned}
0 &= -(a0) + a0 && \text{because } -(a0) \text{ is the additive inverse of } a0 \\
&= -(a0) + (a0 + a0) && \text{because } a0 = a0 + a0 \text{ (proven above)} \\
&= (-(a0) + a0) + a0 && \text{because } + \text{ is associative} \\
&= 0 + a0 && \text{because } -(a0) \text{ is the additive inverse of } a0 \\
&= a0 && \text{because 0 is the additive identity of the field } \mathbb{F}.
\end{aligned}$$

Thus, $a0 = 0$. This proves (a).

Next, we prove (b). Fix $a, b \in \mathbb{F}$ such that $ab = 0$. We may assume that $b \neq 0$, for otherwise we are done. But now b has a multiplicative inverse b^{-1}, and we compute:

$$a = a \cdot 1 = a(bb^{-1}) \overset{(*)}{=} (ab)b^{-1} \overset{(**)}{=} 0b^{-1} \overset{(***)}{=} 0,$$

where (*) follows from the associativity of multiplication, (**) follows from the fact that $ab = 0$, and (***) follows from (a).

It remains to prove (c). Fix $a \in \mathbb{F}$. First, we have that

$$0 \overset{(*)}{=} 0a = (1-1)a = 1a + (-1)a = a + (-1)a,$$

where (*) follows from (a). Consequently,

[31] This statement may require some clarification. Here, $-a$ is the additive inverse of a. On the other hand, $(-1)a$ is the product of -1 (the additive inverse of the multiplicative identity) and a. So, $-a$ is not simply a shorthand for $(-1)a$. The two quantities are indeed equal, but this requires proof!

$$
\begin{aligned}
-a &= -a+0 && \text{because 0 is the additive identity of the field } \mathbb{F} \\
&= -a+(a+(-1)a) && \text{because } 0 = a+(-1)a \text{ (proven above)} \\
&= (-a+a)+(-1)a && \text{because } + \text{ is associative} \\
&= 0+(-1)a && \text{because } -a \text{ is the additive inverse of } a \\
&= (-1)a && \text{because 0 is the additive identity of the field } \mathbb{F}.
\end{aligned}
$$

This proves (c). □

2.4.2 Finite fields

In this subsection, we show that $(\mathbb{Z}_p, +, \cdot)$ is a field for all **prime** numbers p (see Theorem 2.4.3 below). We will use Fermat's Little Theorem, proven in subsection 0.2.2, and restated below for the reader's convenience.

Fermat's Little Theorem. *If $p \in \mathbb{N}$ is a prime number and $a \in \mathbb{Z}_p \setminus \{0\}$, then $a^{p-1} = 1$.*

Theorem 2.4.3. *For every **prime** number p, $(\mathbb{Z}_p, +, \cdot)$ is a field.*

Proof. By Proposition 0.2.11, addition and multiplication are associative and commutative in $\mathbb{Z}_p$, and multiplication is distributive over addition in $\mathbb{Z}_p$. So, $(\mathbb{Z}_p, +, \cdot)$ satisfies axioms 1, 2, and 6 from the definition of a field. Further, $0 := [0]_p$ is the additive identity and $1 := [1]_p$ is the multiplicative identity of $(\mathbb{Z}_p, +, \cdot)$. Moreover, $[0]_p \neq [1]_p$, since $0 \not\equiv 1 \pmod{p}$.[32] Thus, $(\mathbb{Z}_p, +, \cdot)$ satisfies axiom 3 from the definition of a field. Further, for all $a \in \mathbb{Z}$, the additive inverse of $[a]_p$ in $(\mathbb{Z}_p, +, \cdot)$ is $[-a]_p$, and so axiom 4 is satisfied. Finally, by Fermat's Little Theorem, every number $a \in \mathbb{Z}_p \setminus \{0\}$ has a multiplicative inverse, namely, a^{p-2}, and it follows that axiom 5 is satisfied. This proves that $(\mathbb{Z}_p, +, \cdot)$ is indeed a field, which is what we needed to show. □

Remark: For a positive integer n that is **not** prime, $(\mathbb{Z}_n, +, \cdot)$ is not a field. If $n = 1$, then this follows from the fact that $\mathbb{Z}_n = \mathbb{Z}_1$ has only one element, whereas every field has at least two elements (namely, the additive and multiplicative identities, which cannot be equal by axiom 3 of the definition of a field). Now, let us suppose that $n \geq 2$

[32] This follows from the fact that $p \geq 2$.

is composite, say $n = pq$ where $p, q \geq 2$ are integers. Then $[p]_n[q]_n = [pq]_n = [n]_n = 0$. So, if $(\mathbb{Z}_n, +, \cdot)$ were a field, Proposition 2.4.2(b) would imply that at least one of $[p]_n$ and $[q]_n$ is 0, a contradiction.

Finally, we state the following theorem without proof.

Theorem 2.4.4. *Let $n \geq 2$ be an integer. Then there exists a field of size n if and only if n is a power of a prime.*[33] *Moreover, if n is a power of a prime, then up to "isomorphism" (i.e. up to renaming the operations and elements of the field), there is exactly one field of size n, and it is denoted by $\mathbb{F}_n$.*[34]

Proof. Omitted. □

Remark: For a prime number p, we have that $\mathbb{F}_p = \mathbb{Z}_p$. However, if $n = p^m$, where p is a prime number and $m \geq 2$ is an integer, then $\mathbb{F}_n \neq \mathbb{Z}_n$ (this is because $\mathbb{F}_n$ is a field, but by the Remark following the proof of Theorem 2.4.3, $\mathbb{Z}_n$ is **not** a field).

2.4.3 The fraction notation in fields

Let $\mathbb{F}$ be a field. For $a \in \mathbb{F} \setminus \{0\}$, we sometimes use the notation $\frac{1}{a}$ instead of a^{-1} (the multiplicative inverse of a in the field $\mathbb{F}$). For instance, in $\mathbb{Z}_3$, we have $\frac{1}{1} = 1^{-1} = 1$ and $\frac{1}{2} = 2^{-1} = 2$ (because in $\mathbb{Z}_3$, we have that $2 \cdot 2 = 1$). In a similar vein, for scalars $a, b \in \mathbb{F}$ such that $b \neq 0$, we sometimes write $\frac{a}{b}$ instead of $b^{-1}a$. For example, in $\mathbb{Z}_5$, we have that $3^{-1} = 2$ (because $3 \cdot 2 = 1$), and so $\frac{4}{3} = 3^{-1} \cdot 4 = 2 \cdot 4 = 3$. It is sometimes more convenient to use the notation $\frac{1}{a}$ instead of a^{-1}, and $\frac{a}{b}$ instead of $b^{-1}a$. However, when working over a finite field such as $\mathbb{Z}_p$ (for a prime number p), we **never** leave a fraction as a final answer, and instead, we always simplify.

2.4.4 The characteristic of a field

The *characteristic* of a field $\mathbb{F}$ is the smallest positive integer n (if it exists) such that, in the field $\mathbb{F}$, we have that

$$\underbrace{1 + \cdots + 1}_{n} = 0,$$

where the 1's and the 0 are understood to be in the field $\mathbb{F}$. If no such n exists, then $\operatorname{char}(\mathbb{F}) := 0$. Note that fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ all have characteristic 0. On the other hand, for all prime numbers p, we have that $\operatorname{char}(\mathbb{Z}_p) = p$.

Theorem 2.4.5. *The characteristic of any field is either 0 or a prime number.*

[33] "n is a power of a prime" means that there exists some prime number p and a positive integer m such that $n = p^m$.

[34] Technically, the field is $(\mathbb{F}_n, +, \cdot)$, but we typically write just $\mathbb{F}_n$.

Proof. Let $\mathbb{F}$ be a field. We may assume that $\mathrm{char}(\mathbb{F}) \neq 0$, for otherwise we are done. So, $\mathrm{char}(\mathbb{F})$ is a positive integer. By the definition of a field, we have that $1 \neq 0$,[35] and so $\mathrm{char}(\mathbb{F}) \geq 2$. Now, suppose that $\mathrm{char}(\mathbb{F})$ is not prime, and fix integers $p, q \geq 2$ such that $\mathrm{char}(\mathbb{F}) = pq$. Then

$$\Big(\underbrace{1+\cdots+1}_{p}\Big)\Big(\underbrace{1+\cdots+1}_{q}\Big) \;=\; \underbrace{1+\cdots+1}_{pq} \;=\; 0.$$

Since $\mathbb{F}$ is a field, Proposition 2.4.2(b) guarantees that at least one of the numbers $\underbrace{1+\cdots+1}_{p}$ and $\underbrace{1+\cdots+1}_{q}$ is zero. But this is impossible since $0 < p, q < \mathrm{char}(\mathbb{F})$. $\square$

2.4.5 Algebraically closed fields

Polynomials. Given a field $\mathbb{F}$, a *polynomial* with coefficients in $\mathbb{F}$ is an expression of the form

$$p(x) \;=\; a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 + a_0,$$

where n is a non-negative integer, and $a_0, a_1, \ldots, a_{n-1}, a_n$, called the *coefficients* of $p(x)$, are some elements of $\mathbb{F}$. A *constant polynomial* is a polynomial of the form $p(x) = a_0$ for some constant (i.e. fixed element of the field) a_0. The *zero polynomial* is the constant polynomial $p(x) = 0$. The *degree* of a non-zero polynomial $p(x)$, denoted by $\deg\big(p(x)\big)$, is the largest non-negative integer n for which the coefficient in front of x^n is non-zero. As a convention, the *degree* of the zero polynomial is defined to be $-\infty$. Two polynomials are equal if their corresponding coefficients are equal.

Essentially, the polynomials that we study are the same as the ones that you saw in high school, except that the coefficients may belong to an arbitrary field. In particular, we may add, subtract, and multiply polynomials. This is done in the usual way, except that we have to keep in mind which particular field we are working over. For example:

- $\big(x^2 + x\big) + \big(2x + 1\big) = x^2 + 3x + 1$ when the coefficients are considered to be in $\mathbb{R}$;
- $\big(x^2 + x\big) + \big(2x + 1\big) = x^2 + 1$ when the coefficients are considered to be in $\mathbb{Z}_3$.

Roots of polynomials. For a polynomial $p(x)$ with coefficients in a field $\mathbb{F}$, a *root* of $p(x)$ is a number $a \in \mathbb{F}$ such that $p(a) = 0$. For example, the polynomial $p(x) = x^2 - 3x + 2$, seen as a polynomial with coefficients in $\mathbb{R}$, has two real roots, namely, 1 and 2. We could have computed these roots using the familiar quadratic equation, and we can check that they really are roots of our polynomial by computing

[35] Here, 1 and 0 are understood to be in the field $\mathbb{F}$.

$p(1) = 1^2 - 3 \cdot 1 + 2 = 0$ and $p(2) = 2^2 - 3 \cdot 2 + 2 = 0$, where the computation was in $\mathbb{R}$.

Now, consider the polynomial

$$q(x) \quad = \quad x^2 + 1.$$

Does $q(x)$ have a root? If we consider $q(x)$ to be a polynomial with real coefficients and we ask if it has real roots, then the answer is that it has none. If we consider $q(x)$ to be a polynomial with complex coefficients, then we see that it has two complex roots, namely i and $-i$. On the other hand, if we consider $q(x)$ to be a polynomial with coefficients in $\mathbb{Z}_2$, then we see that 1 is a root, since $q(1) = 1^2 + 1 = 0$ (in $\mathbb{Z}_2$). If we consider $q(x)$ to be a polynomial with coefficients in $\mathbb{Z}_3$, and we ask if it has any roots in $\mathbb{Z}_3$, then we see that the answer is "no": none of $0, 1, 2$ is a root, as we can see by direct computation.

Algebraically closed fields. An *algebraically closed field* is a field $\mathbb{F}$ that has the property that every non-constant polynomial with coefficients in $\mathbb{F}$ has a root in $\mathbb{F}$. By the Fundamental Theorem of Algebra (see subsection 0.3.2), $\mathbb{C}$ is an algebraically closed field. On the other hand, $\mathbb{R}$ is not algebraically closed (as our example above demonstrates), and similarly, neither is $\mathbb{Q}$. Moreover, no finite field is algebraically closed. To see this, consider any finite field $\mathbb{F} = \{f_1, \ldots, f_t\}$ ($t \geq 2$), and consider the polynomial

$$p(x) \quad = \quad (x - f_1) \ldots (x - f_t) + 1,$$

which is a polynomial of degree t with coefficients in $\mathbb{F}$. Then for each $i \in \{1, \ldots, t\}$, we have that $p(f_i) = 1$, and consequently, f_i is not a root of $p(x)$. Since $\mathbb{F} = \{f_1, \ldots, f_t\}$, we see that $p(x)$ has no roots in $\mathbb{F}$.

Thus, of the fields that we have seen so far, namely, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}_p$ (where p is a prime number), only the field $\mathbb{C}$ is algebraically closed. Other algebraically closed fields do exist, but they will not be discussed in these lecture notes.

Factoring polynomials into linear terms. It can be shown (though we will not give a formal proof) that any non-constant polynomial with coefficients in an algebraically closed field $\mathbb{F}$ can be factored into linear terms in a unique way. More precisely, if $p(x)$ is a polynomial of degree $n \geq 1$, and with coefficients in an algebraically closed field $\mathbb{F}$, then there exist numbers $a, \alpha_1, \ldots, \alpha_\ell$ in $\mathbb{F}$ such that $a \neq 0$ and such that $\alpha_1, \ldots, \alpha_\ell$ are pairwise distinct, and positive integers $n_1, \ldots, n_\ell$ satisfying $n_1 + \cdots + n_\ell = n$, such that

$$p(x) = a(x - \alpha_1)^{n_1} \ldots (x - \alpha_\ell)^{n_\ell}.$$

Moreover, $a, \alpha_1, \ldots, \alpha_\ell, n_1, \ldots, n_\ell$ are uniquely determined by the polynomial $p(x)$, up to a permutation of the α_i's and the corresponding n_i's. Here, a is the leading coefficient of $p(x)$, i.e. the coefficient in front of x^n. Numbers $\alpha_1, \ldots, \alpha_\ell$ are the roots

of $p(x)$ with *multiplicities* $n_1, \ldots, n_\ell$, respectively. If we think of each α_i as being a root "n_i times" (due to its multiplicity), then we see that the n-th degree polynomial $p(x)$ has exactly n roots in $\mathbb{F}$. This is often summarized as follows: "every n-th degree polynomial (with $n \geq 1$) with coefficients in an algebraically closed field has exactly n roots in that field, when multiplicities are taken into account."

When a field $\mathbb{F}$ is **not** algebraically closed, then some of its non-constant polynomials can be factored into linear terms, while others cannot.

Chapter 3

Vector spaces

3.1 Vector spaces

Let $\mathbb{F}$ be a field with additive identity 0 and multiplicative identity 1. In what follows, we shall refer to elements of $\mathbb{F}$ as *scalars*. A *vector space* (or *linear space*) over the field $\mathbb{F}$ is a set V, together with a binary operation $+$ on V (called *vector addition*) and an operation $\cdot : \mathbb{F} \times V \to V$ (called *scalar multiplication*), satisfying the following axioms:

1. $(V, +)$ is an abelian group; the identity element of $(V, +)$ is denoted by $\mathbf{0}$ ("zero vector"), and for any vector $\mathbf{v} \in V$, the inverse of $\mathbf{v}$ in $(V, +)$ is dented by $-\mathbf{v}$;
2. for all vectors $\mathbf{v} \in V$, we have $1\mathbf{v} = \mathbf{v}$;
3. for all vectors $\mathbf{v} \in V$ and scalars $\alpha, \beta \in \mathbb{F}$, we have $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$;
4. for all vectors $\mathbf{v} \in V$ and scalars $\alpha, \beta \in \mathbb{F}$, we have $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$;
5. for all vectors $\mathbf{u}, \mathbf{v} \in V$ and scalars $\alpha \in \mathbb{F}$, we have $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$.

Terminology: A *real vector space* is a vector space over the field $\mathbb{R}$, and a *complex vector space* is a vector space over the field $\mathbb{C}$.

Example 3.1.1. *Let $\mathbb{F}$ be a field. Then all the following are vector spaces over $\mathbb{F}$ (in each case, vector addition and scalar multiplication are defined in the natural way):*

1. *$\mathbb{F}^n$;*
2. *$\mathbb{F}^{n \times m}$;*
3. *the set of all functions from $\mathbb{F}$ to $\mathbb{F}$;*

4. *the set $\mathbb{P}_{\mathbb{F}}$ of all polynomials (in one varialble, typically x) with coefficients in the field $\mathbb{F}$;*[1]

   - ***Notation:*** *Some texts use the notation $\mathbb{F}[x]$ instead of $\mathbb{P}_{\mathbb{F}}$ (if x is the variable used in the polynomials in question).*

5. *for a non-negative integer n, the set $\mathbb{P}^n_{\mathbb{F}}$ of all polynomials of degree at most n and with coefficients in $\mathbb{F}$.*[2]

Note that each of the cases above, elements of our vector space are considered vectors (even if they do not "look like" vectors, i.e. even if they are matrices, functions, or polynomials).

Remark: In these lecture notes, we will see quite a few examples with polynomials, especially in chapter 4. So, a remark on when two polynomials are equal is in order. For a field $\mathbb{F}$, two polynomials in $\mathbb{P}_{\mathbb{F}}$ are equal precisely when their corresponding coefficients are the same.[3] For instance, polynomials $p_1(x) = x^2 + 2x + 3$ and $p_1(x) = 2x^2 - x + 3$ in $\mathbb{P}_{\mathbb{R}}$ are **not** equal, and we write $p_1(x) \neq p_2(x)$. This is because $p_1(x)$ and $p_2(x)$ do not have the same corresponding coefficients. One might object that $p_1(0) = p_2(0)$ and $p_1(3) = p_2(3)$, and so $p_1(x)$ and $p_2(x)$ are "sometimes equal." This does not matter. The point is that polynomials $p_1(x)$ and $p_2(x)$ are the different **as polynomials** (because they do not have the same corresponding coefficients). For an even more interesting example, consider polynomials $q_1(x) = x^4 + x^3$ and $q_2(x) = x^5 + x$ in $\mathbb{P}_{\mathbb{Z}_2}$. Then $q_1(x) \neq q_2(x)$ (because $q_1(x)$ and $q_2(x)$ do not have the same corresponding coefficients), even though $q_1(0) = q_2(0)$ and $q_1(1) = q_2(1)$, and 0 and 1 are the only elements of $\mathbb{Z}_2$! On the other hand, note that polynomials $r_1(x) = (x+3)^2$ and $r_2(x) = x(x+6) + 9$ in $\mathbb{P}_{\mathbb{R}}$ are equal because if we write them in the standard form,[4] we see that they have the same corresponding coefficients. Indeed, $r_1(x) = x^2 + 6x + 9$ and $r_2(x) = x^2 + 6x + 9$, and so $r_1(x) = r_2(x)$.

If you have studied calculus, here is another example of a vector space.

Example 3.1.2. *The following are real vector spaces (with vector addition and scalar multiplication defined in the usual way):*

1. *the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$;*

2. *the set of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$.*

[1] One could also consider polynomials in more than one variable (say, $x_1, \ldots, x_k$) and with coefficients in $\mathbb{F}$. This, too, is a vector space over $\mathbb{F}$.

[2] The notation $\mathbb{P}^n_{\mathbb{F}}$ is not fully standard (there is no fully standard notation for this), but it is the notation that we will use in these lecture notes.

[3] "Their corresponding coefficients are the same" means that for each non-negative integer k, the two polynomials have the same coefficient in front of x^k. This includes the free coefficient: the coefficient in front of $x^0 = 1$.

[4] That is, in the form $a_n x^n + \cdots + a_1 x + a_0$ for some coefficients $a_0, a_1, \ldots, a_n \in \mathbb{F}$.

We note that for any field $\mathbb{F}$, we have the *trivial* vector space $\{\mathbf{0}\}$ over the field $\mathbb{F}$. In this vector space, vector addition and scalar multiplication are defined in the obvious way: $\mathbf{0}+\mathbf{0}=\mathbf{0}$ and $\alpha\mathbf{0}=\mathbf{0}$ for all scalars $\alpha\in\mathbb{F}$. A vector space is *non-trivial* if it contains at least one non-zero vector.

Proposition 3.1.3. *Let V be a vector space over a field $\mathbb{F}$. Then all the following hold:*

(a) for all $\mathbf{v}\in V$, $0\mathbf{v}=\mathbf{0}$;[5]

(b) for all $\alpha\in\mathbb{F}$, $\alpha\mathbf{0}=\mathbf{0}$;

(c) for all $\mathbf{v}\in V$ and $\alpha\in\mathbb{F}$, if $\alpha\mathbf{v}=\mathbf{0}$, then $\alpha=0$ or $\mathbf{v}=\mathbf{0}$;

(d) for all $\mathbf{v}\in V$, $(-1)\mathbf{v}=-\mathbf{v}$.[6]

Proof. The proof is similar to that of Proposition 2.4.2. We prove (a), and we leave the rest as an exercise. Fix $\mathbf{v}\in V$. Then $0\mathbf{v}=(0+0)\mathbf{v}=0\mathbf{v}+0\mathbf{v}$, and consequently,

$$\begin{aligned}
\mathbf{0} &= -(0\mathbf{v})+0\mathbf{v} \\
&= -(0\mathbf{v})+0\mathbf{v}+0\mathbf{v} && \text{because } 0\mathbf{v}=0\mathbf{v}+0\mathbf{v} \\
&= \mathbf{0}+0\mathbf{v} && \text{because } -(0\mathbf{v})+0\mathbf{v}=\mathbf{0} \\
&= 0\mathbf{v}.
\end{aligned}$$

This proves (a). □

3.1.1 Vector (linear) subspaces

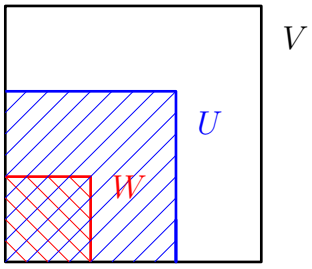
Let V be a vector space over a field $\mathbb{F}$. A *vector subspace* (or *linear subspace* or simply *subspace*) of V is a set $U\subseteq V$ such that U is itself a vector space over $\mathbb{F}$, when equipped with the vector addition and scalar multiplication operations "inherited" from V.[7] This means that we add two vectors of U using the vector addition operation from V, and similar for scalar multiplication. Moreover, U must be "closed under" vector addition and scalar multiplication from V, that is, that for all $\mathbf{u}_1,\mathbf{u}_2\in U$, we have that $\mathbf{u}_1+\mathbf{u}_2\in U$ and that for all $\mathbf{u}\in U$ and $\alpha\in\mathbb{F}$, we have that $\alpha\mathbf{u}\in U$ (where vector addition and scalar multiplication are those from the vector space V).

[5] Here, 0 is the zero of the field $\mathbb{F}$, and $\mathbf{0}$ is the zero vector in V.

[6] Here, -1 is the additive inverse of the multiplicative identity of the field $\mathbb{F}$, and in particular, -1 is a scalar. So, $(-1)\mathbf{v}$ is the product of the scalar -1 and the vector $\mathbf{v}$. On the other hand, $-\mathbf{v}$ is the additive inverse of the vector $\mathbf{v}$.

[7] Note that the field $\mathbb{F}$ must be the same for U and V!

Remark: It is obvious that the subspace relation is transitive. More precisely, for any vector space V over a field $\mathbb{F}$, if U is a subspace of V, and W is a subspace of U, then W is a subspace of V. Informally, we would say that "a subspace of a subspace is a subspace" (see the picture below).



Example 3.1.4. *Let V be a vector space over a field $\mathbb{F}$. Then V is a subspace of itself, and $\{\mathbf{0}\}$ is a subspace of V.*

Terminology: For a vector space V over a field $\mathbb{F}$, the *trivial subspace* of V is the subspace $\{\mathbf{0}\}$. A *non-trivial* subspace of V is one that contains at least one non-zero vector. A subspace U of V is *proper* if $U \subsetneqq V$.

Example 3.1.5. *Let n be a positive integer, and let $\mathbb{F}$ be a field. Then $\mathbb{P}^n_{\mathbb{F}}$ is a subspace of $\mathbb{P}_{\mathbb{F}}$.*

If you have studied calculus, here is another example.

Example 3.1.6. *The real vector space of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ is a subspace of the real vector space of continuous functions from $\mathbb{R}$ to $\mathbb{R}$, which is in turn a subspace of the real vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$.*

Theorem 3.1.7 (below) is a vector space analog of Theorem 2.2.9 (which dealt with groups). Moreover, our proof of Theorem 3.1.7 relies on Theorem 2.2.9.

Theorem 3.1.7. *Let V be a vector space over a field $\mathbb{F}$, and let $U \subseteq V$. Then U is a subspace of V if and only if the following three conditions are satisfied:*

(i) $\mathbf{0} \in U$;[8]

(ii) *U is closed under vector addition, that is, for all $\mathbf{u}, \mathbf{v} \in U$, we have that $\mathbf{u} + \mathbf{v} \in U$;*

(iii) *U is closed under scalar multiplication, that is, for all $\mathbf{u} \in U$ and $\alpha \in \mathbb{F}$, we have that $\alpha\mathbf{u} \in U$.*

[8] Here, $\mathbf{0}$ is the zero vector in the vector space V.

Proof. Suppose first that (i), (ii), and (iii) are satisfied; we must show that U is a subspace of V. By (ii), the restriction of $+$ to $U \times U$ (denoted $+ \restriction (U \times U)$, or just $+$ for simplicity) is a binary operation on U,[9] and by (iii), the restriction of $\cdot$ to $\mathbb{F} \times U$ (denoted by $\cdot \restriction (\mathbb{F} \times U)$, or just $\cdot$ for simplicity) is indeed a function from $\mathbb{F} \times U$ to U. So, U is equipped with both the vector addition operation and the scalar multiplication operation. Next, U satisfies axioms 2-5 from the definition of a vector space because the vector space V satisfies those axioms and because the vector addition and scalar multiplication operations in U are inherited from V. It remains to show that U satisfies axiom 1 from the definition of a vector space, that is, that U is an abelian group under vector addition. Since $(V, +)$ is an abelian group (because V is a vector space), it suffices to show that $(U, +)$ is a subgroup of $(V, +)$.[10] By (i), we have that $\mathbf{0} \in U$, and by (ii), we have that U is closed under vector addition. Moreover, by (iii) and by Proposition 3.1.3(d), for all $\mathbf{u} \in U$, we have that $-\mathbf{u} = (-1)\mathbf{u} \in U$,[11] and so U is closed under additive (vector) inverses. Theorem 2.2.9 now guarantees that $(U, +)$ is a subgroup of $(V, +)$. This proves that U is indeed a subspace of V.

Suppose now that U is a subspace of V; we must show that (i), (ii), and (iii) hold. Since the vector addition and scalar multiplication operations of the vector space U are inherited from the ones for V, we see that (ii) and (iii) hold. Moreover, since U is a vector space, we know that it contains the zero vector, call it $\mathbf{0}_U$.[12] We must show that $\mathbf{0}_U = \mathbf{0}$.[13] Since $\mathbf{0}_U$ is the identity element of $(U, +)$, we see that $\mathbf{0}_U + \mathbf{0}_U = \mathbf{0}_U$. Since $\mathbf{0}_U \in V$ and $\mathbf{0}$ is the identity element of V, we see that $\mathbf{0}_U + \mathbf{0} = \mathbf{0}_U$. So, $\mathbf{0}_U + \mathbf{0}_U = \mathbf{0}_U + \mathbf{0}$. By now adding $-\mathbf{0}_U$ to both sides of the equation,[14] and we obtain $\mathbf{0}_U = \mathbf{0}$. So, (i) holds. □

3.1.2 Linear combinations and linear span

In section 1.4, we introduced the linear span of vectors in $\mathbb{F}^n$ (where $\mathbb{F}$ is a field). Here, we generalize the concept to arbitrary vector spaces, as follows.

Suppose that V is a vector space over a field $\mathbb{F}$. Given vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$, we say that a vector $\mathbf{v} \in V$ is a *linear combination* of $\mathbf{u}_1, \ldots, \mathbf{u}_k$ if there exist scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\mathbf{v} \quad = \quad \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k.$$

[9] In other words, we have that $+ \restriction (U \times U) : U \times U \to U$.

[10] If $(U, +)$ is a subgroup of $(V, +)$, then in particular, $(U, +)$ is a group, and it must be abelian because $(V, +)$ is abelian.

[11] Indeed, by Proposition 3.1.3(d), we have that $-\mathbf{u} = (-1)\mathbf{u}$, and by (iii), we have that $(-1)\mathbf{u} \in U$.

[12] Since U is a subspace of V, we know, in particular, that $(U, +)$ is an abelian group, and consequently, it contains a (unique) identity element. We call this identity element $\mathbf{0}_U$.

[13] Here, $\mathbf{0}$ is the zero vector in V, and i.e. the identity element of the abelian group $(V, +)$. Since $(U, +)$ is an abelian group, it must have an identity element, and we call this identity element $\mathbf{0}_U$. However, could it be that $\mathbf{0}_U \neq \mathbf{0}$, so that (i) potentially fails? We show that this cannot happen. We argue similarly as in the proof of Theorem 2.2.9, only with different notation.

[14] Here, $-\mathbf{0}_U$ is the additive inverse of the vector $\mathbf{0}_U$ in V.

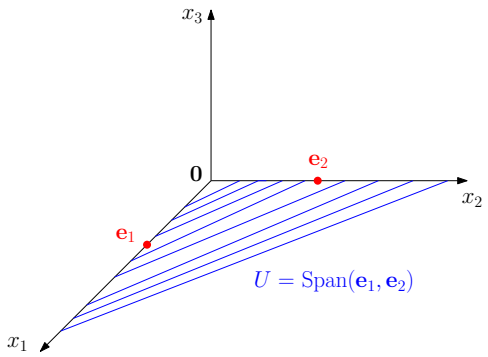
The *linear span* (or simply *span*) of the set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, denoted by $\text{Span}(\{\mathbf{u}_1, \dots, \mathbf{u}_k\})$ or $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$, is the set of all linear combinations of $\mathbf{u}_1, \dots, \mathbf{u}_k$, i.e.

$$\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k) \quad = \quad \big\{\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k \mid \alpha_1, \dots, \alpha_k \in \mathbb{F}\big\}.$$

As a convention, we define the "empty sum" of vectors in V to be $\mathbf{0}$ (the zero vector in V),[15] and consequently, $\text{Span}(\emptyset) = \{\mathbf{0}\}$.

Given a vector space V over a field $\mathbb{F}$, and given vectors $\mathbf{u}_1, \dots, \mathbf{u}_k \in V$, we say that $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a *spanning set* of V, or that that the set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ *spans* V, or that vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ *span* V, provided that $V = \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$. (Note that $\emptyset$ is a spanning set of the trivial vector space $\{\mathbf{0}\}$ over a field $\mathbb{F}$.)

Example 3.1.8. *Consider vectors* $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ *and* $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ *in* $\mathbb{R}^3$. *Then* $\text{Span}(\mathbf{e}_1, \mathbf{e}_2) = \{\begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R}\}$. *So,* $\text{Span}(\mathbf{e}_1, \mathbf{e}_2)$ *is the* x_1x_2*-plane in the Euclidean space* $\mathbb{R}^3$.



Example 3.1.9. *Consider the polynomials* $1, x, x^2$ *in* $\mathbb{P}_\mathbb{R}$. *Then* $\text{Span}(1, x, x^2) = \{a_2x^2 + a_1x + a_0 \mid a_0, a_1, a_2 \in \mathbb{R}\} = \mathbb{P}^2_\mathbb{R}$.

Remark: As we saw in subsection 1.4.4,[16] for a field $\mathbb{F}$ and a matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{F}^{n\times m}$, we have that

$$\text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m) \quad = \quad \big\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m\big\}.$$

Consequently, for all vectors $\mathbf{b} \in \mathbb{F}^n$, we have that $\mathbf{b} \in \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$ if and only if the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent. (See also Examples 1.5.5

[15] An "empty sum" might be the sum $\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k$, where $k = 0$ (and so we do not actually have any $\mathbf{u}_i$'s or α_i's).

[16] See the Remark following Example 1.4.3.

and 1.5.6 for a couple of numerical examples.) Moreover, as Proposition 3.1.10 (below) shows, checking whether a finite set of vectors in $\mathbb{F}^n$ is a spanning set of $\mathbb{F}^n$ is quite straightforward. (However, Proposition 3.1.10 only works for $\mathbb{F}^n$, and not for arbitrary vector spaces.)

Proposition 3.1.10. *Let $\mathbb{F}$ be a field, and let $\mathbf{a}_1, \dots, \mathbf{a}_m$ ($m \geq 1$) be some vectors in $\mathbb{F}^n$. Set $A := \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$. Then the following are equivalent:*

(a) vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ span $\mathbb{F}^n$;

(b) for all $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent;

(c) $\text{rank}(A) = n$ (i.e. A has full row rank).

Proof. By Corollary 1.6.6, (b) and (c) are equivalent. On the other hand, the fact that (a) and (b) are equivalent essentially follows from the fact that

$$\text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m) \quad = \quad \big\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m\big\}.$$

Indeed, we have the following sequence of equivalent statements:

$$\begin{aligned}
\text{vectors } \mathbf{a}_1, \dots, \mathbf{a}_m \text{ span } \mathbb{F}^n \quad &\iff \quad \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m) = \mathbb{F}^n \\
&\iff \quad \big\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m\big\} = \mathbb{F}^n \\
&\iff \quad \forall \mathbf{b} \in \mathbb{F}^n \ \exists \mathbf{x} \in \mathbb{F}^m \text{ s.t. } A\mathbf{x} = \mathbf{b} \\
&\iff \quad \forall \mathbf{b} \in \mathbb{F}^n\text{: } A\mathbf{x} = \mathbf{b} \text{ is consistent.}
\end{aligned}$$

Thus, (a) and (b) are indeed equivalent. This completes the argument. □

Theorem 3.1.11. *Let V be a vector space over a field $\mathbb{F}$, and let $\mathbf{u}_1, \dots, \mathbf{u}_k \in V$ ($k \geq 0$).*[17] *Then all the following hold:*

(a) $\mathbf{u}_1, \dots, \mathbf{u}_k \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$;

(b) $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is a subspace of V;

(c) for all subspaces U of V such that $\mathbf{u}_1, \dots, \mathbf{u}_k \in U$, $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is a subspace of U;

(d) $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is precisely the intersection of all subspaces of V that contain the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$.

[17] If $k = 0$, then $\mathbf{u}_1, \dots, \mathbf{u}_k$ is an empty list of vectors, the set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is empty, and $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k) = \{\mathbf{0}\}$.

Proof. To prove (a), we simply observe that for all $i \in \{1, \dots, k\}$, we have that

$$\mathbf{u}_i \;=\; 0\mathbf{u}_1 + \cdots + 0\mathbf{u}_{i-1} + 1\mathbf{u}_i + 0\mathbf{u}_{i+1} + \cdots + 0\mathbf{u}_k,$$

and so $\mathbf{u}_i \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$.

Next, we prove (b). It suffices to show that $Span(\mathbf{u}_1, \dots, \mathbf{u}_k)$ satisfies (i), (ii) and (iii) from Theorem 3.1.7, that is, that all the following hold:

(i) $\mathbf{0} \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$;

(ii) $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is closed under vector addition, that is, for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$, we have that $\mathbf{v}_1 + \mathbf{v}_2 \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$;

(iii) $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is closed under scalar multiplication, that is, for all vectors $\mathbf{v} \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ and scalars $\alpha \in \mathbb{F}$, we have that $\alpha\mathbf{v} \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$.

For (i), we simply note that $\mathbf{0} = 0\mathbf{u}_1 + \cdots + 0\mathbf{u}_k \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$.

Next, we prove (ii). Fix $\mathbf{v}_1, \mathbf{v}_2 \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$. Then there exist scalars $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k \in \mathbb{F}$ such that $\mathbf{v}_1 = \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k$ and $\mathbf{v}_2 = \beta_1\mathbf{u}_1 + \cdots + \beta_k\mathbf{u}_k$. But now

$$\begin{aligned} \mathbf{v}_1 + \mathbf{v}_2 &= (\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k) + (\beta_1\mathbf{u}_1 + \cdots + \beta_k\mathbf{u}_k) \\ &= (\alpha_1 + \beta_1)\mathbf{u}_1 + \cdots + (\alpha_k + \beta_k)\mathbf{u}_k, \end{aligned}$$

and we deduce that $\mathbf{v}_1 + \mathbf{v}_2 \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$. This proves (ii).

It remains to prove (iii). Fix $\mathbf{v} \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ and $\alpha \in \mathbb{F}$. Since $\mathbf{v} \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$, we see that there exist scalars $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k$. But now

$$\alpha\mathbf{v} \;=\; \alpha\big(\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k\big) \;=\; (\alpha\alpha_1)\mathbf{u}_1 + \cdots + (\alpha\alpha_k)\mathbf{u}_k,$$

and so $\alpha\mathbf{v} \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$. This proves (iii). We have now proven (b).

> **Claim.** For all subspaces U of V such that $\mathbf{u}_1, \dots, \mathbf{u}_k \in U$, we have that $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k) \subseteq U$.

Proof of the Claim. Fix any subspace U of V that contains $\mathbf{u}_1, \dots, \mathbf{u}_k$; we must show that $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k) \subseteq U$. Fix any $\mathbf{v} \in \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$. Then there exist scalars $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k$. Since U is a subspace of V, it satisfies (ii) and (iii) from Theorem 3.1.7. Since $\mathbf{u}_1, \dots, \mathbf{u}_k \in U$, (iii) from Theorem 3.1.7 guarantees that $\alpha_1\mathbf{u}_1, \dots, \alpha_k\mathbf{u}_k \in U$; but then (ii) from Theorem 3.1.7 guarantees that $\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k \in U$, i.e. $\mathbf{v} \in U$. So, $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k) \subseteq U$. ♦

We now prove (c). Fix any subspace U of V such that $\mathbf{u}_1, \dots, \mathbf{u}_k \in U$. By the Claim, we have that $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k) \subseteq U$, and by (a), $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is a

subspace of V. So, U is a vector space, and $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is subset of U that is a vector space in its own right (when equipped with the vector addition and scalar multiplication operations inherited from U).[18] So, by definition, $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is a subspace of U. This proves (c).

It remains to prove (d). By (a) and (b), $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is itself a subspace of V that contains $\mathbf{u}_1, \ldots, \mathbf{u}_k$. So, the intersection of all subspaces of V that contain $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a subset of $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$.[19] On the other hand, by the Claim, $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is a subset of each subspace of V that contains the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$, and consequently, of the intersection of all such subspaces. This proves (d). □

Remark: In some texts, for a vector space V over a field $\mathbb{F}$, and for vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$, the linear span (or simply span) of $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is defined to be the intersection of all subspaces of V that contain $\mathbf{u}_1, \ldots, \mathbf{u}_k$. By Theorem 3.1.11, this definition is equivalent to the one that we gave at the beginning of this subsection.

Rescaling vectors in spanning sets. The following proposition readily follows from the relevant definitions, but is may be useful to state explicitly.

Proposition 3.1.12. *Let V be a vector space over a field $\mathbb{F}$, let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$, and let $\alpha_1, \ldots, \alpha_k \in \mathbb{F} \setminus \{0\}$. Then*

$$\textit{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \quad = \quad \textit{Span}(\alpha_1\mathbf{v}_1, \ldots, \alpha_k\mathbf{v}_k).$$

Proof. We need to prove two inclusions:

(i) $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \subseteq \text{Span}(\alpha_1\mathbf{v}_1, \ldots, \alpha_k\mathbf{v}_k)$;

(ii) $\text{Span}(\alpha_1\mathbf{v}_1, \ldots, \alpha_k\mathbf{v}_k) \subseteq \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.

We prove (i); the proof of (ii) is similar and is left as an exercise. Fix any vector $\mathbf{v} \in \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. Then, by definition, there exist scalars $\beta_1, \ldots, \beta_k \in \mathbb{F}$ such that

$$\mathbf{v} \quad = \quad \beta_1\mathbf{v}_1 + \cdots + \beta_k\mathbf{v}_k.$$

Since scalars $\alpha_1, \ldots, \alpha_k$ are all non-zero, they have multiplicative inverses $\alpha_1^{-1}, \ldots, \alpha_k^{-1}$, respectively. We now have that

$$\mathbf{v} \quad = \quad \beta_1\mathbf{v}_1 + \cdots + \beta_k\mathbf{v}_k \quad = \quad (\beta_1\alpha_1^{-1})(\alpha_1\mathbf{v}_1) + \cdots + (\beta_k\alpha_k^{-1})(\alpha_k\mathbf{v}_k),$$

and so $\mathbf{v} \in \text{Span}(\alpha_1\mathbf{v}_1, \ldots, \alpha_k\mathbf{v}_k)$. This proves (i). □

[18] Technically, both $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ and U are equipped with the vector addition and scalar multiplication operations inherited from V. However, since $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k) \subseteq U \subseteq V$, we in fact have that $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is equipped with the vector addition and scalar multiplication operations inherited from U (where U inherited those two operations from V).

[19] Here, we are using the fact that for all non-empty collections $\mathcal{I}$ of sets, and for every set $A \in \mathcal{I}$, we have that $\bigcap_{X \in \mathcal{I}} X \subseteq A$.

Remark: In Proposition 3.1.12, it is important that the scalars $\alpha_1, \ldots, \alpha_k$ are all non-zero, and indeed, the proposition becomes false without this hypothesis. For example, for the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2$ in $\mathbb{R}^2$, we have that $\text{Span}(\mathbf{e}_1, \mathbf{e}_2) = \mathbb{R}^2$, but $\text{Span}(1\mathbf{e}_1, 0\mathbf{e}_2) = \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \mid x_1 \in \mathbb{R} \right\}$, which is a proper subspace of $\mathbb{R}^2$.

3.1.3 Making new vector (sub)spaces out of old ones

The Cartesian product of two vector spaces. Suppose we are given two vector spaces, say U and W, over a field $\mathbb{F}$.[20] Then the Cartesian product

$$U \times W \quad := \quad \big\{(\mathbf{u}, \mathbf{w}) \mid \mathbf{u} \in U,\ \mathbf{w} \in W\big\}$$

can be turned into a vector space over $\mathbb{F}$ in a natural way. We define vector addition in $U \times W$ by setting

$$(\mathbf{u}_1, \mathbf{w}_1) + (\mathbf{u}_2, \mathbf{w}_2) \quad := \quad (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{w}_1 + \mathbf{w}_2),$$

for all $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{w}_1, \mathbf{w}_2 \in W$, where in the first coordinate ("$\mathbf{u}_1 + \mathbf{u}_2$") we applied addition from the vector space U, and in the second coordinate ("$\mathbf{w}_1 + \mathbf{w}_2$") we applied vector addition from the vector space W. Scalar multiplication in $U \times W$ (with scalars from the field $\mathbb{F}$) is defined in an equally natural way, i.e. by setting

$$\alpha(\mathbf{u}, \mathbf{w}) \quad := \quad (\alpha\mathbf{u}, \alpha\mathbf{w})$$

for all $\alpha \in \mathbb{F}$, $\mathbf{u} \in U$, and $\mathbf{w} \in W$. The zero vector of $U \times W$ is the vector $\mathbf{0}_{U \times W} := (\mathbf{0}_U, \mathbf{0}_W)$, where $\mathbf{0}_U$ is the zero vector of the vector space U, and $\mathbf{0}_W$ is the zero of the vector space W. The additive inverse of a vector $(\mathbf{u}, \mathbf{w})$ in $U \times W$ is the vector $(-\mathbf{u}, -\mathbf{w})$, where $-\mathbf{u}$ (resp. $-\mathbf{w}$) is the additive inverse of $\mathbf{u}$ (resp. $\mathbf{w}$) in the vector space U (resp. W). It is straightforward to verify that all the axioms of a vector space hold for $U \times W$ (with vector addition and scalar multiplication defined as above). Indeed, this simply follows from the fact that those axioms hold for U and W, and the details are left as an exercise.

The intersection and sum of linear subspaces. Suppose that V is a vector space over a field $\mathbb{F}$, and that U and W are subspaces of V. Using Theorem 3.1.7, it can easily be verified that $U \cap W$ is a subspace of V, as is

$$U + W \quad := \quad \big\{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U,\ \mathbf{w} \in W\big\}.$$

The details are left as an exercise.

[20] The field $\mathbb{F}$ must be the same for both U and W! Otherwise, the construction does not work.

3.2 Bases of vector spaces

3.2.1 Linear independence

Given a vector space V over a field $\mathbb{F}$, and given vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$, we say that $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a *linearly independent set*, or that vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are *linearly independent*, if for all scalars $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k \quad = \quad \mathbf{0},$$

we have that $\alpha_1 = \cdots = \alpha_k = 0$. In other words, vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent if and only if the equation $\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$ has only the "trivial solution," i.e. the solution $\alpha_1 = \cdots = \alpha_k = 0$. On the other hand, if vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are not linearly independent, then we say that they are *linearly dependent*, or that the $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a *linearly dependent set*. So, vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly dependent if and only if there exist scalars $\alpha_1, \dots, \alpha_k \in \mathbb{F}$, not all zero, such that $\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$. We note that $\emptyset$ is a linearly independent set in any vector space.

As Proposition 3.2.1 (below) shows, for a field $\mathbb{F}$, it is easy to check if a finite set of vectors in $\mathbb{F}^n$ is linearly independent. Note, however, that linear independence is defined for general vector spaces, and not just for $\mathbb{F}^n$.

Proposition 3.2.1. *Let $\mathbb{F}$ be a field, and let $\mathbf{a}_1, \dots, \mathbf{a}_m$ ($m \geq 1$) be vectors in $\mathbb{F}^n$. Set $A := \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$. Then the following are equivalent:*

(a) vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ are linearly independent;

(b) the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$);

(c) $\operatorname{rank}(A) = m$ (i.e. A has full column rank).

Proof. By Corollary 1.6.5, (b) and (c) are equivalent. Let us show that (a) and (b) are equivalent. We have the following sequence of equivalent statements:

vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ are linearly independent

$\iff$ the equation $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{0}$ has only the trivial solutuon (i.e. the solution $x_1 = \cdots = x_m = 0$)

$\iff$ the equation $\underbrace{\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}}_{=A} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \mathbf{0}$ has only the trivial solutuon (i.e. the solution $x_1 = \cdots = x_m = 0$)

$\iff$ the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$).

Thus, (a) and (b) are equivalent. This completes the argument. □

For linear independence, we have the following analog of Proposition 3.1.12.

Proposition 3.2.2. *Let V be a vector space over a field $\mathbb{F}$, let $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$, and let $\alpha_1, \dots, \alpha_k \in \mathbb{F} \setminus \{0\}$. Then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is linearly independent if and only if the set $\{\alpha_1\mathbf{v}_1, \dots, \alpha_k\mathbf{v}_k\}$ is linearly independent.*

Proof. This readily follows from the definition of linear independence and is left as an exercise. □

3.2.2 Bases of vector spaces: definition and basic properties

A *finite basis* (or simply *basis*) of a vector space V over a field $\mathbb{F}$ is a set $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ of vectors in V that satisfies the following two conditions:

1. $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is linearly independent in V;
2. $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a spanning set of V, i.e. $\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) = V$.

A vector space is *finite-dimensional* if it has a finite basis. A vector space that does not have a finite basis is *infinite-dimensional*.

Not all vector spaces have a finite basis. For example, for any field $\mathbb{F}$, the vector space $\mathbb{P}_{\mathbb{F}}$ (over $\mathbb{F}$) of all polynomials with coefficients in $\mathbb{F}$ is infinite-dimensional (see Proposition 3.2.5). It is, indeed, possible to define a basis more generally, so that it may possibly be an infinite set. This is briefly discussed in subsection 3.2.7. However, with the exception of subsection 3.2.7, these lecture notes deal only with finite bases.

Notation: Suppose that V is a vector-space over a field $\mathbb{F}$. If V is finite-dimensional (i.e. has a finite basis), then we write $\dim(V) < \infty$. On the other hand, if V is infinite-dimensional (i.e. does not have a finite basis), then we write $\dim(V) = \infty$.

Remarks: Suppose that V is a vector space over a field $\mathbb{F}$.

- Obviously, any subset of a linearly independent set of vectors in V is linearly independent. Similarly, any superset of a spanning set of V is a spanning set of V.[21]
- $\{\mathbf{0}\}$ is **not** a linearly independent set in V (because $1 \cdot \mathbf{0} = \mathbf{0}$ and $1 \neq 0$); so, by the previous bullet point, no linearly independent set of vectors in V, and in particular, no basis of V, contains the zero vector.
- $\emptyset$ is a basis of the trivial vector space $\{\mathbf{0}\}$ (over any field $\mathbb{F}$), and in particular, $\{\mathbf{0}\}$ is finite dimensional. In fact, $\emptyset$ is the unique basis of $\{\mathbf{0}\}$ (because, by the previous bullet point, no linearly independent set contains $\mathbf{0}$).

[21] A set A is a *superset* of a set B provided that $B \subseteq A$.

- Suppose we are given vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$, and we are trying to check if $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a spanning set of V, i.e. whether $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ (this is one of the two conditions from the definition of a basis). Obviously, $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) \subseteq V$, and so the only question is whether $V \subseteq \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$. But "$V \subseteq \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$" simply means "every vector in V is a linear combination of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$." So, the second condition from the definition of a basis holds if and only if every vector in V is a linear combination of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$.

- In the literature, there is a bit of ambiguity about whether (finite) bases are sets or **ordered** sets. An "ordered set" is a set in which order and repetitions matter. For instance, $\{1, 2, 3\}$, $\{1, 2, 2, 3\}$, and $\{3, 1, 2\}$ are the same as sets, but they are pairwise distinct as ordered sets. In what follows, we will implicitly treat finite sets (when discussed in the context of linearly independent sets, spanning sets, and bases) as ordered, and in particular, we will care about repetitions. It is important to note that no linearly independent set (and in particular, no basis), may contain more than one copy of the same vector. Indeed, if $\mathbf{v}_1, \dots, \mathbf{v}_k$ is a list of vectors that contains more than one copy of some vector (say, $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$), then we can set $\alpha_i = 1$, $\alpha_j = -1$, and $\alpha_k = 0$ for all $k \in \{1, \dots, n\} \setminus \{i, j\}$, and we get $\alpha_1\mathbf{v}_1 + \dots + \alpha_n\mathbf{v}_n = \mathbf{0}$; so $\mathbf{v}_1, \dots, \mathbf{v}_n$ are not linearly independent.

  - In what follows, if A and B are ordered sets (possibly with repeating elements), then $A \subseteq B$ means that every element of A appears at least as many times in B as in A. Moreover, for $x \in A$, $A \setminus \{x\}$ is the set obtained from A by deleting one copy of x.

Example 3.2.3. *Let $\mathbb{F}$ be a field. Then the standard basis $\mathcal{E}_n = \{\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}$ of $\mathbb{F}^n$ (defined in subsection 1.4.4) is indeed a basis of $\mathbb{F}^n$.*[22]

Example 3.2.4. *Let $\mathbb{F}$ be a field. Then*

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is a basis of $\mathbb{F}^{3\times 2}$.[23]

[22] Let us check this! We first show that $\mathcal{E}_n = \{\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}$ is linearly independent. Fix scalars $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ such that $\alpha_1\mathbf{e}_1^n + \dots + \alpha_n\mathbf{e}_n^n = \mathbf{0}$. Clearly, $\alpha_1\mathbf{e}_1^n + \dots + \alpha_n\mathbf{e}_n^n = \begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T$. So, $\begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T = \mathbf{0}$, and it follows that $\alpha_1 = \dots = \alpha_n = 0$. So, $\mathcal{E}_n = \{\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}$ is linearly independent. Let us now show that $Span(\mathbf{e}_1^n, \dots, \mathbf{e}_n^n) = \mathbb{F}^n$, i.e. that every vector in V is a linear combination of $\mathbf{e}_1^n, \dots, \mathbf{e}_n^n$. Fix any $\mathbf{v} \in V$, and set $\mathbf{v} = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}^T$. But now $\mathbf{v} = v_1\mathbf{e}_1^n + \dots + v_n\mathbf{e}_n^n$, i.e. $\mathbf{v}$ is a linear combination of $\mathbf{e}_1^n, \dots, \mathbf{e}_n^n$. So, $\mathcal{E}_n = \{\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}$ is indeed a basis of $\mathbb{F}^n$.

[23] Proof?

Proposition 3.2.5. *Let $\mathbb{F}$ be a field. Then $\mathbb{P}_{\mathbb{F}}$ is inifnite-dimensional. On the other hand, for all non-negative integers n, $\{1, x, \ldots, x^n\}$ is a basis of $\mathbb{P}^n_{\mathbb{F}}$, and in particular, $\mathbb{P}^n_{\mathbb{F}}$ is finite-dimensional.*

Proof. It is clear that for any non-negative integer n, $\{1, x, \ldots, x^n\}$ is a basis of $\mathbb{P}^n_{\mathbb{F}}$.[24] Let us show that $\mathbb{P}_{\mathbb{F}}$ is infinite-dimensional. We must show that $\mathbb{P}_{\mathbb{F}}$ does not have a (finite) basis. Fix any finite set $\{p_1(x), \ldots, p_k(x)\}$ ($k \geq 0$) of polynomials in $\mathbb{P}_{\mathbb{F}}$; we must show that this set is not a basis of $\mathbb{P}_{\mathbb{F}}$. Let d be any non-negative integer such that $\deg\big(p_i(x)\big) \leq d$ for all $i \in \{1, \ldots, k\}$.[25] Then any linear combination of the polynomials $p_1(x), \ldots, p_k(x)$ is a polynomial of degree at most d, and it follows that $x^{d+1} \notin \mathrm{Span}\big(p_1(x), \ldots, p_k(x)\big)$. So, $\{p_1(x), \ldots, p_k(x)\}$ is not a spanning set of $\mathbb{P}_{\mathbb{F}}$, and consequently, it is not a basis of $\mathbb{P}_{\mathbb{F}}$. This proves that $\mathbb{P}_{\mathbb{F}}$ does not have a finite basis, and consequently, $\mathbb{P}_{\mathbb{F}}$ is infinite-dimensional. □

As Proposition 3.2.6 (below) shows, for a field $\mathbb{F}$, we can easily check if a finite set of vectors in $\mathbb{F}^n$ is a basis of $\mathbb{F}^n$. Note, however, that Proposition 3.2.6 only works for $\mathbb{F}^n$, and not for general vector spaces V.

Proposition 3.2.6. *Let $\mathbb{F}$ be a field, and let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ ($m \geq 1$) be vectors in $\mathbb{F}^n$. Set $A := \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix}$. Then $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is a basis of $\mathbb{F}^n$ if and only if $\mathrm{rank}(A) = n = m$ (i.e. A is a square matrix of full rank). In particular, every basis of $\mathbb{F}^n$ contains exactly n vectors.*

Proof. By Proposition 3.2.1, vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are linearly independent if and only if $\mathrm{rank}(A) = m$, and by Proposition 3.1.10, vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ span $\mathbb{F}^n$ if and only if $\mathrm{rank}(A) = n$. So, $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is a basis of $\mathbb{F}^n$ if and only if $\mathrm{rank}(A) = m = n$. □

Remark: By the Invertible Matrix Theorem (version 1; see subsection 1.11.7), square matrices of full rank are precisely the invertible matrices. So, Proposition 3.2.6 yields another characterizations of invertible matrices: a matrix in $\mathbb{F}^{n \times n}$ (where $\mathbb{F}$ is a field) is invertible if and only if its columns form a basis of $\mathbb{F}^n$.

Remark: By Proposition 3.2.6, every basis of $\mathbb{F}^n$ (where $\mathbb{F}$ is a field) contains exactly n vectors. In fact (see Theorem 3.2.16), if V is **any** finite-dimensional vector space,

[24] Here are the details. Fix a non-negative integer n. To show that $\{1, x, \ldots, x^n\}$ is linearly independent, we fix $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\alpha_0 \cdot 1 + \alpha_1 x + \cdots + \alpha_n x^n = 0$. The 0 on the right-hand-side denotes the zero polynomial, i.e. the polynomial whose coefficients are all zero. So, $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$, and we deduce that $\{1, x, \ldots, x^n\}$ is linearly independent. On the other hand, by definition, for every $p(x) \in \mathbb{P}^n_{\mathbb{F}}$, there exist $a_0, a_1, \ldots, a_n \in \mathbb{F}$ such that $p(x) = a_0 + a_1 x + \cdots + a_n x^n$. But then $p(x) = a_0 \cdot 1 + a_1 x + \cdots + a_n x^n$, and so $p(x)$ is a linear combination of $1, x, \ldots, x^n$. Thus, $\{1, x, \ldots, x^n\}$ is a spanning set of $\mathbb{P}^n_{\mathbb{F}}$. This proves that $\{1, x, \ldots, x^n\}$ is indeed a basis of $\mathbb{P}^n_{\mathbb{F}}$.

[25] For example, we can take $d := \max\{0, \deg\big(p_1(x)\big), \ldots, \deg\big(p_k(x)\big)\}$. (Here, we needed to put 0 into the set so that d would be defined even in the case when $k = 0$: the empty set does not have a minimum!)

then all bases of V are of the same size (i.e. contain exactly the same number of vectors). However, to prove this, we first need to develop some more theory.

Theorem 3.2.7. *Let V be a vector space over a field $\mathbb{F}$, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. Then the following are equivalent:*

(i) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V;

*(ii) for all vectors $\mathbf{v} \in V$, there exist **unique** scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$.*

Proof. Suppose first that (i) holds; we must show that (ii) holds. Fix $\mathbf{v} \in V$. We must show that there exist unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$. Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V, and consequently a spanning set of V, we know that every vector in V is a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. This proves existence. It remains to prove uniqueness. Fix scalars $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$ and $\mathbf{v} = \beta_1\mathbf{v}_1 + \cdots + \beta_n\mathbf{v}_n$. Then

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n \quad = \quad \beta_1\mathbf{v}_1 + \cdots + \beta_n\mathbf{v}_n,$$

and consequently,

$$(\alpha_1 - \beta_1)\mathbf{v}_1 + \cdots + (\alpha_n - \beta_n)\mathbf{v}_n \quad = \quad \mathbf{0}.$$

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent (because it is a basis of V), we deduce that $\alpha_1 - \beta_1 = \cdots = \alpha_n - \beta_n = 0$. So, $\alpha_1 = \beta_1, \ldots, \alpha_n = \beta_n$. This proves uniqueness, and (ii) follows.

Suppose now that (ii) holds; we must show that (i) holds. By (ii), every vector in V is a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and so $V = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. It remains to show that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent. Clearly, the equation $\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$ has a solution, namely $\alpha_1 = \cdots = \alpha_n = 0$; by (ii), this solution is unique, and we deduce that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent. Thus, (i) holds. □

Remark/Notation: Theorem 3.2.7 is one of the main reasons we care about bases. Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ ($n \geq 1$) is a basis of a vector space V over a field $\mathbb{F}$. Then by Theorem 3.2.7, to every vector $\mathbf{v} \in V$, we can associate a unique vector

$$\left[\ \mathbf{v}\ \right]_{\mathcal{B}} \quad := \quad \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

in $\mathbb{F}^n$ such that $\mathbf{v} = \alpha_1\mathbf{b}_1 + \cdots + \alpha_n\mathbf{b}_n$; the vector $\left[\ \mathbf{v}\ \right]_{\mathcal{B}}$ is called the *coordinate vector* of $\mathbf{v}$ associated with the basis $\mathcal{B}$. So, V is in a sense "equivalent" to $\mathbb{F}^n$. The technical term here is "isomorphic": V is "isomorphic" to $\mathbb{F}^n$. We will discuss this more formally in chapter 4 (see Proposition 4.3.1).

Example 3.2.8. *Let $\mathbb{F}$ be a field.*

(a) Consider the basis $\mathcal{E}_n = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of $\mathbb{F}^n$. Then for all $\mathbf{x} \in \mathbb{F}^n$, we have that $\left[\ \mathbf{x}\ \right]_{\mathcal{E}_n} = \mathbf{x}$.[26]

(b) Consider the basis $\mathcal{B} := \{1, x, \ldots, x^n\}$ of $\mathbb{P}^n_{\mathbb{F}}$. Then for all polynomials $p(x) = a_n x^n + \cdots + a_1 x + a_0$ in $\mathbb{P}^n_{\mathbb{F}}$ (where $a_n, \ldots, a_1, a_0 \in \mathbb{F}$), we have that $\left[\ p(x)\ \right]_{\mathcal{B}} = \begin{bmatrix} a_0 & a_1 & \ldots & a_n \end{bmatrix}^T$.

(c) Consider the basis

$$\mathcal{C} := \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

of $\mathbb{F}^{3\times 2}$. Then for all matrices

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,1} \end{bmatrix},$$

in $\mathbb{F}^{3\times 2}$, we have that $\left[\ A\ \right]_{\mathcal{C}} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{2,1} & a_{2,2} & a_{3,1} & a_{3,1} \end{bmatrix}^T$.[27]

Remark: When working with coordinate vectors, we must always **specify which basis we are working with**. This is because the same vector of a given finite-dimensional vector space may have different coordinate vectors with respect to different bases.

Remark: Note that if we change the order of basis elements, then coordinate vectors change. This is, in fact, the main reason for treating bases as **ordered** sets, rather than simply sets. For instance, consider the following two bases of $\mathbb{R}^{2\times 2}$:

- $\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$;
- $\mathcal{B}_2 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$.

[26] Indeed, for any $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$, we have that $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, and so $\left[\ \mathbf{x}\ \right]_{\mathcal{E}_n} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T = \mathbf{x}$.

[27] Indeed, we have that $A = a_{1,1}\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} + a_{1,2}\begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} + a_{2,1}\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} + a_{2,2}\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} + a_{3,1}\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} + a_{3,2}\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$, and it follows that $\left[\ A\ \right]_{\mathcal{C}} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{2,1} & a_{2,2} & a_{3,1} & a_{3,1} \end{bmatrix}^T$.

These two bases are the same except for the order in which matrices appear in them (the second and third matrix get swapped). But note that for any matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

in $\mathbb{R}^{2\times 2}$, we have that $\left[\ A\ \right]_{\mathcal{B}_1} = \begin{bmatrix} a & b & c & d \end{bmatrix}^T$ and $\left[\ A\ \right]_{\mathcal{B}_2} = \begin{bmatrix} a & c & b & d \end{bmatrix}^T$. So, $\left[\ A\ \right]_{\mathcal{B}_1} \neq \left[\ A\ \right]_{\mathcal{B}_2}$ (unless $b = c$).

The following proposition follows immediately from the definition of a coordinate vector, but it is useful to keep in mind.

Proposition 3.2.9. *Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ ($n \geq 1$) be a basis of a vector space V over a field $\mathbb{F}$. Then for all $i \in \{1, \dots, n\}$, we have that $\left[\ \mathbf{b}_i\ \right]_{\mathcal{B}} = \mathbf{e}_i^n$.*

Proof. Fix $i \in \{1, \dots, n\}$. Then

$$\mathbf{b}_i = 0\mathbf{b}_1 + \cdots + 0\mathbf{b}_{i-1} + 1\mathbf{b}_i + 0\mathbf{b}_{i+1} + \cdots + 0\mathbf{b}_n$$

and consequently,

$$\left[\ \mathbf{b}_i\ \right]_{\mathcal{B}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{-th entry}$$

i.e. $\left[\ \mathbf{b}_i\ \right]_{\mathcal{B}} = \mathbf{e}_i^n$. □

Rescaling basis vectors. The following proposition states that if we rescale the vectors of a basis using **non-zero** scalars, then we obtain another basis of the same vector space.

Proposition 3.2.10. *Let V be a vector space over a field $\mathbb{F}$, let $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$, and let $\alpha_1, \dots, \alpha_n \in \mathbb{F} \setminus \{0\}$. Then $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of V if and only if $\{\alpha_1\mathbf{v}_1, \dots, \alpha_k\mathbf{v}_n\}$ is a basis of V.*

Proof. This follows immediately from the definition of a basis and from Propositions 3.1.12 and 3.2.2. □

3.2.3 "Shrinking" a spanning set to a basis

The main goal of this subsection is to show that every spanning set of a vector space has a subset that is a basis of that vector space (see Theorem 3.2.14). We begin with a couple of technical propositions.

Proposition 3.2.11. *Let V be a vector space over a field $\mathbb{F}$, and let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in V$. Set $A := \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$. Then the following hold:*

(a) A is linearly independent if and only if no vector in A is a linear combination of the other vectors in A;[28]

(b) if A is a spanning set of V, and some vector $\mathbf{a}_i \in A$ is a linear combination of the other vectors in A, then $A \setminus \{\mathbf{a}_i\}$ is a spanning set of V.[29]

Proof. We first prove (a). We prove the following equivalent statement: "A is linearly dependent if and only if some vector of A is a linear combination of the other vectors in A."

Suppose first that A is linearly dependent. Then there exist scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$, not all zero, such that $\alpha_1\mathbf{a}_1 + \cdots + \alpha_k\mathbf{a}_k = \mathbf{0}$. Fix an index $i \in \{1, \ldots, k\}$ such that $\alpha_i \neq 0$. Then α_i has a multiplicative inverse α_i^{-1}, and we see that

$$\mathbf{a}_i \;=\; -\alpha_i^{-1}\alpha_1\mathbf{a}_1 - \cdots - \alpha_i^{-1}\alpha_{i-1}\mathbf{a}_{i-1} - \alpha_i^{-1}\alpha_{i+1}\mathbf{a}_{i+1} - \cdots - \alpha_i^{-1}\alpha_1\mathbf{a}_k.$$

So, $\mathbf{a}_i$ is a linear combination of the other vectors in A.

Suppose now that some vector in A is a linear combination of the other vectors in A. Say, $\mathbf{a}_i$ is a linear combination of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_k$. Then there exist scalars $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\mathbf{a}_i \;=\; \alpha_1\mathbf{a}_1 + \cdots + \alpha_{i-1}\mathbf{a}_{i-1} + \alpha_{i+1}\mathbf{a}_{i+1} + \cdots + \alpha_k\mathbf{a}_k.$$

We now set $\alpha_i = -1$, and we observe that

$$\alpha_1\mathbf{a}_1 + \cdots + \alpha_{i-1}\mathbf{a}_{i-1} + \alpha_i\mathbf{a}_i + \alpha_{i+1}\mathbf{a}_{i+1} + \cdots + \alpha_k\mathbf{a}_k \;=\; \mathbf{0}.$$

Since not all of $\alpha_1, \ldots, \alpha_k$ are zero (indeed, $\alpha_i \neq 0$), we see that $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly dependent. This proves (a).

We now prove (b). Assume that A is a spanning set of V, and that some $\mathbf{a}_i \in A$ is a linear combination of the other vectors in A. Then there exist scalars $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\mathbf{a}_i \;=\; \alpha_1\mathbf{a}_1 + \cdots + \alpha_{i-1}\mathbf{a}_{i-1} + \alpha_{i+1}\mathbf{a}_{i+1} + \cdots + \alpha_k\mathbf{a}_k.$$

[28] If A contains more than one copy of the same vector, then we treat each copy as distinct. So, when expressing a vector $\mathbf{v}$ in A as a linear combination of the "other" vectors in A, we are allowed to use any additional copies of $\mathbf{v}$ (if there are any) in that linear combination.

[29] If $\mathbf{a}_i$ appears more than once in A, then $A \setminus \{\mathbf{a}_i\}$ is understood to be the set obtained from A by removing only one copy of $\mathbf{a}_i$.

Now, fix any vector $\mathbf{v} \in V$. We must show that $\mathbf{v}$ is a linear combination of vectors in $A \setminus \{\mathbf{a}_i\} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_k\}$. Since $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is a spanning set of V, we know that there exist scalars $\beta_1, \ldots, \beta_k \in \mathbb{F}$ such that $\mathbf{v} = \beta_1\mathbf{a}_1 + \cdots + \beta_k\mathbf{a}_k$. We now compute:

$$\begin{aligned}
\mathbf{v} &= \beta_1\mathbf{a}_1 + \cdots + \beta_{i-1}\mathbf{a}_{i-1} + \beta_i\mathbf{a}_i + \beta_{i+1}\mathbf{a}_{i+1} + \cdots + \beta_k\mathbf{a}_k \\
&= \beta_1\mathbf{a}_1 + \cdots + \beta_{i-1}\mathbf{a}_{i-1} + \\
&\quad + \beta_i(\alpha_1\mathbf{a}_1 + \cdots + \alpha_{i-1}\mathbf{a}_{i-1} + \alpha_{i+1}\mathbf{a}_{i+1} + \cdots + \alpha_k\mathbf{a}_k) \\
&\quad + \beta_{i+1}\mathbf{a}_{i+1} + \cdots + \beta_k\mathbf{a}_k \\
&= (\beta_1 + \beta_i\alpha_1)\mathbf{a}_1 + \cdots + (\beta_{i-1} + \beta_i\alpha_{i-1})\mathbf{a}_{i-1} + \\
&\quad + (\beta_{i+1} + \beta_i\alpha_{i+1})\mathbf{a}_{i+1} + \cdots + (\beta_k + \beta_i\alpha_k)\mathbf{a}_k.
\end{aligned}$$

So, $\mathbf{v}$ is a linear combination of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_k$, and (b) follows. □

We note that Proposition 3.2.11(a) can be slightly strengthened as follows.

Proposition 3.2.12. *Let V be a vector space over a field $\mathbb{F}$, and let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in V$. Then the set $A := \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly **dependent** if and only if there exists some index $i \in \{1, \ldots, k\}$ such that $\mathbf{a}_i$ is a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$.*[30]

Proof. Exercise. □

Proposition 3.2.13. *Let V be a vector space over a field $\mathbb{F}$, and let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ be a spanning set of V. Let $B' \subseteq B$ be such that every vector in $B \setminus B'$ is a linear combination of vectors in B'. Then B' is a spanning set of V.*

Proof. Choose a set $\widetilde{B}$ such that

- $B' \subseteq \widetilde{B} \subseteq B$,
- $\widetilde{B}$ is a spanning set of V;
- subject to the above, $\widetilde{B}$ is as small as possible.

(The fact that $\widetilde{B}$ exists follows from the fact that $B' \subseteq B \subseteq B$, and B is a spanning set of V.) If $\widetilde{B} = B'$, then we are done. So, assume that $B' \subsetneq \widetilde{B}$, and fix some $\mathbf{v} \in \widetilde{B} \setminus B'$. Then $\mathbf{v}$ is a linear combination of the other vectors in $\widetilde{B}$ (because $\mathbf{v}$ is a linear combination of the vectors in B'), and so by Proposition 3.2.11(b), $\widetilde{B} \setminus \{\mathbf{v}\}$ is a spanning set of V. But now $\widetilde{B} \setminus \mathbf{v}$ contradicts the minimality of $\widetilde{B}$. □

[30] By definition, $\mathbf{0}$ is a linear combination of the empty set/list of vectors, and so if $\mathbf{a}_1 = \mathbf{0}$, then $\mathbf{a}_1$ is, in fact, a linear combination of the empty list $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$.

Our next theorem (Theorem 3.2.14) states that, given any spanning set B of a vector space V over a field $\mathbb{F}$, we can obtain a basis of V by possibly removing some vectors from B. As we shall see later (see Theorem 3.2.19), any linearly independent set in a **finite-dimensional** vector space can be extended to a basis; however, we cannot prove this yet.

Theorem 3.2.14. *Let V be a vector space over a field $\mathbb{F}$, and let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ be a spanning set of V. Then some subset of B is a basis of V.*

Proof. Let $B' \subseteq B$ be a spanning set of V that has as few elements as possible.[31] We claim that B' is a basis of V. It suffices to show that B' is linearly independent. Suppose otherwise. Then Proposition 3.2.11(a) guarantees that some $\mathbf{b} \in B'$ is a linear combination of the other vectors in B'; but then by Proposition 3.2.11(b), $B' \setminus \{\mathbf{b}\}$ is a spanning set of V, contrary to the minimality of B'. □

3.2.4 The Steinitz exchange lemma

In this subsection, we prove a technical result called the Steinitz exchange lemma. It essentially states that if we are given a linearly independent set A and a spanning set B of a vector space V, then we can extend A to a spanning set $A \cup B'$ of V of the same size as B, and moreover, we can choose B' so that it is a subset of B. (So, our spanning set $A \cup B'$ is obtained by adding to A some vectors from B. It is possible that $B' = \emptyset$, but the important point is that we do not add to A any vectors that are not in B.)

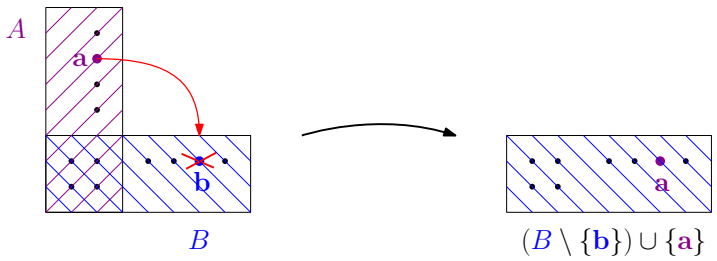
We note that the Steinitz exchange lemma has a number of important corollaries, which we discuss in our next subsection (see subsection 3.2.5). Perhaps the most important of these corollaries is the fact that any two bases of a finite-dimensional vector space are of the same cardinality, i.e. contain the same number of vectors (see Theorem 3.2.16).[32] This will allow us to define the "dimension" of a finite-dimensional vector space as the number of vectors in any basis of that vector space.

The proof of the Steinitz exchange lemma essentially proceeds by induction on $|A \setminus B|$ (i.e. the numbers of vectors that belong to A, but not to B), using Lemma 3.2.15 (a technical lemma stated and proven below).

Lemma 3.2.15. *Let V be a vector space over a field $\mathbb{F}$. Let $\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{b}_1, \ldots, \mathbf{b}_\ell \in V$, and assume that $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are pairwise distinct and that $\mathbf{b}_1, \ldots, \mathbf{b}_\ell$ are pairwise distinct. Assume furthermore that $A := \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is a linearly independent set in V, and that $B := \{\mathbf{b}_1, \ldots, \mathbf{b}_\ell\}$ is a spanning set of V. Then for all $\mathbf{a} \in A \setminus B$, there exists some $\mathbf{b} \in B \setminus A$ such that $(B \setminus \{\mathbf{b}\}) \cup \{\mathbf{a}\}$ is a spanning set of V.*

[31] Let us explain why B' exists. Clearly, B has a subset (namely itself) that is a spanning set of V. Of all subsets of B that span V, we choose B' to be one of minimum size.

[32] We have already proven this for $\mathbb{F}^n$, where $\mathbb{F}$ is a field (see Proposition 3.2.6). However, Theorem 3.2.16 deals with general finite-dimensional vector spaces, and not just $\mathbb{F}^n$.



Proof. We may assume that $A \not\subseteq B$, for otherwise, the lemma is vacuously true.[33] Fix any $\mathbf{a} \in A \setminus B$. Then there exists an index $i \in \{1, \ldots, k\}$ such that $\mathbf{a} = \mathbf{a}_i$. Since $\mathbf{a}_i \in V = \mathrm{Span}(B)$, we know that there exist scalars $\alpha_1, \ldots, \alpha_\ell \in \mathbb{F}$ such that

$$\mathbf{a}_i \quad = \quad \alpha_1 \mathbf{b}_1 + \cdots + \alpha_\ell \mathbf{b}_\ell.$$

> **Claim.** There exists an index $j \in \{1, \ldots, \ell\}$ such that $\alpha_j \neq 0$ and $\mathbf{b}_j \in B \setminus A$.

Proof of the Claim. Suppose otherwise. Then for all $j \in \{1, \ldots, \ell\}$ such that $\alpha_j \neq 0$, we have that $\mathbf{b}_j \in B \cap A \subseteq A \setminus \{\mathbf{a}_i\}$.[34] But now $\mathbf{a}_i$ is a linear combination of the other vectors in the linearly independent set A,[35] contrary to Proposition 3.2.11(a). ♦

Using the Claim, we fix an index $j \in \{1, \ldots, \ell\}$ such that $\alpha_j \neq 0$ and $\mathbf{b}_j \in B \setminus A$. We will show that $(B \setminus \{\mathbf{b}_j\}) \cup \{\mathbf{a}_i\}$ is a spanning set of V (this will complete the proof of the lemma). Since $\mathbf{b}_j \neq \mathbf{a}_i$,[36] we see that $(B \setminus \{\mathbf{b}_j\}) \cup \{\mathbf{a}_i\} = (B \cup \{\mathbf{a}_i\}) \setminus \{\mathbf{b}_j\}$, and we need to show that $(B \cup \{\mathbf{a}_i\}) \setminus \{\mathbf{b}_j\}$ is a spanning set of V. Since B is a spanning set of V, so is $B \cup \{\mathbf{a}_i\}$. In view of Proposition 3.2.11(b), it now suffices to show that $\mathbf{b}_j$ is a linear combination of the other vectors in $B \cup \{\mathbf{a}_i\}$. Since $\mathbf{a}_i = \alpha_1 \mathbf{b}_1 + \cdots + \alpha_\ell \mathbf{b}_\ell$, we see that

$$\alpha_j \mathbf{b}_j \quad = \quad \mathbf{a}_i - \alpha_1 \mathbf{b}_1 - \cdots - \alpha_{j-1} \mathbf{b}_{j-1} - \alpha_{j+1} \mathbf{b}_{j+1} - \cdots - \alpha_\ell \mathbf{b}_\ell.$$

Since $\alpha_j \neq 0$, we know that α_j has a multiplicative inverse α_j^{-1}, and we deduce that

$$\mathbf{b}_j \quad = \quad \alpha_j^{-1} \mathbf{a}_i - \alpha_j^{-1} \alpha_1 \mathbf{b}_1 - \cdots - \alpha_j^{-1} \alpha_{j-1} \mathbf{b}_{j-1} - \alpha_j^{-1} \alpha_{j+1} \mathbf{b}_{j+1} - \cdots - \alpha_j^{-1} \alpha_\ell \mathbf{b}_\ell.$$

So, $\mathbf{b}_j$ is indeed a linear combination of the other vectors in $B \cup \{\mathbf{a}_i\}$, and we are done. □

[33] Indeed, if $A \subseteq B$, then there are no vectors $\mathbf{a} \in A \setminus B$, and so there is nothing to prove.

[34] We are using the fact that $\mathbf{a}_i \in A \setminus B$, and so $B \cap A \subseteq A \setminus \{\mathbf{a}_i\}$.

[35] Indeed, since $\mathbf{a}_i = \alpha_1 \mathbf{b}_1 + \cdots + \alpha_\ell \mathbf{b}_\ell$, we see that $\mathbf{a}_i$ is a linear combination of those $\mathbf{b}_j$'s for which $\alpha_j \neq 0$. But all such $\mathbf{b}_j$'s belong to $A \setminus \{\mathbf{a}_i\}$.

[36] This is because $\mathbf{a}_i \in A$, whereas $\mathbf{b}_j \in B \setminus A$.

The Steinitz exchange lemma. *Let V be a vector space over a field $\mathbb{F}$, let $\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{b}_1, \ldots, \mathbf{b}_\ell \in V$, and assume that $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are pairwise distinct and that $\mathbf{b}_1, \ldots, \mathbf{b}_\ell$ are pairwise distinct. Assume furthermore that $A := \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is a linearly independent set in V, and assume that $B := \{\mathbf{b}_1, \ldots, \mathbf{b}_\ell\}$ is a spanning set of V. Then $k \leq \ell$ (i.e. $|A| \leq |B|$). Moreover, there exists a set $B' \subseteq B \setminus A$ such that $|B'| = |B| - |A| = \ell - k$ and $A \cup B'$ is a spanning set of V.*

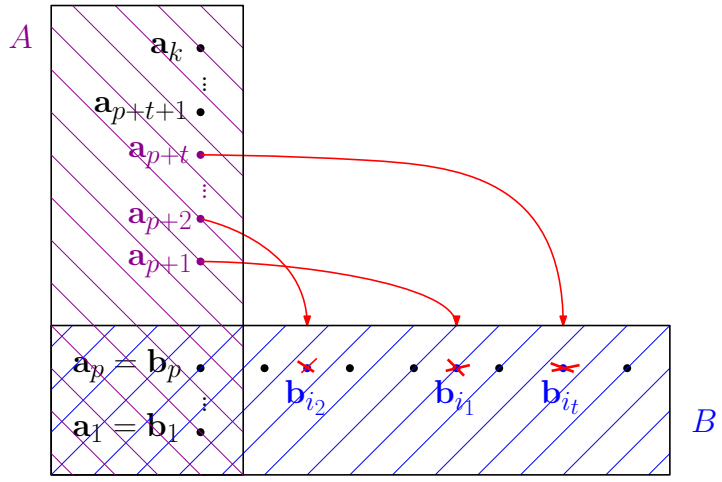
Remark: Since $A \cap B' = \emptyset$ (because $B' \subseteq B \setminus A$), we have that $|A \cup B'| = |A| + |B'| = |A| + \big(|B| - |A|\big) = |B|$. So, the "new" spanning set $A \cup B'$ is of the same size as the "old" spanning set B.

Proof. We may assume that $A \not\subseteq B$, for otherwise, the result is immediate.[37] Set $p := |A \cap B|$.[38] After possibly permuting the elements of A and B, we may assume that the following hold:

- $\mathbf{a}_1 = \mathbf{a}_1, \ldots, \mathbf{a}_p = \mathbf{b}_p$;
- $\{\mathbf{a}_{p+1}, \ldots, \mathbf{a}_k\} \cap \{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} \neq \emptyset$.

We now prove a technical claim.

> **Claim.** For all $t \in \{0, \ldots, k-p\}$, there exist pairwise distinct indices $i_1, \ldots, i_t \in \{p+1, \ldots, \ell\}$ such that
>
> $$\{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \cup \{\mathbf{a}_{p+1}, \ldots, \mathbf{a}_{p+t}\} \cup \Big(\{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} \setminus \{\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_t}\}\Big)$$
>
> is a spanning set of V.



[37] Indeed, if $A \subseteq B$, then $|A| \leq |B|$, and we may set $B' := B \setminus A$.
[38] Since $A \not\subseteq B$, we see that $p < k$.

Proof of the Claim. We proceed by induction on t, using Lemma 3.2.15.

For $t = 0$, we need only show that $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \cup \{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\}$ is a spanning set of V. But note that $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \cup \{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} = B$, and by hypothesis, B is a spanning set of V.

Now, fix some $t \in \{0, \ldots, k-p-1\}$, and assume inductively that the statement is true for t, i.e. that there exist pairwise distinct indices $i_1, \ldots, i_t \in \{p+1, \ldots, \ell\}$ such that

$$B_t \;:=\; \{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \cup \{\mathbf{a}_{p+1}, \ldots, \mathbf{a}_{p+t}\} \cup \Big(\{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} \setminus \{\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_t}\}\Big)$$

is a spanning set of V. Now, $\mathbf{a}_{p+t+1} \in A \setminus B_t$, and so by Lemma 3.2.15, there exists some $\mathbf{b} \in B_t \setminus A$ such that $(B_t \setminus \{\mathbf{b}\}) \cup \{\mathbf{a}_{p+t+1}\}$ is a spanning set of V. Since $B_t \setminus A = \{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} \setminus \{\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_t}\}$, there exists some $i_{t+1} \in \{p+1, \ldots, \ell\} \setminus \{i_1, \ldots, i_t\}$ such that $\mathbf{b} = \mathbf{b}_{i_{t+1}}$. Now $i_1, \ldots, i_t, i_{t+1}$ are pairwise distinct indices in $\{p+1, \ldots, \ell\}$, and

$$\begin{aligned}(B_t \setminus \{\mathbf{b}\}) \cup \{\mathbf{a}_{p+t+1}\} \;&=\; \{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \cup \{\mathbf{a}_{p+1}, \ldots, \mathbf{a}_{p+t}, \mathbf{a}_{p+t+1}\} \cup \\ &\quad \cup \Big(\{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} \setminus \{\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_t}, \mathbf{b}_{i_{t+1}}\}\Big)\end{aligned}$$

is a spanning set of V. This completes the induction. ♦

We now apply the Claim for $t = k - p$, and we get that there exist pairwise distinct indices $i_1, \ldots, i_{k-p} \in \{p+1, \ldots, \ell\}$ such that

$$C \;:=\; \{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \cup \{\mathbf{a}_{p+1}, \ldots, \mathbf{a}_k\} \cup \Big(\{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} \setminus \{\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_{k-p}}\}\Big)$$

is a spanning set of V. But note that $|C| = \ell = |B|$ and $A \subseteq C$. Thus, $|A| \leq |C| = |B|$, and so $k \leq \ell$. Next, set $B' := \{\mathbf{b}_{p+1}, \ldots, \mathbf{b}_\ell\} \setminus \{\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_{k-p}}\}$. Then $B' \subseteq B \setminus A$, $|B'| = (\ell - p) - (k - p) = \ell - k = |B| - |A|$, and $C = A \cup B'$ is a spanning set of V. This completes the argument. □

Remark: For technical reasons (in order to get the set B'), the Steinitz exchange lemma assumes that the sets A and B contain no repetitions.[39] However, if we only care about the "$|A| \leq |B|$" part of the Steinitz exchange lemma (which is what we usually care about), then this assumption is not necessary. Indeed, suppose that V is a vector space over a field $\mathbb{F}$, and suppose that A is a linearly independent set of vectors in V and that B is a spanning set of V (with repetitions allowed). Since A is linearly independent, it contains no repetitions; however, B may possibly contain repetitions. But then we let $\widetilde{B}$ be the set obtained from B by eliminating repetitions. Then $\widetilde{B}$ is still a spanning set of V, and by the Steinitz exchange lemma, we get that $|A| \leq |\widetilde{B}| \leq |B|$.

[39] Actually, it would be possible to state and prove a version of the Steinitz exchange lemma that allows repetitions. However, this would be notationally messy.

3.2.5 The dimension of a finite-dimensional vector space

The following theorem is perhaps the most important corollary of the Steinitz exchange lemma.

Theorem 3.2.16. *Let V be a finite-dimensional vector space over a field $\mathbb{F}$. Then all bases of V are of the same size.*

Proof. We apply the Steinitz exchange lemma twice. Fix bases $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V. Since $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is linearly independent and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a spanning set of V, the Steinitz exchange lemma guarantees that $m \leq n$. On the other hand, since $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a linearly independent set and $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is a spanning set of V, the Steinitz exchange lemma guarantees that $n \leq m$. So, $m = n$. □

The *dimension* of a finite-dimensional vector space V over a field $\mathbb{F}$, denoted by $\dim(V)$, is the number of elements in any basis of V (by Theorem 3.2.16, this is well defined).

Remarks:

- Note that $\dim(\{\mathbf{0}\}) = 0$ (where $\{\mathbf{0}\}$ is understood to be a vector space over an arbitrary field $\mathbb{F}$), because $\emptyset$ is a basis of $\{\mathbf{0}\}$.
- For any field $\mathbb{F}$, we have that $\dim(\mathbb{F}^n) = n$, because the standard basis of $\mathbb{F}^n$ has n elements. We note, however, that the standard basis is not the only basis of $\mathbb{F}^n$ (except in some very special cases; see Proposition 3.2.6).

Theorem 3.2.17 (below) is another important corollary of the Steinitz exchange lemma. It essentially states that if V is a finite-dimensional vector space, then any linearly independent set in V is of size at most $\dim(V)$, and any spanning set of V is of size at most V. Schematically (and informally), we can summarize this is follows:

$$|\text{linearly independent set of } V| \quad \leq \quad \dim(V) \quad \leq \quad |\text{spanning set of } V|.$$

Theorem 3.2.17. *Let V be a finite-dimensional vector space over a field $\mathbb{F}$, and set $n := \dim(V)$. Then both the following hold:*

(a) every linearly independent set of vectors in V has at most n vectors;

(b) every spanning set of V has at least n vectors.

Proof. Fix a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of V. Then B is both a linearly independent set and a spanning set of V. Now, by the Steinitz exchange lemma, the number of vectors in any linearly independent set of V is at most the number of vectors in the spanning set B of V, which is n; so, (a) holds. On the other hand, by the Steinitz exchange lemma, any spanning set of V has at least as many vectors as the linearly independent set B; so, (b) holds. □

By Theorem 3.2.17(a), linearly independent sets in any finite-dimensional vector space have bounded size (bounded above by the dimension of the vector space in question). On the other hand, by Proposition 3.2.18 (below), infinite-dimensional vector spaces have linearly independent sets of arbitrarily large (finite) size. For instance, if $\mathbb{F}$ is a field, then for any positive integer n, $\{1, x, x^2, \ldots, x^n\}$ is a linearly independent set in $\mathbb{P}_{\mathbb{F}}$ (the vector space of all polynomials with coefficients in $\mathbb{F}$).

Proposition 3.2.18. *Let V be an infinite-dimensional vector space over a field $\mathbb{F}$. Then for every non-negative integer n, V has a linearly independent set of size n.*

Proof. We proceed by induction on n. For $n = 0$, we observe that $\emptyset$ is a linearly independent set of size 0 in V. Next, fix a non-negative integer n, and assume that V has a linearly independent set of size n, say $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Then $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is not a spanning set of V, for otherwise, it would be a basis of V, contrary to the fact that V is infinite-dimensional. Thus, $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \subsetneqq V$; fix some $\mathbf{a}_{n+1} \in V \setminus \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. We now claim that $\{\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{a}_{n+1}\}$ is a linearly independent set in V. Fix $\alpha_1, \ldots, \alpha_n, \alpha_{n+1} \in \mathbb{F}$ such that

$$\alpha_1\mathbf{a}_1 + \cdots + \alpha_n\mathbf{a}_n + \alpha_{n+1}\mathbf{a}_{n+1} \quad = \quad \mathbf{0}.$$

If $\alpha_{n+1} \neq 0$, then $\mathbf{a}_{n+1} = -\alpha_{n+1}^{-1}\alpha_1\mathbf{a}_1 - \cdots - \alpha_{n+1}^{-1}\alpha_n\mathbf{a}_n$, contrary to the fact that $\mathbf{a}_{n+1} \notin \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. So, $\alpha_{n+1} = 0$, and it follows that

$$\alpha_1\mathbf{a}_1 + \cdots + \alpha_n\mathbf{a}_n \quad = \quad \mathbf{0}.$$

But since $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a linearly independent set, it follows that $\alpha_1 = \cdots = \alpha_n = 0$. We have now shown that $\alpha_1 = \cdots = \alpha_n = \alpha_{n+1} = 0$, and we deduce that the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{a}_{n+1}\}$ is linearly independent. This completes the induction. □

Our next theorem states that any linearly independent set of vectors in a finite-dimensional vector space can be extended to a basis of that vector space.

Theorem 3.2.19. *Let V be a **finite-dimensional** vector space over a field $\mathbb{F}$, and let $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ be a linearly independent set of vectors in V. Then there exists some basis of V that contains all of $\mathbf{a}_1, \ldots, \mathbf{a}_k$.*

Proof. Set $n := \dim(V)$. By Theorem 3.2.17, any linearly independent set of vectors in V has at most n vectors; in particular, $k \leq n$ (because $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent). Now, let A be a linearly independent set that contains vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$, and subject to that, is of maximum possible size.[40] Set $A =$

[40] Let us explain why A exists. There exists at least one linearly independent set that contains vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$, namely, the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$. On the other hand, all linearly independent sets are of size at most n, and in particular, there is an upper bound on the size of linearly independent sets containing $\mathbf{a}_1, \ldots, \mathbf{a}_k$. So, A exists.

$\{\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_{k+\ell}\}$. We claim that A is a basis of V. Since A is linearly independent, it suffices to show that A is a spanning set of V. Fix $\mathbf{v} \in V$; we must show that $\mathbf{v}$ is a linear combination of vectors in A. If $\mathbf{v} \in A$, then this is immediate.[41] So, assume that $\mathbf{v} \notin A$. Then by the maximality of A, the set $\{\mathbf{v}\} \cup A$ is not linearly independent. So, there exist scalars $\alpha_0, \alpha_1, \ldots, \alpha_{k+\ell} \in \mathbb{F}$, not all zero, such that

$$\alpha_0\mathbf{v} + \alpha_1\mathbf{a}_1 + \cdots + \alpha_{k+\ell}\mathbf{a}_{k+\ell} \quad = \quad \mathbf{0}.$$

If $\alpha_0 = 0$, then at least one of $\alpha_1, \ldots, \alpha_{k+\ell}$ is non-zero and $\alpha_1\mathbf{a}_1 + \cdots + \alpha_{k+\ell}\mathbf{a}_{k+\ell} = \mathbf{0}$, contrary to the fact that A is linearly independent. So, $\alpha_0 \neq 0$, it follows that

$$\mathbf{v} \quad = \quad (-\alpha_0^{-1}\alpha_1)\mathbf{a}_1 + \cdots + (-\alpha_0^{-1}\alpha_{k+\ell})\mathbf{a}_{k+\ell},$$

and we see that $\mathbf{v}$ is a linear combination of vectors in A. This proves that A is a basis of V, and we are done. □

Remark: Suppose that V is a vector space over a field $\mathbb{F}$. By Theorem 3.2.14, any (finite) spanning set of V contains a subset that is a basis of V; in particular, if a vector space has a (finite) spanning set, then it is finite-dimensional.[42] On the other hand, by Theorem 3.2.19, if V is **finite-dimensional**, then any linearly independent set in V can be extended to a basis of V.[43]

Theorems 3.2.14 and 3.2.19 together yield the following corollary.

Corollary 3.2.20. *Let V be a finite-dimensional vector space over a field $\mathbb{F}$, and set $n := dim(V)$. Then both the following hold:*

(a) any linearly independent set of n vectors of V is a basis of V;

(b) any set of n vectors of V that spans V is a basis of V.

Proof. We first prove (a). Let A be any linearly independent set of vectors in V such that $|A| = n$. By Theorem 3.2.19, V has a basis A' such that $A \subseteq A'$. Since $\dim(V) = n$, we see that $|A'| = n$. Since $|A| = n$ and $A \subseteq A'$, it follows that $A = A'$. So, A is a basis of V (because A' is). This proves (a).

It remains to prove (b). Let B be any set of n vectors of V such that $V = \mathrm{Span}(B)$. Then by Theorem 3.2.14, V has a basis B' such that $B' \subseteq B$. Since $\dim(V) = n$, we see that $|B'| = n$. Since $|B| = n$ and $B' \subseteq B$, it follows that $B' = B$. So, B is a basis of V (because B' is). This proves (b). □

[41] Indeed, suppose $\mathbf{v} \in A$. Then there exists an index $i \in \{1, \ldots, k+\ell\}$ such that $\mathbf{v} = \mathbf{a}_i$. Now set $\alpha_i = 1$, and for all $j \in \{1, \ldots, k+\ell\}$, set $\alpha_j = 0$. Then $\mathbf{v} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_{k+\ell}\mathbf{a}_{k+\ell}$, and so $\mathbf{v}$ is linear combination of vectors in A.

[42] It is possible to define infinite spanning sets, and we will very briefly discuss this in subsection 3.2.7. However, except in subsection 3.2.7, we only consider finite spanning sets in these lecture notes.

[43] Actually, Theorem 3.2.19 can be generalized to all vector spaces (including infinite-dimensional ones), as long as we allow infinite bases. However, the proof would be significantly more complicated, and it would involve some relatively advanced set theory (in particular, "Zorn's lemma").

Theorem 3.2.21. *Let V be a finite-dimensional vector space over a field $\mathbb{F}$, and let U be a subspace of V. Then all the following hold:*

(a) U is finite-dimensional;

(b) $dim(U) \leq dim(V)$;

(c) if $dim(U) = dim(V)$, then $U = V$.

Proof. Set $n := \dim(V)$. Since U is a subspace in V, any linearly independent set of vectors in U is also linearly independent in V, and by Theorem 3.2.17(a), any such set contains at most n vectors. Now, let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a linearly independent set of vectors in U of maximum possible size.[44] (Then $k \leq n$.) Let us show that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ spans U. Fix $\mathbf{u} \in U$; we must show that $\mathbf{u}$ is a linear combination of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$. If $\mathbf{u} \in \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, then this is immediate. So, assume that $\mathbf{u} \notin \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. By the maximality of $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, we see that $\{\mathbf{u}, \mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly dependent. So, there exist scalars $\alpha_0, \alpha_1, \ldots, \alpha_k \in \mathbb{F}$, not all zero, such that $\alpha_0\mathbf{u} + \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k = \mathbf{0}$. If $\alpha_0 = 0$, then $\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k = \mathbf{0}$ and at least one of the scalars $\alpha_1, \ldots, \alpha_k$ is non-zero, contrary to the fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent. So, $\alpha_0 \neq 0$, and we deduce that $\mathbf{u} = (-\alpha_0^{-1}\alpha_1)\mathbf{u}_1 + \cdots + (-\alpha_0^{-1}\alpha_k)\mathbf{u}_k$. So, $\mathbf{u} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$, and we deduce that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a spanning set of U. So, $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of U, and it follows that U is finite-dimensional, with $\dim(U) = k$. So, $\dim(U) = k \leq n = \dim(V)$. This proves (a) and (b). It remains to prove (c). Suppose that $\dim(U) = \dim(V)$, i.e. $k = n$. But now $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a linearly independent set of n vectors in V, and so Corollary 3.2.20 guarantees that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of V. So, $U = \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k) = V$, and we are done. □

Warning: Theorem 3.2.21(c) fails if V is infinite-dimensional! Infinite-dimensional vector spaces can have proper subspaces that are infinite-dimensional. For example, $\{p(x) \in \mathbb{P}_{\mathbb{R}} \mid p(0) = 0\}$ is an infinite-dimensional proper subspace of $\mathbb{P}_{\mathbb{R}}$.[45]

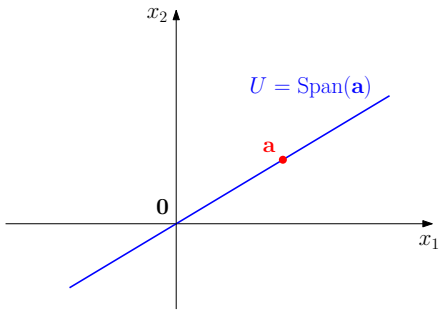
Subspaces of $\mathbb{R}^n$. Let us consider a geometric interpretation of subspaces in $\mathbb{R}^n$.

- The only 0-dimensional subspace of $\mathbb{R}^n$ is $\{\mathbf{0}\}$.[46]
- 1-dimensional subspaces of $\mathbb{R}^n$ are lines though the origin. Indeed, suppose that $\{\mathbf{a}\}$ is a basis of a subspace U of $\mathbb{R}^n$. Then $\mathbf{a} \neq \mathbf{0}$ (by linear independence), and we see that $U = \mathrm{Span}(\mathbf{a})$ is the line through the origin and $\mathbf{a}$. (This is illustrated below for the case of $\mathbb{R}^2$.) So, 1-dimensional subspaces of $\mathbb{R}^n$ essentially look like copies of $\mathbb{R}^1$ inside of $\mathbb{R}^n$.

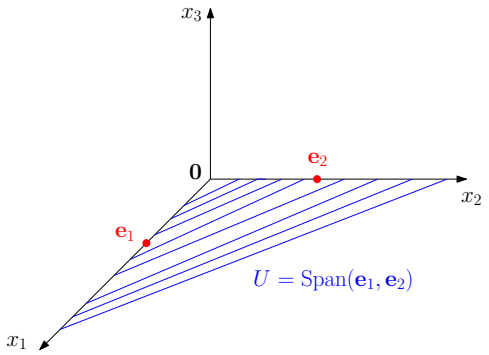
[44] It is possible that $k = 0$. In this case, our linearly independent set is empty.

[45] Details?

[46] This holds for any vector space V (not just $\mathbb{R}^n$), as long as the zero vector is from the vector space V in question. Recall that we defined $\mathrm{Span}(\emptyset) = \{\mathbf{0}\}$, and obviously, $\emptyset$ is linearly independent.



- 2-dimensional subspaces of $\mathbb{R}^n$ are planes through the origin. Indeed, suppose that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a basis of a subspace U of $\mathbb{R}^n$. By linear independence, $\mathbf{a}_1, \mathbf{a}_2$ are both non-zero and are not scalar multiples of each other. So, $U = \text{Span}(\mathbf{a}_1, \mathbf{a}_2)$ is the plane through the origin and through $\mathbf{a}_1$ and $\mathbf{a}_2$. For example, the subspace of $\mathbb{R}^3$ whose basis is $\{\mathbf{e}_1, \mathbf{e}_2\}$ is simply the x_1x_2-plane in $\mathbb{R}^3$ (illustrated below). In general, 2-dimensional subspaces of $\mathbb{R}^n$ look like copies of $\mathbb{R}^2$ inside of $\mathbb{R}^n$ (of course, those copies of $\mathbb{R}^2$, i.e. planes, may possibly be "tilted," i.e. not formed by any two of the coordinate axes of $\mathbb{R}^n$); however, they must all pass through the origin.



- In general, for a positive integer $m \leq n$, an m-dimensional subspace of $\mathbb{R}^n$ looks like a copy of $\mathbb{R}^m$ inside of $\mathbb{R}^n$. Again, our copy of $\mathbb{R}^m$ may possibly be "tilted," i.e. not be formed by any m of the n axes of $\mathbb{R}^n$. However, it must pass through the origin.

3.2.6 On the dimension of some vector spaces obtained from old ones

Recall from subsection 3.1.3 that if U and W are vector spaces over a field $\mathbb{F}$, then $U \times W$ is also a vector space over $\mathbb{F}$ (with vector addition and scalar multiplication defined in a natural way, as explained in subsection 3.1.3).

Proposition 3.2.22. *Let U and W be finite-dimensional vector spaces over a field $\mathbb{F}$. Then the vector space $U \times W$ is finite-dimensional, and moreover,*

$$dim(U \times W) \quad = \quad dim(U) + dim(W).$$

Proof (outline). Let $\mathbf{0}_U$ be the zero vector of the vector space U, and let $\mathbf{0}_W$ be the zero of the vector space W. Set $m := \dim(U)$ and $n := \dim(W)$, and fix a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of U and a basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ of W. It is then straightforward to check that

$$\Big\{(\mathbf{u}_1, \mathbf{0}_W), \ldots, (\mathbf{u}_m, \mathbf{0}_W), (\mathbf{0}_U, \mathbf{w}_1), \ldots, (\mathbf{0}_U, \mathbf{w}_n)\Big\}$$

is a basis of $U \times W$ (the details are left as an exercise), and consequently, $\dim(U \times W) = m + n = \dim(U) + \dim(W)$. □

Recall from subsection 3.1.3 that if V is a vector space over a field $\mathbb{F}$, and U and W are subspaces of V, then $U \cap W$ and $U + W$ are also subspaces of V. Theorem 3.2.23 (below) specifies the relationship between the dimensions of these four subspaces of V.

Theorem 3.2.23. *Let V be a finite-dimensional vector space over a field $\mathbb{F}$, and let U and W be subspaces of V. Then $U \cap W$ and $U + W$ are subspaces of V. Moreover, U, W, $U \cap W$, and $U + W$ are all finite-dimensional and satisfy*

$$dim(U + W) + dim(U \cap W) \quad = \quad dim(U) + dim(W).$$

Proof (outline). The fact that $U \cap W$ and $U + W$ are subspaces of V follows from the discussion in subsection 3.1.3. Since V is finite-dimensional, Theorem 3.2.21 guarantees that all its subspaces are finite dimensional; in particular, U, W, $U \cap W$, and $U + W$ are all finite-dimensional. Set $m := \dim(U)$, $n := \dim(W)$, and $p := \dim(U \cap W)$. Fix a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of $U \cap W$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a linearly independent set in the finite-dimensional vector space U, and so by Theorem 3.2.19, it can be extended to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{u}_1, \ldots, \mathbf{u}_{m-p}\}$ of U.[47] Similarly, $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ can be extended to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{w}_1, \ldots, \mathbf{w}_{n-p}\}$ of W.[48] It is now straightforward to check that

$$\Big\{\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{u}_1, \ldots, \mathbf{u}_{m-p}, \mathbf{w}_1, \ldots, \mathbf{w}_{n-p}\Big\}$$

is a basis of $U + W$ (the details are left as an exercise). So,

$$\dim(U + W) \quad = \quad p + (m - p) + (n - p) \quad = \quad m + n - p.$$

[47] We are using the fact that $\dim(U) = m$, and so to extend the linearly independent set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ to a basis of U, we must add precisely $m - p$ suitably chosen vectors to this set.

[48] We are using the fact that $\dim(W) = n$, and so to extend the linearly independent set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ to a basis of W, we must add precisely $n - p$ suitably chosen vectors to this set.

It now follows that

$$\begin{aligned}\dim(U+W)+\dim(U\cap W) &= (m+n-p)+p\\ &= m+n\\ &= \dim(U)+\dim(W),\end{aligned}$$

which is what we needed to show. □

If V is a vector space over a field $\mathbb{F}$ and U and V are its subspaces such that $U\cap W=\{\mathbf{0}\}$ and $V=U+W$, then we say that V is the *direct sum* of U and W, and we write $V=U\oplus W$. If $V=U\oplus W$ is also finite-dimensional, then Theorem 3.2.23 immediately implies that $\dim(V)=\dim(U)+\dim(W)$.[49] Moreover, we have the following theorem.

Theorem 3.2.24. *Let V be a vector space over a field $\mathbb{F}$, and let U and W be subspaces of V such that $V=U\oplus W$. Then for all $\mathbf{v}\in V$, there exist unique $\mathbf{u}\in U$ and $\mathbf{w}\in W$ such that $\mathbf{v}=\mathbf{u}+\mathbf{w}$.*

Proof. Exercise. □

3.2.7 A very brief introduction to infinite bases

We can define a basis of a vector space in more generality, as follows. Let V be a vector space over a field $\mathbb{F}$, and let $B\subseteq V$. (B may possibly be infinite. However, we do not allow repetitions in B.)

- B is said to be *linearly independent* provided that for all pairwise distinct vectors $\mathbf{v}_1,\dots,\mathbf{v}_k\in V$, and all scalars $\alpha_1,\dots,\alpha_k\in\mathbb{F}$, if $\alpha_1\mathbf{v}_1+\cdots+\alpha_k\mathbf{v}_k=\mathbf{0}$, then $\alpha_1=\cdots=\alpha_k=0$.[50]
- $\mathrm{Span}(B)=\big\{\alpha_1\mathbf{v}_1+\cdots+\alpha_k\mathbf{v}_k \mid \mathbf{v}_1,\dots,\mathbf{v}_k\in B,\ \alpha_1,\dots,\alpha_k\in\mathbb{F}\big\}$.[51]
- B is a basis of V if it satisfies the following two conditions:
  1. B is linearly independent;
  2. $V=\mathrm{Span}(B)$.

With a basis defined in this way, it is possible to show that every vector space has a (possibly infinite) basis. However, the proof uses "Zorn's lemma" (an equivalent of the "Axiom of Choice," which is studied in set theory) and is non-constructive. So,

[49] This is because $\dim(U\cap W)=0$.

[50] So, B is linearly independent if and only if all finite subsets of B are linearly independent.

[51] So, $\mathrm{Span}(B)$ is the set of vectors that can be expressed as a linear combination of finitely many vectors in B.

it is possible to show that every vector space has a basis, but for some vector spaces, we have no idea what a basis might look like. For instance, consider the set of all functions from $\mathbb{R}$ to $\mathbb{R}$; this set is a vector space (over $\mathbb{R}$) and therefore has a basis, but it is not known what a basis of this vector space might look like.

In some cases, though, we can get a "nice" infinite basis. For instance, $\mathbb{P}_{\mathbb{R}}$ has a basis $\{1, x, x^2, x^3, x^4, \dots\}$.

3.3 The column space, row space, and null space of a matrix

3.3.1 The column and row space of a matrix

For a field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n\times m}$, we define the following:

- the *column space* of A, denoted by $\text{Col}(A)$, is the subspace of $\mathbb{F}^n$ spanned by the columns of A;[52]
- the *row space* of A, denoted by $\text{Row}(A)$, is the subspace of $\mathbb{F}^{1\times m}$ spanned by the rows of A.[53]

We can easily relate the column space to the row space using transposes, as the following proposition shows.

Proposition 3.3.1. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$ be a matrix. Then both the following hold:*

(a) $\text{Col}(A) = \left\{\mathbf{u}^T \mid \mathbf{u} \in \text{Row}(A^T)\right\}$;

(b) $\text{Row}(A) = \left\{\mathbf{u}^T \mid \mathbf{u} \in \text{Col}(A^T)\right\}$.

Proof. This is essentially "obvious," since transposes turn rows into columns and vice versa. However, let us give a formal proof of (a). The proof of (b) is similar and is left as an exercise.

Set $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$, so that $A^T = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix}$. By definition, we have that $\text{Col}(A) = \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$ and $\text{Row}(A^T) = \text{Span}(\mathbf{a}_1^T, \dots, \mathbf{a}_m^T)$. Now, to prove (a), we must prove the following two inclusions:

[52] More precisely, if $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$, then $\text{Col}(A) := \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$. The fact that $\text{Col}(A)$ really is a subspace of $\mathbb{F}^n$ follows from Theorem 3.1.11.

[53] More precisely, if $A = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix}$ (i.e. $\mathbf{r}_1, \dots, \mathbf{r}_n$ are the rows of A, appearing in A in that order, from top to bottom), then $\text{Row}(\mathbf{r}_1, \dots, \mathbf{r}_n) := \text{Span}(\mathbf{r}_1, \dots, \mathbf{r}_n)$. The fact that $\text{Row}(A)$ really is a subspace of $\mathbb{F}^{1\times m}$ follows from Theorem 3.1.11.

(1) $\mathrm{Col}(A) \subseteq \{\mathbf{u}^T \mid \mathbf{u} \in \mathrm{Row}(A^T)\}$;

(2) $\{\mathbf{u}^T \mid \mathbf{u} \in \mathrm{Row}(A^T)\} \subseteq \mathrm{Col}(A)$.

We first prove (1). Fix any $\mathbf{u} \in \mathrm{Col}(A)$. Then $\mathbf{u} \in \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_m)$, and so there exist scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that $\mathbf{u} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_m\mathbf{a}_m$. By taking the transpose of both sides, we get

$$\mathbf{u}^T \quad = \quad (\alpha_1\mathbf{a}_1 + \cdots + \alpha_m\mathbf{a}_m)^T \quad = \quad \alpha_1\mathbf{a}_1^T + \cdots + \alpha_m\mathbf{a}_m^T,$$

and it follows that $\mathbf{u}^T \in \mathrm{Span}(\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T) = \mathrm{Row}(A^T)$. Since $\mathbf{u} = (\mathbf{u}^T)^T$, this proves (1).[54]

Let us now prove (2). Fix $\mathbf{u} \in \mathrm{Row}(A^T)$. Then $\mathbf{u} \in \mathrm{Span}(\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T)$, and so there exist scalars $\alpha_1, \ldots, \alpha_m$ such that $\mathbf{u} = \alpha_1\mathbf{a}_1^T + \cdots + \alpha_m\mathbf{a}_m^T$. By taking the transpose of both sides, we get that

$$\begin{aligned} \mathbf{u}^T &= (\alpha_1\mathbf{a}_1^T + \cdots + \alpha_m\mathbf{a}_m^T)^T \\ &= \alpha_1(\mathbf{a}_1^T)^T + \cdots + \alpha_m(\mathbf{a}_m^T)^T \\ &= \alpha_1\mathbf{a}_1 + \cdots + \alpha_m\mathbf{a}_m, \end{aligned}$$

and it follows that $\mathbf{u}^T \in \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_m) = \mathrm{Col}(A)$. This proves (2). □

Proposition 3.3.2. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$ be a matrix. Then both the following hold:*

(a) $Col(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m\}$;

(b) $Row(A) = \{\mathbf{x}A \mid \mathbf{x} \in \mathbb{F}^{1\times n}\}$.[55]

Proof. (a) Set $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix}$. Then

$$\mathrm{Col}(A) \overset{(*)}{=} \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_m) \overset{(**)}{=} \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m\}$$

where (*) follows from the definition of the column space, and (**) was obtained in subsection 1.4.4.[56]

(b) We use part (a) and Proposition 3.3.1(b), as follows:

[54] Indeed, we have shown that any element $\mathbf{u}$ of $\mathrm{Col}(A)$ is the transpose of some element (namely, $\mathbf{u}^T$) of $\mathrm{Row}(A^T)$. So, we have proven (1).

[55] Note that in the expression $\mathbf{x}A$, we have that $\mathbf{x}$ is a **row** vector with n entries.

[56] See the Remark following Example 1.4.3.

$$\begin{aligned}
\mathrm{Row}(A) &= \left\{\mathbf{u}^T \mid \mathbf{u} \in \mathrm{Col}(A^T)\right\} && \text{by Proposition 3.3.1(b)} \\
&= \left\{(A^T\mathbf{x})^T \mid \mathbf{x} \in \mathbb{F}^n\right\} && \text{by (a)} \\
&= \left\{\mathbf{x}^T A \mid \mathbf{x} \in \mathbb{F}^n\right\} \\
&= \left\{\mathbf{x}A \mid \mathbf{x} \in \mathbb{F}^{1\times n}\right\}.
\end{aligned}$$

□

Our main goal in this section is to give a recipe for finding a basis of the column space and the row space of a matrix (see Theorems 3.3.4 and 3.3.9). As we shall see, both of those spaces have dimension precisely $\mathrm{rank}(A)$. We begin with a technical proposition.

Proposition 3.3.3. *Let $\mathbb{F}$ be a field, let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{F}^n$, and let $B \in \mathbb{F}^{n\times n}$ be an invertible matrix. Then both the following hold:*

(a) $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent if and only if $\{B\mathbf{a}_1, \ldots, B\mathbf{a}_k\}$ is linearly independent;

(b) for all $\mathbf{v} \in \mathbb{F}^n$, $\mathbf{v} \in$ Span$(\mathbf{a}_1, \ldots, \mathbf{a}_k)$ if and only if $B\mathbf{v} \in$ Span$(B\mathbf{a}_1, \ldots, B\mathbf{a}_k)$;

Proof. We first prove (a). Suppose first that $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent. We must show that $\{B\mathbf{a}_1, \ldots, B\mathbf{a}_k\}$ is linearly independent. Fix scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_1 B\mathbf{a}_1 + \cdots + \alpha_k B\mathbf{a}_k = \mathbf{0}.$$

Since B is invertible, it has an inverse B^{-1}. By multiplying both sides of the equation above by B^{-1} (on the left), we obtain

$$\alpha_1 \mathbf{a}_1 + \cdots + \alpha_k \mathbf{a}_k = \mathbf{0}.$$

Since $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent, we have that $\alpha_1 = \cdots = \alpha_k = 0$. So, $\{B\mathbf{a}_1, \ldots, B\mathbf{a}_k\}$ is linearly independent.

Suppose, conversely, that $\{B\mathbf{a}_1, \ldots, B\mathbf{a}_k\}$ is linearly independent. We must show that $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent. Fix scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_1 \mathbf{a}_1 + \cdots + \alpha_k \mathbf{a}_k = \mathbf{0}.$$

We now multiply both sides by B (on the left), and we obtain

$$\alpha_1 (B\mathbf{a}_1) + \cdots + \alpha_k (B\mathbf{a}_k) = \mathbf{0}.$$

Since $\{B\mathbf{a}_1, \ldots, B\mathbf{a}_k\}$ is linearly independent, it follows that $\alpha_1 = \cdots = \alpha_k = 0$. So, $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent. This completes the proof of (a).

We now prove (b). Fix $\mathbf{v} \in \mathbb{F}^n$. Suppose first that $\mathbf{v} \in \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_k)$. Then there exist scalars $\alpha_1, \ldots, \alpha_k$ such that $\mathbf{v} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_k\mathbf{a}_k$. By multiplying both sides by B (on the left), we get $B\mathbf{v} = \alpha_1(B\mathbf{a}_1) + \cdots + \alpha_k(B\mathbf{a}_k)$, and so $B\mathbf{v} \in \text{Span}(B\mathbf{a}_1, \ldots, B\mathbf{a}_k)$.

Suppose, conversely, that $B\mathbf{v} \in \text{Span}(B\mathbf{a}_1, \ldots, B\mathbf{a}_k)$. Then there exist scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$B\mathbf{v} \quad = \quad \alpha_1(B\mathbf{a}_1) + \cdots + \alpha_k(B\mathbf{a}_k).$$

Since B is invertible, it has an inverse B^{-1}. We now multiply both sides of the equation by B^{-1} (on the left), and we obtain $\mathbf{v} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_k\mathbf{a}_k$. So, $\mathbf{v} \in \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_k)$. This proves (b). □

Theorem 3.3.4. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$. Then the pivot columns of A form a basis of Col(A). Moreover, dim$\big(Col(A)\big) = rank(A)$.*

Remark/Warning: To get a basis of $\text{Col}(A)$, we need to take the pivot columns of the original matrix A, not of $\text{RREF}(A)$.

Remark: If A has no pivot columns (which can only happen if A is the zero matrix, in which case $\text{Col}(A) = \{\mathbf{0}\}$), then $\emptyset$ is the (unique) basis of $\text{Col}(A)$.

Proof. Since $r := \text{rank}(A)$ is equal to the number of pivot columns of A the first statement implies the second.

It remains to prove the first statement. Set $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix}$. Let $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}$ (with $1 \leq i_1 < \cdots < i_r \leq m$) be the pivot columns of A. We must show that $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}\}$ is a basis of $\text{Col}(A)$.

Set $U := \text{RREF}(A)$. Then $A \sim U$, and so by Theorem 1.11.13, there exists an invertible matrix $B \in \mathbb{F}^{n\times n}$ such that $U = BA = \begin{bmatrix} B\mathbf{a}_1 & \ldots & B\mathbf{a}_m \end{bmatrix}$. But now since $U = \text{RREF}(A)$, we see that all the following hold:

(i) $B\mathbf{a}_{i_1}, \ldots, B\mathbf{a}_{i_r}$ are the pivot columns of U;

(ii) for all $j \in \{1, \ldots, r\}$, we have that $B\mathbf{a}_{i_j} = \mathbf{e}^n_j$;[57]

(iii) in any column of U, only the top r entries may possibly be non-zero (the other entries are all zero).

Clearly, $\{\mathbf{e}^n_1, \ldots, \mathbf{e}^n_r\}$ is a linearly independent set; so, by (ii), $\{B\mathbf{a}_{i_1}, \ldots, B\mathbf{a}_{i_r}\}$ is a linearly independent set. Consequently, by Proposition 3.3.3(a), $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}\}$ is a linearly independent set. Moreover, it is clear that any vector in $\mathbb{F}^n$ in which only the top r entries may possibly be non-zero (and the other entries are all zero), is a linear combination of vectors $\mathbf{e}^n_1, \ldots, \mathbf{e}^n_r$. So, (i), (ii), and (iii) together imply that every column of $U = \begin{bmatrix} B\mathbf{a}_1 & \ldots & B\mathbf{a}_m \end{bmatrix}$ is a linear combination of vectors $B\mathbf{a}_{i_1}, \ldots, B\mathbf{a}_{i_r}$.

[57] As usual, $\mathbf{e}^n_1, \ldots, \mathbf{e}^n_n$ are the standard basis vectors of $\mathbb{F}^n$.

But now by Proposition 3.3.3(b), we see that every column of $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ is a linear combination of vectors $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$. So, by Propositon 3.2.13, $\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}\}$ is a spanning set of $\mathrm{Col}(A)$.[58] It now follows that $\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}\}$ is a basis of $\mathrm{Col}(A)$, and we are done. □

Example 3.3.5. *Consider the matrix*

$$A := \begin{bmatrix} 1 & 2 & -1 & 2 & 1 \\ 2 & 4 & 1 & 3 & 1 \\ 4 & 8 & -1 & 7 & 3 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$*. Find a basis* $\mathcal{A}$ *of Col*(A)*.*

Solution. By performing the "forward" phase of the row reduction algorithm, we get that

$$A \sim \begin{bmatrix} 1 & 2 & -1 & 2 & 1 \\ 0 & 0 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that the pivot columns of A are its first and third column, and by Theorem 3.3.4, those two columns form a basis of $\mathrm{Col}(A)$. So,

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \right\}$$

is a basis of $\mathrm{Col}(A)$.

Warning: Make sure you use the pivot columns of A itself, and **not** of one of its row echelon forms!

Remark: We could also have computed

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & 2 & 0 & 5/3 & 2/3 \\ 0 & 0 & 1 & -1/3 & -1/3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and identified the pivot columns of A that way. However, if all we need to do is identify the pivot columns, then this is not necessary: we can identify the pivot columns from **any** row echelon form of A, not just from its reduced row echelon form. □

[58] By definition, $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ is a spanning set of $\mathrm{Col}(A)$. By what we just showed, every vector in $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ is a linear combination of vectors in $\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}\}$. So, by Propositon 3.2.13 $\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}\}$ is a spanning set of $\mathrm{Col}(A)$.

Example 3.3.6. *Consider the matrix*

$$A = \begin{bmatrix} 1 & 1 & 2 & 2 & 0 \\ 1 & 2 & 1 & 2 & 1 \\ 2 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 2 & 1 \end{bmatrix}$$

and vectors

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 1 \end{bmatrix},$$

all with entries in $\mathbb{Z}_3$.

(a) Find a basis $\mathcal{A}$ *for* $Col(A)$.

(b) Find the coordinate vectors of all the columns of A *with respect to the basis* $\mathcal{A}$.

(c) Determine whether $\mathbf{b} \in Col(A)$*, and if so, compute the coordinate vector* $\left[\ \mathbf{b}\ \right]_{\mathcal{A}}$;

(d) Determine whether $\mathbf{c} \in Col(A)$*, and if so, compute the coordinate vector* $\left[\ \mathbf{c}\ \right]_{\mathcal{A}}$.

Solution. Set $A = \left[\ \mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4 \quad \mathbf{a}_5\ \right]$.[59] We form the matrix

$$\left[\ A\ \vdots\ \mathbf{b} \quad \mathbf{c}\ \right] = \left[\ \mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4\ \vdots\ \mathbf{b} \quad \mathbf{c}\ \right]$$

$$= \left[\begin{array}{ccccc:cc} 1 & 1 & 2 & 2 & 0 & 1 & 2 \\ 1 & 2 & 1 & 2 & 1 & 1 & 0 \\ 2 & 0 & 0 & 1 & 1 & 1 & 2 \\ 1 & 0 & 0 & 2 & 1 & 1 & 1 \end{array}\right].$$

Because we need to compute some coordinate vectors (and not just a basis of $\mathrm{Col}(A)$), we need to find the **reduced** row echelon form of $\left[\ A\ \vdots\ \mathbf{b} \quad \mathbf{c}\ \right]$, not just any row echelon form. By row reducing, we get

$$\mathrm{RREF}\Big(\left[\ A\ \vdots\ \mathbf{b} \quad \mathbf{c}\ \right]\Big) = \left[\begin{array}{ccccc:cc} 1 & 0 & 0 & 2 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right],$$

where for emphasis, we colored the pivot columns to the left of the vertical dotted line red and to the right of the vertical dotted line blue. By focusing on the submatrix

[59] So, $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{a}_4 = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \end{bmatrix}$, and $\mathbf{a}_5 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

to the left of the vertical dotted line, we see that the pivot columns of A are its first, second, and fifth column. So, by Theorem 3.3.4,

$$\mathcal{A} \;:=\; \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5\} \;=\; \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is a basis of $\mathrm{Col}(A)$. We now need to express all the columns of A as a linear combination of the pivot columns, which allows us to form the coordinate vectors of the columns of A in terms of the basis $\mathcal{A}$ of $\mathrm{Col}(A)$. We simply read this off from the submatix $\mathrm{RREF}\big(\begin{bmatrix} A \,\vdots\, \mathbf{b} & \mathbf{c} \end{bmatrix}\big)$ to the left of the vertical dotted line, as follows:

- $\mathbf{a}_1 = 1\mathbf{a}_1 + 0\mathbf{a}_2 + 0\mathbf{a}_5$, and so $\begin{bmatrix} \mathbf{a}_1 \end{bmatrix}_{\mathcal{A}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$;
- $\mathbf{a}_2 = 0\mathbf{a}_1 + 1\mathbf{a}_2 + 0\mathbf{a}_5$, and so $\begin{bmatrix} \mathbf{a}_2 \end{bmatrix}_{\mathcal{A}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$;
- $\mathbf{a}_3 = 0\mathbf{a}_1 + 2\mathbf{a}_2 + 0\mathbf{a}_5$, and so $\begin{bmatrix} \mathbf{a}_3 \end{bmatrix}_{\mathcal{A}} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$;
- $\mathbf{a}_4 = 2\mathbf{a}_1 + 0\mathbf{a}_2 + 0\mathbf{a}_5$, and so $\begin{bmatrix} \mathbf{a}_4 \end{bmatrix}_{\mathcal{A}} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$;
- $\mathbf{a}_5 = 0\mathbf{a}_1 + 0\mathbf{a}_2 + 1\mathbf{a}_5$, and so $\begin{bmatrix} \mathbf{a}_5 \end{bmatrix}_{\mathcal{A}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Remark: Note that our solution roughly follows the recipe from subsection 1.5.1, albeit with some modifications. For each **pivot** column $\mathbf{a}_i$, we simply get that that it is equal to "1 times itself, plus 0 times all the other pivot columns." For each **non-pivot** column $\mathbf{a}_i$, we focus on the submatrix of $\mathrm{RREF}\big(\begin{bmatrix} A \,\vdots\, \mathbf{b} & \mathbf{c} \end{bmatrix}\big)$ consisting of its leftmost i columns, we ignore any non-pivot columns other than $\mathbf{a}_i$ itself, and we express $\mathbf{a}_i$ as a linear combination of the pivot columns to the left of it (for the pivot columns to the right of it, we simply get the weight/scalar 0). For example, for the non-pivot column $\mathbf{a}_4$, we first focus on the first four columns of $\mathrm{RREF}\big(\begin{bmatrix} A \,\vdots\, \mathbf{b} & \mathbf{c} \end{bmatrix}\big)$:

$$\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

we ignore any non-pivot columns other than the fourth column itself,[60] and we express the fourth column as a linear combination of the pivot columns. This yields $\mathbf{a}_4 = 2\mathbf{a}_1 + 0\mathbf{a}_2$. To express $\mathbf{a}_4$ as a linear combination of **all** the pivot columns of A, we simply add the remaining pivot columns with weight 0 at the front, i.e. $\mathbf{a}_4 = 2\mathbf{a}_1 + 0\mathbf{a}_2 + 0\mathbf{a}_5$. Now we can read off the coordinate vector of $\mathbf{a}_4$ with respect to $\mathcal{A}$: $\begin{bmatrix} \mathbf{a}_4 \end{bmatrix}_{\mathcal{A}} = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}^T$.

It remains to deal with vectors $\mathbf{b}$ and $\mathbf{c}$. Here again, we essentially follow the recipe from subsection 1.5.1. Since $\mathbf{b}$ is a pivot column of $\begin{bmatrix} A \,\vdots\, \mathbf{b} & \mathbf{c} \end{bmatrix}$, we see that $\mathbf{b}$ cannot be written as a linear combination of the columns of A, and it follows that $\mathbf{b} \notin \mathrm{Col}(A)$. On the other hand, $\mathbf{c}$ is **not** a pivot column of $\begin{bmatrix} A \,\vdots\, \mathbf{b} & \mathbf{c} \end{bmatrix}$; consequently, $\mathbf{c}$ can indeed be expressed as a linear combination of the columns of A, and in particular, $\mathbf{c} \in \mathrm{Col}(A)$. From the matrix $\mathrm{RREF}\big(\begin{bmatrix} A \,\vdots\, \mathbf{b} & \mathbf{c} \end{bmatrix}\big)$, we read off $\mathbf{c} = 1\mathbf{a}_1 + 1\mathbf{a}_2 + 0\mathbf{a}_5$, which yields $\begin{bmatrix} \mathbf{c} \end{bmatrix}_{\mathcal{A}} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$. □

Using Proposition 3.3.1 and Theorem 3.3.4, we can also compute a basis of the row space of a matrix (see Example 3.3.7 below). However, we will later give another way of computing a basis of the row space (see Theorem 3.3.9), one that implies that the dimension of the row space of a matrix is equal to the rank of that matrix.

Example 3.3.7. *Consider the matrix*

$$A := \begin{bmatrix} 1 & 0 & 0 & 2 & 1 \\ 2 & 0 & 0 & 1 & 2 \\ 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 2 & 1 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_3$. Compute a basis of $\mathrm{Row}(A)$.

Solution. We first take the transpose of A:

$$A^T = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ 1 & 2 & 2 & 1 \end{bmatrix}.$$

[60] In this case, we ignore the third column.

By row reducing, we get

$$\mathrm{RREF}(A^T) = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, the pivot columns of A^T are its first, third, and fourth column. By Theorem 3.3.4, it follows that

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} \right\}$$

is a basis of $\mathrm{Col}(A^T)$. But by Proposition 3.3.1, we have that $\mathrm{Row}(A) = \{\mathbf{u}^T \mid \mathbf{u} \in \mathrm{Col}(A^T)\}$. So, by simply turning columns into rows, we get that

$$\left\{ \begin{bmatrix} 1 & 0 & 0 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 2 & 2 & 1 \end{bmatrix} \right\}$$

is a basis of $\mathrm{Row}(A)$. □

Proposition 3.3.8. *Let $\mathbb{F}$ be a field. Then any two row equivalent matrices in $\mathbb{F}^{n\times m}$ have the same row space.*

Proof. We begin by showing that applying one elementary row operation to a matrix does not alter the row space (see the Claim below).

> **Claim.** Let $A, B \in \mathbb{F}^{n\times m}$ be matrices such that B is obtained from A by performing one elementary row operation. Then $\mathrm{Row}(A) = \mathrm{Row}(B)$.

Proof of the Claim. Set $A = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$ (so, $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the rows of A appearing in that order in A, from top to bottom, and similar for B). By definition, $\mathrm{Row}(A) = \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ and $\mathrm{Row}(B) = \mathrm{Span}(\mathbf{b}_1, \ldots, \mathbf{b}_n)$.

Since B is obtained from A by performing one elementary row operation R, we know that A can be obtained from B by performing one elementary row operation (the one that "undoes" R). So, it is enough to show that $\mathrm{Row}(A) \subseteq \mathrm{Row}(B)$, for then an analogous argument will establish that $\mathrm{Row}(B) \subseteq \mathrm{Row}(A)$, and then the result will follow.

If B is obtained by swapping two rows of A, then obviously, $\mathrm{Row}(A) = \mathrm{Row}(B)$. Next, suppose that B is obtained by multiplying one row of A (say, the i-th row)

by a non-zero scalar $\alpha \in \mathbb{F}$. Then $\mathbf{b}_i = \alpha\mathbf{a}_i$ and $\mathbf{b}_j = \mathbf{a}_j$ for all $j \in \{1, \ldots, n\} \setminus \{i\}$. Now, fix $\mathbf{v} \in \mathrm{Row}(A)$; we must show that $\mathbf{v} \in \mathrm{Row}(B)$. Since $\mathbf{v} \in \mathrm{Row}(A)$, there exist scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_n\mathbf{a}_n$. But now

$$\begin{aligned}
\mathbf{v} &= \alpha_1\mathbf{a}_1 + \cdots + \alpha_{i-1}\mathbf{a}_{i-1} + \alpha_i\mathbf{a}_i + \alpha_{i+1}\mathbf{a}_{i+1} + \cdots + \alpha_n\mathbf{a}_n \\
&= \alpha_1\mathbf{a}_1 + \cdots + \alpha_{i-1}\mathbf{a}_{i-1} + (\alpha_i\alpha^{-1})(\alpha\mathbf{a}_i) + \alpha_{i+1}\mathbf{a}_{i+1} + \cdots + \alpha_n\mathbf{a}_n \\
&= \alpha_1\mathbf{b}_1 + \cdots + \alpha_{i-1}\mathbf{b}_{i-1} + (\alpha_i\alpha^{-1})\mathbf{b}_i + \alpha_{i+1}\mathbf{b}_{i+1} + \cdots + \alpha_n\mathbf{b}_n,
\end{aligned}$$

and so $\mathbf{v} \in \mathrm{Row}(B)$. Thus, $\mathrm{Row}(A) \subseteq \mathrm{Row}(B)$.

Finally, suppose that B is obtained from A by adding a scalar multiple of one row to another row. Then there exist distinct indices $i, j \in \{1, \ldots, n\}$ and a scalar $\alpha \in \mathbb{F}$ such that $\mathbf{b}_j = \mathbf{a}_j + \alpha\mathbf{a}_i$, and $\mathbf{b}_k = \mathbf{a}_k$ for all $k \in \{1, \ldots, n\} \setminus \{j\}$.[61] Now, fix $\mathbf{v} \in \mathrm{Row}(A)$. Then there exist scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_n\mathbf{a}_n$. We now set $\beta_i := \alpha_i - \alpha_j\alpha$, and we set $\beta_k := \alpha_k$ for all $k \in \{1, \ldots, n\} \setminus \{i\}$. Then

$$\beta_i\mathbf{b}_i + \beta_j\mathbf{b}_j = (\alpha_i - \alpha_j\alpha)\mathbf{a}_i + \alpha_j(\mathbf{a}_j + \alpha\mathbf{a}_i) = \alpha_i\mathbf{a}_i + \alpha_j\mathbf{a}_j,$$

whereas $\beta_k\mathbf{b}_k = \alpha_k\mathbf{a}_k$ for all $k \in \{1, \ldots, n\} \setminus \{i, j\}$. Thus,

$$\beta_1\mathbf{b}_1 + \cdots + \beta_n\mathbf{b}_n = \alpha_1\mathbf{a}_1 + \cdots + \alpha_n\mathbf{a}_n = \mathbf{v},$$

and it follows that $\mathbf{v} \in \mathrm{Row}(B)$. Thus, $\mathrm{Row}(A) \subseteq \mathrm{Row}(B)$. ♦

Now, fix row equivalent matrices $A, B \in \mathbb{F}^{n\times m}$. Then there exists a sequence $R_1, \ldots, R_k$ of elementary row operations such that, by starting with A and then successively applying $R_1, \ldots, R_k$ to it, we obtain B. By the Claim, each time we apply an elementary row operation, the row space remains unchanged. So, $\mathrm{Row}(A) = \mathrm{Row}(B)$.[62] □

Theorem 3.3.9. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times m}$, and let U be any matrix in row echelon form that is row equivalent to A.*[63] *Then the non-zero rows of U form a basis of Row(A). Moreover, dim$\big($Row$(A)\big)$ = rank(A).*

Remark: If U has no non-zero rows (which can only happen if A is the zero matrix, in which case $\mathrm{Row}(A) = \{\mathbf{0}\}$, where $\mathbf{0}$ is the zero vector in $\mathbb{F}^{1\times m}$), then $\emptyset$ is a basis of $\mathrm{Row}(A)$.

[61] So, we applied the elementary row operation "$R_j \to R_j + \alpha R_i$."

[62] Technically, we are doing an induction on the number of elementary row operations. (Details?)

[63] It may be that $U = \mathrm{RREF}(A)$, but this assumption is not necessary. U may be any matrix in row echelon form obtained from A via a sequence of elementary row operations. For instance, U may be the matrix obtained from A by performing only the "forward" part of the row reduction algorithm in order to transform A into a matrix in row echelon form.

Proof. Since $r := \text{rank}(A)$ is equal to the number of non-zero rows of U, the first statement implies the second. Moreover, by Proposition 3.3.8, $\text{Row}(A) = \text{Row}(U)$. So, it suffices to show that the non-zero rows of U form a basis of $\text{Row}(U)$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be the non-zero rows of U, appearing in that order (from top to bottom) in U.[64] We must show that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of $\text{Row}(U)$. Clearly, $\text{Row}(U) = \text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$. It remains to show that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a linearly independent set. Fix scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that $\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k = \mathbf{0}$. We must show that $\alpha_1 = \cdots = \alpha_k = 0$. Suppose otherwise, and let $i \in \{1, \ldots, k\}$ be the smallest index such that $\alpha_i \neq 0$. We may assume that the leading entry (i.e. the leftmost non-zero entry) of the row $\mathbf{u}_i$ is in position j. But since U is in row echelon form, the leading entries of $\mathbf{u}_{i+1}, \ldots, \mathbf{u}_k$ are all strictly to the right of the leading entry of $\mathbf{u}_i$, and so their j-th entry is 0. Since $\alpha_1 = \cdots = \alpha_{i-1} = 0$ (by the minimality of i), it follows that the j-th entry of $\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k$ is non-zero,[65] contrary to the fact that $\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k = \mathbf{0}$. □

Example 3.3.10. *Consider the matrix*

$$A = \begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 1 & -2 & 2 & 2 & 1 \end{bmatrix}$$

with entries understood to be in $\mathbb{R}$.

(a) Compute rank(A).

(b) Find a basis of Col(A).

(c) Find a basis of Row(A).

Solution. By performing the "forward" part of the row reduction algorithm, we see that the following matrix is a row echelon form of A:

$$U = \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

(a) The matrix U has three pivot columns, and so $\text{rank}(A) = 3$.

[64] It is possible that $k = 0$. In that case, we have that $U = O_{n\times m}$ (and consequently, $A = O_{n\times m}$), and $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is an empty list of vectors.

[65] Indeed, it is equal to $\alpha_i u_{i,j}$, where $u_{i,j}$ is the i,j-th entry of the matrix U. Since α_i and $u_{i,j}$ are both non-zero, we see that $\alpha_i u_{i,j} \neq 0$.

(b) The pivot columns of U are its first, second, and fifth column. So, the pivot columns of A are its first, second, and fifth column, and so those columns of A form a basis of $\mathrm{Col}(A)$. More precisely, the following is a basis of $\mathrm{Col}(A)$:

$$\left\{ \begin{bmatrix} 0 \\ 3 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -7 \\ -9 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 8 \\ 6 \\ 2 \end{bmatrix} \right\}.$$

(c) The non-zero rows of U form a basis of $\mathrm{Row}(A)$. So, the following is a basis of $\mathrm{Row}(A)$:

$$\Big\{ \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}, \begin{bmatrix} 0 & 2 & -4 & 4 & 2 & -6 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \Big\}.$$

Remark: We could also have computed a basis of $\mathrm{Row}(A)$ using the method described in Example 3.3.7 (i.e. by finding a basis of $\mathrm{Col}(A^T)$ and then taking the transpose of the basis vectors in order to obtain a basis of $\mathrm{Row}(A)$). However, this would have meant having to row reduce twice. Indeed, we would row reduce A in order to find a basis of $\mathrm{Col}(A)$, and then we would row reduce A^T in order to find a basis of $\mathrm{Row}(A)$. On the other hand, the method from Example 3.3.7 has the advantage that it produces a basis of $\mathrm{Row}(A)$ all of whose elements are rows of the original matrix A itself (which was not the case for the basis of $\mathrm{Row}(A)$ that we obtained above). □

Corollary 3.3.11. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times m}$. Then both the following hold:*

(a) $\dim\big(\mathrm{Col}(A)\big) = \dim\big(\mathrm{Row}(A)\big) = \mathrm{rank}(A)$;

(b) $\mathrm{rank}(A) = \mathrm{rank}(A^T)$.

Proof. Part (a) follows immediately from Theorems 3.3.4 and 3.3.9. For (b), we observe that

$$\begin{aligned} \mathrm{rank}(A) &= \dim\big(\mathrm{Col}(A)\big) && \text{by (a)} \\ &= \dim\big(\mathrm{Row}(A^T)\big) && \text{by Proposition 3.3.1} \\ &= \mathrm{rank}(A^T) && \text{by (a),} \end{aligned}$$

and we are done. □

3.3.2 Matrices of full rank

Theorem 3.3.12. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$. Then all the following hold:*

(a) the columns of A are linearly independent if and only if rank$(A) = m$ (i.e. A has full column rank);

(b) the columns of A span $\mathbb{F}^n$ (i.e. Col$(A) = \mathbb{F}^n$) if and only if rank$(A) = n$ (i.e. A has full row rank);

(c) the rows of A are linearly independent if and only if rank$(A) = n$ (i.e. A has full row rank);

(d) the rows of A span $\mathbb{F}^{1\times m}$ (i.e. Row$(A) = \mathbb{F}^{1\times m}$) if and only if rank$(A) = m$ (i.e. A has full column rank).

Proof. Part (a) follows from Proposition 3.2.1, and part (b) follows from Proposition 3.1.10.

Further, by Corollary 3.3.11(b), we have that $\text{rank}(A^T) = \text{rank}(A)$. In particular, A^T has full column rank if and only if A has full row rank,[66] and A^T has full row rank if and only if A has full column rank.[67] So, by applying (a) to A^T, we obtain (c), and by applying (b) to A^T, we obtain (d). □

Recall that a **square** matrix has full rank if and only if it is invertible (see Corollary 1.11.10). For square matrices of full rank, Theorem 3.3.12 yields the following corollary.

Corollary 3.3.13. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a **square** matrix. Then the following are equivalent:*

(a) rank$(A) = n$;

[66] Indeed A^T is an $m \times n$ matrix, and so

$$\begin{aligned} A^T \text{ has full column rank} &\iff \text{rank}(A^T) = n \\ &\overset{(*)}{\iff} \text{rank}(A) = n \\ &\iff A \text{ has full row rank,} \end{aligned}$$

where (*) follows from Corollary 3.3.11(b).

[67] Indeed, A^T is an $m \times n$ matrix, and so

$$\begin{aligned} A^T \text{ has full row rank} &\iff \text{rank}(A^T) = m \\ &\overset{(*)}{\iff} \text{rank}(A) = m \\ &\iff A \text{ has full column rank,} \end{aligned}$$

where (*) follows from Corollary 3.3.11(b).

(b) $\mathrm{rank}(A^T) = n$*;*

(c) the columns of A *are linearly independent;*

(d) the columns of A *span* $\mathbb{F}^n$ *(i.e.* $\mathrm{Col}(A) = \mathbb{F}^n$*);*

(e) the columns of A *form a basis of* $\mathbb{F}^n$*;*

(f) the rows of A *are linearly independent;*

(g) the rows of A *span* $\mathbb{F}^{1\times n}$ *(i.e.* $\mathrm{Row}(A) = \mathbb{F}^{1\times n}$*);*

(h) the rows of A *form a basis of* $\mathbb{F}^{1\times n}$*.*

Proof. By Corollary 3.3.11(b), we have that $\mathrm{rank}(A) = \mathrm{rank}(A^T)$, and so (a) and (b) are equivalent.[68]

The fact that (a) and (c) are equivalent follows from Theorem 3.3.12(a),[69] the fact that (a) and (d) are equivalent follows from Theorem 3.3.12(b),[70], and the fact that that (a) and (e) are equivalent follows from Proposition 3.2.6.

So far, we have shown that (a), (b), (c), (d), and (e) are equivalent.

The equivalence of (a), (c), (d), and (e) applied to A^T yields the equivalence of (b), (f), (g), and (h). This completes the argument. □

3.3.3 The rank of matrix products. Left and right inverses of a matrix

Using the results of subsection 3.3.1 (and also of section 1.11), we can show that multiplying a matrix by an invertible matrix (on the left or the right) leaves the rank unchanged. More precisely, we have the following proposition.

Proposition 3.3.14. *Let* $\mathbb{F}$ *be a field, and let* $A \in \mathbb{F}^{n\times m}$*. Then all the following hold:*

(a) for all invertible matrices $S \in \mathbb{F}^{n\times n}$*, we have that* $\mathrm{rank}(SA) = \mathrm{rank}(A)$*;*

(b) for all invertible matrices $S \in \mathbb{F}^{m\times m}$*, we have that* $\mathrm{rank}(AS) = \mathrm{rank}(A)$*;*

(c) for all invertible matrices $S_1 \in \mathbb{F}^{n\times n}$ *and* $S_2 \in \mathbb{F}^{m\times m}$*, we have that* $\mathrm{rank}(S_1AS_2) = \mathrm{rank}(A)$*.*

Proof. We first prove (a). Fix an invertible matrix $S \in \mathbb{F}^{n\times n}$. By Theorem 1.11.13, A and SA are row equivalent, and so by Proposition 1.6.2, they have the same rank. This proves (a).

[68] Alternatively, the equivalence of (a) and (b) follows from Corollary 1.11.10.
[69] Alternatively, this follows from Proposition 3.2.1.
[70] Alternatively, this follows from Proposition 3.1.10

We now prove (b). Fix an invertible matrix $S \in \mathbb{F}^{m\times m}$. Then by the Invertible Matrix Theorem (version 1; see subsection 1.11.7), S^T is also invertible. We now compute:

$$\begin{aligned}
\operatorname{rank}(AS) &= \operatorname{rank}\big((AS)^T\big) && \text{by Corollary 3.3.11(b)} \\
&= \operatorname{rank}\big(S^T A^T\big) && \text{by Proposition 1.8.1(d)} \\
&= \operatorname{rank}\big(A^T\big) && \text{by (a), since } S^T \text{ is invertible} \\
&= \operatorname{rank}(A) && \text{by Corollary 3.3.11(b).}
\end{aligned}$$

This proves (b).

Finally, for (c), we fix invertible matrices $S_1 \in \mathbb{F}^{n\times n}$ and $S_2 \in \mathbb{F}^{m\times m}$, and we observe that

$$\operatorname{rank}(S_1 A S_2) \overset{(a)}{=} \operatorname{rank}(AS_2) \overset{(b)}{=} \operatorname{rank}(A),$$

and we are done. □

Our next theorem states that the rank of a product of two matrices is no greater than the minimum of the ranks of the two matrices.

Theorem 3.3.15. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times p}$. Then*

$$\textit{rank}(AB) \leq \min\Big\{\textit{rank}(A), \textit{rank}(B)\Big\}.$$

Proof. Set $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_p \end{bmatrix}$. We must show that $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$ and $\operatorname{rank}(AB) \leq \operatorname{rank}(B)$.

We first prove that $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$. By definition, we have that $AB = \begin{bmatrix} A\mathbf{b}_1 & \dots & A\mathbf{b}_p \end{bmatrix}$, and in particular, every column of AB is a linear combination of the columns of A,[71] i.e. every column of AB belongs to $\operatorname{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m) = \operatorname{Col}(A)$. Since $\operatorname{Col}(A)$ is a subspace of $\mathbb{F}^n$ (and is therefore a vector space in its own right), Theorem 3.1.11(b) now guarantees that $\operatorname{Col}(AB) = \operatorname{Span}(A\mathbf{b}_1, \dots, A\mathbf{b}_p)$ is a subspace of $\operatorname{Col}(A)$.[72] Since $\operatorname{Col}(A)$ is finite-dimensional, Theorem 3.2.21 now implies that

$$\dim\big(\operatorname{Col}(AB)\big) \leq \dim\big(\operatorname{Col}(A)\big),$$

and we deduce that

$$\operatorname{rank}(AB) \overset{(*)}{=} \dim\big(\operatorname{Col}(AB)\big) \leq \dim\big(\operatorname{Col}(A)\big) \overset{(*)}{=} \operatorname{rank}(A),$$

[71] Indeed, if $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_m \end{bmatrix}^T$ is any vector in $\mathbb{F}^m$, then $A\mathbf{x} = x_1\mathbf{a}_1 + \dots + x_m\mathbf{a}_m$, which is a linear combination of the columns of A.

[72] Indeed, we have shown that every column of AB belongs to $\operatorname{Col}(A)$, that is, that $A\mathbf{b}_1, \dots, A\mathbf{b}_p \in \operatorname{Col}(A)$. Since $\operatorname{Col}(A)$ is a vector space (because it is a subspace of $\mathbb{F}^n$), Theorem 3.1.11(b) guarantees that $\operatorname{Span}(A\mathbf{b}_1, \dots, A\mathbf{b}_p)$ is a subspace of $\operatorname{Col}(A)$. But $\operatorname{Span}(A\mathbf{b}_1, \dots, A\mathbf{b}_p) = \operatorname{Col}(AB)$. So, $\operatorname{Col}(AB)$ is a subspace of $\operatorname{Col}(A)$.

where both instances of (*) follow from Theorem 3.3.4 (or alternatively, from Corollary 3.3.11(a)).

We have now shown that $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$. A completely analogous argument shows that $\operatorname{rank}(B^T A^T) \leq \operatorname{rank}(B^T)$, and we deduce that

$$\operatorname{rank}(AB) \overset{(*)}{=} \operatorname{rank}\big((AB)^T\big) = \operatorname{rank}(B^T A^T) \leq \operatorname{rank}(B^T) \overset{(*)}{=} \operatorname{rank}(B)$$

where both instances of (*) follow from Corollary 3.3.11(b). □

Corollary 3.3.16. *Let $\mathbb{F}$ be a field, and let $A, B \in \mathbb{F}^{n\times n}$. Then AB is invertible if and only if A and B are both invertible.*

Proof. If A and B are invertible, then Proposition 1.11.8(d) guarantees that AB is invertible. For the other direction, assume that AB is invertible. Then by the Invertible Matrix Theorem (version 1; see subsection 1.11.7), we have that $\operatorname{rank}(AB) = n$, and it suffices to show that $\operatorname{rank}(A) = n$ and $\operatorname{rank}(B) = n$. By Theorem 3.3.15, we have that $n = \operatorname{rank}(AB) \leq \min\big\{\operatorname{rank}(A), \operatorname{rank}(B)\big\}$, and it follows that $\operatorname{rank}(A) \geq n$ and $\operatorname{rank}(B) \geq n$. But since A and B are $n \times n$ matrices, Proposition 1.6.3 now implies that $\operatorname{rank}(A) = n$ and $\operatorname{rank}(B) = n$, and we are done. □

Left and right inverses. Suppose that $A \in \mathbb{F}^{n\times m}$, where $\mathbb{F}$ is some field. A *left inverse* of A is a matrix $B \in \mathbb{F}^{m\times n}$ such that $BA = I_m$, and a *right inverse* of A is a matrix $C \in \mathbb{F}^{m\times n}$ such that $AC = I_n$. Thus, a left inverse (resp. right inverse) of a matrix A is a matrix that we can multiply A by on the left (resp. on the right) in order to obtain the identity matrix of the appropriate size. Consider, for example, matrices

$$A_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 3 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Then $A_1 A_2 = I_2$, and consequently, A_1 is a left inverse of A_2, and A_2 is a right inverse of A_1. Obviously, a matrix need not have a left or a right inverse. For example, zero matrices have no left inverses and no right inverses. On the other hand, a matrix may possibly have more than one left inverse or more than one right inverse.[73] On the other hand, as Corollary 3.3.17 (below) shows, any matrix A that has both a left inverse and a right inverse is in fact invertible (and in particular, square), and moreover, both its left inverse and its right inverse are unique are equal to A^{-1}.

[73] Examples?

Corollary 3.3.17. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times m}$ be a matrix, and assume that $B \in \mathbb{F}^{m\times n}$ is a left inverse of A (i.e. $BA = I_m$) and that $C \in \mathbb{F}^{m\times n}$ is a right inverse of A (i.e. $AC = I_n$). Then A is invertible (and in particular square, i.e. $m = n$), and $B = C = A^{-1}$.*

Proof. First, we have that

$$m \;=\; \operatorname{rank}(I_m) \;\overset{(*)}{=}\; \operatorname{rank}(BA) \;\overset{(**)}{\leq}\; \min\big\{\operatorname{rank}(B), \operatorname{rank}(A)\big\} \;\overset{(***)}{\leq}\; n,$$

where (*) follows from the fact that $BA = I_m$, (**) follows from Theorem 3.3.15, and (***) follows from Proposition 1.6.3 (because A is an $n \times m$ matrix and B is an $m \times n$ matrix). Since $AC = I_n$, an analogous argument establishes that $n \leq m$ (we simply use the fact that $AC = I_n$ instead of $BA = I_m$). So, $m = n$. In particular, we have that $A, B, C \in \mathbb{F}^{n\times n}$, and that $BA = I_m = I_n$ and $AC = I_n$. But now

$$B \;=\; BI_n \;=\; B(\underbrace{AC}_{=I_n}) \;=\; (\underbrace{BA}_{=I_n})C \;=\; I_nC \;=\; C.$$

So, $BA = I_n$ and $AB = AC = I_n$. Thus, A is invertible, and its inverse is $B = C$. This completes the argument. □

Remark: Corollary 3.3.17 is the reason that we defined invertibility only for square matrices. Any reasonable definition of an invertible matrix would entail the existence of both a left and a right inverse for that matrix, and by Corollary 3.3.17, only square matrices can have both a left and a right inverse.

As a corollary of Theorem 3.3.15 for **square** matrices, we get the following.

Corollary 3.3.18. *Let $\mathbb{F}$ be field, and let $A, B \in \mathbb{F}^{n\times n}$ be such that $AB = I_n$ or $BA = I_n$. Then $AB = BA = I_n$, i.e. A and B are both invertible and are each other's inverses.*

Remark: Note that Corollary 3.3.18 implies that if a square matrix A has a left **or** a right inverse B, then B is in fact a "two-sided inverse" of A, i.e. the (ordinary) inverse of A, and in particular, A is invertible.

Proof. By symmetry, we may assume that $AB = I_n$. The argument is analogous in the case when $BA = I_n$ (in that case, we simply swap the roles of A and B). We then have that

$$\begin{aligned} n \;&=\; \operatorname{rank}(I_n) && \\ &=\; \operatorname{rank}(AB) && \text{because } AB = I_n \\ &\leq\; \min\Big\{\operatorname{rank}(A), \operatorname{rank}(B)\Big\} && \text{by Theorem 3.3.15} \end{aligned}$$

$$\begin{array}{rcll} & \leq & \max\Big\{\text{rank}(A), \text{rank}(B)\Big\} & \\ & \leq & n & \text{by Proposition 1.6.3, since} \\ & & & A \text{ and } B \text{ are } n \times n \text{ matrices,} \end{array}$$

and it follows that

$$n \;=\; \min\Big\{\text{rank}(A), \text{rank}(B)\Big\} \;=\; \max\Big\{\text{rank}(A), \text{rank}(B)\Big\} \;=\; n,$$

which in turn implies that $\text{rank}(A) = \text{rank}(B) = n$. But now the Invertible Matrix Theorem (version 1; see subsection 1.11.7) guarantees that A and B are invertible. Since $AB = I_n$, Proposition 1.11.3 now implies that $A^{-1} = B$ and $B^{-1} = A$. □

3.3.4 Extending a linearly independent set of vectors in $\mathbb{F}^n$ to a basis of $\mathbb{F}^n$

Proposition 3.3.19. *Let $\mathbb{F}$ be a field, let $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ be a linearly independent set of vectors in $\mathbb{F}^n$, and let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be any basis of $\mathbb{F}^n$. Then the pivot columns of the matrix $C := \left[\begin{array}{ccc|ccc} \mathbf{a}_1 & \ldots & \mathbf{a}_k & \mathbf{b}_1 & \ldots & \mathbf{b}_n \end{array}\right]$ form a basis of $\mathbb{F}^n$ that extends the linearly independent set $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$.*[74]

Proof. Set $A := \left[\begin{array}{ccc} \mathbf{a}_1 & \ldots & \mathbf{a}_k \end{array}\right]$ and $B := \left[\begin{array}{ccc} \mathbf{b}_1 & \ldots & \mathbf{b}_n \end{array}\right]$, so that $C = \left[\begin{array}{c|c} A & B \end{array}\right]$. Since the columns of B form a basis of $\mathbb{F}^n$, it is clear that $\text{Col}(C) = \mathbb{F}^n$.[75]

Since the columns of the $n \times k$ matrix A are linearly independent, Theorem 3.3.12(a) guarantees that $\text{rank}(A) = k$, and it follows that all columns of A are pivot columns.[76] Next, the pivot columns of C are the pivot columns of A (i.e. all the columns of A, by what we just showed), plus possibly some columns of B; moreover, by Theorem 3.3.4, the pivot columns of C form a basis of $\text{Span}(C) = \mathbb{F}^n$. This proves that the pivot columns of C indeed form a basis of $\mathbb{F}^n$ that extends $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$. □

[74] "Extends" simply means "contains as a subset."

[75] Indeed, we have that

$$\begin{array}{rcll} \mathbb{F}^n & \supseteq & \text{Col}(C) & \text{because } C \in \mathbb{F}^{n \times (k+n)} \\ & = & \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{b}_1, \ldots, \mathbf{b}_n) & \text{by the definition of Col}(C) \\ & \supseteq & \text{Span}(\mathbf{b}_1, \ldots, \mathbf{b}_n) & \\ & = & \mathbb{F}^n & \text{because } \{\mathbf{b}_1, \ldots, \mathbf{b}_n\} \text{ is a basis of } \mathbb{F}^n, \end{array}$$

and it follows that $\text{Col}(C) = \mathbb{F}^n$.

[76] Indeed, since $\text{rank}(A) = k$, we know that A has k pivot columns. Since A has precisely k columns, it follows that all columns of A are pivot columns.

Remark: Proposition 3.3.19 will work for **any** basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of $\mathbb{F}^n$. However, in practice, it is usually easiest to use the standard basis $\mathcal{E}_n = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of $\mathbb{F}^n$.

Example 3.3.20. *Consider the vectors*

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_2$. Extend the linearly independent set $\{\mathbf{a}_1, \mathbf{a}_2\}$ to a basis of $\mathbb{Z}_2^4$. (Assume that the set $\{\mathbf{a}_1, \mathbf{a}_2\}$ is indeed linearly independent.)

Solutution. We apply Proposition 3.3.19 to the linearly independent set $\{\mathbf{a}_1, \mathbf{a}_2\}$ and the standard basis $\mathcal{E}_4 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ of $\mathbb{Z}_2^4$. We form the matrix

$$C := \left[\begin{array}{cc:cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{array}\right] = \left[\begin{array}{cc:cccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right].$$

By row reducing, we obtain

$$\mathrm{RREF}(C) = \left[\begin{array}{cc:cccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right],$$

and we see that the pivot columns of C are the first, second, third, and fifth column. It follows that

$$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{e}_1, \mathbf{e}_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis of $\mathbb{Z}_2^4$ that extends $\{\mathbf{a}_1, \mathbf{a}_2\}$. □

Proposition 3.3.21. *Let $\mathbb{F}$ be a field, let $\mathbf{a}_1, \dots, \mathbf{a}_k \in \mathbb{F}^n$ be arbitrary vectors, and let $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be any basis of $\mathbb{F}^n$. Set $C := \left[\begin{array}{ccc:ccc} \mathbf{a}_1 & \dots & \mathbf{a}_k & \mathbf{b}_1 & \dots & \mathbf{b}_n \end{array}\right]$. Then the pivot columns of C to the left of the vertical dotted line form a basis $\mathcal{A}$ of Span$(\mathbf{a}_1, \dots, \mathbf{a}_k)$, and all the pivot columns of C form a basis $\mathcal{C}$ of $\mathbb{F}^n$ that extends $\mathcal{A}$ (i.e. satisfies $\mathcal{A} \subseteq \mathcal{C}$).*

Proof. Set $A := \left[\begin{array}{ccc} \mathbf{a}_1 & \dots & \mathbf{a}_k \end{array}\right]$ and $B := \left[\begin{array}{ccc} \mathbf{b}_1 & \dots & \mathbf{b}_n \end{array}\right]$, so that $C = \left[\begin{array}{c:c} A & B \end{array}\right]$. By definition, we have that $\mathrm{Col}(A) = \mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_k)$.

First, by Theorem 3.3.4, the pivot columns of C form some basis $\mathcal{C}$ of $\mathrm{Col}(C)$. Next, since $C = \left[\begin{array}{c:c} A & B \end{array}\right]$, and since the columns of B form a basis of $\mathbb{F}^n$, we have that $\mathrm{Col}(C) = \mathbb{F}^n$.[77] So, $\mathcal{C}$ is a basis of $\mathbb{F}^n$. Now, the pivot columns of $C = \left[\begin{array}{c:c} A & B \end{array}\right]$

[77] Here, the argument is the same as in the proof of Proposition 3.3.19.

to the left of the vertical dotted line are precisely the pivot columns of A, and by Theorem 3.3.4, the pivot columns of A form a basis $\mathcal{A}$ of $\mathrm{Col}(A) = \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_k)$. Obviously, $\mathcal{A} \subseteq \mathcal{C}$. This completes the argument. □

Example 3.3.22. *Consider the vectors*

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_3$. Find a basis $\mathcal{A}$ of Span$(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$, and find a basis $\mathcal{C}$ of $\mathbb{Z}_3^4$ that extends $\mathcal{A}$ (i.e. that satisfies $\mathcal{A} \subseteq \mathcal{C}$).

Solution. We apply Proposition 3.3.21 to the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ and the standard basis $\mathcal{E}_4 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ of $\mathbb{Z}_3^4$. We form the matrix

$$\begin{aligned} C &:= \left[\begin{array}{cccc:cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{array}\right] \\ &= \left[\begin{array}{cccc:cccc} 1 & 2 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 2 & 1 & 2 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 2 & 0 & 0 & 0 & 1 \end{array}\right]. \end{aligned}$$

By row reducing, we obtain

$$\mathrm{RREF}(C) = \left[\begin{array}{cccc:cccc} 1 & 2 & 0 & 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array}\right],$$

and we see that the pivot columns of C are the first, third, fifth and sixth column. Since the first and third column are to the left of the vertical dotted line, while the fifth and sixth are to the right, we see that

$$\mathcal{A} := \{\mathbf{a}_1, \mathbf{a}_3\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} \right\}$$

is a basis of $\mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$, and that

$$\mathcal{C} := \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{e}_1, \mathbf{e}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

is a basis of $\mathbb{Z}_3^4$ that extends $\mathcal{A}$. □

3.3.5 The null space of a matrix

For field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n\times m}$, we define the *null space* of A, denoted by $\mathrm{Nul}(A)$, to be the set of all solutions of the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$, i.e.

$$\mathrm{Nul}(A) \quad := \quad \big\{\mathbf{x} \in \mathbb{F}^m \mid A\mathbf{x} = \mathbf{0}\big\}.$$

Notation: In some texts, notation $\mathrm{Ker}(A)$ is used instead of $\mathrm{Nul}(A)$. "Ker" stands for "kernel."

Proposition 3.3.23. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$. Then Nul$(A)$ is a subspace of $\mathbb{F}^m$.*

Proof. We apply Theorem 3.1.7. First, $A\mathbf{0} = \mathbf{0}$, and so $\mathbf{0} \in \mathrm{Nul}(A)$. Next, if $\mathbf{u}, \mathbf{v} \in \mathrm{Nul}(A)$, then

$$\begin{aligned} A(\mathbf{u}+\mathbf{v}) &= A\mathbf{u} + A\mathbf{v} \\ &= \mathbf{0} + \mathbf{0} \qquad \text{because } \mathbf{u}, \mathbf{v} \in \mathrm{Nul}(A) \\ &= \mathbf{0}, \end{aligned}$$

and so $\mathbf{u} + \mathbf{v} \in A$. Finally, if $\mathbf{u} \in \mathrm{Nul}(A)$ and $\alpha \in \mathbb{F}$, then

$$\begin{aligned} A(\alpha\mathbf{u}) &= \alpha(A\mathbf{u}) \\ &= \alpha\mathbf{0} \qquad \text{because } \mathbf{u} \in \mathrm{Nul}(A) \\ &= \mathbf{0}, \end{aligned}$$

and so $\alpha\mathbf{u} \in \mathrm{Nul}(A)$. It now follows from Theorem 3.1.7 that $\mathrm{Nul}(A)$ is a subspace of $\mathbb{F}^n$. □

Terminology: The dimension of $\mathrm{Nul}(A)$ is called the *nullity* of the matrix A.

Proposition 3.3.24. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{m\times n}$. Then the columns of A are linearly independent if and only if Nul$(A) = \{\mathbf{0}\}$.*

Proof. This essentially follows from the definition of $\mathrm{Nul}(A)$ and from Proposition 3.2.1. Indeed, by definition, $\mathrm{Nul}(A)$ is the set of all solutions of the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$; consequently,

$$\begin{aligned} \mathrm{Nul}(A) = \{\mathbf{0}\} \quad &\Longleftrightarrow \quad \text{the homogeneous matrix-vector equation } A\mathbf{x} = \mathbf{0} \\ &\qquad\qquad \text{has only the trivial solution (i.e. the solution } \mathbf{x} = \mathbf{0}) \\ &\overset{(*)}{\Longleftrightarrow} \quad \text{the columns of } A \text{ are linearly independent,} \end{aligned}$$

where (*) follows from Proposition 3.2.1. □

Example 3.3.25. *Let*

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_2$*. Find a basis of* $Nul(A)$*. What is* $dim\big(Nul(A)\big)$*?*

Proof. We begin by finding the general solution of the matrix-vector equation $A\mathbf{x} = \mathbf{0}$.[78] By row reducing, we get

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The general solution of $A\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = \begin{bmatrix} r+t \\ s \\ r \\ s \\ t \end{bmatrix}, \qquad \text{where } r, s, t \in \mathbb{Z}_2,$$

that is,

$$\mathbf{x} = r \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \text{where } r, s, t \in \mathbb{Z}_2.$$

So,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis of $\mathrm{Nul}(A)$, and it follows that $\dim\big(\mathrm{Nul}(A)\big) = 3$. □

Remark: Suppose that for some matrix $A \in \mathbb{F}^{n \times m}$ (where $\mathbb{F}$ is some field), the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, i.e. the solution $\mathbf{x} = \mathbf{0}$. In this case, $\mathrm{Nul}(A) = \{\mathbf{0}\}$, and $\emptyset$ is the (unique) basis of $\mathrm{Nul}(A)$.

[78] As discussed toward the end of subsection 1.3.4 (see Example 1.3.17), when solving a homogeneous linear system, we need only row reduce the coefficient matrix, and not the whole augmented matrix. The same obviously applies to homogeneous matrix-vector equations.

We note that the matrix A from Example 3.3.25 satisfies $\text{rank}(A) = 2$ and $\dim\big(\text{Nul}(A)\big) = 3$. The sum of these two numbers is 5, which is the number of columns
is not an accident. We give a slightly informal proof of the rank-nullity theorem for matrices (however, this proof hopefully provides the right intuition). We will give a fully formal proof of the (more general) rank-nullity theorem for linear functions in chapter 4 (see subsection 4.2.2). As we shall see in subsection 4.2.2, the rank-nullity theorem for linear functions immediately implies the rank-nullity theorem for matrices.

The rank–nullity theorem (matrix version). *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$. Then*

$$\textit{rank}(A) + \textit{dim}\big(\textit{Nul}(A)\big) \;=\; \underbrace{m}_{\substack{= \textit{number of}\\ \textit{columns of } A}} .$$

Proof (outline/informal). By definition, $\text{rank}(A)$ is equal to the number of pivot columns of A. On the other hand, when computing the general solution of $A\mathbf{x} = \mathbf{0}$, the number of free variables is equal to the number of non-pivot columns of A, and the number of vectors in a basis of $\text{Nul}(A)$ is equal to the number of free variables.[79] So, $\dim\big(\text{Nul}(A)\big)$ is equal to the number of non-pivot columns of A. It now follows that $\text{rank}(A) + \dim\big(\text{Nul}(A)\big)$ is equal to the number of columns of A, and we are done. □

Remark: The diagram below (informally) summarizes the idea behind the rank-nullity theorem for matrices $A \in \mathbb{F}^{n\times m}$ (where $\mathbb{F}$ is a field).

$$\underbrace{\text{rank}(A)}_{\substack{= \text{number of}\\ \text{pivot}\\ \text{columns of } A\\ \\ = \text{number of}\\ \text{basic variables}}} \quad + \quad \underbrace{\dim\big(\text{Nul}(A)\big)}_{\substack{= \text{number of}\\ \text{non-pivot}\\ \text{columns of } A\\ \\ = \text{number of}\\ \text{free variables}}} \quad = \quad \underbrace{m}_{\substack{= \text{number of}\\ \text{columns of } A}}$$

Proposition 3.3.26 (below) states that performing elementary row operations on a matrix, or deleting its zero rows, does not alter the null space of the matrix. This follows immediately from the definition of the null space, but is useful to state as a separate proposition for easy future reference (in particular, we will need this in section 6.6).

[79] This last part ("the number of vectors in a basis of $\text{Nul}(A)$ is equal to the number of free variables") is not fully justified, and we omit the full details. Can you convince yourself that this is true?

Proposition 3.3.26. *Let $\mathbb{F}$ be a field.*

(a) If $A, B \in \mathbb{F}^{n\times m}$ are row equivalent matrices, then Nul(A) = Nul(B).

(b) If a matrix C is obtained from a matrix $A \in \mathbb{F}^{n\times m}$ by possibly deleting some zero rows of A, then Nul(A) = Nul(C).

Proof. We first prove (a). Assume that matrices $A, B \in \mathbb{F}^{n\times m}$ are row equivalent. Then matrices $\left[\begin{array}{c:c} A & \mathbf{0} \end{array}\right]$ and $\left[\begin{array}{c:c} B & \mathbf{0} \end{array}\right]$ are row equivalent,[80] and consequently, the matrix-vector equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ are equivalent (i.e. have exactly the same solutions). It follows that

$$\begin{array}{rcll} \text{Nul}(A) & = & \{\mathbf{x} \in \mathbb{F}^m \mid A\mathbf{x} = \mathbf{0}\} & \text{by definition} \\ & = & \{\mathbf{x} \in \mathbb{F}^m \mid B\mathbf{x} = \mathbf{0}\} & \text{because the matrix-vector equations } A\mathbf{x} = \mathbf{0} \text{ and } B\mathbf{x} = \mathbf{0} \text{ are equivalent} \\ & = & \text{Nul}(B) & \text{by definition.} \end{array}$$

This proves (a).

For (b), we note that if a matrix C is obtained from a matrix A by possibly deleting some zero rows of A, then the matrix-vector equations $A\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ are equivalent,[81] and similarly to the above, it follows that

$$\begin{array}{rcll} \text{Nul}(A) & = & \{\mathbf{x} \in \mathbb{F}^m \mid A\mathbf{x} = \mathbf{0}\} & \text{by definition} \\ & = & \{\mathbf{x} \in \mathbb{F}^m \mid C\mathbf{x} = \mathbf{0}\} & \text{because the matrix-vector equations } A\mathbf{x} = \mathbf{0} \text{ and } C\mathbf{x} = \mathbf{0} \text{ are equivalent} \\ & = & \text{Nul}(C) & \text{by definition.} \end{array}$$

This proves (b). □

3.3.6 The Invertible Matrix Theorem (version 2)

In subsection 1.11.7, we stated and proved the Invertible Matrix Theorem (version 1). Using the results of the present section, we can now extend that theorem, as follows.

The Invertible Matrix Theorem (version 2). *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a **square** matrix. Further, let $f : \mathbb{F}^n \to \mathbb{F}^n$ be given by $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$.*[82] *Then the following are equivalent:*

[80] This is "obvious," but it also follows from Proposition 1.3.21(a).

[81] Details?

[82] Since f is a matrix transformation, Proposition 1.10.4 guarantees that f is linear. Moreover, A is the standard matrix of f.

(a) *A is invertible (i.e. A has an inverse);*

(b) *A^T is invertible;*

(c) *$RREF(A) = I_n$;*

(d) *$RREF\big(\left[\begin{array}{c|c} A & I_n \end{array}\right]\big) = \left[\begin{array}{c|c} I_n & B \end{array}\right]$ for some matrix $B \in \mathbb{F}^{n\times n}$;*

(e) *$rank(A) = n$;*

(f) *$rank(A^T) = n$;*

(g) *A is a product of elementary matrices;*

(h) *the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$);*

(i) *there exists some vector $\mathbf{b} \in \mathbb{F}^n$ such that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;*

(j) *for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;*

(k) *for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has at most one solution;*

(l) *for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent;*

(m) *f is one-to-one;*

(n) *f is onto;*

(o) *f is an isomorphism;*

(p) *there exists a matrix $B \in \mathbb{F}^{n\times n}$ such that $BA = I_n$ (i.e. A has a left inverse);*

(q) *there exists a matrix $C \in \mathbb{F}^{n\times n}$ such that $AC = I_n$ (i.e. A has a right inverse);*

(r) *the columns of A are linearly independent;*

(s) *the columns of A span $\mathbb{F}^n$ (i.e. $Col(A) = \mathbb{F}^n$);*

(t) *the columns of A form a basis of $\mathbb{F}^n$;*

(u) *the rows of A are linearly independent;*

(v) *the rows of A span $\mathbb{F}^{1\times n}$ (i.e. $Row(A) = \mathbb{F}^{1\times n}$);*

(w) *the rows of A form a basis of $\mathbb{F}^{1\times n}$;*

(x) *$Nul(A) = \{\mathbf{0}\}$ (i.e. $dim\big(Nul(A)\big) = 0$).*

Proof. The fact that (a)-(o) are equivalent follows from the Invertible Matrix Theorem (version 1; see subsection 1.11.7). By definition, (a) implies (p) and (q). On the other hand, Corollary 3.3.18 guarantees that any one of (p) and (q) implies (a). By Corollary 3.3.13, (r)-(w) are equivalent to each other, as well as to (e). Finally, Proposition 3.3.24 guarantees that (r) and (x) are equivalent. This completes the argument. □

Chapter 4

Linear functions

4.1 Linear functions: definition, examples, and basic properties

We have already studied linear functions from $\mathbb{F}^m$ to $\mathbb{F}^n$, for a field $\mathbb{F}$ (see section 1.10). The concept of a linear function can easily be extended to a more general setting, that of arbitrary vector spaces, as follows.

Given vector spaces U and V over a field $\mathbb{F}$, we say that a function $f : U \to V$ is *linear* provided it satisfies the following two conditions (axioms):

1. for all vectors $\mathbf{u}_1, \mathbf{u}_2 \in U$, we have that $f(\mathbf{u}_1 + \mathbf{u}_2) = f(\mathbf{u}_1) + f(\mathbf{u}_2)$;
2. for all vectors $\mathbf{u} \in U$ and scalars $\alpha \in \mathbb{F}$, we have that $f(\alpha\mathbf{u}) = \alpha f(\mathbf{u})$.

If the linear function f is also a bijection, then we say that it is an *isomorphism*, and that the vector spaces U and V are *isomorphic*. Linear functions are also called *linear transformations*.

Remark: We note that in the definition of a linear function, the two vector spaces (the domain and the codomain of the function) must be over the same field $\mathbb{F}$.

Let us now take a look at a couple of examples (Examples 4.1.1 and 4.1.2). Recall that a *real vector space* is a vector space over the field $\mathbb{R}$.

Example 4.1.1. *Let $\mathbb{P}_\mathbb{R}$ be the real vector space of all polynomials with coefficients in $\mathbb{R}$. Show that the function $D : \mathbb{P}_\mathbb{R} \to \mathbb{P}_\mathbb{R}$ given by*

$$D\big(\sum_{k=0}^{n} a_k x^k\big) \;=\; \sum_{k=1}^{n} k a_k x^{k-1}$$

for all integers $n \geq 0$ and real numbers $a_0, \ldots, a_k$, is linear.

Remark: If you have studied calculus, you will recognize this as the formula for the derivative of polynomial functions.

Solution. We need to check that D satisfies the two axioms from the definition of a linear function.

1. Fix $p(x), q(x) \in \mathbb{P}_{\mathbb{R}}$. Then there exists an integer $n \geq 0$ and real numbers $a_0, \ldots, a_n, b_0, \ldots, b_n$ such that

$$p(x) = \sum_{k=0}^{n} a_k x^k \qquad \text{and} \qquad q(x) = \sum_{k=0}^{n} b_k x^k.$$

Remark: Here, n is some non-negative integer such that $\deg\big(p(x)\big), \deg\big(q(x)\big) \leq n$. The inequality may possibly be strict, i.e. it is possible that $a_n = 0$ or $b_n = 0$.

We now compute:

$$\begin{aligned} D\big(p(x)+q(x)\big) &= D\Big(\big(\sum_{k=0}^{n} a_k x^k\big) + \big(\sum_{k=0}^{n} b_k x^k\big)\Big) \\ &= D\Big(\sum_{k=0}^{n} (a_k + b_k) x^k\Big) \\ &= \sum_{k=1}^{n} k(a_k + b_k) x^{k-1} \\ &= \Big(\sum_{k=1}^{n} k a_k x^{k-1}\Big) + \Big(\sum_{k=1}^{n} k b_k x^{k-1}\Big) \\ &= D\Big(\sum_{k=0}^{n} a_k x^k\Big) + D\Big(\sum_{k=0}^{n} a_k x^k\Big) \\ &= D\big(p(x)\big) + D\big(q(x)\big). \end{aligned}$$

2. Fix $p(x) \in \mathbb{P}_{\mathbb{F}}$ and $\alpha \in \mathbb{F}$. Then there exists an integer $n \geq 0$ and real numbers $a_0, \ldots, a_n$ such that

$$p(x) = \sum_{k=0}^{n} a_k x^k.$$

We now compute:

$$\begin{aligned} D\big(\alpha p(x)\big) &= D\Big(\alpha\big(\sum_{k=0}^{n} a_k x^k\big)\Big) \\ &= D\Big(\sum_{k=0}^{n} (\alpha a_k) x^k\Big) \\ &= \sum_{k=0}^{n} k(\alpha a_k) x^{k-1} \end{aligned}$$

$$\begin{aligned} &= \alpha \sum_{k=0}^{n} k a_k x^{k-1} \\ &= \alpha D\Big(\sum_{k=0}^{n} a_k x^k \Big) \\ &= \alpha D\big(p(x)\big). \end{aligned}$$

From 1. and 2., we conclude that D is linear. □

Here is another example, for those who have studied calculus.

Example 4.1.2. *Let Diff$(\mathbb{R})$ be the real vector space of all differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, and let Func$(\mathbb{R})$ be the real vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$. Show that the function $D : $ Diff$(\mathbb{R}) \to $ Func$(\mathbb{R})$ given by $D(f) = f'$ for all $f \in $ Diff$(\mathbb{R})$ is linear. (As usual, f' denotes the derivative of f.)*

Proof. 1. Fix $f, g \in \mathrm{Diff}(\mathbb{R})$. Then by the properties of the derivative, we have that $D(f+g) = (f+g)' = f' + g' = D(f) + D(g)$.

2. Fix $f \in \mathrm{Diff}(\mathbb{R})$ and $\alpha \in \mathbb{R}$. Then by the properties of the derivative, we have that $D(\alpha f) = (\alpha f)' = \alpha f' = \alpha D(f)$.

From 1. and 2., we conclude that D is linear. □

In the remainder of this section, we prove some basic properties of linear functions. For the most part (though not exclusively), these are generalizations of the results that we proved in section 1.10 for linear functions $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field). As we shall see, most of the results readily generalize to linear functions between arbitrary vectors spaces (over the same field), with one exception: linear functions between general vector spaces do not have standard matrices. It is in fact possible to define the matrix of a linear function between non-trivial,[1] finite-dimensional vectors spaces, but such matrices depend on the particular choice of basis of the domain and codomain (see section 4.5).

4.1.1 Basic properties of linear functions

Theorem 4.1.3. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a function. Then the following are equivalent:*

(i) f is linear;

[1] Recall that a vector space is *trivial* if it only contains the zero vector (i.e. if its dimension is 0), and it is *non-trivial* if it contains at least one non-zero vector (i.e. if its dimension is greater than 0).

(ii) for all vectors $\mathbf{u}_1, \mathbf{u}_2 \in U$ and scalars $\alpha_1, \alpha_2 \in \mathbb{F}$, we have that

$$f(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2) \quad = \quad \alpha_1 f(\mathbf{u}_1) + \alpha_2 f(\mathbf{u}_2).$$

Proof. This easily follows from the definition of a linear function. The details are left as an exercise. □

Proposition 4.1.4. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then for all $\mathbf{u}_1, \mathbf{u}_2 \in U$, we have that*

$$f(\mathbf{u}_1 - \mathbf{u}_2) \quad = \quad f(\mathbf{u}_1) - f(\mathbf{u}_2).$$

Proof. Fix $\mathbf{u}_1, \mathbf{u_2} \in U$. Then

$$f(\mathbf{u}_1 - \mathbf{u}_2) + f(\mathbf{u}_2) \quad = \quad f(\mathbf{u}_1 - \mathbf{u}_2 + \mathbf{u}_2) \quad \overset{(*)}{=} \quad f(\mathbf{u}_1),$$

where (*) follows from the linearity of f. By subtracting $f(\mathbf{u}_2)$ from both sides, we get that $f(\mathbf{u}_1 - \mathbf{u}_2) = f(\mathbf{u}_1) - f(\mathbf{u}_2)$, which is what we needed to show.

Alternatively, we could observe that for all $\mathbf{u}_1, \mathbf{u}_2 \in U$, we have the following:

$$\begin{aligned}
f(\mathbf{u}_1 - \mathbf{u}_2) &= f(\mathbf{u}_1 + (-1)\mathbf{u}_2) && \text{by Proposition 3.1.3(d)} \\
&= f(\mathbf{u}_1) + f((-1)\mathbf{u}_2) && \text{by the linearity of } f \\
&= f(\mathbf{u}_1) + (-1)f(\mathbf{u}_2) && \text{by the linearity of } f \\
&= f(\mathbf{u}_1) - f(\mathbf{u}_2) && \text{by Proposition 3.1.3(d).}
\end{aligned}$$

□

The following proposition is an analogue of Proposition 1.10.1 for general linear functions.

Proposition 4.1.5. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then for all vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in U$ and all scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$, we have that*

$$f\Big(\sum_{i=1}^{k} \alpha_i \mathbf{u}_i\Big) \quad = \quad \sum_{i=1}^{k} \alpha_i f(\mathbf{u}_i),$$

of, written in another way, that

$$f\Big(\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k\Big) \quad = \quad \alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_k f(\mathbf{u}_k).$$

Proof. This follows from the definition of a linear function via an easy induction on k. The details are left as an exercise. □

We remark that our next proposition (Proposition 4.1.6) was already proven in the context of linear functions from $\mathbb{F}^m$ to $\mathbb{F}^n$, where $\mathbb{F}$ is a field (see Proposition 1.10.3). Here, we prove the proposition more generally, with an essentially identical proof.

Proposition 4.1.6. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then $f(\mathbf{0}) = \mathbf{0}$.*[2]

Proof. We observe that

$$f(\mathbf{0}) \overset{(*)}{=} f(0 \cdot \mathbf{0}) \overset{(**)}{=} 0f(\mathbf{0}) \overset{(*)}{=} \mathbf{0},$$

where both instances of (*) follows from Proposition 3.1.3(a), and (**) follows from the fact that f is linear. □

4.1.2 Making new linear functions out of old ones

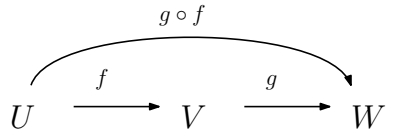
Our first proposition (Proposition 4.1.7 below) is a generalization of Proposition 1.10.13 for linear functions between arbitrary vector spaces. The proof of Proposition 1.10.13 relied on standard matrices. However, linear functions in general need not have an associated standard matrix (for example, the derivative function D from Example 4.1.1, and also from Example 4.1.2, does not have an associated standard matrix). So, our proof of Proposition 4.1.7 does not rely on matrices, and instead, it uses the definition of a linear function.

Proposition 4.1.7. *Let U, V, and W be vector spaces over a field $\mathbb{F}$. Then all the following hold:*

(a) for all linear functions $f, g : U \to V$, the function $f + g$ is linear;[3]

(b) for all linear functions $f : U \to V$ and scalars $\alpha \in \mathbb{F}$, the function $\alpha f : U \to V$ is linear;[4]

(c) for all linear functions $f : U \to V$ and $g : V \to W$, the function $g \circ f$ is linear.[5]



[2]Technically, this means $f(\mathbf{0}_U) = \mathbf{0}_V$, where $\mathbf{0}_U$ is the zero vector in U, and $\mathbf{0}_V$ is the zero vector in V.

[3]As usual, the function $f + g : U \to V$ is defined by $(f + g)(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u})$ for all $\mathbf{u} \in U$.

[4]As usual, the function $\alpha f : U \to V$ is defined by $(\alpha f)(\mathbf{u}) = \alpha\big(f(\mathbf{u})\big)$ for all $\mathbf{u} \in U$.

[5]As usual, the function $g \circ f : U \to W$ is defined by $(g \circ f)(\mathbf{u}) = g\big(f(\mathbf{u})\big)$ for all $\mathbf{u} \in U$.

Proof. We prove (c). The proofs of (a) and (b) are left as an exercise. Fix linear functions $f : U \to V$ and $g : V \to W$. We must show that $g \circ f$ is linear.

1. Fix $\mathbf{u}_1, \mathbf{u}_2 \in U$. Then

$$\begin{aligned}(g \circ f)(\mathbf{u}_1 + \mathbf{u}_2) &= g\big(f(\mathbf{u}_1 + \mathbf{u}_2)\big) \\ &= g\big(f(\mathbf{u}_1) + f(\mathbf{u}_2)\big) && \text{because } f \text{ is linear} \\ &= g\big(f(\mathbf{u}_1)\big) + g\big(f(\mathbf{u}_2)\big) && \text{because } g \text{ is linear} \\ &= (g \circ f)(\mathbf{u}_1) + (g \circ f)(\mathbf{u}_2).\end{aligned}$$

2. Fix $\mathbf{u} \in U$ and $\alpha \in \mathbb{F}$. Then

$$(g \circ f)(\alpha\mathbf{u}) = g\big(f(\alpha\mathbf{u})\big) \overset{(*)}{=} g\big(\alpha f(\mathbf{u})\big) \overset{(**)}{=} \alpha g\big(f(\mathbf{u})\big) = \alpha(g \circ f)(\mathbf{u}),$$

where (*) follows from the fact that f is linear, and (**) follows from the fact that g is linear.

From 1. and 2., we conclude that $g \circ f$ is linear. □

Given vector spaces U and V over a field $\mathbb{F}$, the set of all linear functions from U to V is denoted by $\mathrm{Hom}(U, V)$.[6] We note that $\mathrm{Hom}(U, V)$ is a vector space over $\mathbb{F}$. The vector addition and scalar multiplication operations in $\mathrm{Hom}(U, V)$ are the addition and scalar multiplication of functions; by parts (a) and (b) of Proposition 4.1.7, $\mathrm{Hom}(U, V)$ is indeed closed under the addition and scalar multiplication of functions. The zero vector in $\mathrm{Hom}(U, V)$ is the zero function, i.e. the function $f_0 : U \to V$ given by $f_0(\mathbf{u}) = \mathbf{0}_V$ for all $\mathbf{u} \in U$, where $\mathbf{0}_V$ is the zero of the vector space V.

4.1.3 A remark on infinity

In the context of vector space dimension, we adopt the convention that $n \leq \infty$ for all non-negative integers n, and that $\infty = \infty$. Readers who have some familiarity with infinite cardinals might object to "$\infty = \infty$" on the grounds that "there is more than one kind of infinity, and some infinities are bigger than others." We will not worry about this in these lecture notes. In particular, when we write $\dim(U) = \dim(V)$ for some vector spaces U and V, we mean that either U and V have the same finite dimension, or U and V are both infinite-dimensional. On the other hand, "$\dim(U) < \dim(V)$," in particular, means that U if finite-dimensional (and unless specified otherwise, V may possibly be infinite-dimensional).

[6] Linear functions are sometimes called "homomorphisms," which is where the notation $\mathrm{Hom}(U, V)$ comes from.

4.2 The image and kernel of a linear function. The rank-nullity theorem

We begin with a definition. Suppose we are given a function $f : A \to B$ (not necessarily linear). Then we define the following:

- for all subsets $A' \subseteq A$, the set $f[A'] := \{f(a) \mid a \in A'\}$ is called the *image* of A' under the function f;
- the set $\mathrm{Im}(f) := f[A]$ is called the *image* of f;
- for all subsets $B' \subseteq B$, the set $f^{-1}[B'] := \{a \in A \mid f(a) \in B'\}$ is called the *preimage* of B' under f.

We note that in some texts, the image of f is called the "range of f."

Remark: If $f : A \to B$ is a **bijection**, then it has an inverse function $f^{-1} : B \to A$. In this case, for $B' \subseteq B$, the notation $f^{-1}[B']$ can be interpreted in two ways: as the preimage of B' under f, and as the image of B' under the inverse function f^{-1}. However, in both cases, $f^{-1}[B']$ is the same subset of A, which is why we usually do not need to specify which interpretation we have in mind.

Given a **linear** function $f : U \to V$, where U and V are vector spaces over a field $\mathbb{F}$, the *kernel* of f is defined to be the set

$$\mathrm{Ker}(f) \quad := \quad \{\mathbf{u} \in U \mid f(\mathbf{u}) = \mathbf{0}\}.$$

Note that this means that $\mathrm{Ker}(f) = f^{-1}[\{\mathbf{0}\}]$, i.e. $\mathrm{Ker}(f)$ is the preimage of the set $\{\mathbf{0}\}$ under f. We further note the kernel is only defined for **linear** functions, and not for general functions.

In the case of linear functions from $\mathbb{F}^m$ to $\mathbb{F}^n$ (where $\mathbb{F}$ is a field), Proposition 4.2.1 (below) gives the correspondence between the image and kernel of the linear function on the one hand, and the column and null space of the standard matrix on the other hand. Note, however, that the image and kernel are defined for all linear functions, not just those from $\mathbb{F}^m$ to $\mathbb{F}^n$ (see Example 4.2.2 below).

Proposition 4.2.1. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function, and let $A \in \mathbb{F}^{n\times m}$ be the standard matrix of f. Then both the following hold:*

(a) $Im(f) = Col(A)$;

(b) $Ker(f) = Nul(A)$.

Proof. For (a), we observe that

$$\mathrm{Col}(A) \quad \overset{(*)}{=} \quad \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^m\} \quad \overset{(**)}{=} \quad \{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{F}^m\} \quad = \quad \mathrm{Im}(f),$$

where (*) follows from Proposition 3.3.2(a), and (**) follows from the fact that A is the standard matrix of f.

For (b), we observe that

$$\mathrm{Nul}(A) \;=\; \{\mathbf{x} \in \mathbb{F}^m \mid A\mathbf{x} = \mathbf{0}\} \;\overset{(*)}{=}\; \{\mathbf{x} \in \mathbb{F}^m \mid f(\mathbf{x}) = \mathbf{0}\} \;=\; \mathrm{Ker}(f),$$

where (*) follows from the fact that A is the standard matrix of f. □

Example 4.2.2. *Let $\mathbb{P}_{\mathbb{R}}$ be the real vector space of all polynomials with coefficients in $\mathbb{R}$. Consider the function $D : \mathbb{P}_{\mathbb{R}} \to \mathbb{P}_{\mathbb{R}}$ given by*

$$D(\sum_{k=0}^{n} a_k x^k) \;=\; \sum_{k=1}^{n} k a_k x^{k-1}$$

for all integers $n \geq 0$ and real numbers $a_0, \ldots, a_k$. By Example 4.1.1, D is linear. Clearly, $Ker(D)$ is the set of all constant polynomials, and $Im(D)$ is the set of all polynomials (i.e. $Im(D) = \mathbb{P}_{\mathbb{R}}$).

As an easy corollary of Theorem 3.1.7, we get the following theorem.

Theorem 4.2.3. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then all the following hold:*

(a) for all subspaces U' of U, we have that $f[U']$ is a subspace of V;

(b) $Im(f)$ is a subspace of V;

(c) for all subspaces V' of V, we have that $f^{-1}[V']$ is a subspace of U;

(d) $Ker(f)$ is a subspace of U.

Proof. Since U is a subspace of itself, (a) implies (b). Similarly, since $\mathrm{Ker}(f) = f^{-1}[\{\mathbf{0}\}]$ and $\{\mathbf{0}\}$ is a subspace of V, we have that (c) implies (d). So, it suffices to prove (a) and (c). We prove those two parts using Theorem 3.1.7. To avoid any possible confusion, we will denote the zero vectors of the vector spaces U and V by $\mathbf{0}_U$ and $\mathbf{0}_V$, respectively.

We first prove (a). Fix a subspace U' of U. We must show that $f[U']$ is a subspace of V. Since $f : U \to V$ and $U' \subseteq U$, we have that $f[U'] \subseteq V$. In view of Theorem 3.1.7, it now suffices to prove the following:

(i) $\mathbf{0}_V \in f[U']$;

(ii) for all $\mathbf{v}_1, \mathbf{v}_2 \in f[U']$, we have that $\mathbf{v}_1 + \mathbf{v}_2 \in f[U']$;

(iii) for all $\mathbf{v} \in f[U']$ and $\alpha \in \mathbb{F}$, we have that $\alpha\mathbf{v} \in f[U']$.

We first prove (i). Since U' is a subspace of U, Theorem 3.1.7 guarantees that $\mathbf{0}_U \in U'$. On the other hand, by Proposition 4.1.6, we have that $f(\mathbf{0}_U) = \mathbf{0}_V$, and it follows that $\mathbf{0}_V \in f[U']$. This proves (i).

Next, we prove (ii). Fix $\mathbf{v}_1, \mathbf{v}_2 \in f[U']$; we must show that $\mathbf{v}_1 + \mathbf{v}_2 \in f[U']$. Since $\mathbf{v}_1, \mathbf{v}_2 \in f[U']$, we know that there exist $\mathbf{u}_1, \mathbf{u}_2 \in U'$ such that $\mathbf{v}_1 = f(\mathbf{u}_1)$ and $\mathbf{v}_2 = f(\mathbf{u}_2)$. Since U' is a subspace of U, we have that $\mathbf{u}_1 + \mathbf{u}_2 \in U'$.[7] But now we have that

$$\mathbf{v}_1 + \mathbf{v}_2 \quad = \quad f(\mathbf{u}_1) + f(\mathbf{u}_2) \quad \overset{(*)}{=} \quad f(\mathbf{u}_1 + \mathbf{u}_2) \quad \overset{(**)}{\in} \quad f[U'],$$

where (*) follows from the fact that f is linear, and (**) follows from the fact that $\mathbf{u}_1 + \mathbf{u}_2 \in U'$. This proves (ii).

It remains to prove (iii). Fix $\mathbf{v} \in f[U']$ and $\alpha \in \mathbb{F}$; we must show that $\alpha\mathbf{v} \in f[U']$. Since $\mathbf{v} \in f[U']$, we know that there exists some $\mathbf{u} \in U'$ such that $\mathbf{v} = f(\mathbf{u})$. Since U' is a subspace of U, we have that $\alpha\mathbf{u} \in U'$.[8] But now we have that

$$\alpha\mathbf{v} \quad = \quad \alpha f(\mathbf{u}) \quad \overset{(*)}{=} \quad f(\alpha\mathbf{u}) \quad \overset{(**)}{\in} \quad f[U'],$$

where (*) follows from the fact that f is linear, and (**) follows from the fact that $\alpha\mathbf{u} \in U'$.

We have now proven (i), (ii), and (iii). So, by Theorem 3.1.7, we have that $f[U']$ is a subspace of V. This proves (a).

It remains to prove (c). Fix a subspace V' of V. We must show that $f^{-1}[V']$ is a subspace of U. Since $f : U \to V$ and $V' \subseteq V$, we have that $f^{-1}[V'] \subseteq U$. In view of Theorem 3.1.7, it now suffices to prove the following:

(i) $\mathbf{0}_U \in f^{-1}[V']$;

(ii) for all $\mathbf{u}_1, \mathbf{u}_2 \in f^{-1}[V']$, we have that $\mathbf{u}_1 + \mathbf{u}_2 \in f^{-1}[V']$;

(iii) for all $\mathbf{u} \in f^{-1}[V']$ and $\alpha \in \mathbb{F}$, we have that $\alpha\mathbf{u} \in f^{-1}[V']$.

We first prove (i). Since V' is a subspace of V, Theorem 3.1.7 guarantees that $\mathbf{0}_V \in V'$. On the other hand, by Proposition 4.1.6, we have that $f(\mathbf{0}_U) = \mathbf{0}_V$. So, $f(\mathbf{0}_U) = \mathbf{0}_V \in V'$, and it follows that $\mathbf{0}_U \in f^{-1}[V']$. This proves (i).

Next, we prove (ii). Fix $\mathbf{u}_1, \mathbf{u}_2 \in f^{-1}[V']$; we must show that $\mathbf{u}_1 + \mathbf{u}_2 \in f^{-1}[V']$, i.e. that $f(\mathbf{u}_1 + \mathbf{u}_2) \in V'$. Since $\mathbf{u}_1, \mathbf{u}_2 \in f^{-1}[V']$, we have that $\mathbf{v}_1 := f(\mathbf{u}_1)$ and $\mathbf{v}_2 := f(\mathbf{u}_2)$ belong to V'. Since V' is a subspace of V, we have that $\mathbf{v}_1 + \mathbf{v}_2 \in V'$. We now have that

$$f(\mathbf{u}_1 + \mathbf{u}_2) \quad \overset{(*)}{=} \quad f(\mathbf{u}_1) + f(\mathbf{u}_2) \quad = \quad \mathbf{v}_1 + \mathbf{v}_2 \quad \in \quad V',$$

[7] This actually follows from the definition of a subspace. However, it is also possible to use Theorem 3.1.7.

[8] Again, this follows from the definition of a subspace. Alternatively, we can use Theorem 3.1.7.

where (*) follows from the fact that f is linear. This proves (ii).

It remain to prove (iii). Fix $\mathbf{u} \in f^{-1}[V']$ and $\alpha \in \mathbb{F}$; we must show that $\alpha\mathbf{u} \in f^{-1}[V']$, i.e. that $f(\alpha\mathbf{u}) \in V'$. Since $\mathbf{u} \in f^{-1}[V']$, we know that $\mathbf{v} := f(\mathbf{u})$ belongs to V'. Since V' is a subspace of V, we have that $\alpha\mathbf{v} \in V'$. We now have that

$$f(\alpha\mathbf{u}) \overset{(*)}{=} \alpha f(\mathbf{u}) = \alpha\mathbf{v} \in V',$$

where (*) follows from the fact that f is linear. This proves (iii).

We have now proven (i), (ii), and (iii). So, by Theorem 3.1.7, we have that $f^{-1}[V']$ is a subspace of U. This proves (c). □

4.2.1 One-to-one linear functions and kernel

Theorem 4.2.4. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then f is one-to-one if and only if $Ker(f) = \{\mathbf{0}\}$.*

Proof. To avoid any possible confusion, we denote by $\mathbf{0}_U$ the zero vector of the vector space U, and we denote by $\mathbf{0}_V$ the zero vector of the vector space V. We need to show that f is one-to-one if and only if $\mathrm{Ker}(f) = \{\mathbf{0}_U\}$.

Suppose first that f is one-to-one. By Proposition 4.1.6, we have that $f(\mathbf{0}_U) = \mathbf{0}_V$, and it follows that $\mathbf{0}_U \in \mathrm{Ker}(f)$. It remains to show that $\mathbf{0}_U$ is the **only** element of $\mathrm{Ker}(f)$. So, fix any $\mathbf{u} \in \mathrm{Ker}(f)$. Then $f(\mathbf{u}) = \mathbf{0}_V = f(\mathbf{0}_U)$, and so since f is one-to-one, we have that $\mathbf{u} = \mathbf{0}_U$. This proves that $\mathrm{Ker}(f) = \{\mathbf{0}_U\}$.

Suppose now that $\mathrm{Ker}(f) = \{\mathbf{0}_U\}$. Fix $\mathbf{u}_1, \mathbf{u}_2 \in U$, and assume that $f(\mathbf{u}_1) = f(\mathbf{u}_2)$; we must show that $\mathbf{u}_1 = \mathbf{u}_2$. We note that

$$f(\mathbf{u}_1 - \mathbf{u}_2) \overset{(*)}{=} f(\mathbf{u}_1) - f(\mathbf{u}_2) \overset{(**)}{=} \mathbf{0}_V,$$

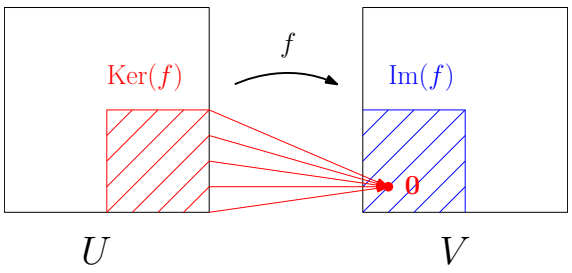
where (*) follows from Proposition 4.1.4, and (**) follows from the fact that $f(\mathbf{u}_1) = f(\mathbf{u}_2)$. So, $\mathbf{u}_1 - \mathbf{u}_2 \in \mathrm{Ker}(f)$. Since $\mathrm{Ker}(f) = \{\mathbf{0}_U\}$, it follows that $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}_U$, and consequently, $\mathbf{u}_1 = \mathbf{u}_2$. This proves that f is one-to-one. □

4.2.2 The rank of a linear function. The rank-nullity theorem

Suppose that U and V are vector spaces over a field $\mathbb{F}$, and that $f : U \to V$ is a linear function. By Theorem 4.2.3, $\mathrm{Im}(f)$ is a subspace of V, and $\mathrm{Ker}(f)$ is a subspace of U. The *rank* of f is defined to be

$$\mathrm{rank}(f) := \dim\big(\mathrm{Im}(f)\big),$$

and the *nullity* of f is $\dim(\mathrm{Ker}(f))$. We note that both the rank and the nullity of f may possibly be infinite.



Proposition 4.2.5. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then rank$(f) \leq$ dim(V).*

Proof. We may assume that $n := \dim(V)$ is finite, for otherwise, this is immediate. By Theorem 4.2.3, $\mathrm{Im}(f)$ is a subspace of V, and so by Theorem 3.2.21, we have that $\dim\big(\mathrm{Im}(f)\big) \leq \dim(V)$, i.e. $\mathrm{rank}(f) \leq \dim(V)$. □

By Proposition 4.2.4, a linear function if one-to-one if and only if its nullity is 0. As our next proposition shows, a linear function with a finite-dimensional codomain is onto if and only if the rank of the linear function is equal to the dimension of its codomain.

Proposition 4.2.6. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Assume that V is **finite-dimensional**. Then f is onto if and only if rank$(f) =$ dim(V).*

Proof. We have the following sequence of equivalent statements:

$$\begin{aligned} f \text{ is onto} &\overset{(*)}{\iff} \mathrm{Im}(f) = V \\ &\overset{(**)}{\iff} \dim\big(\mathrm{Im}(f)\big) = \dim(V) \\ &\overset{(***)}{\iff} \mathrm{rank}(f) = \dim(V), \end{aligned}$$

where (*) follows from the definition of an onto function, (**) follows from Theorem 3.2.21 (since $\mathrm{Im}(f)$ is a subspace of V, and V is finite-dimensional), and (***) follows from the definition of rank. □

Warning: Proposition 4.2.6 only applies to linear functions that have a **finite-dimensional codomain**. Do not apply Proposition 4.2.6 to linear functions with an infinite-dimensional codomain!

In the case of a linear function $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field), there is a natural relationship between the rank and nullity of f and the rank and nullity of the standard matrix of f, as our next proposition shows.

Proposition 4.2.7. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function, and let $A \in \mathbb{F}^{n\times m}$ be the standard matrix of f. Then both the following hold:*

(a) $rank(f) = rank(A)$;

(b) $dim\big(Ker(f)\big) = dim\big(Nul(A)\big)$.

Proof. By Proposition 4.2.1, we have that $\mathrm{Im}(f) = \mathrm{Col}(A)$ and $\mathrm{Ker}(f) = \mathrm{Nul}(A)$. The latter immediately implies (b). For (a), we observe that

$$\mathrm{rank}(f) \;=\; \dim\big(\mathrm{Im}(f)\big) \;=\; \dim\big(\mathrm{Col}(A)\big) \;\overset{(*)}{=}\; \mathrm{rank}(A),$$

where (*) follows from Theorem 3.3.4. □

As we pointed out above, both the rank and the nullity of a linear function $f : U \to V$ (where U and V are vector spaces over a field $\mathbb{F}$) may possibly be infinite. However, as the rank-nullity theorem for linear functions (below) states, if the domain U is finite-dimensional, then both $\mathrm{Im}(f)$ and $\mathrm{Ker}(f)$ are finite-dimensional, and moreover, the sum of their dimensions (i.e. the sum of rank and nullity of f) is precisely $\dim(U)$. We also note that, together with Proposition 4.2.7, the rank-nullity theorem for linear functions immediately implies the rank-nullity theorem for matrices (the details are below).

The rank–nullity theorem (linear function version). *Let U and V be vector spaces over a field $\mathbb{F}$, and assume that U is finite-dimensional. Then every linear function $f : U \to V$ satisfies*

$$rank(f) + dim\big(Ker(f)\big) \;=\; dim(U),$$

and in particular, both $Ker(f)$ and $Im(f)$ are finite-dimensional.

Proof. By Theorem 4.2.3, $\mathrm{Ker}(f)$ is a subspace of U, and $\mathrm{Im}(f)$ is a subspace of V. Next, since U is finite-dimensional, Theorem 3.2.21 guarantees that its subspace $\mathrm{Ker}(f)$ is finite-dimensional and satisfies $\dim\big(\mathrm{Ker}(f)\big) \leq \dim(U)$. Set $k := \dim\big(\mathrm{Ker}(f)\big)$ and $m := \dim(U)$ (so, $k \leq m$). By definition, we have that $\mathrm{rank}(f) = \dim\big(\mathrm{Im}(f)\big)$. Thus, to complete the proof, we need only exhibit a basis of $\mathrm{Im}(f)$ of size $m - k$. Indeed, this will imply $\mathrm{rank}(f) = \dim\big(\mathrm{Im}(f)\big) = m - k$, and the result will follow immediately.[9]

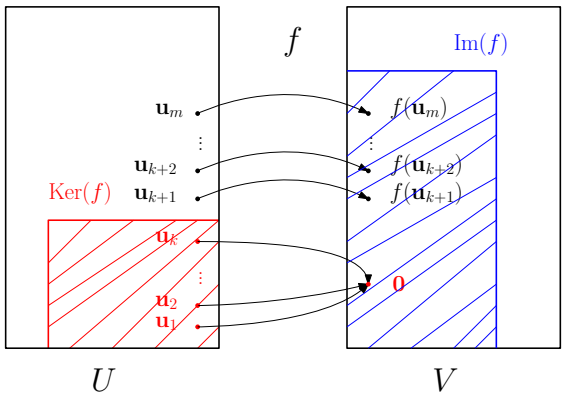
We proceed as follows. Fix a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of $\mathrm{Ker}(f)$.[10] Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a linearly independent set in a finite-dimensional vector space U; so, by Theorem 3.2.19, $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ can be extended to a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$

[9] Indeed, if $\mathrm{rank}(f) = \dim\big(\mathrm{Im}(f)\big) = m - k$, then $\mathrm{Im}(f)$ is finite-dimensional and $\mathrm{rank}(f) + \dim\big(\mathrm{Ker}(f)\big) = (m - k) + k = m = \dim(U)$.

[10] We are using the fact that $\dim\big(\mathrm{Ker}(f)\big) = k$, and so every basis of $\mathrm{Ker}(f)$ has k elements.

of U.[11] We will complete the proof by showing that the $(m-k)$-element set $\{f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)\}$ is a basis of $\mathrm{Im}(f)$.

Before proceeding with the technical details, we give a picture of our set-up (see below). Since $\mathbf{u}_1, \ldots, \mathbf{u}_k$ all belong to $\mathrm{Ker}(f)$, the function f maps them all to $\mathbf{0}$. On the other hand, as we shall prove (see Claim 1), vectors $f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)$ are linearly independent, and therefore, they are pairwise distinct and non-zero. (Thus, the picture below is indeed correct, but have not proven this yet!)



Claim 1. Vectors $f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)$ are linearly independent.

Proof of Claim 1. Fix scalars $\alpha_{k+1}, \ldots, \alpha_m \in \mathbb{F}$ such that

$$\alpha_{k+1} f(\mathbf{u}_{k+1}) + \cdots + \alpha_m f(\mathbf{u}_m) \quad = \quad \mathbf{0}.$$

We must show that $\alpha_{k+1} = \cdots = \alpha_m = 0$. Note that

$$f(\alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_m \mathbf{u}_m) \quad \overset{(*)}{=} \quad \alpha_{k+1} f(\mathbf{u}_{k+1}) + \cdots + \alpha_m f(\mathbf{u}_m) \quad = \quad \mathbf{0},$$

where (*) follows from the fact that f is linear (and more precisely, from Proposition 4.1.5). But now we have that $\alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_m \mathbf{u}_m \in \mathrm{Ker}(f)$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of $\mathrm{Ker}(f)$, we have that $\alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_m \mathbf{u}_m$ is a linear combination of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$, i.e. there exist scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_m \mathbf{u}_m \quad = \quad \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k.$$

But this implies that

$$-\alpha_1 \mathbf{u}_1 - \cdots - \alpha_k \mathbf{u}_k + \alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_m \mathbf{u}_m \quad = \quad \mathbf{0}.$$

Since vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m$ are linearly independent (because they form a basis of U), we deduce that $-\alpha_1 = \cdots = -\alpha_k = \alpha_{k+1} = \cdots = \alpha_m = 0$. In particular, $\alpha_{k+1} = \cdots = \alpha_m = 0$, and it follows that vectors $f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)$ are indeed linearly independent, which is what we needed to show. ♦

[11] We are using the fact that $\dim(U) = m$, and so every basis of U has m elements.

Claim 2. $\mathrm{Im}(f) = \mathrm{Span}\big(f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)\big)$.

Proof of Claim 2. By definition, $f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m) \in \mathrm{Im}(f)$. Since $\mathrm{Im}(f)$ is a subspace of V (and therefore a vector space in its own right), Theorem 3.1.11 guarantees that $\mathrm{Span}\big(f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)\big)$ is a subspace (and in particular, a subset) of $\mathrm{Im}(f)$. It remains to show that $\mathrm{Im}(f) \subseteq \mathrm{Span}\big(f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)\big)$, i.e. that every vector in $\mathrm{Im}(f)$ is a linear combination of the vectors $f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)$.

Fix $\mathbf{v} \in \mathrm{Im}(f)$. Then there exists some $\mathbf{u} \in U$ such that $\mathbf{v} = f(\mathbf{u})$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a basis of U, we know that there exist scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that $\mathbf{u} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_m\mathbf{u}_m$. We now have the following:

$$\begin{aligned}
\mathbf{v} &= f(\mathbf{u}) \\
&= f(\alpha_1\mathbf{u}_1 + \cdots + \alpha_m\mathbf{u}_m) \\
&\overset{(*)}{=} \alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_m f(\mathbf{u}_m) \\
&\overset{(**)}{=} \alpha_{k+1} f(\mathbf{u}_{k+1}) + \cdots + \alpha_m f(\mathbf{u}_m),
\end{aligned}$$

where (*) follows from the fact that f is linear (and more precisely, from Proposition 4.1.5), and (**) follows from the fact that $f(\mathbf{u}_1) = \cdots = f(\mathbf{u}_k) = \mathbf{0}$ (because $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathrm{Ker}(f)$). This proves that $\mathbf{v}$ is indeed a linear combination of the vectors $f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)$, and we are done. ♦

Clearly, Claims 1 and 2 together imply that $\big\{f(\mathbf{u}_{k+1}), \ldots, f(\mathbf{u}_m)\big\}$ is a basis of $\mathrm{Im}(f)$. This completes the argument. □

We are now ready to give a fully formal proof of the rank-nullity theorem for matrices, one that relies on the rank-nullity theorem for linear functions, which we just proved.

The rank–nullity theorem (matrix version). *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times m}$. Then*

$$\textit{rank}(A) + \textit{dim}\big(\textit{Nul}(A)\big) = \underbrace{m}_{\substack{= \textit{number of} \\ \textit{columns of } A}}.$$

Proof. Let $f : \mathbb{F}^m \to \mathbb{F}^n$ be given by $f(\mathbf{u}) = A\mathbf{u}$ for all $\mathbf{u} \in \mathbb{F}^m$. By Proposition 1.10.4, f is linear, and obviously, A is the standard matrix of f. We now have that

$$\mathrm{rank}(A) + \dim\big(\mathrm{Nul}(A)\big) \overset{(*)}{=} \mathrm{rank}(f) + \dim\big(\mathrm{Ker}(f)\big) \overset{(**)}{=} \dim(\mathbb{F}^m) = m,$$

where (*) follows from Proposition 4.2.7, and (**) follows from the rank-nullity theorem for linear functions. □

Dimension considerations. The following is an easy corollary of the rank-nullity theorem for linear functions.

Corollary 4.2.8. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then*

$$\mathit{rank}(f) \quad \leq \quad \min\big\{\mathit{dim}(U), \mathit{dim}(V)\big\}.$$

Remark: By definition, $\mathrm{rank}(f) = \dim\big(\mathrm{Im}(f)\big)$. So, Corollary 4.2.8 states that the dimension of the image of a linear function is at most the dimension of the domain and also at most the dimension of the codomain. We note that in Corollary 4.2.8, vector spaces U and V may possibly be infinite-dimensional.

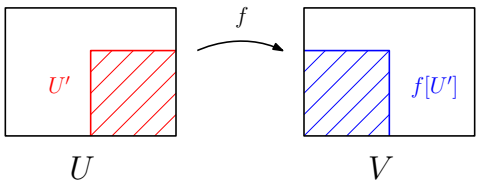
Proof of Corollary 4.2.8. The fact that $\mathrm{rank}(f) \leq \dim(V)$ follows from Proposition 4.2.5. It remains to show that $\mathrm{rank}(f) \leq \dim(U)$. If $\dim(U) = \infty$, then this is immediate. So, let us assume that U is finite-dimensional. Then

$$\mathrm{rank}(f) \quad \leq \quad \mathrm{rank}(f) + \dim\big(\mathrm{Ker}(f)\big) \quad \overset{(*)}{=} \quad \dim(U),$$

where (*) follows from the rank-nullity theorem for linear functions. □

Corollary 4.2.9. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then for any subspace U' of U, we have that*

$$\mathit{dim}\big(f[U']\big) \quad \leq \quad \min\big\{\mathit{dim}(U'), \mathit{dim}(V)\big\}.$$



Proof. Consider the function $f' := f \restriction U'$ (the restriction of f to U').[12] Since U' is a subspace of U and $f : U \to V$ is linear, we have that $f' : U' \to V$ is also linear. So,

$$\dim\big(f[U']\big) \quad = \quad \dim\big(f'[U']\big) \quad = \quad \dim\big(\mathrm{Im}(f')\big) \quad \overset{(*)}{\leq} \quad \min\big\{\dim(U'), \dim(V)\big\},$$

where (*) follows from Corollary 4.2.8. □

Geometric considerations. First of all, recall that subspaces of a Euclidean space $\mathbb{R}^k$ are $\{\mathbf{0}\}$, lines through the origin, planes through the origin, and higher dimensional generalizations. Now, suppose that $f : \mathbb{R}^m \to \mathbb{R}^n$ is a linear function. By Theorem 4.2.3(c), for any subspace U of the domain $\mathbb{R}^m$, we have that $f[U]$ is a subspace of the codomain $\mathbb{R}^n$, and by Corollary 4.2.9, $\dim\big(f[U]\big) \leq \dim(U)$. This

[12] So, $f' : U' \to V$ is given by $f'(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in U$.

implies that f maps $\{\mathbf{0}\}$ onto $\{\mathbf{0}\}$, maps any line through the origin onto either a line through the origin or $\{\mathbf{0}\}$, maps planes through the origin onto either planes through the origin or lines through the origin or $\{\mathbf{0}\}$. Similar remarks apply to higher-dimensional generalizations of subspaces of $\mathbb{R}^m$ and $\mathbb{R}^n$. (Compare these remarks to the discussion in subsection 1.10.2.)

Linear functions between vector spaces of the same finite dimension. By the Invertible Matrix Theorem (see subsection 3.3.6), for a linear function $f : \mathbb{F}^n \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field), the following are equivalent:

- f is one-to-one;
- f is onto;
- f is an isomorphism.

Here, we assumed that the domain and the codomain of f are the same (namely, $\mathbb{F}^n$). Using Theorem 4.2.4 (which states that a linear function is one-to-one if and only if its kernel is $\{\mathbf{0}\}$) and the rank-nullity theorem for linear functions, we can generalize this to linear functions between two vector spaces of the same finite dimension.

Corollary 4.2.10. *Let U and V be **finite-dimensional** vector spaces over a field $\mathbb{F}$, and assume that $dim(U) = dim(V)$. Let $f : U \to V$ be a linear function. Then the following are equivalent:*

(i) f is one-to-one;

(ii) f is onto;

(iii) f is a bijection (and therefore an isomorphism).

Warning: Corollary 4.2.10 only works if U and V (the domain and codomain of our linear function f) are of the same **finite** dimension. Do not attempt to apply the corollary to linear functions between infinite-dimensional vector spaces, or between vector spaces of different dimension.

Proof of Corollary 4.2.10. By definition, (i) and (ii) together are equivalent to (iii). So, it suffices to prove that (i) and (ii) are equivalent. By Theorem 4.2.4, we have that f is one-to-one if and only if $\mathrm{Ker}(f) = \{\mathbf{0}\}$, and by the rank-nullity theorem for linear functions, we have that

$$\mathrm{rank}(f) + \dim\big(\mathrm{Ker}(f)\big) \quad = \quad \dim(U).$$

We now have the following sequence of equivalent statements:

f is one-to-one $\iff$ $\mathrm{Ker}(f) = \{\mathbf{0}\}$ by Theorem 4.2.4

$$
\begin{aligned}
&\iff \dim\big(\mathrm{Ker}(f)\big) = 0 && \\
&\iff \mathrm{rank}(f) = \dim(U) && \text{by the rank-nullity theorem} \\
&\iff \dim\big(\mathrm{Im}(f)\big) = \dim(U) && \text{by the definition of } \mathrm{rank}(f) \\
&\iff \dim\big(\mathrm{Im}(f)\big) = \dim(V) && \text{because } \dim(U) = \dim(V) \\
&\iff \mathrm{Im}(f) = V && \text{by Theorem 3.2.21, since } V \text{ is finite-dimensional} \\
&\iff f \text{ is onto } V. &&
\end{aligned}
$$

So, (i) and (ii) are equivalent. This completes the argument. □

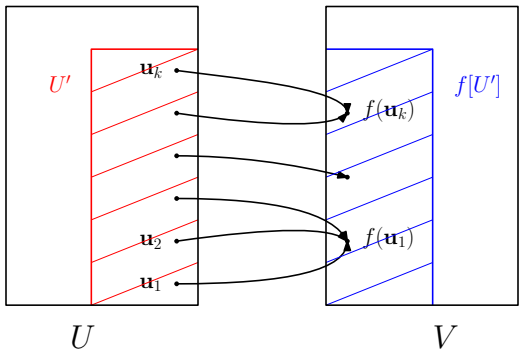
4.2.3 The effect of a linear function on linearly independent and spanning sets

Theorem 4.2.11. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Let $\mathbf{u}_1, \dots, \mathbf{u}_k \in U$, and set $U' := Span(\mathbf{u}_1, \dots, \mathbf{u}_k)$. Then all the following hold:*

(a) U' is a subspace of U, and $f[U']$ is a subspace of V;

(b) $f[U'] = f\big[Span(\mathbf{u}_1, \dots, \mathbf{u}_k)\big] = Span\big(f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)\big)$, i.e. vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ span $f[U'] = f\big[Span(\mathbf{u}_1, \dots, \mathbf{u}_k)\big]$;

(c) $dim\big(f[U']\big) \leq dim(U') \leq k$.



Proof. The fact that U' is a subspace of U follows immediately from Theorem 3.1.11, and the fact that $f[U']$ is a subspace of V follows from Theorem 4.2.3(a). This proves (a).

For (b), we have the following:

$$\begin{aligned}
\mathrm{Span}\big(f(\mathbf{u}_1),\dots,f(\mathbf{u}_k)\big) &= \big\{\alpha_1 f(\mathbf{u}_1)+\cdots+\alpha_k f(\mathbf{u}_k) \mid \alpha_1,\dots,\alpha_k\in\mathbb{F}\big\}\\
&\overset{(*)}{=} \Big\{f\big(\alpha_1\mathbf{u}_1+\cdots+\alpha_k\mathbf{u}_k\big) \mid \alpha_1,\dots,\alpha_k\in\mathbb{F}\Big\}\\
&\overset{(**)}{=} \big\{f(\mathbf{u}) \mid \mathbf{u}\in\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_k)\big\}\\
&= f\big[\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_k)\big]\\
&= f[U'],
\end{aligned}$$

where (*) follows from the linearity of the f (and more precisely, from Proposition 4.1.5), and (**) follows from the fact that, by definition, $\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_k) = \big\{\alpha_1\mathbf{u}_1+\cdots+\alpha_k\mathbf{u}_k \mid \alpha_1,\dots,\alpha_k\in\mathbb{F}\big\}$.

It remains to prove (c). By hypothesis, $\{\mathbf{u}_1,\dots,\mathbf{u}_k\}$ is a spanning set of U' of size k. So, by Theorem 3.2.14, some subset of that spanning set, say $\{\mathbf{u}_{i_1},\dots,\mathbf{u}_{i_m}\}$ (with $1\le i_1<\cdots<i_m\le k$) is a basis of U'. So, $\dim(U')=m\le k$. But now $\{\mathbf{u}_{i_1},\dots,\mathbf{u}_{i_m}\}$ is a spanning set of U'. So, by part (b) applied to the set $\{\mathbf{u}_{i_1},\dots,\mathbf{u}_{i_m}\}$ (rather than to $\{\mathbf{u}_1,\dots,\mathbf{u}_k\}$), we get that $\{f(\mathbf{u}_{i_1}),\dots,f(\mathbf{u}_{i_m})\}$ is a spanning set of $f[U']$. We now apply Theorem 3.2.14 again, and we deduce that some subset of $\{f(\mathbf{u}_{i_1}),\dots,f(\mathbf{u}_{i_m})\}$ is a basis of $f[U']$, and consequently, $\dim\big(f[U']\big)\le m$. This proves (c). □

As an easy corollary of Theorem 4.2.11 for the case when the vectors $\mathbf{u}_1,\dots,\mathbf{u}_k$ span the domain U, we obtain the following.

Corollary 4.2.12. *Let U and V be vector spaces over a field $\mathbb{F}$, let $f: U\to V$ be a linear function, and let $\{\mathbf{u}_1,\dots,\mathbf{u}_k\}$ be a spanning set of U. Then $Im(f) = Span\big(f(\mathbf{u}_1),\dots,f(\mathbf{u}_k)\big)$ and $rank(f) = dim\Big(Span\big(f(\mathbf{u}_1),\dots,f(\mathbf{u}_k)\big)\Big)\le k$.*

Proof. By hypothesis, $U=\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_k)$. So, by Theorem 4.2.11(b), we have that $\mathrm{Im}(f)=f[U]=\mathrm{Span}\big(f(\mathbf{u}_1),\dots,f(\mathbf{u}_k)\big)$, and by Theorem 4.2.11(c), we have that $\mathrm{rank}(f)=\dim\big(\mathrm{Im}(f)\big)=\dim\big(f[U]\big)\le k$.[13] □

[13]Here, the fact that $\mathrm{rank}(f)=\dim\big(\mathrm{Im}(f)\big)=\dim\big(f[U]\big)$ follows from the appropriate definitions. The fact that $\dim\big(f[U]\big)\le k$ follows from Theorem 4.2.11(c).

Theorem 4.2.13. *Let U and V be vector spaces over a field $\mathbb{F}$, let $f : U \to V$ be a linear function, and let $\mathbf{u}_1, \dots, \mathbf{u}_k \in U$. Then all the following hold:*

(a) if f is one-to-one and vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ are linearly independent in U, then vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ are linearly independent in V;

(b) if vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ are linearly independent in V, then vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ are linearly independent in U;

(c) if f is onto and vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ span U, then vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ span V;

(d) if f is one-to-one and vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ span V, then vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ span U.

Remark: Schematically (and informally), Theorem 4.2.13 can be summarized as shown in the diagram below.

$$f : U \xrightarrow{\text{linear}} V$$

(a)-(b)	$\mathbf{u}_1, \dots, \mathbf{u}_k$ are linearly independent	$\overset{\text{if } f \text{ is 1-1}}{\Longrightarrow}$ $\underset{\text{always}}{\Longleftarrow}$	$f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ are linearly independent
(c)-(d)	$\mathbf{u}_1, \dots, \mathbf{u}_k$ span U	$\overset{\text{if } f \text{ is onto}}{\Longrightarrow}$ $\underset{\text{if } f \text{ is 1-1}}{\Longleftarrow}$	$f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ span V

Proof of Theorem 4.2.13. Part (c) essentially follows from Corollary 4.2.12, so let us prove it first. Assume that f is onto and that vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ span U. But then

$$V \overset{(*)}{=} \operatorname{Im}(f) \overset{(**)}{=} \operatorname{Span}\big(f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)\big),$$

where (*) follows from the fact that f is onto, and (**) follows from Corollary 4.2.12. So, vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ span V. This proves (c).

Next, we prove (d). Assume that f is one-to-one and that vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ span V; we must show that vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ span U, i.e. that any vector in U can be written as a linear combination of the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$. Fix $\mathbf{u} \in U$. Since $f(\mathbf{u}) \in V$ and since vectors $f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)$ span V, we know that there exist scalars $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ such that $f(\mathbf{u}) = \alpha_1 f(\mathbf{u}_1) + \dots + \alpha_k f(\mathbf{u}_k)$. But now

$$f(\mathbf{u}) \;=\; \alpha_1 f(\mathbf{u}_1) + \dots + \alpha_k f(\mathbf{u}_k) \overset{(*)}{=} f(\alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k),$$

where (*) follows from the fact that f is linear (and more precisely, from Proposition 4.1.5). Since f is one-to-one, we deduce that $\mathbf{u} = \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k$, and so $\mathbf{u}$ is indeed a linear combination of the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$. This proves (d).

We now prove (a). Assume that f is one-to-one and that vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are linearly independent in U. We must show that vectors $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_k)$ are linearly independent in V. Fix scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_k f(\mathbf{u}_k) \quad = \quad \mathbf{0}.$$

We must show that $\alpha_1 = \cdots = \alpha_k = 0$. First, since f is linear (and more precisely, by Proposition 4.1.5), we have that $\alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_k f(\mathbf{u}_k) = f(\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k)$. So, $f(\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k) = \mathbf{0}$, and it follows that $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k \in \mathrm{Ker}(f)$. But since f is one-to-one, Theorem 4.2.4 guarantees that $\mathrm{Ker}(f) = \{\mathbf{0}\}$, and we deduce that $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k = \mathbf{0}$. Since vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are linearly independent, we deduce that $\alpha_1 = \cdots = \alpha_k = 0$. This proves (a).

It remains to prove (b). Assume that vectors $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_k)$ are linearly independent in V. We must show that vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are linearly independent in U. Fix scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k \quad = \quad \mathbf{0}.$$

We must show that $\alpha_1 = \cdots = \alpha_k = 0$. We have the following:

$$\alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_k f(\mathbf{u}_k) \quad \overset{(*)}{=} \quad \alpha_1 f(\mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k) \quad \overset{(**)}{=} \quad f(\mathbf{0}) \quad \overset{(***)}{=} \quad \mathbf{0},$$

where (*) follows from the fact that f is linear (and more precisely, from Proposition 4.1.5), (**) follows from the fact that $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k = \mathbf{0}$, and (***) follows from Proposition 4.1.6. But now since vectors $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_k)$ are linearly independent, we now have that $\alpha_1 = \cdots = \alpha_k = 0$. This proves (b). □

Dimension considerations. As we know, for any function $f : A \to B$, where A and B are finite sets, the following hold:

- if f is one-to-one, then $|A| \leq |B|$;
- if f is onto, then $|A| \geq |B|$;
- if f is a bijection, then $|A| = |B|$.

(Actually, the above is true even if we allow A and B to be infinite, but to make sense of the statement, we would need infinite cardinals. We omit the details.) In the case of **linear** functions, Theorem 4.2.14 (below) gives us a very similar statement, only involving dimension (rather than cardinality) of the domain and codomain. We note that Theorem 4.2.14 is an easy corollary of Theorem 4.2.13. We also note that the vector spaces U and V from the statement of Theorem 4.2.14 may possibly be infinite-dimensional.

Theorem 4.2.14. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be a linear function. Then all the following hold:*

(a) if f is one-to-one, then $dim(U) \leq dim(V)$;

(b) if f is onto, then $dim(U) \geq dim(V)$;

(c) if f is an isomorphism, then $dim(U) = dim(V)$.

Proof. Obviously, (a) and (b) together imply (c). So, it is enough to prove (a) and (b).

(a) We prove the contrapositive: we assume that $\dim(U) > \dim(V)$ (and in particular, $\dim(V)$ is finite), and we prove that f is **not** one-to-one. Set $n := \dim(V)$. Since $\dim(U) > \dim(V)$, we know that U has a linearly independent set of size greater than n. (Indeed, if U is finite-dimensional, then any one of its bases is a linearly independent set of size $\dim(U) > n$, and if U is infinite-dimensional, then Proposition 3.2.18 guarantees that U has linearly independent sets of any finite size.) So, fix a linearly independent set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of U, with $k > n$. Since $\dim(V) = n$, Theorem 3.2.17(a) guarantees that the set $\big\{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_k)\big\}$ is linearly dependent. But now Theorem 4.2.13(a) guarantees that f is not one-to-one.

(b) Assume that f is onto; we must show that $\dim(U) \geq \dim(V)$. We may assume that $n := \dim(U)$ is finite, for otherwise, we are done. We must show that $\dim(V) \leq n$. Fix any basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of U. In particular, vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ span U, and so since f is onto, Theorem 4.2.13(c) guarantees that vectors $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_n)$ span V. But then by Theorem 3.2.14, some subset of $\big\{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_n)\big\}$ is a basis of V, and it follows that $\dim(V) \leq n$. □

4.2.4 Computing bases of the images and preimages of subspaces under linear functions

In this subsection, we consider linear functions $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field). By Theorem 4.2.3, for every subspace U of the domain $\mathbb{F}^m$, $f[U]$ is a subspace of the codomain $\mathbb{F}^n$, and for every subspace V of the codomain $\mathbb{F}^n$, $f^{-1}[V]$ is a subspace of the domain $\mathbb{F}^m$. We would like to compute bases of $f[U]$ and $f^{-1}[V]$.

Computing a basis of the image of a subspace of the domain of a linear function. Proposition 4.2.15 (below) is an easy (and computationally useful) corollary of Theorem 4.2.11.

Proposition 4.2.15. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function, let $A \in \mathbb{F}^{n \times m}$ be the standard matrix of f, let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{F}^m$ ($k \geq 1$), and set $U := Span(\mathbf{u}_1, \ldots, \mathbf{u}_k)$. Then*

$$f[U] \quad = \quad Col\Big(A \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_k \end{bmatrix}\Big),$$

and moreover, the pivot columns of the matrix $A\begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_k \end{bmatrix}$ *form a basis of* $f[U]$.

Proof. First, we compute:

$$\begin{aligned} f[U] &= f\big[\mathrm{Span}(\mathbf{u}_1, \dots, \mathbf{u}_k)\big] \\ &\overset{(*)}{=} \mathrm{Span}\big(f(\mathbf{u}_1), \dots, f(\mathbf{u}_k)\big) \\ &\overset{(**)}{=} \mathrm{Col}\Big(\begin{bmatrix} f(\mathbf{u}_1) & \dots & f(\mathbf{u}_k) \end{bmatrix}\Big) \\ &\overset{(***)}{=} \mathrm{Col}\Big(\begin{bmatrix} A\mathbf{u}_1 & \dots & A\mathbf{u}_k \end{bmatrix}\Big) \\ &\overset{(****)}{=} \mathrm{Col}\Big(A\begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_k \end{bmatrix}\Big), \end{aligned}$$

where (*) follows from Theorem 4.2.11(b), (**) follows from the definition of the column space, and (***) follows from the fact that A is the standard matrix of f, and (****) follows from the definition of matrix multiplication. By Theorem 3.3.4, the pivot columns of a matrix form a basis of the column space of that matrix, and the result follows. □

Example 4.2.16. *Let* $f : \mathbb{Z}_2^5 \to \mathbb{Z}_2^4$ *be the linear function whose standard matrix is*

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix},$$

and consider the vectors

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{u}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{u}_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

in $\mathbb{Z}_2^5$. *Set* $U := \mathit{Span}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4)$. *Find a basis of* $f[U]$.

Solution. Our goal is to find the pivot columns of the matrix $A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 \end{bmatrix}$, since by Proposition 4.2.15, those columns form a basis of $f[U]$. First, by multiplying matrices, we obtain

$$A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}\Big(A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 \end{bmatrix}\Big) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

As we can see, the pivot columns of $A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 \end{bmatrix}$ are its first and fourth column. Therefore,

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis of $f[U]$. □

Example 4.2.17. *Let $f : \mathbb{R}^4 \to \mathbb{R}^3$ be the linear function whose standard matrix is*

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 2 & 4 & 1 & 4 \\ 1 & 2 & 1 & 1 \end{bmatrix},$$

and consider the vectors

$$\mathbf{u}_1 = \begin{bmatrix} -5 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -3 \\ 0 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1 \\ -1 \\ 2 \\ 1 \end{bmatrix}$$

in $\mathbb{R}^4$. Set $U := Span(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$. Find a basis of $f[U]$.

Solution. Our goal is to find the pivot columns of the matrix $A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$, since by Proposition 4.2.15, those columns form a basis of $f[U]$. We compute:

$$A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 2 & 4 & 1 & 4 \\ 1 & 2 & 1 & 1 \end{bmatrix}\begin{bmatrix} -5 & -3 & -1 \\ 1 & 0 & -1 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So, $A\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ is a zero matrix, and consequently, it has no pivot columns. It follows that $\emptyset$ is a basis of $f[U]$.

Remark: So, we effectively got that $f[U] = \{\mathbf{0}\}$. The only basis of the trivial vector space $\{\mathbf{0}\}$ is the empty basis, i.e. $\emptyset$. □

Computing a basis of the preimage of a subspace of the codomain of a linear function.

Proposition 4.2.18. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function, let $A \in \mathbb{F}^{n\times m}$ be the standard matrix of f, let $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{F}^n$ ($k \geq 1$), and set $V := Span(\mathbf{v}_1, \dots, \mathbf{v}_k)$ Then*

$$f^{-1}[V] = \left\{\mathbf{x} \in \mathbb{F}^m \mid \exists \mathbf{y} \in \mathbb{F}^k \text{ s.t. } \left[\begin{array}{c|ccc} A & \mathbf{v}_1 & \dots & \mathbf{v}_k \end{array}\right]\begin{bmatrix} \mathbf{x} \\ \hline \mathbf{y} \end{bmatrix} = \mathbf{0}\right\}$$

$$= \left\{\mathbf{x} \in \mathbb{F}^m \mid \exists \mathbf{y} \in \mathbb{F}^k \text{ s.t. } \begin{bmatrix} \mathbf{x} \\ \hline \mathbf{y} \end{bmatrix} \in Nul\Big(\left[\begin{array}{c|ccc} A & \mathbf{v}_1 & \dots & \mathbf{v}_k \end{array}\right]\Big)\right\}.$$

Proof. Set $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$. Then for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_m \end{bmatrix}^T$ in $\mathbb{F}^m$, we have the following sequence of equivalent statements:

$$\begin{aligned}
& \mathbf{x} \in f^{-1}[V] \\
\iff\quad & f(\mathbf{x}) \in \underbrace{\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)}_{=V} \\
\overset{(*)}{\iff}\quad & A\mathbf{x} \in \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) \\
\iff\quad & \underbrace{x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m}_{=A\mathbf{x}} \in \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) \\
\overset{(**)}{\iff}\quad & \exists \alpha_1, \dots, \alpha_k \in \mathbb{F} \text{ s.t. } x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k \\
\iff\quad & \exists \alpha_1, \dots, \alpha_k \in \mathbb{F} \text{ s.t. } x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m - \alpha_1\mathbf{v}_1 - \cdots - \alpha_k\mathbf{v}_k = \mathbf{0} \\
\overset{(***)}{\iff}\quad & \exists y_1, \dots, y_k \in \mathbb{F} \text{ s.t. } x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m + y_1\mathbf{v}_1 + \cdots + y_k\mathbf{v}_k = \mathbf{0}
\end{aligned}$$

$$\iff \quad \exists y_1, \ldots, y_k \in \mathbb{F} \text{ s.t. } \left[\begin{array}{ccc:ccc} \mathbf{a}_1 & \ldots & \mathbf{a}_m & \mathbf{v}_1 & \ldots & \mathbf{v}_k \end{array}\right] \left[\begin{array}{c} x_1 \\ \vdots \\ \hdashline x_m \\ y_1 \\ \vdots \\ y_k \end{array}\right] = \mathbf{0}$$

$$\iff \quad \exists \mathbf{y} \in \mathbb{F}^k \text{ s.t. } \left[\begin{array}{ccc:ccc} A & & & \mathbf{v}_1 & \ldots & \mathbf{v}_k \end{array}\right] \left[\begin{array}{c} \mathbf{x} \\ \hdashline \mathbf{y} \end{array}\right] = \mathbf{0}$$

$$\overset{(****)}{\iff} \quad \exists \mathbf{y} \in \mathbb{F}^k \text{ s.t. } \left[\begin{array}{c} \mathbf{x} \\ \hdashline \mathbf{y} \end{array}\right] \in \mathrm{Nul}\Big(\left[\begin{array}{ccc:ccc} A & & & \mathbf{v}_1 & \ldots & \mathbf{v}_k \end{array}\right] \Big),$$

where (*) follows from the fact that A is the standard matrix of f, (**) follows from the definition of span, (***) follows by performing the substitution $y_i := -\alpha_i$ for all $i \in \{1, \ldots, k\}$, and (****) follows from the definition of the null space. The result is now immediate. □

Example 4.2.19. *Consider the linear function $f : \mathbb{R}^4 \to \mathbb{R}^5$ whose standard matrix is*

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & -4 & 0 \\ -2 & -3 & -6 & 1 \\ 4 & 0 & 0 & 0 \\ 2 & -1 & -2 & 0 \end{bmatrix},$$

and consider the following vectors in $\mathbb{R}^5$:

- $\mathbf{v}_1 = \begin{bmatrix} -1 & 6 & 9 & -4 & 1 \end{bmatrix}^T$;
- $\mathbf{v}_2 = \begin{bmatrix} 2 & 2 & -2 & 8 & 5 \end{bmatrix}$;
- $\mathbf{v}_3 = \begin{bmatrix} 0 & 0 & 0 & -1 & 0 \end{bmatrix}^T$;
- $\mathbf{v}_4 = \begin{bmatrix} 0 & -2 & -3 & -1 & -1 \end{bmatrix}^T$;
- $\mathbf{v}_5 = \begin{bmatrix} 0 & -1 & -2 & 1 & 0 \end{bmatrix}^T$;
- $\mathbf{v}_6 = \begin{bmatrix} -3 & -1 & 2 & -11 & -6 \end{bmatrix}^T$.

Set $V := Span(\mathbf{v}_1, \ldots, \mathbf{v}_6)$. Find a basis of $f^{-1}[V]$.

Solution. We apply Proposition 4.2.18. We first form the matrix

$$C := \left[\begin{array}{c|cccccc} A & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 & \mathbf{v}_6 \end{array}\right]$$

$$= \left[\begin{array}{cccc|cccccc} 1 & 0 & 0 & 0 & -1 & 2 & 0 & 0 & 0 & -3 \\ 0 & -2 & -4 & 0 & 6 & 2 & 0 & -2 & -1 & -1 \\ -2 & -3 & -6 & 1 & 9 & -2 & 0 & -3 & -2 & 2 \\ 4 & 0 & 0 & 0 & -4 & 8 & -1 & -1 & 1 & -11 \\ 2 & -1 & -2 & 0 & 1 & 5 & 0 & -1 & 0 & -6 \end{array}\right],$$

and we find the general solution of the matrix-vector equation

$$\underbrace{\left[\begin{array}{c|cccccc} A & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 & \mathbf{v}_6 \end{array}\right]}_{=C} \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{y} \end{array}\right] = \mathbf{0},$$

where the vector $\mathbf{x}$ has four entries (because A has four columns) and the vector $\mathbf{y}$ has six entries (because we have six vectors $\mathbf{v}_1, \ldots, \mathbf{v}_6$). By row reducing, we obtain

$$\mathrm{RREF}(C) = \left[\begin{array}{cccc|cccccc} 1 & 0 & 0 & 0 & -1 & 2 & 0 & 0 & 0 & -3 \\ 0 & 1 & 2 & 0 & -3 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & -1 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right].$$

So, the general solution of our matrix-vector equation is

$$\left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{y} \end{array}\right] = \left[\begin{array}{c} q - 2r + 3t \\ -2p + 3q + r - s \\ p \\ 2q + r + 2t \\ \hline q \\ r \\ -s \\ s \\ -t \\ t \end{array}\right], \qquad \text{where } p, q, r, s, t \in \mathbb{R}.$$

But as per Proposition 4.2.18, we only need $\mathbf{x}$! So, we simply ignore the part below the horizontal dotted line, and we obtain

$$\mathbf{x} = \left[\begin{array}{c} q - 2r + 3t \\ -2p + 3q + r - s \\ p \\ 2q + r + 2t \end{array}\right], \qquad \text{where } p, q, r, s, t \in \mathbb{R}.$$

By separating parameters, we obtain

$$\mathbf{x} \;=\; p\begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + q\begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix} + r\begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix} + s\begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 3 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \qquad \text{where } p,q,r,s,t \in \mathbb{R}.$$

In view of Proposition 4.2.18, we now have that

$$\begin{aligned} f^{-1}[V] &= \Big\{ p\begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + q\begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix} + r\begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix} + s\begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 3 \\ 0 \\ 0 \\ 2 \end{bmatrix} \mid p,q,r,s,t \in \mathbb{R} \Big\} \\ &= \operatorname{Span}\Big(\begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 0 \\ 2 \end{bmatrix} \Big) \\ &= \operatorname{Col}\Big(\underbrace{\begin{bmatrix} 0 & 1 & -2 & 0 & 3 \\ -2 & 3 & 1 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 2 \end{bmatrix}}_{=:B} \Big). \end{aligned}$$

We note that the five vectors that we obtained in the second-to-last line above are not necessarily linearly independent,[14] and so to find an actual basis of $f^{-1}[V]$, we row reduce the matrix B and use Theorem 3.3.4. Indeed, Theorem 3.3.4 guarantees that the pivot columns of B form a basis of $\operatorname{Col}(B) = f^{-1}[V]$. By row reducing, we obtain

$$\operatorname{RREF}(B) \;=\; \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 7/5 \\ 0 & 0 & 1 & 0 & -4/5 \\ 0 & 0 & 0 & 1 & 17/5 \end{bmatrix}.$$

Thus, the pivot columns of B are its leftmost four columns, and those four columns

[14] In fact, we can immediately see that they are not linearly independent: no five vectors in $\mathbb{R}^4$ are linearly independent (by Theorem 3.2.17(a)). More generally, though, the reason our computation does not necessarily yield linearly independent vectors is because we "cut off" the entries below the vertical dotted line.

form a basis of $f^{-1}[V]$. So, our final answer is that

$$\left\{ \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

is a basis of $f^{-1}[V]$. □

Example 4.2.20. *Consider the linear function $f : \mathbb{Z}_3^5 \to \mathbb{Z}_3^5$ whose standard matrix is*

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 0 & 2 \\ 1 & 1 & 0 & 0 & 2 \\ 2 & 1 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 & 2 \end{bmatrix},$$

and consider the following vectors in $\mathbb{Z}_3^5$:

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_5 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}.$$

Set $V := Span(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5)$. Find a basis of $f^{-1}[V]$.

Solution. We apply Proposition 4.2.18. We first form the matrix

$$C := \left[\begin{array}{c|ccccc} A & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 \end{array}\right]$$

$$= \left[\begin{array}{ccccc|ccccc} 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 2 & 1 \\ 2 & 1 & 2 & 0 & 2 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 2 & 1 & 0 & 1 & 1 & 2 \\ 2 & 1 & 2 & 1 & 1 & 0 & 2 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \end{array}\right],$$

and we find the general solution of the matrix-vector equation

$$\underbrace{\left[\begin{array}{c|ccccc} A & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 \end{array}\right]}_{=C} \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{y} \end{array}\right] = \mathbf{0},$$

where $\mathbf{x}$ has five entries (because A has five columns) and $\mathbf{y}$ also has five entries (because we have five vectors $\mathbf{v}_1, \ldots, \mathbf{v}_5$). By row reducing, we obtain

$$\mathrm{RREF}(C) = \left[\begin{array}{ccccc|ccccc} 1 & 0 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \end{array}\right].$$

So, the general solution of our matrix-vector equation is

$$\left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{y} \end{array}\right] = \left[\begin{array}{c} p + 2r + 2t \\ 2p + q + t \\ p \\ q \\ q \\ \hline r + s + 2t \\ r \\ s + 2t \\ s \\ t \end{array}\right], \qquad \text{where } p, q, r, s, t \in \mathbb{Z}_3.$$

For $\mathbf{x}$, we get

$$\mathbf{x} = \left[\begin{array}{c} p + 2r + 2t \\ 2p + q + t \\ p \\ q \\ q \end{array}\right], \qquad \text{where } p, q, r, t \in \mathbb{Z}_3.$$

Remark: The parameter s does not appear in the vector $\mathbf{x}$, and so from this point on, that parameter plays no role in our solution.

By separating parameters, we get

$$\mathbf{x} = p\left[\begin{array}{c} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{array}\right] + q\left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{array}\right] + r\left[\begin{array}{c} 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{array}\right] + t\left[\begin{array}{c} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{array}\right], \qquad \text{where } p, q, r, t \in \mathbb{Z}_3.$$

In view of Proposition 4.2.18, we now have that

$$f^{-1}[V] = \left\{ p\begin{bmatrix}1\\2\\1\\0\\0\end{bmatrix} + q\begin{bmatrix}0\\1\\0\\1\\1\end{bmatrix} + r\begin{bmatrix}2\\0\\0\\0\\0\end{bmatrix} + t\begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix} \mid p,q,r,t \in \mathbb{Z}_3 \right\}$$

$$= \operatorname{Span}\Big(\begin{bmatrix}1\\2\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\1\\1\end{bmatrix}, \begin{bmatrix}2\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix} \Big)$$

$$= \operatorname{Col}\Big(\underbrace{\begin{bmatrix}1&0&2&2\\2&1&0&1\\1&0&0&0\\0&1&0&0\\0&1&0&0\end{bmatrix}}_{=:B} \Big).$$

By row reducing, we get

$$\operatorname{RREF}(B) = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{bmatrix}.$$

So, all four columns of B are pivot columns, and by Theorem 3.3.4, the pivot columns of B form a basis of $\operatorname{Col}(B)$. So,

$$\left\{ \begin{bmatrix}1\\2\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\1\\1\end{bmatrix}, \begin{bmatrix}2\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix} \right\}$$

is a basis of $f^{-1}[V]$. □

Example 4.2.21. *Consider the linear function $f : \mathbb{Z}_2^3 \to \mathbb{Z}_2^5$ whose standard matrix is*

$$A = \begin{bmatrix}1&1&0\\1&1&0\\1&1&1\\0&1&1\\0&1&1\end{bmatrix},$$

and consider the following vectors in $\mathbb{Z}_2^5$*:*

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Set $V := \mathit{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$*. Find a basis of* $f^{-1}[V]$*.*

Solution. We apply Proposition 4.2.18. We first form the matrix

$$C \;:=\; \left[\begin{array}{c|ccc} A & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{array}\right] \;=\; \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{array}\right],$$

and we find the general solution of the matrix-vector equation

$$\underbrace{\left[\begin{array}{c|cccc} A & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{array}\right]}_{=C} \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{y} \end{array}\right] \;=\; \mathbf{0},$$

where $\mathbf{x}$ has three entries (because A has three columns) and $\mathbf{y}$ also has three entries (because we have three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$). By row reducing, we obtain

$$\mathrm{RREF}(C) \;=\; \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right].$$

So, the general solution of our matrix-vector equation is

$$\left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{y} \end{array}\right] \;=\; \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline t \\ t \\ t \end{array}\right], \qquad \text{where } t \in \mathbb{Z}_2,$$

and in particular, $\mathbf{x} = \mathbf{0}$. In view of Proposition 4.2.18, we now have that $f^{-1}[V] = \{\mathbf{0}\}$, and consequently, $\emptyset$ is the (unique) basis of $f^{-1}[V]$. □

4.3 Linear functions and bases

We now consider a particularly important type of linear function, which is in fact an isomorphism. Recall that a non-trivial vector space is one that contains at least one non-zero vector, or equivalently, one that has strictly positive (possibly infinite) dimension. Now, suppose that V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and that $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V. By Theorem 3.2.7, every vector of V can be written as linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in a unique way, that is, for all vectors $\mathbf{v} \in V$, there exist unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$\mathbf{v} \;:=\; \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n,$$

and the *coordinate vector* of $\mathbf{v}$ with respect to the basis $\mathcal{B}$ is defined to be

$$\left[\ \mathbf{v}\ \right]_{\mathcal{B}} \;:=\; \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

As our next proposition shows, $\left[\ \cdot\ \right]_{\mathcal{B}} : V \to \mathbb{F}^n$ is an isomorphism. In fact, it is the single most important isomorphism that we will encounter in these lecture notes. It essentially allows us to "translate" vectors of an n-dimensional vector space ($n \neq 0$) into vectors in $\mathbb{F}^n$. We will see some numerical examples that rely on coordinate vectors in subsection 4.4.3, after we have developed some more theory.

Proposition 4.3.1. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V. Then $\left[\ \cdot\ \right]_{\mathcal{B}} : V \to \mathbb{F}^n$ is an isomorphism.*

Proof. We start by proving that $\left[\ \cdot\ \right]_{\mathcal{B}}$ is linear.

1. Fix $\mathbf{x}, \mathbf{y} \in V$. We must show that $\left[\ \mathbf{x}+\mathbf{y}\ \right]_{\mathcal{B}} = \left[\ \mathbf{x}\ \right]_{\mathcal{B}} + \left[\ \mathbf{y}\ \right]_{\mathcal{B}}$. Set $\left[\ \mathbf{x}\ \right]_{\mathcal{B}} = \left[\ \alpha_1 \ \ \ldots \ \ \alpha_n\ \right]^T$ and $\left[\ \mathbf{y}\ \right]_{\mathcal{B}} = \left[\ \beta_1 \ \ \ldots \ \ \beta_n\ \right]^T$. Then $\mathbf{x} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$ and $\mathbf{y} = \beta_1\mathbf{v}_1 + \cdots + \beta_n\mathbf{v}_n$; consequently,

$$\mathbf{x}+\mathbf{y} \;=\; (\alpha_1+\beta_1)\mathbf{v}_1 + \cdots + (\alpha_n+\beta_n)\mathbf{v}_n,$$

and so $\left[\ \mathbf{x}+\mathbf{y}\ \right]_{\mathcal{B}} = \left[\ \alpha_1+\beta_1 \ \ \ldots \ \ \alpha_n+\beta_n\ \right]^T$. We now have that

$$\left[\ \mathbf{x}+\mathbf{y}\ \right]_{\mathcal{B}} \;=\; \begin{bmatrix} \alpha_1+\beta_1 \\ \vdots \\ \alpha_n+\beta_n \end{bmatrix} \;=\; \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \;=\; \left[\ \mathbf{x}\ \right]_{\mathcal{B}} + \left[\ \mathbf{y}\ \right]_{\mathcal{B}}.$$

2. Fix $\mathbf{x} \in V$ and $\alpha \in \mathbb{F}$. Set $\left[\ \mathbf{x}\ \right]_{\mathcal{B}} = \left[\ \alpha_1 \ \ \ldots \ \ \alpha_n\ \right]^T$. Then $\mathbf{x} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$; consequently, $\alpha\mathbf{x} = \alpha\alpha_1\mathbf{v}_1 + \cdots + \alpha\alpha_n\mathbf{v}_n$, and so $\left[\ \alpha\mathbf{x}\ \right]_{\mathcal{B}} =$

$\begin{bmatrix} \alpha\alpha_1 & \dots & \alpha\alpha_n \end{bmatrix}^T$. We now have that

$$\left[\alpha\mathbf{x} \right]_{\mathcal{B}} = \begin{bmatrix} \alpha\alpha_1 \\ \vdots \\ \alpha\alpha_n \end{bmatrix} = \alpha \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \alpha \left[\mathbf{x} \right]_{\mathcal{B}}.$$

By 1. and 2., f is linear.

It remains to show that $\left[\cdot \right]_{\mathcal{B}}$ is a bijection, i.e. that it is one-to-one and onto $\mathbb{F}^n$. Since V and $\mathbb{F}^n$ are both n dimensional, Corollary 4.2.10 guarantees that f is one-to-one if and only if f is onto $\mathbb{F}^n$. So, it is enough to show that f is onto $\mathbb{F}^n$. Fix $\begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T \in \mathbb{F}^n$. Set $\mathbf{v} := \alpha_1\mathbf{v}_1 + \dots + \alpha_n\mathbf{v}_n$. Then $\left[\mathbf{v} \right]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T$. So, $\left[\cdot \right]_{\mathcal{B}}$ is onto $\mathbb{F}^n$. This completes the argument. □

Theorem 4.3.2 (below) is one of the main reasons we care about bases. It essentially states that, given vector spaces U and V over a field $\mathbb{F}$, where U is finite-dimensional with a basis $\mathcal{B}$, we can uniquely determine a linear function $f : U \to V$ by specifying what the basis vectors from $\mathcal{B}$ get mapped to (and we get to determine arbitrarily what vectors of V those basis vectors get mapped to). We note that Theorem 4.3.2 can, in fact, be generalized to infinite-dimensional domains U, but this would require working with infinite bases, and we omit the details.

Theorem 4.3.2. *Let U and V be vector spaces over a field $\mathbb{F}$, and assume that U is finite-dimensional. Let $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be a basis of U, and let $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$.*[15] *Then there exists a unique linear function $f : U \to V$ such that $f(\mathbf{u}_1) = \mathbf{v}_1, \dots, f(\mathbf{u}_n) = \mathbf{v}_n$. Moreover, if the vector space U is non-trivial (i.e. $n \neq 0$), then this unique linear function $f : U \to V$ satisfies the following: for all $\mathbf{u} \in U$, we have that*

$$f(\mathbf{u}) = \alpha_1\mathbf{v}_1 + \dots + \alpha_n\mathbf{v}_n,$$

where $\left[\mathbf{u} \right]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T$. On the other hand, if U is trivial (i.e. $U = \{\mathbf{0}\}$),[16] *then $f : U \to V$ is given by $f(\mathbf{0}) = \mathbf{0}$.*

Proof. Suppose first that the vector space U is trivial, i.e. $n = 0$ and $U = \{\mathbf{0}\}$. Then the function $f : U \to V$ given by $f(\mathbf{0}) = \mathbf{0}$ is obviously linear, and moreover, it vacuously satisfies $f(\mathbf{u}_1) = \mathbf{v}_1, \dots, f(\mathbf{u}_n) = \mathbf{v}_n$ (because $n = 0$, and so both $\mathbf{u}_1, \dots, \mathbf{u}_n$ and $\mathbf{v}_1, \dots, \mathbf{v}_n$ are empty lists of vectors). The uniqueness of f follows from Proposition 4.1.6.

From now on, we assume that the vector space U is non-trivial, i.e. that $n \neq 0$. We must prove the existence and the uniqueness of the linear function f satisfying the required properties.

[15] Here, $\mathbf{v}_1, \dots, \mathbf{v}_n$ are arbitrary vectors in V. They are not necessarily pairwise distinct.

[16] Note that in this case, we have that $n = 0$ and $\mathcal{B} = \emptyset$.

Existence. Let $f : U \to V$ be defined as in the statement of the theorem, i.e. for all $\mathbf{u} \in U$, we set

$$f(\mathbf{u}) \quad = \quad \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n,$$

where $\left[\begin{array}{c} \mathbf{u} \end{array} \right]_{\mathcal{B}} = \left[\begin{array}{ccc} \alpha_1 & \ldots & \alpha_n \end{array} \right]^T$. Note that this means that for all $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$, we have that

$$f(\alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n) \quad = \quad \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n.$$

Let us show that f is linear and satisfies $f(\mathbf{u}_1) = \mathbf{v}_1, \ldots, f(\mathbf{u}_n) = \mathbf{v}_n$. For the latter, we note that for all $i \in \{1, \ldots, n\}$, we have that

$$\begin{aligned} f(\mathbf{u}_i) &= f(0\mathbf{u}_1 + \cdots + 0\mathbf{u}_{i-1} + 1\mathbf{u}_i + 0\mathbf{u}_{i+1} + \cdots + 0\mathbf{u}_n) \\ &= 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{i-1} + 1\mathbf{v}_i + 0\mathbf{v}_{i+1} + \cdots + 0\mathbf{v}_n \\ &= \mathbf{v}_i. \end{aligned}$$

This proves that $f(\mathbf{u}_1) = \mathbf{v}_1, \ldots, f(\mathbf{u}_n) = \mathbf{v}_n$.

Let us now show that f is linear. We verify that f satisfies the two axioms from the definition of a linear function.

1. Fix $\mathbf{x}, \mathbf{y} \in U$. We must show that $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$. Set $\left[\begin{array}{c} \mathbf{x} \end{array} \right]_{\mathcal{B}} = \left[\begin{array}{ccc} \alpha_1 & \ldots & \alpha_n \end{array} \right]^T$ and $\mathbf{y} = \left[\begin{array}{ccc} \beta_1 & \ldots & \beta_n \end{array} \right]^T$. We then have that $\left[\begin{array}{c} \mathbf{x}+\mathbf{y} \end{array} \right]_{\mathcal{B}} = \left[\begin{array}{ccc} \alpha_1 + \beta_1 & \ldots & \alpha_n + \beta_n \end{array} \right]^T$,[17] and we see that

$$\begin{aligned} f(\mathbf{x}+\mathbf{y}) &\overset{(*)}{=} (\alpha_1 + \beta_1)\mathbf{v}_1 + \cdots + (\alpha_n + \beta_n)\mathbf{v}_n \\ &= (\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n) + (\beta_1 \mathbf{v}_1 + \cdots + \beta_n \mathbf{v}_n) \\ &\overset{(**)}{=} f(\mathbf{x}) + f(\mathbf{y}), \end{aligned}$$

where both (*) and (**) follow from the construction of f.

2. Fix $\mathbf{u} \in U$ and $\alpha \in \mathbb{F}$. We must show that $f(\alpha\mathbf{u}) = \alpha f(\mathbf{u})$. Set $\left[\begin{array}{c} \mathbf{u} \end{array} \right]_{\mathcal{B}} =$

[17] Indeed, since $\left[\begin{array}{c} \mathbf{x} \end{array} \right]_{\mathcal{B}} = \left[\begin{array}{ccc} \alpha_1 & \ldots & \alpha_n \end{array} \right]^T$ and $\mathbf{y} = \left[\begin{array}{ccc} \beta_1 & \ldots & \beta_n \end{array} \right]^T$, we have that $\mathbf{x} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n$ and $\mathbf{y} = \beta_1 \mathbf{u}_1 + \cdots + \beta_n \mathbf{u}_n$, and consequently,

$$\begin{aligned} \mathbf{x}+\mathbf{y} &= (\alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n) + (\beta_1 \mathbf{u}_1 + \cdots + \beta_n \mathbf{u}_n) \\ &= (\alpha_1 + \beta_1)\mathbf{u}_1 + \cdots + (\alpha_n + \beta_n)\mathbf{u}_n, \end{aligned}$$

and so $\left[\begin{array}{c} \mathbf{x}+\mathbf{y} \end{array} \right]_{\mathcal{B}} = \left[\begin{array}{ccc} \alpha_1 + \beta_1 & \ldots & \alpha_n + \beta_n \end{array} \right]^T$.

$\begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T$. Then $\begin{bmatrix} \alpha\mathbf{u} \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \alpha\alpha_1 & \dots & \alpha\alpha_n \end{bmatrix}^T$,[18] and we see that

$$f(\alpha\mathbf{u}) \overset{(*)}{=} (\alpha\alpha_1)\mathbf{v}_1 + \cdots + (\alpha\alpha_n)\mathbf{v}_n = \alpha(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) \overset{(**)}{=} \alpha f(\mathbf{u}),$$

where both (*) and (**) follow from the construction of f.

By 1. and 2., we see that f is linear. This completes the proof of existence.

Uniqueness. Let $f_1, f_2 : U \to V$ be linear functions that satisfy $f_1(\mathbf{u}_1) = \mathbf{v}_1, \dots, f_1(\mathbf{u}_n) = \mathbf{v}_n$ and $f_2(\mathbf{u}_1) = \mathbf{v}_1, \dots, f_2(\mathbf{u}_n) = \mathbf{v}_n$. We must show that $f_1 = f_2$. Fix $\mathbf{u} \in U$. We must show that $f_1(\mathbf{u}) = f_2(\mathbf{u})$. Set $\begin{bmatrix} \mathbf{u} \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T$. Then

$$\begin{aligned}
f_1(\mathbf{u}) &= f_1(\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n) \\
&= \alpha_1 f_1(\mathbf{u}_1) + \cdots + \alpha_n f_1(\mathbf{u}_n) && \text{by the linearity of } f_1 \text{ (and more precisely, by Proposition 4.1.5)} \\
&= \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n && \text{because } f_1(\mathbf{u}_1) = \mathbf{v}_1, \dots, f_1(\mathbf{u}_n) = \mathbf{v}_n \\
&= \alpha_1 f_2(\mathbf{u}_1) + \cdots + \alpha_n f_2(\mathbf{u}_n) && \text{because } f_2(\mathbf{u}_1) = \mathbf{v}_1, \dots, f_2(\mathbf{u}_n) = \mathbf{v}_n \\
&= f_2(\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n) && \text{by the linearity of } f_2 \text{ (and more precisely, by Proposition 4.1.5)} \\
&= f_2(\mathbf{u}).
\end{aligned}$$

Thus, $f_1 = f_2$. This proves uniqueness. □

Corollary 4.3.3. *Let U and V be vector spaces over a field $\mathbb{F}$, and assume that U is finite-dimensional. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ be a linearly independent set of vectors in U, and let $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$.*[19] *Then there exists a linear function $f : U \to V$ such that $f(\mathbf{u}_1) = \mathbf{v}_1, \dots, f(\mathbf{u}_k) = \mathbf{v}_k$. Moreover, if V is non-trivial, then this linear function f is unique if and only if $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a basis of U.*

Remark: If V is trivial (i.e. $V = \{\mathbf{0}\}$, and consequently $\mathbf{v}_1 = \cdots = \mathbf{v}_k = \mathbf{0}$), then there exists exactly one **function** from U to V, this function maps all elements of U to $\mathbf{0}$, and obviously, it is linear.

[18] Indeed, since $\begin{bmatrix} \mathbf{u} \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \alpha_1 & \dots & \alpha_n \end{bmatrix}^T$, we have that $\mathbf{u} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$. Consequently, $\alpha\mathbf{u} = (\alpha\alpha_1)\mathbf{u}_1 + \cdots + (\alpha\alpha_n)\mathbf{u}_n$, and so $\begin{bmatrix} \alpha\mathbf{u} \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \alpha\alpha_1 & \dots & \alpha\alpha_n \end{bmatrix}^T$.

[19] Here, $\mathbf{v}_1, \dots, \mathbf{v}_k$ are arbitrary vectors in V. They are not necessarily pairwise distinct.

Proof of Corollary 4.3.3 (outline). Using Theorem 3.2.19, we extend $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ to a basis of U, and then we apply Theorem 4.3.2. The details are left as an exercise. $\square$

4.4 Isomorphisms

Recall that, for vector spaces U and V over a field $\mathbb{F}$, a function $f : U \to V$ is an *isomorphism* if it is linear and a bijection.

Vector spaces U and V (over the same field $\mathbb{F}$) are *isomorphic*, and we write $U \cong V$, if there exits an isomorphism $f : U \to V$.

4.4.1 Basic properties of isomorphisms

Proposition 4.4.1 (below) generalizes Proposition 1.10.20 to isomorphisms between arbitrary vector spaces. The proof is essentially identical to that of Proposition 1.10.20.

Proposition 4.4.1. *Let U and V be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be an isomorphism. Then $f^{-1} : V \to U$ is also an isomorphism.*

$$U \underset{f^{-1}}{\overset{f}{\rightleftarrows}} V$$

Proof. Since $f : U \to V$ is an isomorphism, it is, in particular, a bijection; consequently, f has an inverse $f^{-1} : V \to U$, which is also a bijection. So, to show that f^{-1} is an isomorphism, it suffices to show that f^{-1} is linear.

First, fix $\mathbf{v}_1, \mathbf{v}_2 \in V$. We must show that $f^{-1}(\mathbf{v}_1 + \mathbf{v}_2) = f^{-1}(\mathbf{v}_1) + f^{-1}(\mathbf{v}_2)$. Set $\mathbf{u}_1 := f^{-1}(\mathbf{v}_1)$ and $\mathbf{u}_2 := f^{-1}(\mathbf{v}_2)$, so that $f(\mathbf{u}_1) = \mathbf{v}_1$ and $f(\mathbf{u}_2) = \mathbf{v}_2$. Then

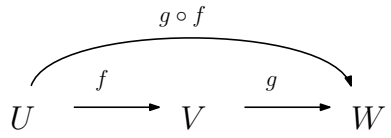
$$\begin{aligned}
f^{-1}(\mathbf{v}_1 + \mathbf{v}_2) &= f^{-1}\big(f(\mathbf{u}_1) + f(\mathbf{u}_2)\big) \\
&= f^{-1}\big(f(\mathbf{u}_1 + \mathbf{u}_2)\big) && \text{because } f \text{ is linear} \\
&= (f^{-1} \circ f)(\mathbf{u}_1 + \mathbf{u}_2) \\
&= \mathrm{Id}_U(\mathbf{u}_1 + \mathbf{u}_2) \\
&= \mathbf{u}_1 + \mathbf{u}_2 \\
&= f^{-1}(\mathbf{v}_1) + f^{-1}(\mathbf{v}_2).
\end{aligned}$$

Next, fix $\mathbf{v} \in V$ and $\alpha \in \mathbb{F}$. We must show that $f^{-1}(\alpha\mathbf{v}) = \alpha f^{-1}(\mathbf{v})$. Set $\mathbf{u} := f^{-1}(\mathbf{v})$, so that $f(\mathbf{u}) = \mathbf{v}$. Then

$$\begin{aligned} f^{-1}(\alpha\mathbf{v}) &= f^{-1}\big(\alpha f(\mathbf{u})\big) \\ &= f^{-1}\big(f(\alpha\mathbf{u})\big) && \text{because } f \text{ is linear} \\ &= (f^{-1}\circ f)(\alpha\mathbf{u}) \\ &= \mathrm{Id}_U(\alpha\mathbf{u}) \\ &= \alpha\mathbf{u} \\ &= \alpha f^{-1}(\mathbf{v}). \end{aligned}$$

We have now proven that f^{-1} linear. This completes the argument. □

Proposition 4.4.2. *Let U, V, and W be vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ and $g : V \to W$ be isomorphisms. Then $g \circ f : U \to W$ is an isomorphism.*



Proof. Since $f : U \to V$ and $g : V \to W$ are linear functions (because they are isomorphisms), Proposition 4.1.7 guarantees that their composition $g \circ f : U \to W$ is also linear. Since $f : U \to V$ and $g : V \to W$ are bijections, Proposition 1.10.17 guarantees that $g \circ f : U \to W$ is also a bijection. So, $g \circ f : U \to W$ is linear and a bijection, i.e. it is an isomorphism. □

Theorem 4.4.3. *Let U, V, and W be vector spaces over a field $\mathbb{F}$. Then all the following hold:*

(a) $U \cong U$;

(b) if $U \cong V$, then $V \cong U$;

(c) if $U \cong V$ and $V \cong W$, then $U \cong W$.

Proof. (a) Clearly, $\mathrm{Id}_U : U \to U$ (the identity function on U) is an isomorphism. So, $U \cong U$.

(b) Suppose that $U \cong V$. Then there exists an isomorphism $f : U \to V$. But then by Proposition 4.4.1, $f^{-1} : V \to U$ is also an isomorphism. So, $V \cong U$.

(c) Suppose that $U \cong V$ and $V \cong W$. Then there exist isomorphisms $f : U \to V$ and $g : V \to W$. But then by Proposition 4.4.2, $g \circ f : U \to W$ is an isomorphism. So, $U \cong W$. □

The final result of this subsection (Theorem 4.4.4 below) essentially states that isomorphisms map linearly independent sets to linearly independent set, spanning sets to spanning sets, and bases to bases. As we shall see, it is an easy corollary of Theorem 4.2.13.

Theorem 4.4.4. *Let U and V be vector spaces over a field $\mathbb{F}$, let $f : U \to V$ be an isomorphism, and let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in U$. Then all the following hold:*

(a) vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are linearly independent in U if and only if vectors $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_k)$ are linearly independent in V;

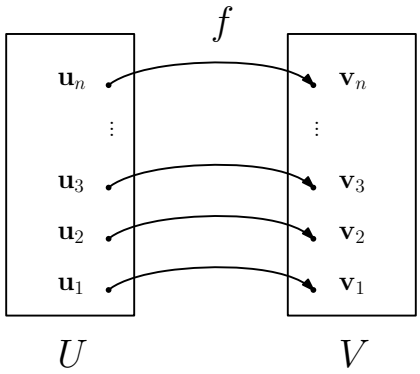
(b) vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ span U if and only if vectors $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_k)$ span V;

(c) $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of U if and only if $\{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_k)\}$ is a basis of V.

Proof. Since f is an isomorphism, it is, by definition, a linear function that is both one-to-one and onto. Thus, (a) follows from Theorem 4.2.13(a-b), and (b) follows from Theorem 4.2.13(c-d). Finally, since a basis of a vector space is simply a linearly independent set of vectors that spans that vector space, parts (a) and (b) together imply (c). □

Proposition 4.4.5 (below) is a converse of sorts of Theorem 4.4.4(c). It essentially states that any linear function that (injectively) maps a basis onto a basis is an isomorphism.

Proposition 4.4.5. *Let U and V be finite-dimensional vector spaces over a field $\mathbb{F}$. Assume that $dim(U) = dim(V) =: n$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis of U, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V. Then there exists a unique linear function $f : U \to V$ such that $f(\mathbf{u}_1) = \mathbf{v}_1, \ldots, f(\mathbf{u}_n) = \mathbf{v}_n$. Moreover, this linear function f is an isomorphism.*



Proof. The existence and uniqueness of the linear function f follows from Theorem 4.3.2. We need to show that the linear function f is in fact an isomorphism. But by hypothesis, U and V are finite-dimensional vector spaces satisfying

$\dim(U) = \dim(V)$, and so by Corollary 4.2.10, it is enough to show that f is onto. Fix $\mathbf{v} \in V$. Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V, we know that there exist scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$. But now

$$\begin{aligned} f(\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n) &\overset{(*)}{=} \alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_n f(\mathbf{u}_n) \\ &= \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n \\ &= \mathbf{v}, \end{aligned}$$

where (*) follows from the linearity of f (and more precisely, from Proposition 4.1.5). So, f is onto, and we are done. □

4.4.2 Isomorphism and dimension

By Theorem 4.2.14(c), any two isomorphic vector spaces have the same dimension. Theorem 4.4.6 (below) guarantees that, in the case of **finite-dimensional** vector spaces, the converse is also true: any two vector spaces (over the same field) that have the same finite dimension are isomorphic. We give two proofs (both of them quite short) of this result. One of the proofs relies on coordinate vectors and Theorem 4.4.3, whereas the other one relies on Proposition 4.4.5.

Theorem 4.4.6. *Let U and V be **finite-dimensional** vector spaces over a field $\mathbb{F}$. Then U and V are isomorphic if and only if $\dim(U) = \dim(V)$.*

Warning: This theorem is only true for finite-dimensional vector spaces, and it becomes false for infinite-dimensional ones.

Proof#1. If U and V are isomorphic, then Theorem 4.2.14(c) guarantees that $\dim(U) = \dim(V)$.[20] Suppose, conversely, that $\dim(U) = \dim(V) =: n$. Fix any basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of U and any basis $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ of V. By Proposition 4.3.1, $\left[\ \cdot\ \right]_{\mathcal{B}} : U \to \mathbb{F}^n$ and $\left[\ \cdot\ \right]_{\mathcal{C}} : V \to \mathbb{F}^n$ are both isomorphisms, and consequently, $U \cong \mathbb{F}^n$ and $V \cong \mathbb{F}^n$. But now Theorem 4.4.3 guarantees that $U \cong V$. □

Proof#2. If U and V are isomorphic, then Theorem 4.2.14(c) guarantees that $\dim(U) = \dim(V)$. Suppose, conversely, that $\dim(U) = \dim(V) =: n$. Fix a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of U and a basis $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ of V. Then by Proposition 4.4.5, there exists a unique linear function $f : U \to V$ such that $f(\mathbf{b}_1) = \mathbf{c}_1, \ldots, f(\mathbf{b}_n) = \mathbf{c}_n$, and moreover, this linear function f is an isomorphism. So, U and V are isomorphic. □

We complete this subsection with a technical proposition that slightly generalizes Theorem 4.2.14(c).

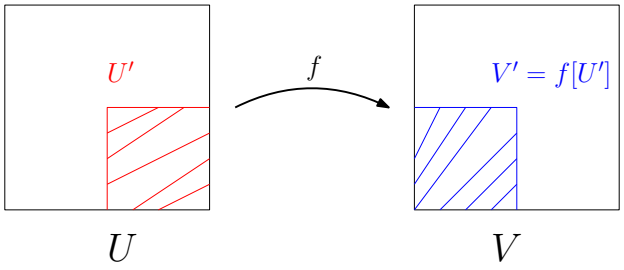
[20] Indeed, if $U \cong V$, then by definition, there exists an isomorphism $f : U \to V$, and so by Theorem 4.2.14(c), we have that $\dim(U) = \dim(V)$.

Proposition 4.4.7. *Let U and V be a vector spaces over a field $\mathbb{F}$, and let $f : U \to V$ be an isomorphism, and let $U' \subseteq U$. Then U' is a subspace of U if and only if $V' := f[U']$ is a subspace of V. Moreover, in this case,*[21] *all the following hold:*

(a) the function $f' : U' \to V'$ given by $f'(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in U'$ is an isomorphism;[22]

(b) $U' \cong V'$;

(c) $dim(U') = dim(V')$.[23]



Proof. First of all, since f is an isomorophism (and in particular, a bijection), we see that $f^{-1}[V'] = U'$. Further, since $f : U \to V$, Proposition 4.4.1 guarantees that $f^{-1} : V \to U$ is an isomorphism. Now, if U' is a subspace of U, then Theorem 4.2.3 guarantees that $f[U'] = V'$ is a subsapce of V'. On the other hand, if V' is a subspace of V, then Theorem 4.2.3 implies that $f^{-1}[V'] = U'$ is a subspace of U.

From now on, we assume that U' is a subspace of U and V' is a subspace of V. We construct the function $f' : U' \to V'$ by setting $f'(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in V'$, and in part (a). Since $f : U \to V$ is an isomorphism, and since U' and V' are subspaces of U and V, respectively, we see that $f' : U' \to V'$ is also an isomorphism,[24] and consequently, $U' \cong V'$. So, (a) and (b) hold. Part (c) follows from part (a) and from Theorem 4.2.14(c) applied to U', V', and f'. □

4.4.3 An application of isomorphisms: transforming polynomials and matrices into vectors

By Theorem 4.4.6, for all positive integers n and fields $\mathbb{F}$, every n-dimensional vector space V over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$. Moreover, by Proposition 4.3.1, given any basis $\mathcal{B}$ of such a vector space V, the coordinate function $\left[\ \cdot\ \right]_{\mathcal{B}} : V \to \mathbb{F}^n$ is an isomorphism. This is useful because we have developed powerful computational tools for vectors in $\mathbb{F}^n$. By using isomorphisms, we can reduce problems of computing in an arbitrary

[21] That is: if U' is a subspace of U and V' is a subspace of V.

[22] So, we constructed f' by restricting both the domain and the codomain of f. This is well defined because for all $\mathbf{u} \in U'$, we have that $f(\mathbf{u}) \in f[U'] = V'$.

[23] So, either U' and V' have the same finite dimension, or they are both infinite-dimensional.

[24] This follows from the definition of an isomorphism. Details?

n-dimensional vector space to problems of computing in $\mathbb{F}^n$, which we know how to do in many cases.

Remark: When working with coordinate vectors, we must always **specify the basis** that we are working with (i.e. with respect to which the coordinate vectors are computed). Choosing a different basis will, in general, produce different coordinate vectors. For instance, consider the real vector space $\mathbb{P}^2_{\mathbb{R}}$ of all polynomials of degree at most 2 and with coefficients in $\mathbb{R}$. There are two "obvious" bases to chose for $\mathbb{P}^2_{\mathbb{R}}$, namely $\mathcal{A}_1 = \{1, x, x^2\}$ and $\mathcal{A}_2 = \{x^2, x, 1\}$. For a polynomial $p(x) = a_2x^2 + a_1x + a_0$ (with $a_0, a_1, a_2 \in \mathbb{R}$), we have

$$\left[\ p(x)\ \right]_{\mathcal{A}_1} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \qquad \text{and} \qquad \left[\ p(x)\ \right]_{\mathcal{A}_2} = \begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix}.$$

As we can see, the coordinate vectors are different (whenever $a_0 \neq a_2$), which is why we have to be careful to specify what basis we are working with.

First of all, using Proposition 4.3.1 and Theorem 4.4.4, we can "translate" Propositions 3.1.10, 3.2.1, and 3.2.6 into statements for arbitrary non-trivial, finite-dimensional vector spaces, as follows.

Proposition 4.4.8. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis of V. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ ($m \geq 1$) be some vectors in V, and for all $i \in \{1, \ldots, n\}$, set $\mathbf{a}_i := \left[\ \mathbf{v}_i\ \right]_{\mathcal{B}}$. Set $A := \left[\ \mathbf{a}_1 \ \ \ldots \ \ \mathbf{a}_m\ \right]$. Then all the following hold:*

(a) $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly independent set in V if and only if rank$(A) = m$ (i.e. A has full column rank);

(b) $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a spanning set of V if and only if rank$(A) = n$ (i.e. A has full row rank);

(c) $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a basis of V if and only if rank$(A) = n = m$ (i.e. A is a square matrix of full rank).

Proof. By Proposition 4.3.1, $\left[\ \cdot\ \right]_{\mathcal{B}} : V \to \mathbb{F}^n$ is an isomorphism. So, Theorem 4.4.4 guarantees that the following hold:

(a') $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly independent set in V if and only if $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is a linearly independent set in $\mathbb{F}^n$;

(b') $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a spanning set of V if and only if $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is a spanning set of $\mathbb{F}^n$;

(c') $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a basis of V if and only if $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is a basis of $\mathbb{F}^n$.

Now (a) follows from (a') and Proposition 3.2.1; (b) follows from (b') and Proposition 3.1.10; and (c) follows from (c') and Proposition 3.2.6. □

Example 4.4.9. *Consider the following sets of polynomials (with coefficients understood to be in $\mathbb{R}$):*

(a) $\mathcal{A} = \{x^2 + x, x^3 + 1, x, x^2 + 1\}$;

(b) $\mathcal{B} = \{3x^3 + 2x^2 + x + 1, 6x^3 + 4x^2 + 5x + 6, 5x + 6, 2x + 2\}$;

(c) $\mathcal{C} = \{x^3 + 1, x^3 + x^2, x^2 + x, x + 1, 1, x\}$;

(d) $\mathcal{D} = \{x^3, 2x^2 + 3x, 4x^3 + 5x + 6\}$.

For each of the four sets above, determine whether

- *it is linearly independent in $\mathbb{P}^3_{\mathbb{R}}$;*
- *it spans $\mathbb{P}^3_{\mathbb{R}}$;*
- *it is a basis of $\mathbb{P}^3_{\mathbb{R}}$.*

Solution. In what follows, we will use the basis $\mathcal{P} = \{1, x, x^2, x^3\}$ of $\mathbb{P}^3_{\mathbb{R}}$.

(a) We set

- $\mathbf{a}_1 := \left[\ x^2 + x\ \right]_{\mathcal{P}} = \left[\ 0 \quad 1 \quad 1 \quad 0\ \right]^T$;
- $\mathbf{a}_2 := \left[\ x^3 + 1\ \right]_{\mathcal{P}} = \left[\ 1 \quad 0 \quad 0 \quad 1\ \right]^T$;
- $\mathbf{a}_3 := \left[\ x\ \right]_{\mathcal{P}} = \left[\ 0 \quad 1 \quad 0 \quad 0\ \right]^T$;
- $\mathbf{a}_4 := \left[\ x^2 + 1\ \right]_{\mathcal{P}} = \left[\ 1 \quad 0 \quad 1 \quad 0\ \right]^T$;

Further, we set

$$A \ := \ \left[\ \mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4\ \right] \ = \ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

By row reducing, we get that $\mathrm{RREF}(A) = I_4$, and consequently, $\mathrm{rank}(A) = 4$. So, by Proposition 4.4.8, $\mathcal{A}$ is a basis of $\mathbb{P}^3_{\mathbb{R}}$, and in particular, it is a linearly independent set in $\mathbb{P}^3_{\mathbb{R}}$, as well as a spanning set of $\mathbb{P}^3_{\mathbb{R}}$.

(b) We set

- $\mathbf{b}_1 := \left[\ 3x^3 + 2x^2 + x + 1\ \right]_{\mathcal{P}} = \left[\ 1 \quad 1 \quad 2 \quad 3\ \right]^T$;

- $\mathbf{b}_2 := \left[\ 6x^3+4x^2+5x+6\ \right]_{\mathcal{P}} = \left[\ 6 \quad 5 \quad 4 \quad 6\ \right]^T$;
- $\mathbf{b}_3 := \left[\ 5x+6\ \right]_{\mathcal{P}} = \left[\ 6 \quad 5 \quad 0 \quad 0\ \right]^T$;
- $\mathbf{b}_4 := \left[\ 2x+2\ \right]_{\mathcal{P}} = \left[\ 2 \quad 2 \quad 0 \quad 0\ \right]^T$.

Further, we set

$$B \ := \ \left[\ \mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \quad \mathbf{b}_4\ \right] \ = \ \begin{bmatrix} 1 & 6 & 6 & 2 \\ 1 & 5 & 5 & 2 \\ 2 & 4 & 0 & 0 \\ 3 & 6 & 0 & 0 \end{bmatrix}.$$

By row reducing, we get that

$$\mathrm{RREF}(B) \ = \ \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and consequently, $\mathrm{rank}(B) = 3$. So, by Proposition 4.4.8, $\mathcal{B}$ is not a linearly independent set in $\mathbb{P}^3_{\mathbb{R}}$, is not a spanning set of $\mathbb{P}^3_{\mathbb{R}}$, and is not a basis of $\mathbb{P}^3_{\mathbb{R}}$.

(c) We set

- $\mathbf{c}_1 := \left[\ x^3+1\ \right]_{\mathcal{P}} = \left[\ 1 \quad 0 \quad 0 \quad 1\ \right]^T$;
- $\mathbf{c}_2 := \left[\ x^3+x^2\ \right]_{\mathcal{P}} = \left[\ 0 \quad 0 \quad 1 \quad 1\ \right]^T$;
- $\mathbf{c}_3 := \left[\ x^2+x\ \right]_{\mathcal{P}} = \left[\ 0 \quad 1 \quad 1 \quad 0\ \right]^T$;
- $\mathbf{c}_4 := \left[\ x+1\ \right]_{\mathcal{P}} = \left[\ 1 \quad 1 \quad 0 \quad 0\ \right]^T$;
- $\mathbf{c}_5 := \left[\ 1\ \right]_{\mathcal{P}} = \left[\ 1 \quad 0 \quad 0 \quad 0\ \right]^T$;
- $\mathbf{c}_6 := \left[\ x\ \right]_{\mathcal{P}} = \left[\ 0 \quad 1 \quad 0 \quad 0\ \right]^T$.

Further, we set

$$C \ := \ \left[\ \mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3 \quad \mathbf{c}_4 \quad \mathbf{c}_5 \quad \mathbf{c}_6\ \right] \ = \ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By row reducing, we get that

$$\mathrm{RREF}(C) \;=\; \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix},$$

and consequently, $\mathrm{rank}(C) = 4$. So, by Proposition 4.4.8, $\mathcal{C}$ is not linearly independent, it is spanning set of $\mathbb{P}^3_{\mathbb{R}}$, and it is not a basis of $\mathbb{P}^3_{\mathbb{R}}$.

Remark: Since $\dim(\mathbb{P}^3_{\mathbb{R}}) = 4$ (because $\mathcal{P} = \{1, x, x^2, x^3\}$ is a basis of $\mathbb{P}^3_{\mathbb{R}}$), and since $\mathcal{C}$ contains six vectors (polynomials), Theorem 3.2.17 guarantees that $\mathcal{C}$ is not a linearly independent set in $\mathbb{P}^3_{\mathbb{R}}$, and consequently, it is not a basis of $\mathbb{P}^3_{\mathbb{R}}$. However, to determine whether $\mathcal{C}$ spans $\mathbb{P}^3_{\mathbb{R}}$, we did in fact have to row reduce.

(d) We set

- $\mathbf{d}_1 := \left[\; x^3 \;\right]_{\mathcal{P}} = \left[\; 0 \quad 0 \quad 0 \quad 1 \;\right]^T$;
- $\mathbf{d}_2 := \left[\; 2x^2 + 3x \;\right]_{\mathcal{P}} = \left[\; 0 \quad 3 \quad 2 \quad 0 \;\right]^T$;
- $\mathbf{d}_3 := \left[\; 4x^3 + 5x + 6 \;\right]_{\mathcal{P}} = \left[\; 6 \quad 5 \quad 0 \quad 4 \;\right]^T$.

Further, we set

$$D \;:=\; \left[\; \mathbf{d}_1 \quad \mathbf{d}_2 \quad \mathbf{d}_3 \;\right] \;=\; \begin{bmatrix} 0 & 0 & 6 \\ 0 & 3 & 5 \\ 0 & 2 & 0 \\ 1 & 0 & 4 \end{bmatrix}.$$

By row reducing, we get that

$$\mathrm{RREF}(D) \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and consequently, $\mathrm{rank}(D) = 3$. So, by Proposition 4.4.8, $\mathcal{D}$ is linearly independent, but it is not a spanning set of $\mathbb{P}^3_{\mathbb{R}}$, and it is not a basis of $\mathbb{P}^3_{\mathbb{R}}$.

Remark: Since $\dim(\mathbb{P}^3_{\mathbb{R}}) = 4$ (because $\mathcal{P} = \{1, x, x^2, x^3\}$ is a basis of $\mathbb{P}^3_{\mathbb{R}}$), but $\mathcal{D}$ contains only three vectors (polynomials), Theorem 3.2.17 guarantees that $\mathcal{D}$ is not a spanning set of $\mathbb{P}^3_{\mathbb{R}}$, and consequently, it is not a basis of $\mathbb{P}^3_{\mathbb{R}}$. However, to determine whether $\mathcal{D}$ is linearly independent, we had to row reduce. □

Example 4.4.10. *Consider the following sets of matrices (with coefficients understood to be in $\mathbb{Z}_3$):*

(a) $\mathcal{A} = \left\{ \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix} \right\}$;

(b) $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \right\}$.

For each of the two sets above, determine whether

- *it is linearly independent in* $\mathbb{Z}_3^{2\times 2}$*;*
- *it spans* $\mathbb{Z}_3^{2\times 2}$*;*
- *it is a basis of* $\mathbb{Z}_3^{2\times 2}$*.*

Solution. In what follows, we will use the basis

$$\mathcal{M} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

of $\mathbb{Z}_3^{2\times 2}$.

(a) We set

- $\mathbf{a}_1 := \left[\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \right]_{\mathcal{M}} = \begin{bmatrix} 1 & 2 & 0 & 0 \end{bmatrix}^T$;
- $\mathbf{a}_2 := \left[\begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} \right]_{\mathcal{M}} = \begin{bmatrix} 0 & 1 & 0 & 2 \end{bmatrix}^T$;
- $\mathbf{a}_3 := \left[\begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix} \right]_{\mathcal{M}} = \begin{bmatrix} 0 & 0 & 2 & 1 \end{bmatrix}^T$;
- $\mathbf{a}_4 := \left[\begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix} \right]_{\mathcal{M}} = \begin{bmatrix} 2 & 0 & 1 & 0 \end{bmatrix}^T$.

Further, we set

$$A := \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 2 & 1 & 0 \end{bmatrix}.$$

By row reducing, we get that

$$\operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and consequently, $\operatorname{rank}(A) = 3$. So, by Proposition 4.4.8, $\mathcal{A}$ is not a linearly independent set in $\mathbb{Z}_2^{2\times 2}$, is not a spanning set of $\mathbb{Z}_2^{2\times 2}$, and is not a basis of $\mathbb{Z}_2^{2\times 2}$.

(b) We set

- $\mathbf{b}_1 := \left[\left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right] \right]_{\mathcal{M}} = \left[\begin{array}{cccc} 1 & 1 & 0 & 1 \end{array} \right]^T$;
- $\mathbf{b}_2 := \left[\left[\begin{array}{cc} 1 & 1 \\ 0 & 2 \end{array} \right] \right]_{\mathcal{M}} = \left[\begin{array}{cccc} 1 & 1 & 0 & 2 \end{array} \right]^T$;
- $\mathbf{b}_3 := \left[\left[\begin{array}{cc} 2 & 1 \\ 0 & 2 \end{array} \right] \right]_{\mathcal{M}} = \left[\begin{array}{cccc} 2 & 1 & 0 & 2 \end{array} \right]^T$;
- $\mathbf{b}_4 := \left[\left[\begin{array}{cc} 2 & 0 \\ 1 & 1 \end{array} \right] \right]_{\mathcal{M}} = \left[\begin{array}{cccc} 2 & 0 & 1 & 1 \end{array} \right]^T$.

Further, we set

$$B := \left[\begin{array}{cccc} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{array} \right] = \left[\begin{array}{cccc} 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 \end{array} \right].$$

By row reducing, we get that $\mathrm{RREF}(B) = I_4$, and consequently, $\mathrm{rank}(B) = 4$. So, by Proposition 4.4.8, $\mathcal{B}$ is a basis of $\mathbb{Z}_3^{2\times 2}$, and in particular, it is a linearly independent set in $\mathbb{Z}_3^{2\times 2}$, as well as a spanning set of $\mathbb{Z}_3^{2\times 2}$. □

Finding a basis of the span of polynomials or matrices. We now present three very similar examples, each dealing with finding a basis of the span of a set of polynomials or matrices. In Example 4.4.11, we give full theoretical justification of each step (citing the appropriate theorems and propositions). Examples 4.4.12 and 4.4.13 are very similar, but we do not justify the steps in as much detail. When solving problems by yourself, you should aim for the level of detail given in Examples 4.4.12 and 4.4.13.

Example 4.4.11. *Consider the following polynomials in $\mathbb{P}_{\mathbb{Z}_2}$:*

- $p_1(x) = x^3 + x + 1$;
- $p_2(x) = x^3 + x^2 + 1$;
- $p_3(x) = x^2 + 1$;
- $p_4(x) = x + 1$;
- $p_5(x) = x^2$;
- $p_6(x) = x^3 + 1$.

*Set $U := \mathrm{Span}\big(p_1(x), \ldots, p_6(x)\big)$. Find a basis $\mathcal{B}$ of U. What is $\dim(U)$? For each $i \in \{1, \ldots, 6\}$ such that $p_i(x)$ is **not** in the basis $\mathcal{B}$, express $p_i(x)$ as a linear combination of the basis vectors in $\mathcal{B}$.*

Solution. Note that polynomials $p_1(x), \ldots, p_6(x)$ are all of degree at most 3, and they all belong to $\mathbb{P}^3_{\mathbb{Z}_2}$. Thus, $U = \mathrm{Span}\big(p_1(x), \ldots, p_6(x)\big)$ is a subspace of $\mathbb{P}^3_{\mathbb{Z}_2}$. We

know that

$$\mathcal{A} \;=\; \{1, x, x^2, x^3\}$$

is a basis of $\mathbb{P}^3_{\mathbb{Z}_2}$ and (by Proposition 4.3.1) that $\left[\ \cdot\ \right]_{\mathcal{A}} : \mathbb{P}^3_{\mathbb{Z}_2} \to \mathbb{Z}_2^4$ is an isomorphism. Next, by Theorem 4.2.11, the image of U under $\left[\ \cdot\ \right]_{\mathcal{A}}$ is precisely $V := \mathrm{Span}\Big(\left[\ p_1(x)\ \right]_{\mathcal{A}}, \ldots, \left[\ p_6(x)\ \right]_{\mathcal{A}}\Big)$, and moreover, Proposition 4.4.7 guarantees that when we restrict the domain of $\left[\ \cdot\ \right]_{\mathcal{A}}$ to U and the codomain to V,[25] we obtain an isomorphism. So, we first solve the problem for the coordinate vectors $\left[\ p_1(x)\ \right]_{\mathcal{A}}, \ldots, \left[\ p_6(x)\ \right]_{\mathcal{A}}$ and the subspace $V = \mathrm{Span}\Big(\left[\ p_1(x)\ \right]_{\mathcal{A}}, \ldots, \left[\ p_6(x)\ \right]_{\mathcal{A}}\Big)$ of V, and then using the fact that $\left[\ \cdot\ \right]_{\mathcal{A}}$ is an isomorphism, we "translate" the solution back to $p_1(x), \ldots, p_6(x)$ and $U = \mathrm{Span}\big(p_1(x), \ldots, p_6(x)\big)$.

We first read off the coordinate vectors of our six polynomials with respect to the basis $\mathcal{A}$:

- $\left[\ p_1(x)\ \right]_{\mathcal{A}} = \begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix}^T$;
- $\left[\ p_2(x)\ \right]_{\mathcal{A}} = \begin{bmatrix} 1 & 0 & 1 & 1 \end{bmatrix}^T$;
- $\left[\ p_3(x)\ \right]_{\mathcal{A}} = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}^T$;
- $\left[\ p_4(x)\ \right]_{\mathcal{A}} = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^T$;
- $\left[\ p_5(x)\ \right]_{\mathcal{A}} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^T$;
- $\left[\ p_6(x)\ \right]_{\mathcal{A}} = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}^T$.

We now form the matrix

$$A \;=\; \begin{bmatrix} \left[\ p_1(x)\ \right]_{\mathcal{A}} & \cdots & \left[\ p_6(x)\ \right]_{\mathcal{A}} \end{bmatrix} \;=\; \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and by row reducing, we obtain the following (pivot columns are in red, and non-pivot columns are in blue):

$$\mathrm{RREF}(A) \;=\; \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Since the pivot columns of A are its first, second, third, and fifth column, Theorem 3.3.4 guarantees that

$$\mathcal{C} \;:=\; \Big\{\left[\ p_1(x)\ \right]_{\mathcal{A}}, \left[\ p_2(x)\ \right]_{\mathcal{A}}, \left[\ p_3(x)\ \right]_{\mathcal{A}}, \left[\ p_5(x)\ \right]_{\mathcal{A}}\Big\}$$

is a basis of $\mathrm{Col}(A) = \mathrm{Span}\Big(\left[\ p_1(x)\ \right]_{\mathcal{A}}, \left[\ p_1(x)\ \right]_{\mathcal{A}}, \ldots, \left[\ p_6(x)\ \right]_{\mathcal{A}}\Big) = V$. As we pointed out above, the function obtained from $\left[\ \cdot\ \right]_{\mathcal{A}} : \mathbb{P}_{\mathbb{Z}_2} \to \mathbb{Z}_2^4$ by restricting the domain to $U = \mathrm{Span}\big(p_1(x), \ldots, p_6(x)\big)$ and the codomain to $V =$

[25] This is well defined because the image of U under $\left[\ \cdot\ \right]_{\mathcal{A}}$ is precisely V.

$\text{Span}\Big(\left[\ p_1(x)\ \right]_{\mathcal{A}}, \ldots, \left[\ p_6(x)\ \right]_{\mathcal{A}} \Big)$ is an isomorphism. So, since $\mathcal{C}$ is a basis of V, Theorem 4.4.4(c) guarantees that

$$\mathcal{B} \ := \ \Big\{ p_1(x), p_2(x), p_3(x), p_5(x) \Big\}$$

is a basis of U. Since U has a four-vector basis, we see that $\dim(U) = 4$.

It remains to express $p_4(x)$ and $p_6(x)$ as a linear combination of the polynomials in the basis $\mathcal{B}$. First, we see from the matrix $\text{RREF}(A)$ that the following hold:

- $\left[\ p_4(x)\ \right]_{\mathcal{A}} = \left[\ p_1(x)\ \right]_{\mathcal{A}} + \left[\ p_2(x)\ \right]_{\mathcal{A}} + \left[\ p_3(x)\ \right]_{\mathcal{A}}$;
- $\left[\ p_6(x)\ \right]_{\mathcal{A}} = \left[\ p_2(x)\ \right]_{\mathcal{A}} + \left[\ p_5(x)\ \right]_{\mathcal{A}}$.

But now

$$\begin{aligned} \left[\ p_4(x)\ \right]_{\mathcal{A}} &= \left[\ p_1(x)\ \right]_{\mathcal{A}} + \left[\ p_2(x)\ \right]_{\mathcal{A}} + \left[\ p_3(x)\ \right]_{\mathcal{A}} \\ &\overset{(*)}{=} \left[\ p_1(x) + p_2(x) + p_3(x)\ \right]_{\mathcal{A}} \end{aligned}$$

and

$$\left[\ p_6(x)\ \right]_{\mathcal{A}} = \left[\ p_2(x)\ \right]_{\mathcal{A}} + \left[\ p_5(x)\ \right]_{\mathcal{A}} \overset{(*)}{=} \left[\ p_2(x) + p_5(x)\ \right]_{\mathcal{A}},$$

where both instances of (*) follow from the fact that $\left[\ \cdot\ \right]_{\mathcal{A}}$ is linear (because it is an isomorphism). But $\left[\ \cdot\ \right]_{\mathcal{A}}$ is also one-to-one (again, because it is an isomorphism); it follows that

- $p_4(x) = p_1(x) + p_2(x) + p_3(x)$,
- $p_6(x) = p_2(x) + p_5(x)$,

and we are done.

Optional: It is not a bad idea to check whether our expressions for $p_4(x)$ and $p_6(x)$ are correct (to make sure we did not miscompute). So, we compute:

$$\begin{aligned} p_1(x) + p_2(x) + p_3(x) &= (x^3 + x + 1) + (x^3 + x^2 + 1) + (x^2 + 1) \\ &= (x^3 + x^3) + (x^2 + x^2) + x + (1 + 1 + 1) \\ &= x + 1 \\ &= p_4(x) \end{aligned}$$

and

$$\begin{aligned} p_2(x) + p_5(x) &= (x^3 + x^2 + 1) + x^2 \\ &= x^3 + (x^2 + x^2) + 1 \\ &= x^3 + 1 \\ &= p_6(x). \end{aligned}$$

As we can see, our expressions for $p_4(x)$ and $p_6(x)$ are correct. □

Example 4.4.12. *Consider the following polynomials in $\mathbb{P}_{\mathbb{Z}_3}$:*

- $p_1(x) = x^4 + 2$;
- $p_2(x) = x^3 + x^2$;
- $p_3(x) = x^4 + x^3 + x^2 + 2$;
- $p_4(x) = 2x^4 + x^3 + x^2 + 1$;
- $p_5(x) = 2x + 1$.

*Set $U := Span\big(p_1(x), \ldots, p_5(x)\big)$. Find a basis $\mathcal{B}$ of U. What is dim(U)? For each $i \in \{1, \ldots, 5\}$ such that $p_i(x)$ is **not** in the basis $\mathcal{B}$, express $p_i(x)$ as a linear combination of the basis vectors in $\mathcal{B}$.*

Solution. Note that polynomials $p_1(x), \ldots, p_5(x)$ are all of degree at most 4, and they all belong to $\mathbb{P}^4_{\mathbb{Z}_3}$. Thus, $U = \text{Span}\big(p_1(x), \ldots, p_5(x)\big)$ is a subspace of $\mathbb{P}^4_{\mathbb{Z}_3}$. We know that

$$\mathcal{A} = \{1, x, x^2, x^3, x^4\}$$

is a basis of $\mathbb{P}^4_{\mathbb{Z}_3}$. The coordinate vectors of $p_1(x), \ldots, p_5(x)$ with respect to the basis $\mathcal{A}$ are as follows:

- $\big[\ p_1(x)\ \big]_{\mathcal{A}} = \begin{bmatrix} 2 & 0 & 0 & 0 & 1 \end{bmatrix}^T$;
- $\big[\ p_2(x)\ \big]_{\mathcal{A}} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \end{bmatrix}^T$;
- $\big[\ p_3(x)\ \big]_{\mathcal{A}} = \begin{bmatrix} 2 & 0 & 1 & 1 & 1 \end{bmatrix}^T$;
- $\big[\ p_4(x)\ \big]_{\mathcal{A}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 2 \end{bmatrix}^T$;
- $\big[\ p_5(x)\ \big]_{\mathcal{A}} = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \end{bmatrix}^T$.

We form the matrix

$$A = \Big[\ \big[\ p_1(x)\ \big]_{\mathcal{A}} \ \ \ldots \ \ \big[\ p_5(x)\ \big]_{\mathcal{A}}\ \Big] = \begin{bmatrix} 2 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 2 & 0 \end{bmatrix},$$

and by row reducing, we obtain the following (pivot columns are in red, and non-pivot columns are in blue):

$$\text{RREF}(A) \quad = \quad \begin{bmatrix} 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We see that the pivot columns of A are its first, second, and fifth column. Therefore,

$$\mathcal{C} \quad := \quad \Big\{ \big[\ p_1(x)\ \big]_{\mathcal{A}}, \big[\ p_2(x)\ \big]_{\mathcal{A}}, \big[\ p_5(x)\ \big]_{\mathcal{A}} \Big\}.$$

is a basis of $\text{Col}(A) = \text{Span}\Big(\big[\ p_1(x)\ \big]_{\mathcal{A}}, \ldots, \big[\ p_5(x)\ \big]_{\mathcal{A}} \Big)$. Consequently,

$$\mathcal{B} \quad = \quad \Big\{ p_1(x), p_2(x), p_5(x) \Big\}$$

is a basis of $U = \text{Span}\big(p_1(x), \ldots, p_5(x)\big)$, and it follows that $\dim(U) = 3$.

It remains to express $p_3(x)$ and $p_4(x)$ as a linear combination of the vectors (polynomials) in $\mathcal{B}$. First, we have that

- $\big[\ p_3(x)\ \big]_{\mathcal{A}} \overset{(*)}{=} \big[\ p_1(x)\ \big]_{\mathcal{A}} + \big[\ p_2(x)\ \big]_{\mathcal{A}} \overset{(**)}{=} \big[\ p_1(x) + p_2(x)\ \big]_{\mathcal{A}}$,
- $\big[\ p_4(x)\ \big]_{\mathcal{A}} \overset{(*)}{=} 2\big[\ p_1(x)\ \big]_{\mathcal{A}} + \big[\ p_2(x)\ \big]_{\mathcal{A}} \overset{(**)}{=} \big[\ 2p_1(x) + p_2(x)\ \big]_{\mathcal{A}}$,

where both instances of (*) were obtained from the matrix $\text{RREF}(A)$, and both instances of (**) follow from the fact that $\big[\ \cdot\ \big]_{\mathcal{A}} : \mathbb{P}^4_{\mathbb{Z}_3} \to \mathbb{Z}^5_3$ is linear (because it is an isomorphism). Since $\big[\ \cdot\ \big]_{\mathcal{A}}$ is also one-to-one (again, because it is an isomorphism), we get that

- $p_3(x) = p_1(x) + p_2(x)$,
- $p_4(x) = 2p_1(x) + p_2(x)$,

and we are done.

Optional: Let us check that our expressions for $p_3(x)$ and $p_4(x)$ are correct. We compute:

$$p_1(x) + p_2(x) \quad = \quad (x^4 + 2) + (x^3 + x^2) \quad = \quad x^4 + x^3 + x^2 + 2 \quad = \quad p_3(x)$$

and

$$2p_1(x) + p_2(x) \quad = \quad 2(x^4 + 2) + (x^3 + x^2) \quad = \quad 2x^4 + x^3 + x^2 + 1 \quad = \quad p_4(x).$$

As we can see, our expressions for $p_3(x)$ and $p_4(x)$ are correct. □

Example 4.4.13. *Consider the following matrices in* $\mathbb{R}^{2\times 2}$*:*

- $M_1 = \begin{bmatrix} -2 & 1 \\ 3 & -2 \end{bmatrix}$;
- $M_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$;
- $M_3 = \begin{bmatrix} 5 & -2 \\ -6 & 6 \end{bmatrix}$;
- $M_4 = \begin{bmatrix} -3 & 0 \\ 0 & -6 \end{bmatrix}$;
- $M_5 = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix}$;
- $M_6 = \begin{bmatrix} 7 & -2 \\ -9 & 10 \end{bmatrix}$.

Set $U := Span\Big(M_1(x), \dots, M_6(x)\Big)$. *Find a basis* $\mathcal{B}$ *for* U. *What is* $dim(U)$*? For each* $i \in \{1, \dots, 6\}$ *such that* $M_i(x)$ *is* ***not*** *in the basis* $\mathcal{B}$, *express* $M_i(x)$ *as a linear combination of the basis vectors in* $\mathcal{B}$.

Solution. Consider the basis

$$\mathcal{A} := \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

of $\mathbb{R}^{2\times 2}$. We form the coordinate vectors

- $[\ M_1\ ]_{\mathcal{A}} = \begin{bmatrix} -2 \\ 1 \\ 3 \\ -2 \end{bmatrix}$;
- $[\ M_2\ ]_{\mathcal{A}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}$;
- $[\ M_3\ ]_{\mathcal{A}} = \begin{bmatrix} 5 \\ -2 \\ -6 \\ 6 \end{bmatrix}$;
- $[\ M_4\ ]_{\mathcal{A}} = \begin{bmatrix} -3 \\ 0 \\ 0 \\ -6 \end{bmatrix}$;
- $[\ M_5\ ]_{\mathcal{A}} = \begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \end{bmatrix}$;
- $[\ M_6\ ]_{\mathcal{A}} = \begin{bmatrix} 7 \\ -2 \\ -9 \\ 10 \end{bmatrix}$.

We now form the matrix

$$A := \begin{bmatrix} [\ M_1\ ]_{\mathcal{A}} & \dots & [\ M_6\ ]_{\mathcal{A}} \end{bmatrix} = \begin{bmatrix} -2 & 1 & 5 & -3 & 0 & 7 \\ 1 & 0 & -2 & 0 & 0 & -2 \\ 3 & 0 & -6 & 0 & 3 & -9 \\ -2 & 2 & 6 & -6 & 0 & 10 \end{bmatrix},$$

and by row reducing, we obtain the following (pivot columns are in red, and non-pivot columns are in blue):

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & 0 & -2 & 0 & 0 & -2 \\ 0 & 1 & 1 & -3 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We see that the pivot columns of A are its first, second, and fifth column. Therefore,

$$\mathcal{C} := \Big\{ [\ M_1\ ]_{\mathcal{A}}, [\ M_2\ ]_{\mathcal{A}}, [\ M_5\ ]_{\mathcal{A}} \Big\}$$

is a basis of $\mathrm{Col}(A) = \mathrm{Span}\Big(\left[\ M_1\ \right]_{\mathcal{A}}, \ldots, \left[\ M_6\ \right]_{\mathcal{A}} \Big)$. Consequently,

$$\mathcal{B} \ := \ \Big\{ M_1, M_2, M_5 \Big\}$$

is a basis of $U = \mathrm{Span}\Big(M_1, \ldots, M_6 \Big)$, and it follows that $\dim(U) = 3$.

It remains to express M_3, M_4, M_6 as a linear combination of the vectors (matrices) in $\mathcal{C}$. First, we have that

- $\left[\ M_3\ \right]_{\mathcal{A}} \overset{(*)}{=} -2 \left[\ M_1\ \right]_{\mathcal{A}} + \left[\ M_2\ \right]_{\mathcal{A}} \overset{(**)}{=} \left[\ -2M_1 + M_2\ \right]_{\mathcal{A}}$,
- $\left[\ M_4\ \right]_{\mathcal{A}} \overset{(*)}{=} -3 \left[\ M_2\ \right]_{\mathcal{A}} \overset{(**)}{=} \left[\ -3M_2\ \right]_{\mathcal{A}}$,
- $\left[\ M_6\ \right]_{\mathcal{A}} \overset{(*)}{=} -2 \left[\ M_1\ \right]_{\mathcal{A}} + 3 \left[\ M_2\ \right]_{\mathcal{A}} - \left[\ M_5\ \right]_{\mathcal{A}} \overset{(**)}{=} \left[\ -2M_1 + 3M_2 - M_5\ \right]_{\mathcal{A}}$,

where all three instances of (*) can be read off from $\mathrm{RREF}(A)$, and all three instances of (**) follow from the fact that $\left[\ \cdot\ \right]_{\mathcal{A}} : \mathbb{R}^{2\times 2} \to \mathbb{R}^4$ is linear (because it is an isomorphism). But since $\left[\ \cdot\ \right]_{\mathcal{A}} : \mathbb{R}^{2\times 2} \to \mathbb{R}^4$ is also one-to-one (again, because it is an isomorphism), we now get that

- $M_3 = -2M_1 + M_2$,
- $M_4 = -3M_2$,
- $M_6 = -2M_1 + 3M_2 - M_5$,

and we are done.

Optional: Let us check that our expressions for M_3, M_4, M_6 are correct.

First, for M_3, we compute:

$$-2M_1 + M_2 \ = \ -2 \begin{bmatrix} -2 & 1 \\ 3 & -2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \ = \ \begin{bmatrix} 5 & -2 \\ -6 & 6 \end{bmatrix} \ = \ M_3,$$

and we see that our expression for M_3 is correct.

Next, for M_4, we compute:

$$-3M_2 \ = \ -3 \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \ = \ \begin{bmatrix} -3 & 0 \\ 0 & -6 \end{bmatrix} \ = \ M_4,$$

and we see that our expression for M_4 is correct.

Finally, for M_6, we compute:

$$\begin{aligned} -2M_1 + 3M_2 - M_5 \ &= \ -2 \begin{bmatrix} -2 & 1 \\ 3 & -2 \end{bmatrix} + 3 \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix} \\ &= \ \begin{bmatrix} 7 & -2 \\ -9 & 10 \end{bmatrix} \\ &= \ M_6, \end{aligned}$$

and we see that our expression for M_6 is correct. □

Extending a basis of a subspace to a basis of the whole vector space. Our next example is similar to Example 3.3.22 from subsection 3.3.4, only we have polynomials instead of column vectors. We will use coordinate vectors to transform polynomials into column vectors, and we will rely on Proposition 3.3.21.

Example 4.4.14. *Consider the following polynomials in* $\mathbb{P}^3_{\mathbb{Z}_3}$*:*

- $p_1(x) = x^3 + 1$*;*
- $p_2(x) = 2x^3 + 2$*;*
- $p_3(x) = x^2 + 2x + 1$
- $p_4(x) = 2x^3 + x^2 + 2x$*.*

Find a basis $\mathcal{B}_U$ *of* $U := Span\big(p_1(x), p_2(x), p_3(x), p_4(x)\big)$*, extend it to a basis* $\mathcal{B}$ *of* $\mathbb{P}^3_{\mathbb{Z}_3}$*, and for each* $i \in \{1, 2, 3, 4\}$ *such that* $p_i(x) \notin \mathcal{B}$*, express* $p_i(x)$ *as a linear combination of the basis vectors in* $\mathcal{B}$*.*

Solution. We know that

$$\mathcal{A} \;:=\; \{1, x, x^2, x^3\}$$

is a basis of $\mathbb{P}^3_{\mathbb{Z}_3}$, and we let V be the image of U under the isomorphism $\big[\ \cdot\ \big]_{\mathcal{A}}$. Further, we consider the standard basis

$$\mathcal{E}_4 \;=\; \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\} \;=\; \Big\{ \big[\ 1\ \big]_{\mathcal{A}}, \big[\ x\ \big]_{\mathcal{A}}, \big[\ x^2\ \big]_{\mathcal{A}}, \big[\ x^3\ \big]_{\mathcal{A}} \Big\}$$

of $\mathbb{Z}_3^4$. We now form the 4×8 matrix C whose columns are the coordinate vectors of the polynomials

$$p_1(x), \quad p_2(x), \quad p_3(x), \quad p_4(x), \quad 1, \quad x, \quad x^2, \quad x^3$$

with respect to the basis $\mathcal{A}$. Here is the matrix C explicitly (with tiny font so that it would fit on the page):

$$C \;:=\; \Big[\ \big[\ p_1(x)\ \big]_{\mathcal{A}} \quad \big[\ p_2(x)\ \big]_{\mathcal{A}} \quad \big[\ p_3(x)\ \big]_{\mathcal{A}} \quad \big[\ p_4(x)\ \big]_{\mathcal{A}} \;\vdots\; \big[\ 1\ \big]_{\mathcal{A}} \quad \big[\ x\ \big]_{\mathcal{A}} \quad \big[\ x^2\ \big]_{\mathcal{A}} \quad \big[\ x^3\ \big]_{\mathcal{A}}\ \Big].$$

We then have that

$$C \;=\; \left[\begin{array}{cccc:cccc} 1 & 2 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right].$$

By row reducing, we obtain

$$\mathrm{RREF}(C) \;=\; \left[\begin{array}{cccc:cccc} 1 & 2 & 0 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right],$$

and we see that the pivot columns of C are its first, third, fifth, and sixth column. By Proposition 3.3.21, the pivot columns of C to the left of the vertical dotted line form a basis of V, and all the pivot columns of C together form a basis of $\mathbb{Z}_3^4$. So,

$$\left\{ \left[\ p_1(x)\ \right]_{\mathcal{A}}, \left[\ p_3(x)\ \right]_{\mathcal{A}} \right\}$$

is a basis of V, whereas

$$\left\{ \left[\ p_1(x)\ \right]_{\mathcal{A}}, \left[\ p_3(x)\ \right]_{\mathcal{A}}, \left[\ 1\ \right]_{\mathcal{A}}, \left[\ x\ \right]_{\mathcal{A}} \right\}$$

is a basis of $\mathbb{Z}_3^4$ that extends our basis of V. Since $\left[\ \cdot\ \right]_{\mathcal{A}}$ is an isomorphism, we see that

$$\mathcal{B}_U \ := \ \Big\{ p_1(x), p_3(x) \Big\}$$

is a basis of U, and that

$$\mathcal{B} \ := \ \Big\{ p_1(x), p_3(x), 1, x \Big\}$$

is a basis of $\mathbb{P}^3_{\mathbb{Z}_3}$ that extends our basis $\mathcal{B}_U$ of U. Finally, we can read off from $\mathrm{RREF}(C)$ that

- $p_2(x) = 2p_1(x)$,
- $p_4(x) = 2p_1(x) + p_3(x)$,

and we are done. □

Example 4.4.15. *Consider the following matrices in $\mathbb{Z}_3^{2\times 2}$:*

$$M_1 = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \qquad M_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad M_3 = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}, \qquad M_4 = \begin{bmatrix} 2 & 0 \\ 2 & 2 \end{bmatrix}.$$

Find a basis $\mathcal{B}_U$ of $U := Span(M_1, M_2, M_3, M_4)$, extend it to a basis $\mathcal{B}$ of $\mathbb{Z}_3^{2\times 2}$, and for each $i \in \{1, 2, 3, 4\}$ such that $M_i \notin \mathcal{B}$, express M_i as a linear combination of the basis vectors in $\mathcal{B}$.

Solution. We proceed similarly as in Example 4.4.14. Set

$$A_1 := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad A_2 := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad A_3 := \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \qquad A_4 := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

We know that

$$\mathcal{A} \ := \ \{A_1, A_2, A_3, A_4\}$$

is a basis of $\mathbb{Z}_3^{2\times 2}$, and we let V be the image of U under the isomorphism $\left[\ \cdot\ \right]_{\mathcal{A}}$. Further, we consider the standard basis

$$\mathcal{E}_4 \quad = \quad \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\} \quad = \quad \left\{ \left[\ A_1\ \right]_{\mathcal{A}}, \left[\ A_2\ \right]_{\mathcal{A}}, \left[\ A_3\ \right]_{\mathcal{A}}, \left[\ A_4\ \right]_{\mathcal{A}} \right\}$$

of $\mathbb{Z}_3^4$. We now form the 4×8 matrix C whose columns are the coordinate vectors of the matrices

$$M_1, \quad M_2, \quad M_3, \quad M_4, \quad A_1, \quad A_2, \quad A_3, \quad A_4$$

with respect to the basis $\mathcal{A}$. Here is the matrix C explicitly (as in Example 4.4.14, the font is tiny so that the matrix would fit on the page):

$$C \quad := \quad \left[\ \left[\ M_1\ \right]_{\mathcal{A}} \quad \left[\ M_2\ \right]_{\mathcal{A}} \quad \left[\ M_3\ \right]_{\mathcal{A}} \quad \left[\ M_4\ \right]_{\mathcal{A}} \ \vdots\ \left[\ A_1\ \right]_{\mathcal{A}} \quad \left[\ A_2\ \right]_{\mathcal{A}} \quad \left[\ A_3\ \right]_{\mathcal{A}} \quad \left[\ A_4\ \right]_{\mathcal{A}}\ \right].$$

We then have that

$$C \quad = \quad \left[\begin{array}{cccc:cccc} 2 & 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 2 & 0 & 0 & 0 & 1 \end{array}\right]$$

By row reducing, we obtain

$$\mathrm{RREF}(C) \quad = \quad \left[\begin{array}{cccc:cccc} 1 & 2 & 0 & 2 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right],$$

and we see that the pivot columns of C are its first, third, sixth, and seventh column. By Proposition 3.3.21, the pivot columns of C to the left of the vertical dotted line form a basis of V, and all the pivot columns of C together form a basis of $\mathbb{Z}_4$. So,

$$\left\{ \left[\ M_1\ \right]_{\mathcal{A}}, \left[\ M_3\ \right]_{\mathcal{A}} \right\}$$

is a basis of V, whereas

$$\left\{ \left[\ M_1\ \right]_{\mathcal{A}}, \left[\ M_3\ \right]_{\mathcal{A}}, \left[\ A_2\ \right]_{\mathcal{A}}, \left[\ A_3\ \right]_{\mathcal{A}} \right\}$$

is a basis of $\mathbb{Z}_3^4$ that extends our basis of V. Since $\left[\ \cdot\ \right]_{\mathcal{A}}$ is an isomorphism, we see that

$$\mathcal{B}_U \quad := \quad \{M_1, M_3\}$$

is a basis of U, and that

$$\mathcal{B} \quad := \quad \{M_1, M_3, A_2, A_3\}$$

is a basis of $\mathbb{Z}_3^{2\times 2}$ that extends our basis $\mathcal{B}_U$ of U. Finally, we can read off from $\mathrm{RREF}(C)$ that

- $M_2 = 2M_1$,
- $M_4 = 2M_1 + M_3$,

and we are done. □

4.5 Matrices of linear functions between non-trivial, finite-dimensional vector spaces

As we have already mentioned, linear functions between general vector spaces do not have standard matrices. However, as our next theorem shows, we can associate a matrix to a linear function between non-trivial,[26] finite-dimensional vector spaces, provided we have first specified a basis of the domain and a basis of the codomain.

Theorem 4.5.1. *Let U and V be non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis of U, let $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be a basis of V, and let $f : U \to V$ be a linear function. Then exists a unique matrix in $\mathbb{F}^{n\times m}$, denoted by ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ and called the* matrix of f with respect to $\mathcal{B}$ and $\mathcal{C}$, *such that for all $\mathbf{u} \in U$, we have that*

$$ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\ [\ \mathbf{u}\ ]_{\mathcal{B}} \ = \ [\ f(\mathbf{u})\ ]_{\mathcal{C}}\,. $$

Moreover, the matrix ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is given by

$$ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}} \ = \ [\ [\ f(\mathbf{b}_1)\ ]_{\mathcal{C}} \ \ \ldots \ \ [\ f(\mathbf{b}_m)\ ]_{\mathcal{C}}\ ]\,. $$

Proof. **Existence.** Fix $\mathbf{u} \in U$. We must show that

$$ [\ [\ f(\mathbf{b}_1)\ ]_{\mathcal{C}} \ \ \ldots \ \ [\ f(\mathbf{b}_m)\ ]_{\mathcal{C}}\ ]\ [\ \mathbf{u}\ ]_{\mathcal{B}} \ = \ [\ f(\mathbf{u})\ ]_{\mathcal{C}}\,. $$

Set $[\ \mathbf{u}\ ]_{\mathcal{B}} = [\ \beta_1 \ \ \ldots \ \ \beta_m\ ]^T$, so that $\mathbf{u} = \beta_1\mathbf{b}_1 + \cdots + \beta_m\mathbf{b}_m$. We then compute:

$$
\begin{aligned}
& [\ [\ f(\mathbf{b}_1)\ ]_{\mathcal{C}} \ \ \ldots \ \ [\ f(\mathbf{b}_m)\ ]_{\mathcal{C}}\ ]\ [\ \mathbf{u}\ ]_{\mathcal{B}} \\
= \ & [\ [\ f(\mathbf{b}_1)\ ]_{\mathcal{C}} \ \ \ldots \ \ [\ f(\mathbf{b}_m)\ ]_{\mathcal{C}}\ ] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \\
= \ & \beta_1\,[\ f(\mathbf{b}_1)\ ]_{\mathcal{C}} + \cdots + \beta_m\,[\ f(\mathbf{b}_m)\ ]_{\mathcal{C}} \\
\overset{(*)}{=} \ & [\ \beta_1 f(\mathbf{b}_1) + \cdots + \beta_m f(\mathbf{b}_m)\ ]_{\mathcal{C}} \\
\overset{(**)}{=} \ & [\ f\big(\beta_1\mathbf{b}_1 + \cdots + \beta_m\mathbf{b}_m\big)\ ]_{\mathcal{C}} \\
= \ & [\ f(\mathbf{u})\ ]_{\mathcal{C}}\,,
\end{aligned}
$$

[26] Recall that a vector space V over a field $\mathbb{F}$ is *non-trivial* if it contains at least one non-zero vector, or equivalently, if $\dim(V) > 0$.

where (*) follows from the fact that $\left[\ \cdot\ \right]_{\mathcal{C}} : V \to \mathbb{F}^n$ is an isomorphism (and in particular, a linear function), and (**) follows from the fact that f is linear.[27]

Uniqueness. Fix any matrix $A = \left[\ \mathbf{a}_1 \ \ \ldots \ \ \mathbf{a}_m\ \right]$ in $\mathbb{F}^{n\times m}$ that has the property that for all $\mathbf{u} \in U$, we have that $A\left[\ \mathbf{u}\ \right]_{\mathcal{B}} = \left[\ f(\mathbf{u})\ \right]_{\mathcal{C}}$. We must show that $A = \left[\ \left[\ f(\mathbf{b}_1)\ \right]_{\mathcal{C}} \ \ \ldots \ \ \left[\ f(\mathbf{b}_m)\ \right]_{\mathcal{C}}\ \right]$. We prove this by showing that the two matrices have the same corresponding columns, that is, that $\mathbf{a}_i = \left[\ f(\mathbf{b}_i)\ \right]_{\mathcal{C}}$ for all indices $i \in \{1, \ldots, m\}$. Indeed, for all $i \in \{1, \ldots, m\}$, we have the following:

$$\begin{aligned}
\mathbf{a}_i &= A\mathbf{e}_i^m && \text{by Proposition 1.4.4}\\
&= A\left[\ \mathbf{b}_i\ \right]_{\mathcal{B}} && \text{because } \left[\ \mathbf{b}_i\ \right]_{\mathcal{B}} = \mathbf{e}_i^m \text{ (by Proposition 3.2.9)}\\
&= \left[\ f(\mathbf{b}_i)\ \right]_{\mathcal{C}} && \text{by the choice of } A.
\end{aligned}$$

This proves that $A = \left[\ \left[\ f(\mathbf{b}_1)\ \right]_{\mathcal{C}} \ \ \ldots \ \ \left[\ f(\mathbf{b}_m)\ \right]_{\mathcal{C}}\ \right]$, and we are done. □

Remark: Suppose that U and V are non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$, that $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ are bases of U and V, respectively, and that $f : U \to V$ is a linear function, as in Theorem 4.5.1. Then the uniqueness part of Theorem 4.5.1 guarantees that if $A \in \mathbb{F}^{n\times m}$ is **any** matrix that satisfies the property that for all $\mathbf{u} \in U$, we have that

$$A\left[\ \mathbf{u}\ \right]_{\mathcal{B}} = \left[\ f(\mathbf{u})\ \right]_{\mathcal{C}},$$

then we in fact have that $A = {}_{\mathcal{C}}\left[\ f\ \right]_{\mathcal{B}}$. We will use this observation repeatedly (see the proofs of Theorem 4.5.3, Theorem 4.5.4, and Lemma 4.5.8).

Remark: Note that matrices of the form ${}_{\mathcal{C}}\left[\ f\ \right]_{\mathcal{B}}$ are generalizations of standard matrices. Indeed, if $\mathbb{F}$ is a field and $f : \mathbb{F}^m \to \mathbb{F}^n$ is a linear function, then the matrix ${}_{\mathcal{E}_n}\left[\ f\ \right]_{\mathcal{E}_m}$ is precisely the standard matrix of f. (As usual, $\mathcal{E}_m = \{\mathbf{e}_1^m, \ldots, \mathbf{e}_m^m\}$ is the standard basis of $\mathbb{F}^m$, and $\mathcal{E}_n = \{\mathbf{e}_1^n, \ldots, \mathbf{e}_n^n\}$ is the standard basis of $\mathbb{F}^n$.)[28]

[27] Technically, both (*) and (**) also rely on Proposition 4.1.5.

[28] This is "obvious," but here are the details. Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear function, and let A be the standard matrix of f. Then for all $\mathbf{u} \in \mathbb{F}^m$, we have the following:

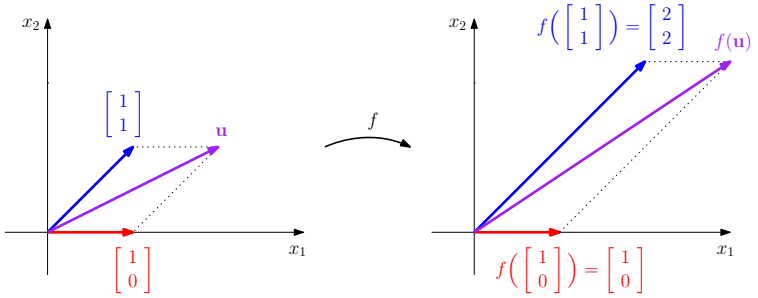
$$A\left[\ \mathbf{u}\ \right]_{\mathcal{E}_m} \overset{(*)}{=} A\mathbf{u} \overset{(**)}{=} f(\mathbf{u}) \overset{(*)}{=} \left[\ f(\mathbf{u})\ \right]_{\mathcal{E}_n},$$

where both instances of (*) follow from Example 3.2.8(a) (or alternatively, from Proposition 3.2.9), and (**) follows from the fact that A is the standard matrix of f. But now the uniqueness part of Theorem 4.5.1 guarantees that $A = {}_{\mathcal{E}_n}\left[\ f\ \right]_{\mathcal{E}_m}$, i.e. ${}_{\mathcal{E}_n}\left[\ f\ \right]_{\mathcal{E}_m}$ is the standard matrix of f.

Example 4.5.2. *Consider the basis* $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ *of* $\mathbb{R}^2$*, and consider the unique linear function* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *that satisfies the following:*

- $f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$,
- $f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Compute the matrix ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.



Remark: The fact that $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is a basis of $\mathbb{R}^2$ follows from the fact that $\operatorname{rank}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) = 2$ and from the Invertible Matrix Theorem (see subsection 3.3.6). The existence and uniqueness of the linear function f follows from Theorem 4.3.2.

Solution. Using the formula from Theorem 4.5.1, we compute:

$$\begin{aligned} {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}} &= \left[\left[f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \right]_{\mathcal{B}} \quad \left[f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \right]_{\mathcal{B}} \right] \\ &= \left[\left[\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right]_{\mathcal{B}} \quad \left[\begin{bmatrix} 2 \\ 2 \end{bmatrix} \right]_{\mathcal{B}} \right] \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}. \end{aligned}$$

□

By Proposition 4.1.7, sums, scalar multiples, and compositions of linear functions are linear (as long as the domains and codomains are compatible). Theorem 4.5.3 (below) describes the matrices of sums, scalar multiples, and compositions or linear functions, and it is a generalization of Proposition 1.10.13.

Theorem 4.5.3. *Let U, V, and W be non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis of U, let $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be a basis of V, and let $\mathcal{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_p\}$ be a basis of W. Then all the following hold:*

(a) for all linear functions $f, g : U \to V$, the function $f+g$ is linear,[29] *and moreover,*

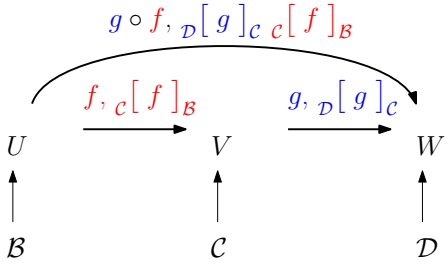
$$ {}_{\mathcal{C}}[\ f+g\ ]_{\mathcal{B}} \ = \ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}} + {}_{\mathcal{C}}[\ g\ ]_{\mathcal{B}}; $$

(b) for all linear functions $f : U \to V$ and scalars $\alpha \in U$, the function αf is linear,[30] *and moreover,*

$$ {}_{\mathcal{C}}[\ \alpha f\ ]_{\mathcal{B}} \ = \ \alpha\, {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}; $$

(c) for all linear functions $f : U \to V$ and $g : V \to W$, the function $g \circ f$ is linear,[31] *and moreover,*

$$ {}_{\mathcal{D}}[\ g \circ f\ ]_{\mathcal{B}} \ = \ {}_{\mathcal{D}}[\ g\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}. $$



Proof. We prove (c). The proofs of (a) and (b) are left as an exercise. The fact that $g \circ f$ is linear follows from Proposition 4.1.7(c). It remains to show that ${}_{\mathcal{D}}[\ g \circ f\ ]_{\mathcal{B}} = {}_{\mathcal{D}}[\ g\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$.

Claim. For all $\mathbf{u} \in U$, we have that

$$ \Big({}_{\mathcal{D}}[\ g\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)\ [\ \mathbf{u}\ ]_{\mathcal{B}} \ = \ [\ (g \circ f)(\mathbf{u})\ ]_{\mathcal{D}}. $$

Proof of the Claim. For all $\mathbf{u} \in U$, we have the following:

$$ \begin{aligned} \Big({}_{\mathcal{D}}[\ g\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)\ [\ \mathbf{u}\ ]_{\mathcal{B}} \ &= \ {}_{\mathcal{D}}[\ g\ ]_{\mathcal{C}}\ \Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\ [\ \mathbf{u}\ ]_{\mathcal{B}}\Big) \\ &= \ {}_{\mathcal{D}}[\ g\ ]_{\mathcal{C}}\ \ [\ f(\mathbf{u})\ ]_{\mathcal{C}} \\ &= \ [\ g\big(f(\mathbf{u})\big)\ ]_{\mathcal{D}} \\ &= \ [\ (g \circ f)(\mathbf{u})\ ]_{\mathcal{D}}. \end{aligned} $$

[29] As usual, the function $f+g : U \to V$ is defined by $(f+g)(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u})$ for all $\mathbf{u} \in U$.
[30] As usual, the function $\alpha f : U \to V$ is defined by $(\alpha f)(\mathbf{u}) = \alpha(f(\mathbf{u}))$ for all $\mathbf{u} \in U$.
[31] As usual, the function $g \circ f : U \to W$ is defined by $(g \circ f)(\mathbf{u}) = g(f(\mathbf{u}))$ for all $\mathbf{u} \in U$.

This proves the Claim. ♦

The Claim and the uniqueness part of Theorem 4.5.1 now imply that

$$ {}_{\mathcal{D}}[\ g \circ f\ ]_{\mathcal{B}} \ = \ {}_{\mathcal{D}}[\ g\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}, $$

which is what we needed to show. □

We can use matrices of linear functions to determine various properties of those linear functions. Notably, we have Theorem 4.5.4 below. We remark that parts (a) and (b) of Theorem 4.5.4 generalize Proposition 4.2.7, part (c) essentially follows from Theorem 4.2.4,[32] parts (d) and (e) generalize Theorem 1.10.18, and parts (f) and (g) generalize (one part of) Theorem 1.11.9.

Theorem 4.5.4. *Let U and V be non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis of U, let $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be a basis of V, and let $f : U \to V$ be a linear function.*[33] *Then all the following hold:*

(a) $rank(f) = rank\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)$;

(b) $dim\big(Ker(f)\big) = dim\Big(Nul\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)\Big)$;

(c) f *is one-to-one if and only if* $Nul\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = \{\mathbf{0}\}$;

(d) f *is one-to-one if and only if* $rank\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = m$ *(i.e. the matrix* ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ *has full column rank);*

(e) f *is onto if and only if* $rank\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = n$ *(i.e. the matrix* ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ *has full row rank);*

(f) f *is an isomorphism if and only if the matrix* ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ *is invertible (and in particular, square);*

(g) if f *is an isomorphism, then* ${}_{\mathcal{B}}[\ f^{-1}\ ]_{\mathcal{C}} = \Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)^{-1}$.

Proof. We first prove (a). By Theorem 4.5.1, we have that

$$ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}} \ = \ \big[\ [\ f(\mathbf{b}_1)\ ]_{\mathcal{C}} \ \ \ldots \ \ [\ f(\mathbf{b}_m)\ ]_{\mathcal{C}}\ \big]. $$

We now compute:

[32] Actually, it follows from Theorem 4.2.4 and Theorem 4.5.4(b).

[33] Note that this means that $\dim(U) = m$, $\dim(V) = n$, and ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}} \in \mathbb{F}^{n \times m}$.

$$\begin{aligned}
\operatorname{rank}(f) &\overset{(*)}{=} \dim\Big(\operatorname{Span}\big(f(\mathbf{b}_1),\dots,f(\mathbf{b}_m)\big)\Big)\\
&\overset{(**)}{=} \dim\left(\operatorname{Span}\Big(\big[\ f(\mathbf{b}_1)\ \big]_{\mathcal{C}},\dots,\big[\ f(\mathbf{b}_m)\ \big]_{\mathcal{C}}\Big)\right)\\
&= \dim\left(\operatorname{Col}\Big(\big[\ \big[\ f(\mathbf{b}_1)\ \big]_{\mathcal{C}}\ \ \dots\ \ \big[\ f(\mathbf{b}_m)\ \big]_{\mathcal{C}}\ \big]\Big)\right)\\
&= \dim\left(\operatorname{Col}\Big({}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}\Big)\right)\\
&\overset{(***)}{=} \operatorname{rank}\Big({}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}\Big),
\end{aligned}$$

where (*) follows from the fact that $\mathcal{B} = \{\mathbf{b}_1,\dots,\mathbf{b}_m\}$ is a spanning set of U (because it is a basis of U) and from Corollary 4.2.12, (**) follows from the fact that $\big[\ \cdot\ \big]_{\mathcal{C}} : V \to \mathbb{F}^n$ is an isomorphism and from Propositions 4.2.11(b) and 4.4.7(c),[34] and (***) follows from Theorem 3.3.4. This proves (a).

For (b), we first observe that

$$\begin{aligned}
\operatorname{rank}(f) + \dim\big(\operatorname{Ker}(f)\big) &\overset{(*)}{=} \dim(U)\\
&= m\\
&\overset{(**)}{=} \operatorname{rank}\Big({}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}\Big) + \dim\Big(\operatorname{Nul}\Big({}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}\Big)\Big)
\end{aligned}$$

where (*) follows from the rank-nullity theorem for linear functions, and (**) follows from the rank-nullity theorem for matrices (since ${}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}$ is an $n \times m$ matrix). But by (a), we have that $\operatorname{rank}(f) = \operatorname{rank}\Big({}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}\Big)$. Therefore, $\dim\big(\operatorname{Ker}(f)\big) = \dim\Big(\operatorname{Nul}\Big({}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}\Big)\Big)$. This proves (b).

For (c), we have the following sequence of equivalent statements:

$$f \text{ is one-to-one} \quad \overset{(*)}{\Longleftrightarrow} \quad \operatorname{Ker}(f) = \{\mathbf{0}\}$$

[34] Let us explain this in more detail. Since $\big[\ \cdot\ \big]_{\mathcal{C}} : V \to \mathbb{F}^n$ is an isomorphism, Proposition 4.4.7(c) guarantees that $V' := \operatorname{Span}\big(f(\mathbf{b}_1),\dots,f(\mathbf{b}_m)\big)$ and the image of V' under $\big[\ \cdot\ \big]_{\mathcal{C}}$ have the same dimension. On the other hand, by Theorem 4.2.11(b), the image of V' under $\big[\ \cdot\ \big]_{\mathcal{C}}$ is precisely equal to $\operatorname{Span}\Big(\big[\ f(\mathbf{b}_1)\ \big]_{\mathcal{C}},\dots,\big[\ f(\mathbf{b}_m)\ \big]_{\mathcal{C}}\Big)$. This justifies (**).

$$\iff \quad \dim\big(\mathrm{Ker}(f)\big) = 0$$

$$\overset{(**)}{\iff} \quad \dim\Big(\mathrm{Nul}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)\Big) = 0$$

$$\iff \quad \mathrm{Nul}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = \{\mathbf{0}\},$$

where (*) follows from Theorem 4.2.4, and (**) follows from part (b).

For (d), we have the following sequence of equivalent statements:

$$f \text{ is one-to-one} \quad \overset{(*)}{\iff} \quad \mathrm{Nul}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = \{\mathbf{0}\}$$

$$\iff \quad \dim\left(\mathrm{Nul}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)\right) = 0$$

$$\overset{(**)}{\iff} \quad \mathrm{rank}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = m,$$

where (*) follows from part (c), and (**) follows from the rank-nullity theorem for matrices.[35]

For (e), we have the following sequence of equivalent statements:

$$f \text{ is onto} \quad \overset{(*)}{\iff} \quad \mathrm{rank}(f) = \dim(V)$$

$$\overset{(**)}{\iff} \quad \mathrm{rank}(f) = n$$

$$\overset{(***)}{\iff} \quad \mathrm{rank}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = n,$$

where (*) follows from Proposition 4.2.6, (**) follows from the fact that $\dim(V) = n$, and (***) follows from part (a).

Next, we prove (f). Suppose first that f is an isomorphism. Then by Theorem 4.2.14(c) (or alternatively, by Theorem 4.4.6), we have that $\dim(U) = \dim(V)$, i.e. $m = n$. In particular, ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is an $n \times n$ matrix. Next, since f is an isomorphism (and in particular, an onto linear function), part (d) guarantees that

[35] Indeed, since ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is an $n \times m$ matrix, the rank-nullity theorem for matrices guarantees that

$$\mathrm{rank}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) + \dim\left(\mathrm{Nul}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)\right) \quad = \quad m,$$

and (**) immediately follows.

$\operatorname{rank}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = n$. But now the Invertible Matrix Theorem (see subsection 1.11.7 or 3.3.6) guarantees that ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is invertible.

Suppose, conversely, that ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is invertible. In particular, ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is a square matrix, and it follows that $m = n$ (because ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is an $n \times m$ matrix). Now, since ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is an invertible $n \times n$ matrix, the Invertible Matrix Theorem (see subsection 1.11.7 or 3.3.6) implies that $\operatorname{rank}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big) = n$. But now parts (d) and (e) guarantee that f is one-to-one and onto,[36] and consequently, a bijection. Since f is also linear (by hypothesis), it follows that f is an isomorphism. This proves (f).

It remains to prove (g). Suppose that f is an isomorphism. Then by Theorem 4.2.14(c) (or alternatively, by Theorem 4.4.6), we have that $m = \dim(U) = \dim(V) = n$. Consequently, ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is an $n \times n$ matrix. Moreover, by (f), the matrix ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$ is invertible. But now for all $\mathbf{v} \in V$, we have the following:

$$\begin{aligned}
\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)^{-1}\ [\ \mathbf{v}\ ]_{\mathcal{C}} &= \Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)^{-1}\ \Big[\ f\Big(f^{-1}(\mathbf{v})\Big)\ \Big]_{\mathcal{C}} \\
&= \Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)^{-1}\ \Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\ [\ f^{-1}(\mathbf{v})\ ]_{\mathcal{B}}\Big) \\
&= \Big(\underbrace{\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)^{-1}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}}_{=I_n}\Big)\ [\ f^{-1}(\mathbf{v})\ ]_{\mathcal{B}} \\
&= [\ f^{-1}(\mathbf{v})\ ]_{\mathcal{B}}.
\end{aligned}$$

The uniqueness part of Theorem 4.5.1 now implies that

$${}_{\mathcal{B}}[\ f^{-1}\ ]_{\mathcal{C}} = \Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}\Big)^{-1}.$$

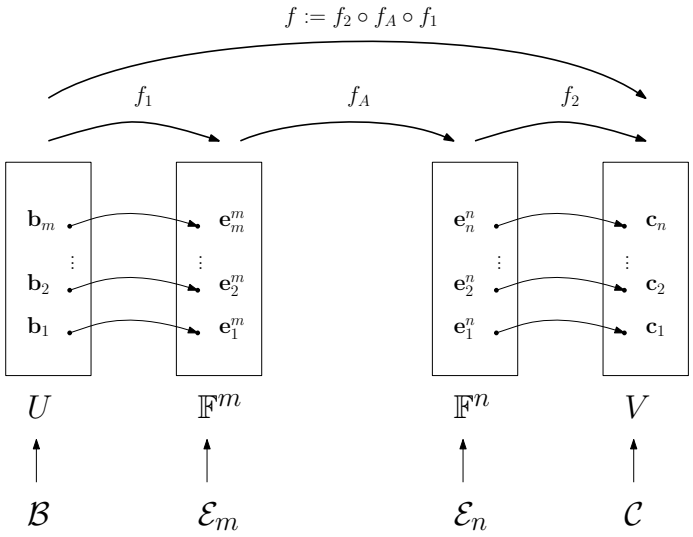
This proves (f). □

Suppose that U and V are non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$, that $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ is a basis of U, and that $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ is a basis of V. By Theorem 4.5.1, to every linear function $f : U \to V$, we can associate a unique matrix $A \in \mathbb{F}^{n\times m}$ (which we denoted by ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$) such that for all $\mathbf{u} \in U$, we have that $A\,[\ \mathbf{u}\ ]_{\mathcal{B}} = [\ f(\mathbf{u})\ ]_{\mathcal{C}}$. How about the converse? Is it true that for every matrix $A \in \mathbb{F}^{n\times m}$, there exists a linear function $f : U \to V$ such that $A = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{B}}$? As our next proposition shows, this is indeed true, but the proof is not completely obvious: it relies on Theorems 4.3.2, 4.5.1, and 4.5.3(c).

[36] We are also using the fact that $m = n$, and so part (d) applies.

Proposition 4.5.5. *Let U and V be non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$, let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis of U, and let $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be a basis of V. Then for every matrix $A \in \mathbb{F}^{n \times m}$, there exists a unique linear function $f : U \to V$ such that $A = {}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}$.*

Proof. **Existence.** As usual, $\mathcal{E}_m = \{\mathbf{e}_1^m, \ldots, \mathbf{e}_m^m\}$ is the standard basis of $\mathbb{F}^m$, and $\mathcal{E}_n = \{\mathbf{e}_1^n, \ldots, \mathbf{e}_n^n\}$ is the standard basis of $\mathbb{F}^n$. Using Theorem 4.3.2, we let $f_1 : U \to \mathbb{F}^m$ be the unique linear function such that $f_1(\mathbf{b}_1) = \mathbf{e}_1^m, \ldots, f_1(\mathbf{b}_m) = \mathbf{e}_m^m$, and we let $f_2 : \mathbb{F}^n \to V$ be the unique linear function such that $f_2(\mathbf{e}_1^n) = \mathbf{c}_1, \ldots, f_2(\mathbf{e}_n^n) = \mathbf{c}_n$. (Note that Proposition 4.4.5 guarantees that f_1 and f_2 are actually isomorphisms, but we will not actually use this fact.) Next, let $f_A : \mathbb{F}^m \to \mathbb{F}^n$ be given by $f_A(\mathbf{u}) = A\mathbf{u}$ for all $\mathbf{u} \in U$; then f_A is linear (by Proposition 1.10.4), and moreover, A is the standard matrix of f_A, i.e. $A = {}_{\mathcal{E}_n}\big[\ f_A\ \big]_{\mathcal{E}_m}$. Finally, set $f := f_2 \circ f_A \circ f_1$ (see the diagram below). Our goal is to show that $A = {}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}$.



First, we have that

$$
\begin{aligned}
{}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}} &\overset{(*)}{=} {}_{\mathcal{C}}\big[\ f_2 \circ f_A \circ f_1\ \big]_{\mathcal{B}} \\
&\overset{(**)}{=} {}_{\mathcal{C}}\big[\ f_2\ \big]_{\mathcal{E}_n}\ {}_{\mathcal{E}_n}\big[\ f_A\ \big]_{\mathcal{E}_m}\ {}_{\mathcal{E}_m}\big[\ f_1\ \big]_{\mathcal{B}} \\
&\overset{(***)}{=} {}_{\mathcal{C}}\big[\ f_2\ \big]_{\mathcal{E}_n}\ A\ {}_{\mathcal{E}_m}\big[\ f_1\ \big]_{\mathcal{B}},
\end{aligned}
$$

where (*) follows from the fact that $f = f_2 \circ f_A \circ f_1$, (**) follows from Theorem 4.5.3(c), and (***) follows from the fact that $A = {}_{\mathcal{E}_n}\big[\ f_A\ \big]_{\mathcal{E}_m}$. It is now enough to show that ${}_{\mathcal{E}_m}\big[\ f_1\ \big]_{\mathcal{B}} = I_m$ and ${}_{\mathcal{C}}\big[\ f_2\ \big]_{\mathcal{E}_n} = I_n$, for it will then follow that

$$ {}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}} \quad = \quad \underbrace{{}_{\mathcal{C}}\big[\ f_2\ \big]_{\mathcal{E}_n}}_{=I_n} \ A \ \underbrace{{}_{\mathcal{E}_m}\big[\ f_1\ \big]_{\mathcal{B}}}_{=I_m} \quad = \quad I_n A I_m \quad = \quad A, $$

which is what we need. For ${}_{\mathcal{E}_m}\big[\ f_1\ \big]_{\mathcal{B}}$, we compute:

$$ \begin{aligned} {}_{\mathcal{E}_m}\big[\ f_1\ \big]_{\mathcal{B}} &\overset{(*)}{=} \Big[\ \big[\ f_1(\mathbf{b}_1)\ \big]_{\mathcal{E}_m} \ \cdots \ \big[\ f_1(\mathbf{b}_m)\ \big]_{\mathcal{E}_m}\ \Big] \\ &\overset{(**)}{=} \Big[\ \big[\ \mathbf{e}_1^m\ \big]_{\mathcal{E}_m} \ \cdots \ \big[\ \mathbf{e}_m^m\ \big]_{\mathcal{E}_m}\ \Big] \\ &\overset{(***)}{=} \big[\ \mathbf{e}_1^m \ \cdots \ \mathbf{e}_m^m\ \big] \quad = \quad I_m, \end{aligned} $$

where (*) follows from Theorem 4.5.1, (**) follows from the construction of f_1, and (***) follows from Example 3.2.8(a) (or alternatively, from Proposition 3.2.9). The computation for ${}_{\mathcal{C}}\big[\ f_2\ \big]_{\mathcal{E}_n}$ is similar and is left as an exercise. This proves existence.

Uniqueness. Suppose that $f, g : U \to V$ are linear functions such that ${}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}} = A$ and ${}_{\mathcal{C}}\big[\ g\ \big]_{\mathcal{B}} = A$. We must show that $f = g$. First of all, note that for all indices $i \in \{1, \ldots, m\}$, we have that

$$ \big[\ f(\mathbf{b}_i)\ \big]_{\mathcal{C}} \quad = \quad \underbrace{{}_{\mathcal{C}}\big[\ f\ \big]_{\mathcal{B}}}_{=A} \ \big[\ \mathbf{b}_i\ \big]_{\mathcal{B}} \quad = \quad \underbrace{{}_{\mathcal{C}}\big[\ g\ \big]_{\mathcal{B}}}_{=A} \ \big[\ \mathbf{b}_i\ \big]_{\mathcal{B}} \quad = \quad \big[\ g(\mathbf{b}_i)\ \big]_{\mathcal{C}}, $$

and consequently, $f(\mathbf{b}_i) = g(\mathbf{b}_i)$ (because $\big[\ \cdot\ \big]_{\mathcal{C}} : V \to \mathbb{F}^n$ is an isomorphism and therefore one-to-one). But now since $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ is a basis of U and $f, g : U \to V$ are linear, the uniqueness part of Theorem 4.3.2 guarantees that $f = g$. □

Remark: Suppose that U and V are non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$, and recall from subsection 4.1.2 that $\mathrm{Hom}(U, V)$, the set of all linear functions from U to V, is a vector space over the field $\mathbb{F}$ (vector addition and scalar multiplication in this vector space are the usual addition and scalar multiplication of functions). Set $m := \dim(U)$ and $n := \dim(V)$, and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases of U and V, respectively. By Theorem 4.5.1 and Proposition 4.5.5, ${}_{\mathcal{C}}\big[\ \cdot\ \big]_{\mathcal{B}} : \mathrm{Hom}(U, V) \to \mathbb{F}^{n \times m}$ is a bijection,[37] and by Theorem 4.5.3(a-b), it is also a linear function. So, ${}_{\mathcal{C}}\big[\ \cdot\ \big]_{\mathcal{B}} : \mathrm{Hom}(U, V) \to \mathbb{F}^{n \times m}$ is in fact an isomorphism. By Theorem 4.2.14(c), it follows that $\dim\Big(\mathrm{Hom}(U, V)\Big) = \dim(\mathbb{F}^{n \times m}) = nm$.

[37] The fact that ${}_{\mathcal{C}}\big[\ \cdot\ \big]_{\mathcal{B}} : \mathrm{Hom}(U, V) \to \mathbb{F}^{n \times m}$ is a well-defined function follows from Theorem 4.5.1. The fact that this function is one-to-one and onto (i.e. a bijection) follows from Proposition 4.5.5.

4.5.1 Change of basis (transition) matrices

Recall that for any set X, the function $\mathrm{Id}_X : X \to X$ is defined by $\mathrm{Id}_X(x) = x$ for all $x \in X$; the function Id_X is called the *identity function* on X.

Given a non-trivial, finite-dimensional vector space V over a field $\mathbb{F}$, and bases $\mathcal{B}$ and $\mathcal{C}$ of V, we call the matrix ${}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}}$ the *change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$* or the *transition matrix from $\mathcal{B}$ to $\mathcal{C}$*.

Proposition 4.5.6. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases of V. Then the change of basis matrix ${}_{\mathcal{C}}\big[\ Id_V\ \big]_{\mathcal{B}}$ satisfies the property that*

$$ {}_{\mathcal{C}}\big[\ Id_V\ \big]_{\mathcal{B}}\ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} \ = \ \big[\ \mathbf{v}\ \big]_{\mathcal{C}}\,. $$

for all vectors $\mathbf{v} \in V$. Moreover, this matrix is given by the formula

$$ {}_{\mathcal{C}}\big[\ Id_V\ \big]_{\mathcal{B}} \ = \ \big[\ \big[\ \mathbf{b}_1\ \big]_{\mathcal{C}} \ \ \ldots \ \ \big[\ \mathbf{b}_n\ \big]_{\mathcal{C}}\ \big]\,. $$

Proof. The first statement follows straight from the definition of a change of basis matrix; indeed, for all vectors $\mathbf{v} \in V$, we have that

$$ {}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}}\ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} \ = \ \big[\ \mathrm{Id}_V(\mathbf{v})\ \big]_{\mathcal{C}} \ = \ \big[\ \mathbf{v}\ \big]_{\mathcal{C}}\,. $$

For the second statement, we observe that

$$ \begin{aligned} {}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}} &\overset{(*)}{=} \big[\ \big[\ \mathrm{Id}_V(\mathbf{b}_1)\ \big]_{\mathcal{C}} \ \ \ldots \ \ \big[\ \mathrm{Id}_V(\mathbf{b}_m)\ \big]_{\mathcal{C}}\ \big] \\ &= \big[\ \big[\ \mathbf{b}_1\ \big]_{\mathcal{C}} \ \ \ldots \ \ \big[\ \mathbf{b}_m\ \big]_{\mathcal{C}}\ \big] \end{aligned} $$

where (*) follows from Theorem 4.5.1. □

Proposition 4.5.7. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases of V. Then the change of basis matrices ${}_{\mathcal{C}}\big[\ Id_V\ \big]_{\mathcal{B}}$ and ${}_{\mathcal{B}}\big[\ Id_V\ \big]_{\mathcal{C}}$ are invertible, and moreover, they are each other's inverses.*

Proof. Clearly, $\mathrm{Id}_V : V \to V$ is an isomorphism, and so by Theorem 4.5.4(f), matrices ${}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}}$ and ${}_{\mathcal{B}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{C}}$ are both invertible. Moreover,

$$ {}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}} \ \overset{(*)}{=} \ {}_{\mathcal{C}}\big[\ \mathrm{Id}_V^{-1}\ \big]_{\mathcal{B}} \ \overset{(**)}{=} \ \Big({}_{\mathcal{B}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{C}}\Big)^{-1}, $$

where (*) follows from the fact that $\mathrm{Id}_V^{-1} = \mathrm{Id}_V$, and (**) follows from Theorem 4.5.4(g). This completes the argument. □

For the special case of $\mathbb{F}^n$ (where $\mathbb{F}$ is a field), we get a nice formula for change of basis matrices (see Theorem 4.5.9 below). First, we need a lemma.

Lemma 4.5.8. *Let $\mathbb{F}$ be a field, let $\mathcal{E}_n = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{F}^n$, and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be any basis of $\mathbb{F}^n$. Set $B := \left[\begin{array}{ccc} \mathbf{b}_1 & \ldots & \mathbf{b}_n \end{array}\right]$. Then B is invertible, and moreover,*

$$ {}_{\mathcal{E}_n}\left[\; Id_{\mathbb{F}^n} \;\right]_{\mathcal{B}} \;=\; B \qquad \text{and} \qquad {}_{\mathcal{B}}\left[\; Id_{\mathbb{F}^n} \;\right]_{\mathcal{E}_n} \;=\; B^{-1}. $$

Proof. Let us first prove that ${}_{\mathcal{E}_n}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{B}} = B$. In view of the uniqueness part of Theorem 4.5.1, it suffices to show that for all $\mathbf{v} \in \mathbb{F}^n$, we have that $B\left[\; \mathbf{v} \;\right]_{\mathcal{B}} = \left[\; \mathbf{v} \;\right]_{\mathcal{E}_n}$. So, fix a vector $\mathbf{v} \in \mathbb{F}^n$, and set $\left[\; \mathbf{v} \;\right]_{\mathcal{B}} = \left[\begin{array}{ccc} \beta_1 & \ldots & \beta_n \end{array}\right]^T$, so that $\mathbf{v} = \beta_1\mathbf{b}_1 + \cdots + \beta_n\mathbf{b}_n$. Then

$$ B\left[\; \mathbf{v} \;\right]_{\mathcal{B}} \;=\; \left[\begin{array}{ccc} \mathbf{b}_1 & \ldots & \mathbf{b}_n \end{array}\right] \left[\begin{array}{c} \beta_1 \\ \vdots \\ \beta_n \end{array}\right] \;=\; \sum_{i=1}^{n} \beta_i \mathbf{b}_i \;=\; \mathbf{v} \;\overset{(*)}{=}\; \left[\; \mathbf{v} \;\right]_{\mathcal{E}_n}, $$

where (*) follows from Proposition 3.2.9.[38] This proves that ${}_{\mathcal{E}_n}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{B}} = B$. The fact that B is invertible and that ${}_{\mathcal{B}}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{E}_n} = B^{-1}$ now follows from Proposition 4.5.7. □

Theorem 4.5.9. *Let $\mathbb{F}$ be a field, and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be two bases of $\mathbb{F}^n$. Set $B := \left[\begin{array}{ccc} \mathbf{b}_1 & \ldots & \mathbf{b}_n \end{array}\right]$ and $C := \left[\begin{array}{ccc} \mathbf{c}_1 & \ldots & \mathbf{c}_n \end{array}\right]$. Then the matrix ${}_{\mathcal{C}}\left[\; Id_{\mathbb{F}^n} \;\right]_{\mathcal{B}}$ is invertible, and it is given by the formula*

$$ {}_{\mathcal{C}}\left[\; Id_{\mathbb{F}^n} \;\right]_{\mathcal{B}} \;=\; C^{-1}B. $$

Proof. The fact that ${}_{\mathcal{C}}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{B}}$ is invertible follows from Proposition 4.5.7. To prove that the formula for ${}_{\mathcal{C}}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{B}}$ is correct, we observe that

$$ {}_{\mathcal{C}}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{B}} \;=\; {}_{\mathcal{C}}\left[\; \mathrm{Id}_{\mathbb{F}^n} \circ \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{B}} \;\overset{(*)}{=}\; {}_{\mathcal{C}}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{E}_n} \; {}_{\mathcal{E}_n}\left[\; \mathrm{Id}_{\mathbb{F}^n} \;\right]_{\mathcal{B}} \;\overset{(**)}{=}\; C^{-1}B, $$

where (*) follows from Theorem 4.5.3, and (**) follows from Lemma 4.5.8. □

Proposition 4.5.10 (below) is an immediate corollary of Theorem 4.5.3(c). We state it as a separate proposition because it is used particularly often for computation.

Proposition 4.5.10. *Let U and V be non-trivial, finite-dimensional vector spaces over a field $\mathbb{F}$, let $\mathcal{B}_1$ and $\mathcal{B}_2$ be bases of U, let $\mathcal{C}_1$ and $\mathcal{C}_2$ be bases of V, and let $f : U \to V$ be a linear function. Then*

$$ \begin{aligned} {}_{\mathcal{C}_2}\left[\; f \;\right]_{\mathcal{B}_2} \;&=\; {}_{\mathcal{C}_2}\left[\; Id_V \circ f \circ Id_U \;\right]_{\mathcal{B}_2} \\ &=\; {}_{\mathcal{C}_2}\left[\; Id_V \;\right]_{\mathcal{C}_1} \; {}_{\mathcal{C}_1}\left[\; f \;\right]_{\mathcal{B}_1} \; {}_{\mathcal{B}_1}\left[\; Id_U \;\right]_{\mathcal{B}_2}. \end{aligned} $$

[38] Alternatively, it follows from Example 3.2.8(a).

Proof. This follows immediately from Theorem 4.5.3(c). □

Let us now return to the linear function f from Example 4.5.2: we would like to compute the standard matrix of this linear function.

Example 4.5.11. *Consider the basis* $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ *of* $\mathbb{R}^2$*, and consider the unique linear function* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *that satisfies the following:*

- $f\Big(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\Big) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$;
- $f\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Compute the standard matrix of the linear function f*.*

Solution. In Example 4.5.2, we saw that ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Now, we set $B := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$,[39] and we compute $B^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$. Then the standard matrix of f is

$$\begin{aligned}
{}_{\mathcal{E}_2}[\ f\ ]_{\mathcal{E}_2} &= {}_{\mathcal{E}_2}[\ \mathrm{Id}_{\mathbb{R}^2}\ ]_{\mathcal{B}}\ {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}\ {}_{\mathcal{B}}[\ \mathrm{Id}_{\mathbb{R}^2}\ ]_{\mathcal{E}_2} && \text{by Proposition 4.5.10} \\
&= B\ {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}\ B^{-1} && \text{by Lemma 4.5.8} \\
&= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}.
\end{aligned}$$

Optional: Let us check that our answer is correct. Indeed, we have that

- $\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = f\Big(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\Big)$;
- $\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = f\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big)$.

So, our answer is correct. □

[39] So, the columns of B are the vectors of the basis $\mathcal{B}$, arranged from left to right in the order in which they appear in $\mathcal{B}$.

A characterization of change of basis matrices. We complete this subsection with Proposition 4.5.12 (below), which essentially states that change of basis matrices are precisely the invertible matrices.

Proposition 4.5.12. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ be a matrix, and let V be any n-dimensional vector space over the field $\mathbb{F}$. Then the following are equivalent:*

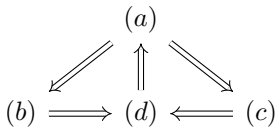
(a) A is invertible;

(b) for all bases $\mathcal{B}$ of V, there exists a basis $\mathcal{C}$ of V such that $A = {}_{\mathcal{C}}\big[\ Id_V\ \big]_{\mathcal{B}}$;

(c) for all bases $\mathcal{C}$ of V, there exists a basis $\mathcal{B}$ of V such that $A = {}_{\mathcal{C}}\big[\ Id_V\ \big]_{\mathcal{B}}$;

(d) there exist bases $\mathcal{B}$ and $\mathcal{C}$ of V such that $A = {}_{\mathcal{C}}\big[\ Id_V\ \big]_{\mathcal{B}}$.

Proof. Clearly, it is enough to prove the implications shown in the diagram below.



Since V has at least one n-element basis (because $\dim(V) = n$), we see that (b) implies (d), and that (c) implies (d).[40] Further, by Proposition 4.5.7, (d) implies (a). It remains to show that (a) implies (b) and (c). We prove the former; the proof of the latter is similar and is left as an exercise.

So, assume that (a) is true; we must prove (b). Fix any basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of V;[41] we must construct a basis $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ of V such that $A = {}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}}$. Using Proposition 4.5.5, we let $f : V \to V$ be the (unique) linear function such that $A = {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. Since A is invertible, Theorem 4.5.4(f) guarantees that f is an isomorphism. Then by Proposition 4.4.1, $f^{-1} : V \to V$ is also an isomorphism. For each index $i \in \{1, \ldots, n\}$, we set $\mathbf{c}_i := f^{-1}(\mathbf{b}_i)$. Since $f^{-1} : V \to V$ is an isomorphism and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis of V, Theorem 4.4.4(c) implies that $\big\{f^{-1}(\mathbf{b}_1), \ldots, f^{-1}(\mathbf{b}_n)\big\} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\} =: \mathcal{C}$ is also a basis of V.

Now, we claim that $A = {}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}}$. First, we note that

$$ {}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}} \;=\; {}_{\mathcal{C}}\big[\ f^{-1}\circ f\ \big]_{\mathcal{B}} \overset{(*)}{=} {}_{\mathcal{C}}\big[\ f^{-1}\ \big]_{\mathcal{B}}\ \underbrace{{}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}}_{=A} \;=\; {}_{\mathcal{C}}\big[\ f^{-1}\ \big]_{\mathcal{B}}\ A, $$

where (*) follows from Theorem 4.5.3(c). It now suffices to show that ${}_{\mathcal{C}}\big[\ f^{-1}\ \big]_{\mathcal{B}} = I_n$, for it will then immediately follow that $A = {}_{\mathcal{C}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}}$, which is what we need. We compute:

[40] We need the fact that V has at least one n-element basis, since that means that (b) and (c) are not just "vacuously true" (due to there not being any bases of V).

[41] Since $\dim(V) = n$, all bases of V have n elements.

$$\begin{aligned} {}_{\mathcal{C}}\left[\ f^{-1}\ \right]_{\mathcal{B}} &\overset{(*)}{=} \left[\ \left[\ f^{-1}(\mathbf{b}_1)\ \right]_{\mathcal{C}} \ \ldots\ \left[\ f^{-1}(\mathbf{b}_n)\ \right]_{\mathcal{C}}\ \right] \\ &= \left[\ \left[\ \mathbf{c}_1\ \right]_{\mathcal{C}} \ \ldots\ \left[\ \mathbf{c}_n\ \right]_{\mathcal{C}}\ \right] \\ &\overset{(**)}{=} \left[\ \mathbf{e}_1^n \ \ldots\ \mathbf{e}_n^n\ \right] = I_n, \end{aligned}$$

where (*) follows from Theorem 4.5.1, and (**) follows from Proposition 3.2.8. This proves (b), and we are done. □

4.5.2 Similar matrices

Let $\mathbb{F}$ be a field. Given matrices $A, B \in \mathbb{F}^{n\times n}$, we say that A is *similar* to B if there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$. By Proposition 4.5.13 (below), matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$.

Proposition 4.5.13. *Let $\mathbb{F}$ be a field. Then all the following hold:*

(a) for all matrices $A \in \mathbb{F}^{n\times n}$, A is similar to A;

(b) for all matrices $A, B \in \mathbb{F}^{n\times n}$, if A is similar to B, then B is similar to A;

(c) for all matrices $A, B, C \in \mathbb{F}^{n\times n}$, if A is similar to B and B is similar to C, then A is similar to C.

Proof. (a) Fix a matrix $A \in \mathbb{F}^{n\times n}$. Then $A = I_n^{-1}AI_n$, and it follows that A is similar to itself.

(b) Fix a matrices $A, B \in \mathbb{F}^{n\times n}$, and assume that A is similar to B. Then there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$. But then $A = PBP^{-1} = (P^{-1})^{-1}BP^{-1}$, and it follows that B is similar to A.

(c) Fix matrices $A, B, C \in \mathbb{F}^{n\times n}$, and assume that A is similar to B and that B is similar to C. Then there exist invertible matrices $P, Q \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$ and $C = Q^{-1}BQ$. But now

$$\begin{aligned} C &= Q^{-1}BQ \\ &= Q^{-1}(P^{-1}AP)Q \\ &= (Q^{-1}P^{-1})A(PQ) \\ &= (PQ)^{-1}A(PQ), \end{aligned}$$

and it follows that A is similar to C. □

Remark: By Proposition 4.5.13(b), the similarity relation on $\mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) is symmetric. Consequently, we may speak of matrices $A, B \in \mathbb{F}^{n\times n}$ as being similar or not being similar **to each other**. In particular, in what follows, we will often write something like "let $A, B \in \mathbb{F}^{n\times n}$ be similar matrices." This means that A is similar to B and vice versa.

Proposition 4.5.14. *Let $\mathbb{F}$ be a field, and let $A, B \in \mathbb{F}^{n\times n}$ be similar matrices, say $B = P^{-1}AP$ for some invertible matrix $P \in \mathbb{F}^{n\times n}$. Then A is invertible if and only if B is invertible, and in this case, $B^{-1} = P^{-1}A^{-1}P$ and $A^{-1} = PB^{-1}P^{-1}$.*

Proof. Since $B = P^{-1}AP$, we have that $A = PBP^{-1}$. Since P and P^{-1} are invertible,[42] Proposition 1.11.8(e) guarantees that A is invertible if and only if B is invertible.[43] Suppose now that A and B are invertible. Then

$$B^{-1} = (P^{-1}AP)^{-1} \overset{(*)}{=} P^{-1}A^{-1}(P^{-1})^{-1} \overset{(**)}{=} P^{-1}A^{-1}P,$$

where (*) follows from Proposition 1.11.8(e), and (**) follows from Proposition 1.11.8(b). But now since $B^{-1} = P^{-1}A^{-1}P$, we immediately get that $A^{-1} = PB^{-1}P^{-1}$. This completes the argument. □

Proposition 4.5.15. *Let $\mathbb{F}$ be a field, and let $A, B \in \mathbb{F}^{n\times n}$ be similar matrices, say $B = P^{-1}AP$ for some invertible matrix $P \in \mathbb{F}^{n\times n}$. Then for all non-negative integers m, we have that $B^m = P^{-1}A^mP$, and in particular, A^m and B^m are similar. Moreover, if A and B are invertible,[44] then we in fact have that $B^m = P^{-1}A^mP$ for all integers m.*

Proof. We first prove that $B^m = P^{-1}A^mP$ for all non-negative integers m. We proceed by induction on m. For $m = 0$, we note that $B^0 = I_n$ and $P^{-1}A^0P = P^{-1}I_nP = P^{-1}P = I_n$, and so $B^0 = P^{-1}A^0P$. Now, fix a non-negative integer m, and assume inductively that $B^m = P^{-1}A^mP$. We then have that

$$\begin{aligned} B^{m+1} &= B^mB \\ &\overset{(*)}{=} (\underbrace{P^{-1}A^mP}_{=B^m})(\underbrace{P^{-1}AP}_{=B}) \\ &= P^{-1}A^m(\underbrace{PP^{-1}}_{=I_n})AP \end{aligned}$$

[42] Since P is invertible, Proposition 1.11.8(b) guarantees that P^{-1} is also invertible, and moreover, that $(P^{-1})^{-1} = P$.

[43] Indeed, if A is invertible, then we have that P^{-1}, A, P are all invertible, and so by Proposition 1.11.8(e), $B = P^{-1}AP$ is invertible. Similarly, if B is invertible, then we have that P, B, P^{-1} are all invertible, and so by Proposition 1.11.8(e), $A = PBP^{-1}$ is invertible.

[44] By Proposition 4.5.14, A is invertible if and only if B is invertible.

$$= \quad P^{-1}A^mAP$$

$$= \quad P^{-1}A^{m+1}P,$$

where in (*) we used the induction hypothesis (for the fact that $B^m = P^{-1}A^mP$), plus the fact that $B = P^{-1}AP$ (by hypothesis). This completes the induction.

Assume now that A and B are invertible. By Proposition 4.5.14, we have that $B^{-1} = P^{-1}A^{-1}P$. But now by an argument completely analogous to the above,[45] we get that for all nonegative integers m, we have that $(B^{-1})^m = P^{-1}(A^{-1})^mP$, that is, $B^{-m} = P^{-1}A^{-m}P$. Combined with the above, this implies that $B^m = P^{-1}A^mP$ for all integers m. □

Our next theorem essentially states that two $n \times n$ matrices are similar if and only if they represent the same linear function from an n-dimensional vector space to itself, but possibly with respect to different bases.

Theorem 4.5.16. *Let $\mathbb{F}$ be a field, let $B, C \in \mathbb{F}^{n\times n}$ be matrices, and let V be an n-dimensional vector space over the field $\mathbb{F}$. Then the following are equivalent:*

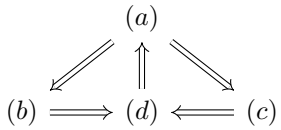
(a) B and C are similar;

(b) for all bases $\mathcal{B}$ of V and linear functions $f : V \to V$ such that $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$, there exists a basis $\mathcal{C}$ of V such that $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$;

(c) for all bases $\mathcal{C}$ of V and linear functions $f : V \to V$ such that $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$, there exists a basis $\mathcal{B}$ of V such that $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$;

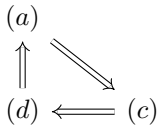
(d) there exist bases $\mathcal{B}$ and $\mathcal{C}$ of V and a linear function $f : V \to V$ such that $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ and $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$.

Proof. Clearly, it is enough to prove the implications shown in the diagram below.



But since matrix similarity in $\mathbb{F}^{n\times n}$ is symmetric (by Proposition 4.5.13(b)), the proofs of the implications "(a) $\Longrightarrow$ (b)" and "(a) $\Longrightarrow$ (c)" are completely analogous, as are the proofs of the implications "(b) $\Longrightarrow$ (d)" and "(c) $\Longrightarrow$ (d)." So, it is enough to prove the implications shown in the diagram below.

[45] We simply replace A and B with A^{-1} and B^{-1}, respectively, in the argument above.



First, we assume (a) and prove (c). Assume that $\mathcal{C}$ is a basis of V and that $f : V \to V$ is a linear function such that $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$. We must show that there exists a basis $\mathcal{B}$ of V such that $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$. By (a), matrices B and C are similar, which by definition means that there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}CP$. Since P is invertible, Proposition 4.5.12 guarantees that there exists a basis $\mathcal{B}$ of V such that $P = {}_{\mathcal{C}}[\ \mathrm{Id}_V\ ]_{\mathcal{B}}$.[46] But now we have that

$$\begin{aligned}
B &= P^{-1}CP \\
&= \left({}_{\mathcal{C}}[\ \mathrm{Id}_V\ ]_{\mathcal{B}}\right)^{-1}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ \mathrm{Id}_V\ ]_{\mathcal{B}} \\
&= {}_{\mathcal{B}}[\ \mathrm{Id}_V\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}\ {}_{\mathcal{C}}[\ \mathrm{Id}_V\ ]_{\mathcal{B}} && \text{by Proposition 4.5.7} \\
&= {}_{\mathcal{B}}[\ \mathrm{Id}_V \circ f \circ \mathrm{Id}_V\ ]_{\mathcal{B}} && \text{by Theorem 4.5.3(c)} \\
&= {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}.
\end{aligned}$$

This proves (c).

Next, we assume (c) and prove (d). Since V is an n-dimensional vector space, it has a basis $\mathcal{C}$ of size n. Next, by Proposition 4.5.5, there exists a (unique) linear function $f : V \to V$ such that $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$. But then by (c), there exists a basis $\mathcal{B}$ of V such that $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$. This proves (d).[47]

Finally, we assume (d) and prove (a). Using (d), we fix bases $\mathcal{B}$ and $\mathcal{C}$ of V and a linear function $f : V \to V$ such that $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ and $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$. Set $P := {}_{\mathcal{B}}[\ \mathrm{Id}_V\ ]_{\mathcal{C}}$. By Proposition 4.5.7, P is invertible and satisfies $P^{-1} = {}_{\mathcal{C}}[\ \mathrm{Id}_V\ ]_{\mathcal{B}}$. We now compute:

$$\begin{aligned}
P^{-1}BP &= {}_{\mathcal{C}}[\ \mathrm{Id}_V\ ]_{\mathcal{B}}\ {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}\ {}_{\mathcal{B}}[\ \mathrm{Id}_V\ ]_{\mathcal{C}} \\
&\overset{(*)}{=} {}_{\mathcal{C}}[\ \mathrm{Id}_V \circ f \circ \mathrm{Id}_V\ ]_{\mathcal{C}}
\end{aligned}$$

[46] We are relying on the "(a) $\Longrightarrow$ (c)" implication of Proposition 4.5.12.

[47] **Remark:** The implication "(c) $\Longrightarrow$ (d)" may seem trivial, but in fact it is not! To get this implication, we need to make sure that (c) is not just "vacuously true" due to there not existing any $\mathcal{C}$ and f such that $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$. The existence of the basis $\mathcal{C}$ follows immediately from dimension considerations, but the existence of a linear function $f : V \to V$ such that $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$ only follows from the not entirely trivial Proposition 4.5.5.

$$= \quad {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$$
$$= \quad C,$$

where (*) follows from Theorem 4.5.3(c). So, B and C are similar. This proves (a), and we are done. □

Corollary 4.5.17. *Let $\mathbb{F}$ be a field, and let $B, C \in \mathbb{F}^{n\times n}$ be similar matrices. Then rank(B) = rank(C).*

Proof. This follows immediately from the definition of matrix similarity and from Proposition 3.3.14(c).[48] However, let us give a different proof, one relying on Theorem 4.5.16 (in order to illustrate how Theorem 4.5.16 can be used).

Since B and C are similar, Theorem 4.5.16 guarantees that there exist bases $\mathcal{B}$ and $\mathcal{C}$ of $\mathbb{F}^n$ and a linear function $f : \mathbb{F}^n \to \mathbb{F}^n$ such that $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ and $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$.[49] But then

$$\begin{aligned}
\operatorname{rank}(B) &= \operatorname{rank}\Big({}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}\Big) && \text{because } B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}} \\
&= \operatorname{rank}(f) && \text{by Theorem 4.5.4(a)} \\
&= \operatorname{rank}\Big({}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}\Big) && \text{by Theorem 4.5.4(a)} \\
&= \operatorname{rank}(C) && \text{because } C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}},
\end{aligned}$$

and we are done. □

4.5.3 Checking the existence and uniqueness of linear functions with certain specifications: examples with polynomials and matrices

In this subsection, we give a few examples similar to those from subsection 1.10.4, except that we do not work with linear functions of the form $f : \mathbb{F}^m \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field), but with linear functions between various other vector spaces (in particular, our examples involve matrices and polynomials). All examples in this subsection are of the following general form: we are given vector spaces U and V over some field $\mathbb{F}$ (where at least U is non-trivial and finite-dimensional), and we are asked whether there exists a linear function $f : U \to V$ that maps certain specified vectors from U to certain specified vectors from V, and if so, whether this linear function f is

[48] Details?

[49] Here, we are using the "(a) $\Longrightarrow$ (d)" part of Theorem 4.5.16, with $V := \mathbb{F}^n$.

unique. In some examples, we are further asked to determine various properties of such a linear function f (for example, we may need to determine whether f is an isomorphism). Our solutions will be of two types: one type relies on matrices of linear functions with respect to the most natural bases of the domain and codomain (natural for the vector spaces in question, with no regard to the particular linear function f),[50] whereas the other type relies on Theorem 4.3.2 or Corollary 4.3.3,[51] and may in addition use matrices of linear functions with respect to convenient bases (convenient for the linear function f in question), as well as change of basis matrices. In most cases, the first type of solution is simpler. However, the second type may be better when, in addition to checking the existence and uniqueness of a linear function with specifications of the type mentioned above, we also need our linear function to satisfy various other properties (e.g. to have some particular rank). Our solution to Example 4.5.18 uses the first method, whereas our solution to Example 4.5.19 uses the second method. We give two solutions (illustrating the two methods) for each of the Examples 4.5.20 and 4.5.21. Finally, our solution to Example 4.5.22 uses a combination of the two methods.

Remark: When working with matrices of linear functions with respect to particular bases, or when working with coordinate vectors, we must always **explicitly specify the bases** that we are working with. If bases are not explicitly specified, then our proof/solution is at best incomplete, and at worst incorrect.

Example 4.5.18. *Consider the following matrices with entries in $\mathbb{Z}_2$:*

- $M_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$;
- $M_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$;
- $M_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$;
- $M_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$;
- $M_5 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$;
- $M_6 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$;
- $M_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$;
- $M_8 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Further, consider the following polynomials with coefficients in $\mathbb{Z}_2$:

[50] This type of solution is a straightforward generalization of the solutions from subsection 1.10.4.

[51] Recall that Theorem 4.3.2 essentially states that a linear function with a finite-dimensional domain can be fully determined by fixing some basis of the domain, and then specifying the images of the basis vectors. The images of the basis vectors can be **any** vectors from the codomain (there are no restrictions on how we may choose those vectors from the codomain), but once those images have been specified, the linear function is fully determined.

- $p_1(x) = x^3 + x^2 + x + 1$;
- $p_2(x) = x^4 + x^2 + x + 1$;
- $p_3(x) = x^5 + x^4 + x^2 + 1$;
- $p_4(x) = x^3$;
- $p_5(x) = x^5 + x^2 + 1$;
- $p_6(x) = x^5 + x^4 + x^2 + x$;
- $p_7(x) = x^2 + x$;
- $p_8(x) = x^5 + x$.

(a) Prove that there exists a unique linear function $f : \mathbb{Z}_2^{2\times 3} \to \mathbb{P}^5_{\mathbb{Z}_2}$ that satisfies the property that $f(M_i) = p_i(x)$ for all indices $i \in \{1, \ldots, 8\}$.

(b) Find $rank(f)$ and $dim\big(Ker(f)\big)$.

(c) Is f one-to-one? Is it onto? Is it an isomorphism?

(d) Find a formula for the linear function f, that is, fill in the blank in the following:

$$f\Big(\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}\Big) \;=\; \underline{\qquad\qquad\qquad\qquad}$$

$$\forall a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3} \in \mathbb{Z}_2.$$

(e) If f is an isomorphism, then find a formula for f^{-1}, that is, fill in the blank in the following:

$$f^{-1}\Big(a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0\Big) \;=\; \underline{\qquad\qquad\qquad\qquad}$$

$$\forall a_0, a_1, a_2, a_3, a_4, a_5 \in \mathbb{Z}_2.$$

Solution. In our solution, we will use the basis

$$\mathcal{M} \;:=\; \Big\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Big\}$$

of $\mathbb{Z}_2^{2\times 3}$, and the basis $\mathcal{P} := \big\{1, x, x^2, x^3, x^4, x^5\big\}$ of $\mathbb{P}^5_{\mathbb{Z}_2}$.

(a) We proceed as in subsection 1.10.4, except that instead of solving for the standard matrix of our linear function (which does not exist, since we are working with vector spaces of matrices and polynomials), we will solve for the matrix ${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}$. We need our linear function f to satisfy $f(M_i) = p_i(x)$ for all indices $i \in \{1, \ldots, 8\}$, and consequently, our (unknown) matrix ${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}$ should satisfy

$${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}} \ \big[\ M_i\ \big]_{\mathcal{M}} \;=\; \big[\ p_i(x)\ \big]_{\mathcal{P}}$$

for all indices $i \in \{1, \ldots, 8\}$. This is equivalent to

$$ {}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}} \underbrace{\begin{bmatrix} [\ M_1\ ]_{\mathcal{M}} & \dots & [\ M_8\ ]_{\mathcal{M}} \end{bmatrix}}_{=:M} = \underbrace{\begin{bmatrix} [\ p_1(x)\ ]_{\mathcal{P}} & \dots & [\ p_8(x)\ ]_{\mathcal{P}} \end{bmatrix}}_{=:P}. $$

Here, matrices M and P can easily be computed (see below), whereas the matrix ${}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}}$ is the unknown that we need to solve for. We proceed as in subsection 1.9.2. We first take the transpose of both sides of the equation above, and we obtain

$$ M^T \Big({}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}} \Big)^T = P^T, $$

which we solve for $\Big({}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}} \Big)^T$. We form the matrix

$$ \begin{bmatrix} M^T & P^T \end{bmatrix} = \left[\begin{array}{c:c} [\ M_1\ ]_{\mathcal{M}}^T & [\ p_1(x)\ ]_{\mathcal{P}}^T \\ \vdots & \vdots \\ [\ M_8\ ]_{\mathcal{M}}^T & [\ p_8(x)\ ]_{\mathcal{P}}^T \end{array}\right] $$

$$ = \left[\begin{array}{cccccc:cccccc} 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right], $$

and we row reduce to obtain

$$ \mathrm{RREF}\Big(\begin{bmatrix} M^T & P^T \end{bmatrix} \Big) = \left[\begin{array}{cccccc:cccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]. $$

We now read off the (unique) solution for $\Big({}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}} \Big)^T$:

$$\Big({}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}\Big)^T \quad = \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

By taking the transpose, we obtain the (unique) solution for the matrix ${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}$:

$${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}} \quad = \quad \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The existence and uniqueness of the matrix ${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}$ guarantees the existence and uniqueness of the linear function $f : \mathbb{Z}_2^{2\times 3} \to \mathbb{P}^5_{\mathbb{Z}_2}$ that satisfies the property that $f(M_i) = p_i(x)$ for all indices $i \in \{1, \ldots, 8\}$.

Remark: In the above, the existence and uniqueness of the matrix ${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}$ implied the existence and uniqueness of the linear function f with the specifications from the statement of the example. If we had obtained more than one solution for the matrix ${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}$, this would have implied that a linear function f with the given specifications exists, but is not unique. On the other hand, if there had been no solutions for ${}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}$, this would have meant that no linear function f with the given specifications exists.

(b) By row reducing, we see that $\mathrm{RREF}\big({}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}\big) = I_6$. Consequently,

$$\mathrm{rank}(f) \quad \overset{(*)}{=} \quad \mathrm{rank}\Big({}_{\mathcal{P}}\big[\ f\ \big]_{\mathcal{M}}\Big) \quad = \quad 6,$$

where (*) follows from Theorem 4.5.4(a). On the other hand, by the rank-nullity theorem, we have that $\mathrm{rank}(f) + \dim\big(\mathrm{Ker}(f)\big) = \dim(\mathbb{Z}_2^{2\times 3})$, and it follows that

$$\dim\big(\mathrm{Ker}(f)\big) \quad = \quad \dim(\mathbb{Z}_2^{2\times 3}) - \mathrm{rank}(f) \quad = \quad 6 - 6 \quad = \quad 0.$$

(c) Since $\dim\big(\mathrm{Ker}(f)\big) = 0$, Theorem 4.2.4 guarantees that f is one-to-one. Since $\mathrm{rank}(f) = 6 = \dim(\mathbb{Z}_2^{2\times 3})$, Proposition 4.2.6 guarantees that f is onto.[52] Since the linear function f is one-to-one and onto, it is an isomorphism.

[52] Alternatively, since the domain and the codomain of the linear function f have the same finite dimension, and since f is one-to-one, Corollary 4.2.10 guarantees that f is also onto and an isomorphism.

(d) Using the matrix ${}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}}$, we can easily read off the formula for f, as follows. For $a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3} \in \mathbb{Z}_2$, we compute:

$$\left[\ f\Big(\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}\Big)\ \right]_{\mathcal{P}} = {}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}} \left[\ \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}\ \right]_{\mathcal{M}}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1,1} \\ a_{1,2} \\ a_{1,3} \\ a_{2,1} \\ a_{2,2} \\ a_{2,3} \end{bmatrix}$$

$$= \begin{bmatrix} a_{2,1} \\ a_{2,2} + a_{2,3} \\ a_{2,1} + a_{2,2} \\ a_{1,3} \\ a_{1,1} + a_{2,1} \\ a_{1,1} + a_{1,2} + a_{2,3} \end{bmatrix}$$

$$= \left[\ \begin{pmatrix} (a_{1,1} + a_{1,2} + a_{2,3})x^5 + \\ +(a_{1,1} + a_{2,1})x^4 + a_{1,3}x^3 + \\ +(a_{2,1} + a_{2,2})x^2 + \\ +(a_{2,2} + a_{2,3})x + a_{2,1} \end{pmatrix}\ \right]_{\mathcal{P}}.$$

Since $\left[\ \cdot\ \right]_{\mathcal{P}}$ is an isomorphism (and in particular, one-to-one), we deduce that

$$f\Big(\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}\Big) = \begin{cases} (a_{1,1} + a_{1,2} + a_{2,3})x^5 + \\ +(a_{1,1} + a_{2,1})x^4 + a_{1,3}x^3 + \\ +(a_{2,1} + a_{2,2})x^2 + \\ +(a_{2,2} + a_{2,3})x + a_{2,1} \end{cases}$$

for all $a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3} \in \mathbb{Z}_2$. This is the formula that we needed.

(e) As we saw in part (c), f is an isomorphism. Let us find a formula for f^{-1}. First, we have that

$${}_{\mathcal{M}}[\ f^{-1}\ ]_{\mathcal{P}} \overset{(*)}{=} \Big({}_{\mathcal{P}}[\ f\ ]_{\mathcal{M}}\Big)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix},$$

where (*) follows from Theorem 4.5.4(g). We now proceed similarly as in part (d). For all $a_0, a_1, a_2, a_3, a_4, a_5 \in \mathbb{Z}_2$, we have the following:

$$\begin{aligned}
&\Big[\ f^{-1}\Big(a_5x^5+a_4x^4+a_3x^3+a_2x^2+a_1x+a_0\Big)\ \Big]_{\mathcal{M}}\\
=\ &{}_{\mathcal{M}}\Big[\ f^{-1}\ \Big]_{\mathcal{P}}\ \Big[\ a_5x^5+a_4x^4+a_3x^3+a_2x^2+a_1x+a_0\ \Big]_{\mathcal{P}}\\
=\ &\begin{bmatrix}1&0&0&0&1&0\\0&1&1&0&1&1\\0&0&0&1&0&0\\1&0&0&0&0&0\\1&0&1&0&0&0\\1&1&1&0&0&0\end{bmatrix}\begin{bmatrix}a_0\\a_1\\a_2\\a_3\\a_4\\a_5\end{bmatrix}\\
=\ &\begin{bmatrix}a_0+a_4\\a_1+a_2+a_4+a_5\\a_3\\a_0\\a_0+a_2\\a_0+a_1+a_2\end{bmatrix}\\
=\ &\Big[\begin{bmatrix}a_0+a_4&a_1+a_2+a_4+a_5&a_3\\a_0&a_0+a_2&a_0+a_1+a_2\end{bmatrix}\Big]_{\mathcal{M}}.
\end{aligned}$$

Since $\big[\ \cdot\ \big]_{\mathcal{M}}$ is an isomorphism (and in particular, one-to-one), it follows that

$$\begin{aligned}
&f^{-1}\Big(a_5x^5+a_4x^4+a_3x^3+a_2x^2+a_1x+a_0\Big)\\
=\ &\begin{bmatrix}a_0+a_4&a_1+a_2+a_4+a_5&a_3\\a_0&a_0+a_2&a_0+a_1+a_2\end{bmatrix}
\end{aligned}$$

for all $a_0, a_1, a_2, a_3, a_4, a_5 \in \mathbb{Z}_2$. This is the formula for f^{-1} that we needed.

Optional: Because it is easy to miscompute, it is a good idea to check our formulas for f and f^{-1}. Let us first check our formula for f. For each index $i \in \{1, \ldots, 8\}$, we compute $f(M_i)$ using the formula that we obtained in part (d), and we check that we do indeed get $f(M_i) = p_i(x)$. (If for some $i \in \{1, \ldots, 8\}$, we get that $f(M_i) \neq p_i(x)$, it means that we made a mistake somewhere.) Here, we only do the computation for $i = 1$ in order to demonstrate the general principle. The rest is similar routine computation. So, for $i = 1$, we compute:

$$\begin{aligned}
f(M_1) &= f\Big(\begin{bmatrix}1&0&1\\1&0&1\end{bmatrix}\Big)\\
&= (1+0+1)x^5+(1+1)x^4+1x^3+(1+0)x^2+(0+1)x+1\\
&= x^3+x^2+x+1 \ = \ p_1(x),
\end{aligned}$$

which is what we were supposed to get.

We check our formula for f^{-1} in a similar way. For each index $i \in \{1, \ldots, 8\}$, we compute $f^{-1}\big(p_i(x)\big)$ using the formula that we obtained in part (d), and we check that we do indeed get $f^{-1}\big(p_i(x)\big) = M_i$. (If for some $i \in \{1, \ldots, 8\}$, we get that $f^{-1}\big(p_i(x)\big) \neq M_i$, it means that we made a mistake somewhere.) Once again, we only do the computation for $i = 1$ in order to demonstrate the general principle. The rest is similar routine computation. So, for $i = 1$, we compute:

$$\begin{aligned} f^{-1}\big(p_1(x)\big) &= f^{-1}\big(x^3 + x^2 + x + 1\big) \\ &= \begin{bmatrix} 1+0 & 1+1+0+0 & 1 \\ 1 & 1+1 & 1+1+1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = M_1, \end{aligned}$$

which is what we were supposed to get. Alternatively, having verified the formula for f, we can verify the formula for f^{-1} by checking that $f^{-1} \circ f = \mathrm{Id}_{\mathbb{Z}_2^{2\times 3}}$ and that $f \circ f^{-1} = \mathrm{Id}_{\mathbb{P}^5_{\mathbb{Z}_2}}$ (using our formulas for f and f^{-1}). □

Example 4.5.19. *Consider the following polynomials with coefficients in $\mathbb{Z}_2$:*

- $p_1(x) = 1$;
- $p_2(x) = x + 1$;
- $p_3(x) = x^2 + x + 1$;
- $p_4(x) = x^3 + x^2 + x + 1$;
- $q_1(x) = x^4$;
- $q_2(x) = x^3 + x^2$;
- $q_3(x) = x^2 + 1$;
- $q_4(x) = x$.

(a) Prove that there exists a unique linear function $f : \mathbb{P}^3_{\mathbb{Z}_2} \to \mathbb{P}_{\mathbb{Z}_2}$ that satisfies the property that $f\big(p_i(x)\big) = q_i(x)$ for all indices $i \in \{1, 2, 3, 4\}$.

(b) Compute rank(f) and dim$\big(Ker(f)\big)$.

(c) If f one-to-one?

> ***Remark:*** *f is a linear function from a finite-dimensional vector space to an infinite-dimensional vector space, and so (by Theorem 4.2.14(b)) it is not onto, and therefore, it is not an isomorphism.*

(d) Find a formula for the linear function f, that is, fill in the blank in the following:

$$f\big(a_3x^3 + a_2x^2 + a_1x + a_0\big) = \_\_\_\_\_ \qquad \forall a_0, a_1, a_2, a_3 \in \mathbb{Z}_2.$$

Solution. In principle, we could proceed similarly as in Example 4.5.18. However, let us present a different approach, one that relies on Theorem 4.3.2 and on change of basis matrices. In what follows, we consider the basis $\mathcal{P}_3 = \{1, x, x^2, x^3\}$ for $\mathbb{P}^3_{\mathbb{Z}_2}$, and the basis $\mathcal{P}_4 = \{1, x, x^2, x^3, x^4\}$ for $\mathbb{P}^4_{\mathbb{Z}_2}$.

(a) It suffices to show that $\mathcal{B} := \big\{p_1(x), p_2(x), p_3(x), p_4(x)\big\}$ is a basis of $\mathbb{P}^3_{\mathbb{Z}_2}$, for Theorem 4.3.2 will then imply that there exist a unique linear function $f : \mathbb{P}^3_{\mathbb{Z}_2} \to \mathbb{P}_{\mathbb{Z}_2}$ satisfying $f\big(p_i(x)\big) = q_i(x)$ for all $i \in \{1, 2, 3, 4\}$, which is what we need to show. We form the matrix

$$\begin{aligned} B \;&:=\; \Big[\; \big[\; p_1(x) \;\big]_{\mathcal{P}_3} \quad \big[\; p_2(x) \;\big]_{\mathcal{P}_3} \quad \big[\; p_3(x) \;\big]_{\mathcal{P}_3} \quad \big[\; p_4(x) \;\big]_{\mathcal{P}_3} \;\Big] \\ &=\; \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

The matrix B is in row echelon form, and we immediately see that $\operatorname{rank}(B) = 4$. So, B is a square matrix of full rank, and so by Proposition 4.4.8(c), $\mathcal{B} = \big\{p_1(x), p_2(x), p_3(x), p_4(x)\big\}$ is a basis of $\mathbb{P}^3_{\mathbb{Z}_2}$, which is what we needed to show.

(b) Since polynomials $q_1(x), q_2(x), q_3(x), q_4(x)$ are all of degree at most four, we see that $\operatorname{Im}(f)$ is a subspace of $\mathbb{P}^4_{\mathbb{Z}_2}$.[53] So, let us temporarily consider f as having $\mathbb{P}^4_{\mathbb{Z}_2}$ for its codomain, so that we can form a suitable matrix of f,[54] as follows:

$$\begin{aligned} {}_{\mathcal{P}_4}\big[\; f \;\big]_{\mathcal{B}} \;&=\; \Big[\; \big[\; f\big(p_1(x)\big) \;\big]_{\mathcal{P}_4} \quad \cdots \quad \big[\; f\big(p_4(x)\big) \;\big]_{\mathcal{P}_4} \;\Big] \\ &=\; \Big[\; \big[\; q_1(x) \;\big]_{\mathcal{P}_4} \quad \big[\; q_2(x) \;\big]_{\mathcal{P}_4} \quad \big[\; q_3(x) \;\big]_{\mathcal{P}_4} \quad \big[\; q_4(x) \;\big]_{\mathcal{P}_4} \;\Big] \end{aligned}$$

[53] Here is a fully rigorous justification. Since polynomials $p_1(x), p_2(x), p_3(x), p_4(x)$ span $\mathbb{P}^3_{\mathbb{Z}_2}$ (because they form a basis of $\mathbb{P}^3_{\mathbb{Z}_2}$), we have that

$$\operatorname{Im}(f) \;\overset{(*)}{=}\; \operatorname{Span}\Big(f\big(p_1(x)\big), \ldots, f\big(p_4(x)\big)\Big) \;\overset{(**)}{=}\; \operatorname{Span}\big(q_1(x), \ldots, q_4(x)\big),$$

where (*) follows from Corollary 4.2.12, and (**) follows from the fact that $f\big(p_i(x)\big) = q_i(x)$ for all $i \in \{1, 2, 3, 4\}$. Since $q_1(x), \ldots, q_4(x) \in \mathbb{P}^4_{\mathbb{Z}_2}$, Theorem 3.1.11 now guarantees that $\operatorname{Im}(f)$ is a subspace of $\mathbb{P}^4_{\mathbb{Z}_2}$.

[54] Technically, we consider the linear function $f' : \mathbb{P}^3_{\mathbb{Z}_2} \to \mathbb{P}^4_{\mathbb{Z}_2}$ given by $f'(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{P}^3_{\mathbb{Z}_2}$, i.e. f' is the function obtained by restricting the codomain $\mathbb{P}_{\mathbb{Z}_2}$ of f to the subspace $\mathbb{P}^4_{\mathbb{Z}_2}$ of $\mathbb{P}_{\mathbb{Z}_2}$, which we can do since $\operatorname{Im}(f)$ is a subspace of $\mathbb{P}^4_{\mathbb{Z}_2}$. Obviously, f and f' have the same rank and kernel, and so we can compute with f' rather than f. However, in order to simplify notation, we continue using the notation f for our function with restricted codomain.

$$= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}\Big({}_{\mathcal{P}_4}\big[\ f\ \big]_{\mathcal{B}}\Big) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and so $\mathrm{rank}({}_{\mathcal{P}_4}\big[\ f\ \big]_{\mathcal{B}}) = 4$. Therefore, by Theorem 4.5.4(a), we have that $\mathrm{rank}(f) = 4$. For the kernel, we compute:

$$\dim\big(\mathrm{Ker}(f)\big) \overset{(*)}{=} \dim\big(\mathbb{P}^3_{\mathbb{Z}_2}\big) - \mathrm{rank}(f) = 4 - 4 = 0,$$

where (*) follows from the rank-nullity theorem.

(c) Since $\dim\big(\mathrm{Ker}(f)\big) = 0$, Theorem 4.2.4 guarantees that f is one-to-one.

(d) As in part (b), we will temporarily consider the codomain of f to be $\mathbb{P}^4_{\mathbb{Z}_2}$, so that we can compute a suitable matrix. The matrix that we need is ${}_{\mathcal{P}_4}\big[\ f\ \big]_{\mathcal{P}_3}$; once we have computed this matrix, we will easily be able to read off the formula for f. We have already computed the matrix ${}_{\mathcal{P}_4}\big[\ f\ \big]_{\mathcal{B}}$, and so we can compute as follows:

$$\begin{aligned} {}_{\mathcal{P}_4}\big[\ f\ \big]_{\mathcal{P}_3} &= {}_{\mathcal{P}_4}\Big[\ f \circ \mathrm{Id}_{\mathbb{P}^3_{\mathbb{Z}_2}}\ \Big]_{\mathcal{P}_3} \\ &= {}_{\mathcal{P}_4}\big[\ f\ \big]_{\mathcal{B}}\ {}_{\mathcal{B}}\Big[\ \mathrm{Id}_{\mathbb{P}^3_{\mathbb{Z}_2}}\ \Big]_{\mathcal{P}_3} && \text{by Theorem 4.5.3(c)} \\ &= {}_{\mathcal{P}_4}\big[\ f\ \big]_{\mathcal{B}}\ \Big({}_{\mathcal{P}_3}\Big[\ \mathrm{Id}_{\mathbb{P}^3_{\mathbb{Z}_2}}\ \Big]_{\mathcal{B}}\Big)^{-1} && \text{by Proposition 4.5.7.} \end{aligned}$$

We compute the change of basis matrix ${}_{\mathcal{P}_3}\Big[\ \mathrm{Id}_{\mathbb{P}^3_{\mathbb{Z}_2}}\ \Big]_{\mathcal{B}}$ as follows:

$${}_{\mathcal{P}_3}\Big[\ \mathrm{Id}_{\mathbb{P}^3_{\mathbb{Z}_2}}\ \Big]_{\mathcal{B}} \overset{(*)}{=} \Big[\ \big[\ p_1(x)\ \big]_{\mathcal{P}_4}\ \ \big[\ p_2(x)\ \big]_{\mathcal{P}_4}\ \ \big[\ p_3(x)\ \big]_{\mathcal{P}_4}\ \ \big[\ p_4(x)\ \big]_{\mathcal{P}_4}\ \Big]$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where (*) follows from Proposition 4.5.6 (or alternatively, from Theorem 4.5.1). We can find the inverse of this matrix by routine computation:

$$\Big({}_{\mathcal{P}_3}\Big[\ \mathrm{Id}_{\mathbb{P}^3_{\mathbb{Z}_2}} \ \Big]_{\mathcal{B}} \Big)^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We can now compute

$${}_{\mathcal{P}_4}\big[\ f \ \big]_{\mathcal{P}_3} = {}_{\mathcal{P}_4}\big[\ f \ \big]_{\mathcal{B}} \Big({}_{\mathcal{P}_3}\Big[\ \mathrm{Id}_{\mathbb{P}^3_{\mathbb{Z}_2}} \ \Big]_{\mathcal{B}} \Big)^{-1} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

Now, for all $a_0, a_1, a_2, a_3 \in \mathbb{Z}_2$, we have the following:

$$\begin{aligned} & \big[\ f\big(a_3x^3 + a_2x^2 + a_1x + a_0\big) \ \big]_{\mathcal{P}_4} \\ = \ & {}_{\mathcal{P}_4}\big[\ f \ \big]_{\mathcal{P}_3} \ \big[\ a_3x^3 + a_2x^2 + a_1x + a_0 \ \big]_{\mathcal{P}_3} \\ = \ & \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_2 + a_3 \\ a_3 \\ a_1 + a_3 \\ a_1 + a_2 \\ a_0 + a_1 \end{bmatrix} \\ = \ & \big[\ (a_0 + a_1)x^4 + (a_1 + a_2)x^3 + (a_1 + a_3)x^2 + a_3x + (a_2 + a_3) \ \big]_{\mathcal{P}_4}. \end{aligned}$$

Since $\big[\ \cdot \ \big]_{\mathcal{P}_4}$ is an isomorphism (and in particular, one-to-one), it follows that

$$\begin{aligned} f(a_3x^3 + a_2x^2 + a_1x + a_0) = \ & (a_0 + a_1)x^4 + (a_1 + a_2)x^3 + (a_1 + a_3)x^2 + \\ & + a_3x + (a_2 + a_3) \end{aligned}$$

for all $a_0, a_1, a_2, a_3, a_4 \in \mathbb{Z}_2$.

Optional: We can check that our formula for f is correct by verifying that it indeed satisfies the property that $f\big(p_i(x)\big) = q_i(x)$ for all indices $i \in \{1, 2, 3, 4\}$. Here, we

only compute this for $i = 4$ in order to demonstrate the general principle. The rest is similar routine computation. So, for $i = 4$, we compute:

$$\begin{aligned} f\big(p_4(x)\big) &= f(x^3 + x^2 + x + 1) \\ &= (1+1)x^4 + (1+1)x^3 + (1+1)x^2 + 1x + (1+1) \\ &= x \quad = \quad q_4(x), \end{aligned}$$

which is what we were supposed to get. □

Example 4.5.20. *Consider the following matrices with entries in $\mathbb{Z}_2$:*

- $M_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$;
- $M_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$;
- $M_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$;
- $M_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$;
- $N_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$;
- $N_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$;
- $N_3 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$;
- $N_4 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.

Determine if there exists a linear function $f : \mathbb{Z}_2^{2\times 2} \to \mathbb{Z}_2^{2\times 3}$ such that $f(M_i) = N_i$ for all $i \in \{1, 2, 3, 4\}$. If such a linear function f exists, determine if it is unique, and if it is not, determine the number of such linear functions f.

Remark: In this particular case, it is not very hard to see that f does not exist. Indeed, we can see that $M_3 = M_1 + M_2$, and so any linear function $f : \mathbb{Z}_2^{2\times 2} \to \mathbb{Z}_2^{2\times 3}$ satisfying $f(M_1) = N_1$ and $f(M_2) = N_2$ must also satisfy

$$f(M_3) \quad = \quad f(M_1 + M_2) \quad \overset{(*)}{=} \quad f(M_1) + f(M_2) \quad = \quad N_1 + N_2 \quad \neq \quad N_3,$$

where (*) follows from the linearity of f. However, to illustrate the general principle (which we can use in those situations when the non-existence of the function in question is not quite so obvious, and also when the function with the given specifications does in fact exist), we give two different solutions. The first solution relies on matrices of linear functions with respect to convenient bases,[55] and the second one relies on Corollary 4.3.3.[56]

[55] This solution is similar to our solution of Example 4.5.18(a).
[56] This solution is similar to our solution of Example 4.5.19(a).

Solution#1. We proceed as in our solution to Example 4.5.18(a). We set

$$A_1 := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad A_3 := \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad A_4 := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

and we further set

$$B_1 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_2 := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_3 := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$B_4 := \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad B_5 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad B_6 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In our solution, we will use the basis $\mathcal{A} := \{A_1, A_2, A_3, A_4\}$ of $\mathbb{Z}_2^{2\times 2}$ and the basis $\mathcal{B} := \{B_1, B_2, B_3, B_4, B_5, B_6\}$ of $\mathbb{Z}_2^{2\times 3}$. Instead of directly solving for the linear function $f : \mathbb{Z}_2^{2\times 2} \to \mathbb{Z}_2^{2\times 3}$ satisfying $f(M_i) = N_i$ for all $i \in \{1, 2, 3, 4\}$, we will solve for the matrix ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{A}}$ in $\mathbb{Z}_2^{6\times 4}$ satisfying

$${}_{\mathcal{B}}[\ f\ ]_{\mathcal{A}} \ [\ M_i\ ]_{\mathcal{A}} \quad = \quad [\ N_i\ ]_{\mathcal{B}}$$

for all $i \in \{1, 2, 3, 4\}$. This is equivalent to

$${}_{\mathcal{B}}[\ f\ ]_{\mathcal{A}} \underbrace{[\ [\ M_1\ ]_{\mathcal{A}} \ \cdots \ [\ M_4\ ]_{\mathcal{A}}\ ]}_{=:M} \quad = \quad \underbrace{[\ [\ N_1\ ]_{\mathcal{A}} \ \cdots \ [\ N_4\ ]_{\mathcal{A}}\ ]}_{=:N}.$$

Matrices M and N can easily be computed, whereas the matrix ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{A}}$ is the unknown that we need to solve for. We proceed as in subsection 1.9.2. We first take the transpose of both sides of the equation above, and we obtain

$$M^T \Big({}_{\mathcal{B}}[\ f\ ]_{\mathcal{A}} \Big)^T \quad = \quad N^T,$$

which we solve for $\Big({}_{\mathcal{B}}[\ f\ ]_{\mathcal{A}} \Big)^T$. We form the matrix

$$\left[\begin{array}{c:c} M^T & N^T \end{array}\right] \quad = \quad \left[\begin{array}{c:c} [\ M_1\ ]_{\mathcal{A}}^T & [\ N_1\ ]_{\mathcal{B}}^T \\ \vdots & \vdots \\ [\ M_4\ ]_{\mathcal{A}}^T & [\ N_4\ ]_{\mathcal{B}}^T \end{array}\right]$$

$$= \quad \left[\begin{array}{cccc:cccccc} 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \end{array}\right],$$

and we row reduce to obtain

$$\mathrm{RREF}\Big(\left[\begin{array}{c|c} M^T & N^T \end{array}\right]\Big) \quad = \quad \left[\begin{array}{cccc|cccccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{array}\right].$$

Because of the fourth row of $\mathrm{RREF}\Big(\left[\begin{array}{c|c} M^T & N^T \end{array}\right]\Big)$,[57] we see that the equation $M^T\Big({}_{\mathcal{B}}\left[\ f\ \right]_{\mathcal{A}}\Big)^T = N^T$ has no solutions for $\Big({}_{\mathcal{B}}\left[\ f\ \right]_{\mathcal{A}}\Big)^T$. Consequently, the equation ${}_{\mathcal{B}}\left[\ f\ \right]_{\mathcal{A}}\ M = N$ has no solutions for ${}_{\mathcal{B}}\left[\ f\ \right]_{\mathcal{A}}$. This implies that there is no linear function $f : \mathbb{Z}_2^{2\times 2} \to \mathbb{Z}_2^{2\times 3}$ satisfying $f(M_i) = N_i$ for all $i \in \{1, 2, 3, 4\}$. □

Solution#2. We proceed similarly as in our solution to Example 4.5.19(a), except that instead of using Theorem 4.3.2, we will use Corollary 4.3.3. Our first goal is to find a basis of $U := \mathrm{Span}(M_1, M_2, M_3, M_4)$, and to express those M_i's that do not belong to this basis as linear combinations of the basis vectors (matrices). We proceed similarly as in subsection 4.4.3.[58]

Consider the basis

$$\mathcal{A} \ := \ \left\{ \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 0 \\ 1 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array}\right] \right\}$$

of $\mathbb{Z}_2^{2\times 2}$. We now form the matrix

$$A \ := \ \left[\ \left[\ M_1\ \right]_{\mathcal{A}} \ \ldots \ \left[\ M_4\ \right]_{\mathcal{A}}\ \right] \ = \ \left[\begin{array}{cccc} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}(A) \quad = \quad \left[\begin{array}{cccc} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

We see that the pivot columns of A are its first, second, and fourth column, and so $\mathcal{B} := \{M_1, M_2, M_4\}$ is a basis of $U = \mathrm{Span}(M_1, M_2, M_3, M_4)$. Moreover, we see from $\mathrm{RREF}(A)$ that $M_3 = M_1 + M_2$.

[57] The point is that the fourth row is of the form $\left[\begin{array}{c|ccc} 0 & \ldots & 0 & * \ \ldots \ * \end{array}\right]$, where at least one of the $*$'s is non-zero.

[58] See Examples 4.4.11, 4.4.12, and 4.4.13.

Now, since $\mathcal{B} = \{M_1, M_2, M_4\}$ is a linearly independent set in $\mathbb{Z}_2^{2\times 2}$,[59] Corollary 4.3.3 guarantees that there exists a linear function $f : \mathbb{Z}_2^{2\times 2} \to \mathbb{Z}_2^{2\times 3}$ such that $f(M_i) = N_i$ for all $i \in \{1, 2, 4\}$. Any such f must further satisfy

$$f(M_3) \overset{(*)}{=} f(M_1 + M_2) \overset{(**)}{=} f(M_1) + f(M_2) \overset{(***)}{=} N_1 + N_2 \neq N_3,$$

where (*) follows from the fact that $M_3 = M_1 + M_2$, (**) follows from the linearity of f, and (***) follows from the fact that $f(M_i) = N_i$ for all $i \in \{1, 2, 4\}$.

We can now conclude that there is no linear function $f : \mathbb{Z}_2^{2\times 2} \to \mathbb{Z}_2^{2\times 3}$ satisfying $f(M_i) = N_i$ for all $i \in \{1, 2, 3, 4\}$. □

Example 4.5.21. *Consider the following polynomials and vectors, the former with coefficients in $\mathbb{Z}_3$, and the latter with entries in $\mathbb{Z}_3$:*

- $p_1(x) = x^2 + x + 1$;
- $p_2(x) = 2x + 1$;
- $p_3(x) = 2x^2 + 1$;
- $p_4(x) = x + 2$;
- $p_5(x) = x^2 + 2x$;
- $\mathbf{v}_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$;
- $\mathbf{v}_2 = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$;
- $\mathbf{v}_3 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$;
- $\mathbf{v}_4 = \begin{bmatrix} 1 & 2 \end{bmatrix}^T$;
- $\mathbf{v}_5 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$.

Determine if there exists a linear function $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ such that $f\big(p_i(x)\big) = \mathbf{v}_i$ for all $i \in \{1, \ldots, 5\}$. If such a linear function f exists, determine if it is unique, and if it is not, determine the number of such linear functions f.

Remark: As in the case of Example 4.5.20, we give two solutions: the first one relies on matrices of linear functions with respect to convenient bases, and the second one relies on Corollary 4.3.3.

Solution#1. In our solution, we will use the basis $\mathcal{P} = \{1, x, x^2\}$ of $\mathbb{P}^2_{\mathbb{Z}_3}$, and the standard basis $\mathcal{E}_2 = \{\mathbf{e}_1, \mathbf{e}_2\}$ of $\mathbb{Z}_3^2$. Instead of directly solving for the linear function $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big) = \mathbf{v}_i$ for all $i \in \{1, \ldots, 5\}$, we will solve for the matrix ${}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}}$ in $\mathbb{Z}_3^{2\times 3}$ satisfying

$${}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}} \ \big[\ p_i(x)\ \big]_{\mathcal{P}} = \underbrace{\big[\ \mathbf{v}_i\ \big]_{\mathcal{E}_2}}_{=\mathbf{v}_i}$$

[59] Since $\mathcal{B}$ is a basis of the subspace U of $\mathbb{Z}_2^{2\times 2}$, we know that $\mathcal{B}$ is, in particular, a linearly independent set in $\mathbb{Z}_2^{2\times 2}$.

for all $i \in \{1, \ldots, 5\}$. This is equivalent to

$$ {}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}} \underbrace{\begin{bmatrix} \big[\ p_1(x)\ \big]_{\mathcal{P}} & \ldots & \big[\ p_5(x)\ \big]_{\mathcal{P}} \end{bmatrix}}_{=:P} \;=\; \underbrace{\begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_5 \end{bmatrix}}_{=:M}. $$

Matrices P and M can easily be computed, whereas the matrix ${}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}}$ is the unknown that we need to solve for. We take the transpose of both sides of the equation above, and we obtain

$$ P^T \Big({}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}} \Big)^T \;=\; M^T, $$

which we solve for $\Big({}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}} \Big)^T$. We form the matrix

$$ \left[\ P^T \,\middle|\, M^T\ \right] \;=\; \left[\begin{array}{c|c} \big[\ p_1(x)\ \big]_{\mathcal{P}}^T & \mathbf{v}_1^T \\ \vdots & \vdots \\ \big[\ p_5(x)\ \big]_{\mathcal{P}}^T & \mathbf{v}_5^T \end{array}\right] \;=\; \left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 1 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 0 & 2 & 1 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 0 & 2 & 1 & 1 & 0 \end{array}\right], $$

and by row reducing, we obtain

$$ \mathrm{RREF}\Big(\left[\ P^T \,\middle|\, M^T\ \right] \Big) \;=\; \left[\begin{array}{ccc|cc} 1 & 0 & 2 & 1 & 1 \\ 0 & 1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]. $$

We can now read off the general solution for $\Big({}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}} \Big)^T$:

$$ \Big({}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}} \Big)^T \;=\; \begin{bmatrix} t_1 + 1 & t_2 + 1 \\ t_1 + 2 & t_2 \\ t_1 & t_2 \end{bmatrix}, \qquad \text{where } t_1, t_2 \in \mathbb{Z}_3. $$

By taking the transpose, we obtain the general solution for ${}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}}$, as follows:

$$ {}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}} \;=\; \begin{bmatrix} t_1 + 1 & t_1 + 2 & t_1 \\ t_2 + 1 & t_2 & t_2 \end{bmatrix}, \qquad \text{where } t_1, t_2 \in \mathbb{Z}_3. $$

Since we got more than one solution for the matrix ${}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}}$, we deduce that there exists a linear function $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big) = \mathbf{v}_i$ for all $i \in \{1, \ldots, 5\}$, but that such a linear function f is **not** unique.

It remains to determine the exact number of linear functions f with the given

specifications. Our general solution for the matrix ${}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}}$ has two parameters, each of which can take any value from $\mathbb{Z}_3$. So, the number of solutions for ${}_{\mathcal{E}_2}\big[\ f\ \big]_{\mathcal{P}}$ is $3^2 = 9$, and consequently, there are exactly nine linear functions $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big) = \mathbf{v}_i$ for all $i \in \{1, \ldots, 5\}$. □

Solution#2. Our first goal is to find a basis of $U := \text{Span}\big(p_1(x), \ldots, p_5(x)\big)$, and to express those $p_i(x)$'s that do not belong to this basis as linear combinations of the basis vectors (polynomials). Consider the basis $\mathcal{P} = \{1, x, x^2\}$ of $\mathbb{P}^2_{\mathbb{Z}_3}$. We form the matrix

$$P \ := \ \big[\ \big[\ p_1(x)\ \big]_{\mathcal{P}} \ \ \ldots \ \ \big[\ p_5(x)\ \big]_{\mathcal{P}}\ \big] \ = \ \begin{bmatrix} 1 & 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 1 & 2 \\ 1 & 0 & 2 & 0 & 1 \end{bmatrix},$$

and by row reducing, we obtain

$$\text{RREF}(P) \ = \ \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We see from RREF(P) that the pivot columns of P are its first and second column. So, $\mathcal{B} := \big\{p_1(x), p_2(x)\big\}$ is a basis of $U = \text{Span}\big(p_1(x), \ldots, p_5(x)\big)$. Moreover, we see from RREF(P) that all the following hold:

- $p_3(x) = 2p_1(x) + 2p_2(x)$;
- $p_4(x) = 2p_2(x)$;
- $p_5(x) = p_1(x) + 2p_2(x)$.

Now, $\mathcal{B} = \big\{p_1(x), p_2(x)\big\}$ is a linearly independent set in $\mathbb{P}^2_{Z_3}$,[60] but since $\mathcal{B}$ contains only two vectors (polynomials) and $\dim(\mathbb{P}^2_{\mathbb{Z}_3}) = 3$, we see that $\mathcal{B}$ is **not** a basis of $\mathbb{P}^2_{\mathbb{Z}_3}$. Moreover, $\mathbb{Z}_3^2$ is non-trivial. Corollary 4.3.3 now guarantees that there exists a linear function $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ that satisfies $f\big(p_i(x)\big) = \mathbf{v}_i$ for each $i \in \{1, 2\}$, but that such a linear function f is not unique.

However, we are not done yet! Since $p_3(x), p_4(x), p_5(x)$ are linear combinations of the polynomials $p_1(x), p_2(x)$, any linear function $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ that satisfies $f\big(p_i(x)\big) = \mathbf{v}_i$ for each $i \in \{1, 2\}$ has a fully determined output for $f\big(p_i(x)\big)$ for all $i \in \{3, 4, 5\}$,[61] and we need to check whether those values are the ones from the statement of the example. Let us compute. For all linear functions $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big) = \mathbf{v}_i$ for each $i \in \{1, 2\}$, the following hold:

[60] This is because $\mathcal{B}$ is a basis of the subspace $U = \text{Span}\big(p_1(x), \ldots, p_5(x)\big)$ of $\mathbb{P}^2_{\mathbb{Z}_3}$.
[61] This output is fully determined by the values of $f\big(p_1(x)\big)$ and $f\big(p_2(x)\big)$.

$$\begin{aligned}
f\big(p_3(x)\big) &= f\big(2p_1(x)+2p_2(x)\big) && \text{because } p_3(x)=2p_1(x)+2p_2(x)\\
&= 2f\big(p_1(x)\big)+2f\big(p_2(x)\big) && \text{because } f \text{ is linear}\\
&= 2\mathbf{v}_1+2\mathbf{v}_2 && \text{because } f\big(p_i(x)\big)=\mathbf{v}_i \ \forall i\in\{1,2\}\\
&= \mathbf{v}_3 && \text{because } \mathbf{v}_3=2\mathbf{v}_1+2\mathbf{v}_2;
\end{aligned}$$

$$\begin{aligned}
f\big(p_4(x)\big) &= f\big(2p_2(x)\big) && \text{because } p_4(x)=2p_2(x)\\
&= 2f\big(p_2(x)\big) && \text{because } f \text{ is linear}\\
&= 2\mathbf{v}_2 && \text{because } f\big(p_i(x)\big)=\mathbf{v}_i \ \forall i\in\{1,2\}\\
&= \mathbf{v}_4 && \text{because } \mathbf{v}_4=2\mathbf{v}_2;
\end{aligned}$$

$$\begin{aligned}
f\big(p_5(x)\big) &= f\big(p_1(x)+2p_2(x)\big) && \text{because } p_5(x)=p_1(x)+2p_2(x)\\
&= f\big(p_1(x)\big)+2f\big(p_2(x)\big) && \text{because } f \text{ is linear}\\
&= \mathbf{v}_1+2\mathbf{v}_2 && \text{because } f\big(p_i(x)\big)=\mathbf{v}_i \ \forall i\in\{1,2\}\\
&= \mathbf{v}_5 && \text{because } \mathbf{v}_5=\mathbf{v}_1+2\mathbf{v}_2.
\end{aligned}$$

So, we got that $f\big(p_i(x)\big)=\mathbf{v}_i$ for all $i\in\{3,4,5\}$, which is consistent with the specifications from the statement of the example. We can now conclude that there exists a linear function $f:\mathbb{P}^2_{\mathbb{Z}_3}\to\mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big)=\mathbf{v}_i$ for all $i\in\{1,\dots,5\}$, but that such a linear function f is **not** unique.

It remains to determine the number of linear functions $f:\mathbb{P}^2_{\mathbb{Z}_3}\to\mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big)=\mathbf{v}_i$ for all $i\in\{1,\dots,5\}$. We saw above that $\mathcal{B}=\big\{p_1(x),p_2(x)\big\}$ is a basis of $U=\mathrm{Span}\big(p_1(x),\dots,p_5(x)\big)$. In particular, $\mathcal{B}$ is a linearly independent set in the 3-dimensional vector space $\mathbb{P}^2_{\mathbb{Z}_3}$. So, by Theorem 3.2.19, $\mathcal{B}$ can be extended to some basis $\mathcal{C}=\big\{p_1(x),p_2(x),q(x)\big\}$ of $\mathbb{P}^2_{\mathbb{Z}_3}$.[62] By Theorem 4.3.2, for all $\mathbf{v}\in\mathbb{Z}_3^2$, there exists a unique linear function $f_{\mathbf{v}}:\mathbb{P}^2_{\mathbb{Z}_3}\to\mathbb{Z}_3^2$ that satisfies $f_{\mathbf{v}}\big(p_1(x)\big)=\mathbf{v}_1$, $f_{\mathbf{v}}\big(p_2(x)\big)=\mathbf{v}_2$, and $f_{\mathbf{v}}\big(q(x)\big)=\mathbf{v}$, and by our argument above, this function $f_{\mathbf{v}}$ also satisfies $f_{\mathbf{v}}\big(p_i(x)\big)=\mathbf{v}_i$ for all $i\in\{3,4,5\}$.[63] Thus, the number of linear functions $f:\mathbb{P}^2_{\mathbb{Z}_3}\to\mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big)=\mathbf{v}_i$ for all $i\in\{1,\dots,5\}$ is equal to the number of

[62] Since $\dim(\mathbb{P}^2_{\mathbb{Z}_3})=3$, any basis of $\mathbb{P}^2_{\mathbb{Z}_3}$ contains exactly three vectors (polynomials).

[63] Indeed, we showed above that any linear function $f:\mathbb{P}^2_{\mathbb{Z}_3}\to\mathbb{Z}_3^2$ that satisfies $f\big(p_1(x)\big)=\mathbf{v}_1$ and $f\big(p_2(x)\big)=\mathbf{v}_2$, in fact satisfies $f\big(p_i(x)\big)=\mathbf{v}_i$ for all $i\in\{1,\dots,5\}$.

vectors $\mathbf{v}$ in $\mathbb{Z}_3^2$, which is 9. So, our final answer is that the there are exactly nine linear functions $f : \mathbb{P}^2_{\mathbb{Z}_3} \to \mathbb{Z}_3^2$ satisfying $f\big(p_i(x)\big) = \mathbf{v}_i$ for all $i \in \{1, \ldots, 5\}$. □

Our final example of this subsection (see Example 4.5.22 below) is slightly more complicated. As we shall see, the linear function satisfying the specifications from that example is not unique. Moreover, we will need to construct examples of linear functions satisfying those specifications, and also having all possible ranks (that is, possible subject to the constraints of the example). Our solution will combine methods from the solutions of Examples 4.5.18 and 4.5.19 (above), and it will also use some other theoretical results that we have obtained so far in these lecture notes.

Example 4.5.22. *Consider the following polynomials with coefficients in $\mathbb{Z}_3$:*

- $p_1(x) = x^3 + 1$;
- $p_2(x) = 2x^3 + 2$;
- $p_3(x) = x^2 + 2x + 1$
- $p_4(x) = 2x^3 + x^2 + 2x$.

Further, consider the following matrices with entries understood to be in $\mathbb{Z}_3$:

- $M_1 = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$;
- $M_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$;
- $M_3 = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}$;
- $M_4 = \begin{bmatrix} 2 & 0 \\ 2 & 2 \end{bmatrix}$.

Prove that there exists a linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ such that $f\big(p_i(x)\big) = M_i$ for all indices $i \in \{1, 2, 3, 4\}$. What are all the possible ranks that such a linear function f can have? For each possible rank, find a formula for one linear function f that has that rank (and satisfies the specifications above).[64] *Can f be an isomorphism? If so, find a formula for one such isomorphism,*[65] *and also for its inverse.*

Solution. In our solution, we will use the basis $\mathcal{P} = \{1, x, x^2, x^3\}$ of $\mathbb{P}^3_{\mathbb{Z}_3}$, as well as the basis $\mathcal{M} = \{A_1, A_2, A_3, A_4\}$ of $\mathbb{Z}_3^{2\times 2}$, where

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad A_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \qquad A_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

First of all, we saw in our solution to Example 4.4.14 that $\mathcal{B}_U := \big\{p_1(x), p_3(x)\big\}$ is a basis of $U := \operatorname{Span}\big(p_1(x), p_2(x), p_3(x), p_4(x)\big)$, and that $\mathcal{B} := \big\{p_1(x), p_3(x), 1, x\big\}$ is a basis of $\mathbb{P}^3_{\mathbb{Z}_3}$ that extends $\mathcal{B}_U$. We also saw that

[64] There may be more than one correct answer. However, we are not asked to find all possible correct answers: any one will do.

[65] Once again, there may be more than one correct answer. We are not asked to find all possible answers: any one will do (provided it exists to begin with).

- $p_2(x) = 2p_1(x)$,
- $p_4(x) = 2p_1(x) + p_3(x)$.

Now, $\mathcal{B}_U$ is a linearly independent set in $\mathbb{P}^3_{\mathbb{Z}_3}$. So, by Corollary 4.3.3, there exists a linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies $f\big(p_1(x)\big) = M_1$ and $f\big(p_3(x)\big) = M_3$. Moreover, since the linearly independent set $\mathcal{B}_U$ is not a basis of the domain $\mathbb{P}^3_{\mathbb{Z}_3}$,[66] and since the codomain $\mathbb{Z}_3^{2\times 2}$ is non-trivial, Corollary 4.3.3 guarantees that such a linear function f is **not** unique.

> **Claim.** If a linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ satisfies $f\big(p_1(x)\big) = M_1$ and $f\big(p_3(x)\big) = M_3$, then it in fact satisfies $f\big(p_i(x)\big) = M_i$ for all $i \in \{1, 2, 3, 4\}$.

Proof of the Claim. Fix a linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies $f\big(p_1(x)\big) = M_1$ and $f\big(p_3(x)\big) = M_3$. Then f also satisfies the following:

$$
\begin{array}{rcll}
f\big(p_2(x)\big) & = & f\big(2p_1(x)\big) & \text{because } p_2(x) = 2p_1(x) \\
& = & 2f\big(p_1(x)\big) & \text{because } f \text{ is linear} \\
& = & 2M_1 & \text{because } f\big(p_1(x)\big) = M_1 \\
& = & M_2 & \text{because } M_2 = 2M_1;
\end{array}
$$

$$
\begin{array}{rcll}
f\big(p_4(x)\big) & = & f\big(2p_1(x) + p_3(x)\big) & \text{because } p_4(x) = 2p_1(x) + p_3(x) \\
& = & 2f\big(p_1(x)\big) + f\big(p_3(x)\big) & \text{because } f \text{ is linear} \\
& = & 2M_1 + M_3 & \text{because } f\big(p_1(x)\big) = M_1 \\
& & & \text{and } f(\big(p_3(x)\big) = M_3 \\
& = & M_4 & \text{because } M_4 = 2M_1 + M_3.
\end{array}
$$

This proves the Claim. ♦

Now, we proved above that there exists a linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies $f\big(p_1(x)\big) = M_1$ and $f\big(p_3(x)\big) = M_3$, and we saw that such an f is not unique. In view of the Claim, it now follows that there exists a linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies $f\big(p_i(x)\big) = M_i$ for all indices $i \in \{1, 2, 3, 4\}$, and that such a linear function f is **not** unique.

[66] This is because $\dim(\mathbb{P}^3_{\mathbb{Z}_3}) = 4$, but $\mathcal{B}_U$ only contains two elements.

Remark: Using Corollary 4.3.3, we were able to show that there exists a linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies $f\big(p_1(x)\big) = M_1$ and $f\big(p_3(x)\big) = M_3$. However, we could not just forget about $p_2(x)$ and $p_4(x)$! Since $p_2(x)$ and $p_4(x)$ are linear combinations of the vectors (polynomials) $p_1(x)$ and $p_3(x)$, and since f is linear, the values of $f\big(p_2(x)\big)$ and $f\big(p_4(x)\big)$ are fully determined by the values of $f\big(p_1(x)\big)$ and $f\big(p_3(x)\big)$. We had to check (see the Claim above) that the values of $f\big(p_2(x)\big)$ and $f\big(p_4(x)\big)$ determined by the values of $f\big(p_1(x)\big)$ and $f\big(p_3(x)\big)$ are those prescribed by the statement of the example, namely, $f\big(p_2(x)\big) = M_2$ and $f\big(p_4(x)\big) = M_4$. If it had turned out that $f\big(p_2(x)\big) \neq M_2$ or $f\big(p_4(x)\big) \neq M_4$, then we would have concluded that no linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ satisfying $f\big(p_i(x)\big) = M_i$ for all $i \in \{1, 2, 3, 4\}$ exists. However, now that we have proven the existence of f (and have proven the Claim above), polynomials $p_2(x), p_4(x)$ and matrices M_2, M_4 play no further role in our computation; this is essentially because $p_2(x), p_4(x)$ do not belong to our basis $\mathcal{B}_U$ of $U = \mathrm{Span}\big(p_1(x), p_2(x), p_3(x), p_4(x)\big)$.

We now find the general solution for the matrix ${}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}}$, where $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ is a linear function that satisfies $f\big(p_1(x)\big) = M_1$ and $f\big(p_3(x)\big) = M_3$, and consequently (by the Claim), also satisfies $f\big(p_2(x)\big) = M_2$ and $f\big(p_4(x)\big) = M_4$. The matrix ${}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}}$ must satisfy

- ${}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \ \big[\ p_1(x)\ \big]_{\mathcal{P}} = \big[\ M_1\ \big]_{\mathcal{M}}$;
- ${}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \ \big[\ p_3(x)\ \big]_{\mathcal{P}} = \big[\ M_3\ \big]_{\mathcal{M}}$.

This is equivalent to

$$ {}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \underbrace{\big[\ \big[\ p_1(x)\ \big]_{\mathcal{P}} \ \big[\ p_3(x)\ \big]_{\mathcal{P}}\ \big]}_{=:P} = \underbrace{\big[\ \big[\ M_1\ \big]_{\mathcal{M}} \ \big[\ M_3\ \big]_{\mathcal{M}}\ \big]}_{=:M}. $$

We now take the transpose of both sides of the equation above, and we get

$$ P^T\Big({}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}}\Big)^T = M^T. $$

We now form the matrix

$$ \big[\ P^T \,\vdots\, M^T\ \big] = \left[\begin{array}{c:c} \big[\ p_1(x)\ \big]^T_{\mathcal{P}} & \big[\ M_1\ \big]^T_{\mathcal{M}} \\ \big[\ p_3(x)\ \big]^T_{\mathcal{P}} & \big[\ M_3\ \big]^T_{\mathcal{M}} \end{array}\right] = \left[\begin{array}{cccc:cccc} 1 & 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 & 2 & 2 \end{array}\right], $$

and by row reducing, we obtain

$$ \mathrm{RREF}\Big(\big[\ P^T \,\vdots\, M^T\ \big]\Big) = \left[\begin{array}{cccc:cccc} 1 & 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 & 1 & 1 \end{array}\right]. $$

From the matrix above, we can read off

$$\Big({}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}}\Big)^T \quad = \quad \begin{bmatrix} 2t_1+2 & 2t_2 & 2t_3 & 2t_4 \\ s_1+2t_1+1 & s_2+2t_2 & s_3+2t_3+1 & s_4+2t_4+1 \\ s_1 & s_2 & s_3 & s_4 \\ t_1 & t_2 & t_3 & t_4 \end{bmatrix},$$

where $s_1, s_2, s_3, s_4, t_1, t_2, t_3, t_4 \in \mathbb{Z}_3$. By taking the transpose, we get

$${}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \quad = \quad \begin{bmatrix} 2t_1+2 & s_1+2t_1+1 & s_1 & t_1 \\ 2t_2 & s_2+2t_2 & s_2 & t_2 \\ 2t_3 & s_3+2t_3+1 & s_3 & t_3 \\ 2t_4 & s_4+2t_4+1 & s_4 & t_4 \end{bmatrix},$$

where $s_1, s_2, s_3, s_4, t_1, t_2, t_3, t_4 \in \mathbb{Z}_3$. Now, we know that

$$\operatorname{rank}(f) \overset{(*)}{=} \operatorname{rank}\Big({}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}}\Big) \overset{(**)}{=} \operatorname{rank}\Big(\Big({}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}}\Big)^T\Big).$$

where (*) follows from Theorem 4.5.4(a), and (**) follows from Corollary 3.3.11. We could now try to row reduce the matrix ${}_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}}$ (or its transpose) in order to identify what rank it has for various values of our parameters $s_1, s_2, s_3, s_4, t_1, t_2, t_3, t_4$. However, because there are so many parameters, this would be fairly messy.[67] Instead, we will use Theorem 4.3.2 and Corollary 4.2.12. Since $\mathcal{B} = \{p_1(x), p_3(x), 1, x\}$ is a basis of the domain $\mathbb{P}^3_{\mathbb{Z}_3}$, we have that

$$\begin{aligned} \operatorname{rank}(f) &\overset{(*)}{=} \dim\Big(\operatorname{Span}\big(f(p_1(x)), f(p_3(x)), f(1), f(x)\big)\Big) \\ &\overset{(**)}{=} \dim\Big(\operatorname{Span}\big(M_1, M_3, f(1), f(x)\big)\Big) \end{aligned}$$

where (*) follows from Corollary 4.2.12, and (**) follows from the fact that $f\big(p_1(x)\big) = M_1$ and $f\big(p_3(x)\big) = M_3$.

By Theorem 3.2.14, we know that some subset of $\{M_1, M_3, f(1), f(x)\}$ will form a basis of $\operatorname{Span}\big(M_1, M_3, f(1), f(x)\big)$, and the size of that basis determines the dimension of $\operatorname{Span}\big(M_1, M_3, f(1), f(x)\big)$, and consequently, it determines $\operatorname{rank}(f)$. So, the question is how many vectors (matrices) out of $M_1, M_3, f(1), f(x)$ can possibly be linearly independent. Since $\mathcal{B} = \{p_1(x), p_3(x), 1, x\}$ is a basis of the domain $\mathbb{P}^3_{\mathbb{Z}_3}$, Theorem 4.3.2 guarantees that we can choose the values of $f(1)$ and $f(x)$ arbitrarily.[68] The question is how those choices modify our rank.

We now proceed as follows. We set $V := \operatorname{Span}(M_1, M_3)$, we find a subset of $\{M_1, M_3\}$ that is a basis $\mathcal{C}_V$ of V, and we extend $\mathcal{C}_V$ to a basis $\mathcal{C}$ of $\mathbb{Z}_3^{2\times 2}$. We know that $\mathcal{M} = \{A_1, A_2, A_3, A_4\}$ is a basis of $\mathbb{Z}_3^{2\times 2}$. So, we form the matrix

[67] Try and see!

[68] More precisely, Theorem 4.3.2 guarantees that for all matrices $N_1, N_2 \in \mathbb{Z}_3^{2\times 2}$, there exists a unique linear function $f : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ such that $f\big(p_1(x)\big) = M_1$, $f\big(p_3(x)\big) = M_3$, $f(1) = N_1$, $f(x) = N_2$. By the Claim, any such f will in fact satisfy $f\big(p_i(x)\big) = M_i$ for all $i \in \{1, 2, 3, 4\}$.

$$C = \left[\left[M_1 \right]_{\mathcal{M}} \left[M_3 \right]_{\mathcal{M}} \middle| \left[A_1 \right]_{\mathcal{M}} \left[A_2 \right]_{\mathcal{M}} \left[A_3 \right]_{\mathcal{M}} \left[A_4 \right]_{\mathcal{M}} \right]$$
$$= \left[\begin{array}{cc|cccc} 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right],$$

and by row reducing, we obtain

$$\operatorname{RREF}(C) = \left[\begin{array}{cc|cccc} 1 & 0 & 2 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right].$$

We see that the pivot columns of C are its first, second, fourth, and fifth column. So, $\mathcal{C}_V := \{M_1, M_3\}$ is a basis of $V = \operatorname{Span}(M_1, M_3)$, and $\mathcal{C} := \{M_1, M_3, A_2, A_3\}$ is a basis of $\mathbb{Z}_3^{2\times 2}$ that extends $\mathcal{C}_V$. We now have that

$$2 \overset{(*)}{\leq} \underbrace{\dim\Big(\operatorname{Span}\big(M_1, M_3, f(1), f(x)\big)\Big)}_{=\operatorname{rank}(f)} \overset{(**)}{\leq} 4,$$

where (*) follows from the fact that M_1 and M_3 are linearly independent and from Theorem 3.2.17(a),[69] and (**) follows from Theorem 3.2.14.[70] We claim that the possible ranks for f are precisely $2, 3, 4$. The inequality above proves that there are no other possible values for $\operatorname{rank}(f)$. To show that each of those ranks is indeed a possibility, we will exhibit a linear function f satisfying the specifications from the statement of the example, and having that rank. We will do this by varying the values of $f(1)$ and $f(x)$, which Theorem 4.3.2 allows us to do freely. We will first create the functions in question, and we will compute their formulas later. As usual, $O_{2\times 2} = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right]$.

Rank 2: Since $\mathcal{B} = \{p_1(x), p_2(x), 1, x\}$ is a basis of $\mathbb{P}^3_{\mathbb{Z}_3}$, Theorem 4.3.2 guarantees that there exists a unique linear function $f_2 : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies the following:

- $f_2\big(p_1(x)\big) = M_1$,
- $f_2\big(p_3(x)\big) = M_3$,
- $f_2(1) = O_{2\times 2}$,
- $f_2(x) = O_{2\times 2}$.

[69] The fact that M_1, M_3 are linearly independent follows from the fact that $\mathcal{C}_V = \{M_1, M_3\}$ is a basis of $V = \operatorname{Span}(M_1, M_3)$.

[70] Indeed, by Theorem 3.2.14, some subset of the spanning set $\big\{M_1, M_3, f(1), f(x)\big\}$ of the vector space $\operatorname{Span}\big(M_1, M_3, f(1), f(x)\big)$ is a basis of that vector space. Obviously, such a basis contains at most four vectors (because our spanning set contains only four vectors). So, $\dim\Big(\operatorname{Span}\big(M_1, M_3, f(1), f(x)\big)\Big) \leq 4$.

By the Claim, our linear function f_2 in fact satisfies $f_2\big(p_i(x)\big) = M_i$ for all $i \in \{1,2,3,4\}$. But now

$$\text{rank}(f_2) \quad = \quad \dim\Big(\text{Span}\big(M_1, M_3, O_{2\times 2}, O_{2\times 2}\big)\Big) \quad \overset{(*)}{=} \quad 2,$$

where (*) follows from the fact that $\{M_1, M_3\}$ is a basis of $\text{Span}\big(M_1, M_3, O_{2\times 2}, O_{2\times 2}\big)$.[71]

Remark: To get $\text{rank}(f_2) = 2$, we did not necessarily have to set $f_2(1) = O_{2\times 2}$ and $f_2(x) = O_{2\times 2}$. We simply needed $f_2(1)$ and $f_2(x)$ to be linear combinations of M_1 and M_3.[72] However, the simplest possible choice was to set $f_2(1) = O_{2\times 2}$ and $f_2(x) = O_{2\times 2}$. This also makes the computation of the formula for f_2 (below) a little bit easier.

Rank 3: Since $\mathcal{B} = \{p_1(x), p_2(x), 1, x\}$ is a basis of $\mathbb{P}^3_{\mathbb{Z}_3}$, Theorem 4.3.2 guarantees that there exists a unique linear function $f_3 : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies the following:

- $f_3\big(p_1(x)\big) = M_1$,
- $f_3\big(p_3(x)\big) = M_3$,
- $f_3(1) = A_2$,
- $f_3(x) = O_{2\times 2}$.

By the Claim, our linear function f_3 in fact satisfies $f_3\big(p_i(x)\big) = M_i$ for all $i \in \{1,2,3,4\}$. We then have that

$$\text{rank}(f_3) \quad = \quad \dim\Big(\text{Span}\big(M_1, M_3, A_2, O_{2\times 2}\big)\Big) \quad \overset{(*)}{=} \quad 3,$$

where (*) follows from the fact that $\{M_1, M_3, A_2\}$ is a basis of $\text{Span}\big(M_1, M_3, A_2, O_{2\times 2}\big)$.[73]

Rank 4: Since $\mathcal{B} = \{p_1(x), p_2(x), 1, x\}$ is a basis of $\mathbb{P}^3_{\mathbb{Z}_3}$, Theorem 4.3.2 guarantees that there exists a unique linear function $f_4 : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ that satisfies the following:

- $f_4\big(p_1(x)\big) = M_1$,
- $f_4\big(p_3(x)\big) = M_3$,
- $f_4(1) = A_2$,
- $f_4(x) = A_3$.

By the Claim, our linear function f_4 in fact satisfies $f_4\big(p_i(x)\big) = M_i$ for all $i \in \{1,2,3,4\}$. We then have that

[71] This is because $O_{2\times 2}$ is a linear combination of M_1, M_3, and consequently (by Proposition 3.2.13), we have that $\text{Span}(M_1, M_3, O_{2\times 2}, O_{2\times 2}) = \text{Span}(M_1, M_3)$. Since M_1, M_3 are linearly independent, it follows that $\{M_1, M_3\}$ is indeed a basis of $\text{Span}\big(M_1, M_3, O_{2\times 2}, O_{2\times 2}\big)$.

[72] So, we could just as well have set $f_2(1) = M_1$ and $f_2(x) = M_3$.

[73] Let us justify this. First of all, $O_{2\times 2}$ is a linear combination of the matrices M_1, M_3, A_2, and so (by Proposition 3.2.13), we have that $\text{Span}\big(M_1, M_3, A_2, O_{2\times 2}\big) = \text{Span}\big(M_1, M_3, A_2\big)$. On the other hand, since $\mathcal{C} = \{M_1, M_3, A_2, A_3\}$ is a basis of $\mathbb{Z}_3^{2\times 2}$, we see that M_1, M_2, A_2 are linearly independent. So, $\{M_1, M_2, A_2\}$ is indeed a basis of $\text{Span}\big(M_1, M_3, A_2, O_{2\times 2}\big)$.

$$\operatorname{rank}(f_4) \quad = \quad \dim\Big(\operatorname{Span}\big(M_1, M_3, A_2, A_3\big)\Big) \quad \overset{(*)}{=} \quad \dim(\mathbb{Z}_3^{2\times 2}) \quad = \quad 4,$$

where (*) follows from the fact that $\mathcal{C} = \{M_1, M_3, A_2, A_3\}$ is a basis of $\mathbb{Z}_3^{2\times 2}$.

Let us now compute the formulas for our linear functions f_2, f_3, f_4. We proceed as follows. Let f be one of our functions f_2, f_3, f_4. Then as we saw above, there exist parameters $s_1, s_2, s_3, s_4, t_1, t_2, t_3, t_4 \in \mathbb{Z}_3$ such that

$$_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \quad = \quad \begin{bmatrix} 2t_1 + 2 & s_1 + 2t_1 + 1 & s_1 & t_1 \\ 2t_2 & s_2 + 2t_2 & s_2 & t_2 \\ 2t_3 & s_3 + 2t_3 + 1 & s_3 & t_3 \\ 2t_4 & s_4 + 2t_4 + 1 & s_4 & t_4 \end{bmatrix}.$$

We also have that

- $_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \ \big[\ 1\ \big]_{\mathcal{P}} = \big[\ f(1)\ \big]_{\mathcal{M}}$,
- $_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \ \big[\ x\ \big]_{\mathcal{P}} = \big[\ f(x)\ \big]_{\mathcal{M}}$,

which is equivalent to

$$_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \ \big[\ \big[\ 1\ \big]_{\mathcal{P}} \ \ \big[\ x\ \big]_{\mathcal{P}}\ \big] \quad = \quad \big[\ \big[\ f(1)\ \big]_{\mathcal{M}} \ \ \big[\ f(x)\ \big]_{\mathcal{M}}\ \big].$$

But note that

$$\big[\ \big[\ 1\ \big]_{\mathcal{P}} \ \ \big[\ x\ \big]_{\mathcal{P}}\ \big] \quad = \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

So, by matrix multiplication, we obtain

$$_{\mathcal{M}}\big[\ f\ \big]_{\mathcal{P}} \ \big[\ \big[\ 1\ \big]_{\mathcal{P}} \ \ \big[\ x\ \big]_{\mathcal{P}}\ \big] \quad = \quad \begin{bmatrix} 2t_1 + 2 & s_1 + 2t_1 + 1 \\ 2t_2 & s_2 + 2t_2 \\ 2t_3 & s_3 + 2t_3 + 1 \\ 2t_4 & s_4 + 2t_4 + 1 \end{bmatrix},$$

and consequently,

$$\big[\ \big[\ f(1)\ \big]_{\mathcal{M}} \ \ \big[\ f(x)\ \big]_{\mathcal{M}}\ \big] \quad = \quad \begin{bmatrix} 2t_1 + 2 & s_1 + 2t_1 + 1 \\ 2t_2 & s_2 + 2t_2 \\ 2t_3 & s_3 + 2t_3 + 1 \\ 2t_4 & s_4 + 2t_4 + 1 \end{bmatrix}.$$

We can now compute the matrices $_{\mathcal{M}}\big[\ f_2\ \big]_{\mathcal{P}}$, $_{\mathcal{M}}\big[\ f_3\ \big]_{\mathcal{P}}$, and $_{\mathcal{M}}\big[\ f_4\ \big]_{\mathcal{P}}$, by plugging in the values for $f_i(1)$ and $f_i(x)$ (for $i \in \{2, 3, 4\}$) into the equation above, and solving for the parameters $s_1, s_2, s_3, s_4, t_1, t_2, t_3, t_4$. Once we have computed these matrices, we can easily compute the formulas for the linear functions f_2, f_3, f_4.

Formula for f_2. For f_2, we have that

$$\left[\ \left[\ f_2(1)\ \right]_{\mathcal{M}}\ \ \left[\ f_2(x)\ \right]_{\mathcal{M}}\ \right] \ = \ \left[\ \left[\ O_{2\times 2}\ \right]_{\mathcal{M}}\ \ \left[\ O_{2\times 2}\ \right]_{\mathcal{M}}\ \right] \ = \ \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

So, we get

$$\begin{bmatrix} 2t_1+2 & s_1+2t_1+1 \\ 2t_2 & s_2+2t_2 \\ 2t_3 & s_3+2t_3+1 \\ 2t_4 & s_4+2t_4+1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

which yields

$$\begin{array}{ll} s_1=1 & t_1=2 \\ s_2=0 & t_2=0 \\ s_3=2 & t_3=0 \\ s_4=2 & t_4=0 \end{array}$$

and consequently,

$${}_{\mathcal{M}}\left[\ f_2\ \right]_{\mathcal{P}} \ = \ \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}.$$

Using the procedure described in the solution of Examples 4.5.18 and 4.5.19, we obtain the following formula for f_2:

$$f_2\Big(a_3x^3+a_2x^2+a_1x+a_0\Big) \ = \ \begin{bmatrix} a_2+2a_3 & 0 \\ 2a_2 & 2a_2 \end{bmatrix}$$

for all $a_0, a_1, a_2, a_3 \in \mathbb{Z}_3$.

Formula for f_3. For f_3, we have that

$$\left[\ \left[\ f_3(1)\ \right]_{\mathcal{M}}\ \ \left[\ f_3(x)\ \right]_{\mathcal{M}}\ \right] \ = \ \left[\ \left[\ A_2\ \right]_{\mathcal{M}}\ \ \left[\ O_{2\times 2}\ \right]_{\mathcal{M}}\ \right] \ = \ \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

So, we get

$$\begin{bmatrix} 2t_1+2 & s_1+2t_1+1 \\ 2t_2 & s_2+2t_2 \\ 2t_3 & s_3+2t_3+1 \\ 2t_4 & s_4+2t_4+1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

which yields

$$\begin{array}{ll} s_1 = 1 & t_1 = 2 \\ s_2 = 2 & t_2 = 2 \\ s_3 = 2 & t_3 = 0 \\ s_4 = 2 & t_4 = 0 \end{array}$$

and consequently,

$${}_{\mathcal{M}}\big[\ f_3\ \big]_{\mathcal{P}} \ = \ \begin{bmatrix} 0 & 0 & 1 & 2 \\ 1 & 0 & 2 & 2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}.$$

Using the procedure described in the solution of Examples 4.5.18 and 4.5.19, we obtain the following formula for f_3:

$$f_3\Big(a_3x^3 + a_2x^2 + a_1x + a_0\Big) \ = \ \begin{bmatrix} a_2 + 2a_3 & a_0 + 2a_2 + 2a_3 \\ 2a_2 & 2a_2 \end{bmatrix}$$

for all $a_0, a_1, a_2, a_3 \in \mathbb{Z}_3$.

Formula for f_4. For f_4, we have that

$$\big[\ \big[\ f_4(1)\ \big]_{\mathcal{M}} \ \ \big[\ f_4(x)\ \big]_{\mathcal{M}}\ \big] \ = \ \big[\ \big[\ A_2\ \big]_{\mathcal{M}} \ \ \big[\ A_3\ \big]_{\mathcal{M}}\ \big] \ = \ \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

So, we get

$$\begin{bmatrix} 2t_1 + 2 & s_1 + 2t_1 + 1 \\ 2t_2 & s_2 + 2t_2 \\ 2t_3 & s_3 + 2t_3 + 1 \\ 2t_4 & s_4 + 2t_4 + 1 \end{bmatrix} \ = \ \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

which yields

$$\begin{array}{ll} s_1 = 1 & t_1 = 2 \\ s_2 = 2 & t_2 = 2 \\ s_3 = 0 & t_3 = 0 \\ s_4 = 2 & t_4 = 0 \end{array}$$

and consequently,

$${}_{\mathcal{M}}\big[\ f_4\ \big]_{\mathcal{P}} \ = \ \begin{bmatrix} 0 & 0 & 1 & 2 \\ 1 & 0 & 2 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}.$$

Using the procedure described in the solution of Examples 4.5.18 and 4.5.19, we obtain the following formula for f_4:

$$f_4\Big(a_3x^3 + a_2x^2 + a_1x + a_0\Big) \;=\; \begin{bmatrix} a_2 + 2a_3 & a_0 + 2a_2 + 2a_3 \\ a_1 & 2a_2 \end{bmatrix}$$

for all $a_0, a_1, a_2, a_3 \in \mathbb{Z}_3$.

Finally, we note that our linear function $f_4 : \mathbb{P}^3_{\mathbb{Z}_3} \to \mathbb{Z}_3^{2\times 2}$ (above) is an isomorphism. Indeed, since $\operatorname{rank}(f_4) = 4 = \dim(\mathbb{Z}_3^{2\times 2})$, Proposition 4.2.6 guarantees that f_4 is onto. Since the domain and codomain of f_4 have the same finite dimension, Corollary 4.2.10 guarantees that f_4 is in fact an isomorphism. It remains to find a formula for f_4^{-1}. First of all, we have that

$${}_{\mathcal{P}}\big[\; f_4^{-1} \;\big]_{\mathcal{M}} \overset{(*)}{=} \Big({}_{\mathcal{M}}\big[\; f_4 \;\big]_{\mathcal{P}}\Big)^{-1} \overset{(**)}{=} \begin{bmatrix} 2 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 2 & 0 & 0 & 2 \end{bmatrix},$$

where (*) follows from Theorem 4.5.4(g), and (**) is obtained via routine computation. Finally, using the procedure described in the solution of Examples 4.5.18 and 4.5.19, we obtain the following formula for f_4^{-1}:

$$f_4^{-1}\Big(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\Big) \;=\; (2a + 2d)x^3 + 2dx^2 + cx + (2a + b + d)$$

for all $a, b, c, d \in \mathbb{Z}_3$. □

Chapter 5

Affine subspaces and affine functions

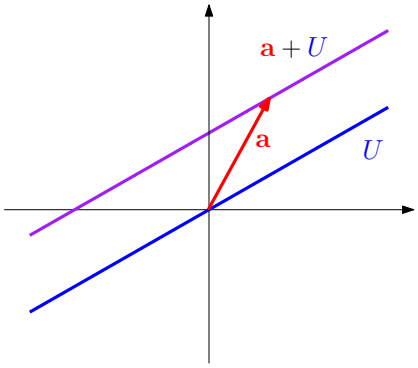
Terminology: So far, we have typically referred to vector/linear subspaces of a vector space simply as "subspaces." In this chapter, we will study a generalization of linear subspaces, called "affine subspaces." To avoid any confusion, in this chapter, we will not use the term "subspace" and will instead always write either "linear subspace" or "affine subspace." However, in subsequent chapters, we will again use the term "subspace" to mean "linear subspace."

5.1 Affine subspaces

An *affine subspace* of a vector space V over a field $\mathbb{F}$ is any set of the form

$$\mathbf{a} + U \quad := \quad \{\mathbf{a} + \mathbf{u} \mid \mathbf{u} \in U\},$$

where $\mathbf{a}$ is a vector in V and U is a linear subspace of V.



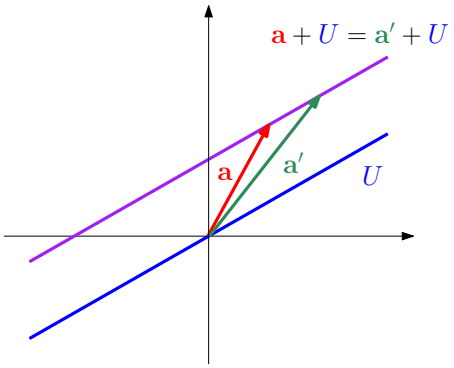
Thus, an affine subspace of V is obtained by shifting a linear subspace U of V by some vector $\mathbf{a}$.

Remark: For a vector space V over a field $\mathbb{F}$:

- every linear subspace U of V is also an affine subspace of V, since $U = \mathbf{0} + U$;[1]
- V is an affine subspace of itself (because V is a linear subspace of itself);
- for every vector $\mathbf{a} \in V$, $\{\mathbf{a}\}$ is an affine subspace of V, since $\{\mathbf{a}\} = \mathbf{a} + \{\mathbf{0}\}$ and $\{\mathbf{0}\}$ is a linear subspace of V.

Geometric considerations. As we know, linear subspaces of $\mathbb{R}^n$ are $\{\mathbf{0}\}$, lines through the origin, planes through the origin, and higher dimensional generalizations. So, affine subspaces of $\mathbb{R}^n$ are $\{\mathbf{a}\}$ (for any vector $\mathbf{a} \in \mathbb{R}^n$), lines, planes, and higher dimensional generalizations (these lines, planes, and higher dimensional generalizations may, but need not, pass through the origin).

As Theorem 5.1.1 (below) states, for an affine subspace $M = \mathbf{a} + U$ of a vector space V over a field $\mathbb{F}$ (where $\mathbf{a}$ is a vector and U a linear subspace of V), the vector $\mathbf{a}$ need not be unique (indeed, it can be any vector in M),[2] but the linear subspace U is unique (it depends only on M, and not on the vector $\mathbf{a}$).[3]



Theorem 5.1.1. *Let V be a vector space over a field $\mathbb{F}$, and let $M = \mathbf{a} + U$ be an affine subspace of V, where $\mathbf{a}$ is a vector and U a linear subspace of V. Then all the following hold:*

(a) $\mathbf{a} \in M$ (and in particular, $M \neq \emptyset$);

(b) for all $\mathbf{a}' \in M$, we have that $M = \mathbf{a}' + U$;

(c) for all vectors $\mathbf{a}'$ and linear subspaces U' of V such that $M = \mathbf{a}' + U'$, we have that $U' = U$;

[1] Moreover, as we shall see, linear subspaces of V are precisely those affine subspaces of V that contain $\mathbf{0}$ (see Corollary 5.1.2).

[2] This follows from Theorem 5.1.1(b)

[3] This follows from Theorem 5.1.1(c).

(d) for all $\mathbf{b} \in V \setminus M$*, we have that* $M \cap (\mathbf{b} + U) = \emptyset$*.*

Proof. (a) Since U is a linear subspace of V, Theorem 3.1.7 guarantees that $\mathbf{0} \in U$, and consequently, $\mathbf{a} = \mathbf{a} + \mathbf{0} \in \mathbf{a} + U = M$.

(b) Fix $\mathbf{a}' \in M$. Since $\mathbf{a}' \in M = \mathbf{a} + U$, there exists some $\mathbf{u}' \in U$ such that $\mathbf{a}' = \mathbf{a} + \mathbf{u}'$. Now, we must show that $M = \mathbf{a}' + U$.

Let us first show that $M \subseteq \mathbf{a}' + U$. Fix $\mathbf{x} \in M$. Since $M = \mathbf{a} + U$, there exists some $\mathbf{u} \in U$ such that $\mathbf{x} = \mathbf{a} + \mathbf{u}$. Then $\mathbf{x} = \mathbf{a} + \mathbf{u} = (\mathbf{a}' - \mathbf{u}') + \mathbf{u} = \mathbf{a}' + (\mathbf{u} - \mathbf{u}')$. Since $\mathbf{u}, \mathbf{u}' \in U$, and U is a linear subspace of V, we have that $\mathbf{u} - \mathbf{u}' \in U$; so, $\mathbf{x} = \mathbf{a}' + (\mathbf{u} - \mathbf{u}') \in \mathbf{a}' + U$. This proves that $M \subseteq \mathbf{a}' + U$.

Let us now show that $\mathbf{a}' + U \subseteq M$. Fix $\mathbf{u} \in U$; we must show that $\mathbf{a}' + \mathbf{u} \in M$. But note that $\mathbf{a}' + \mathbf{u} = \mathbf{a} + \mathbf{u}' + \mathbf{u}$. Since $\mathbf{u}', \mathbf{u} \in U$, and U is a linear subspace of V, we have that $\mathbf{u}' + \mathbf{u} \in U$; consequently, $\mathbf{a}' + \mathbf{u} = \mathbf{a} + \mathbf{u}' + \mathbf{u} \in \mathbf{a} + U = M$. This proves that $\mathbf{a}' + U \subseteq M$.

We have now shown that $M = \mathbf{a} + U$, which is what we needed.

(c) Fix a vector $\mathbf{a}'$ and a linear subspace U' of V such that $M = \mathbf{a}' + U'$. By (a), we have that $\mathbf{a}' \in M$, and so by (b), we have that $M = \mathbf{a}' + U$. So, $\mathbf{a}' + U' = \mathbf{a}' + U$, and we deduce that $U' = U$.[4]

(d) Fix $\mathbf{b} \in V \setminus M$. We must show that $M \cap (\mathbf{b} + U) = \emptyset$. Suppose otherwise, and fix $\mathbf{x} \in M \cap (\mathbf{b} + U)$. Since $\mathbf{x} \in M = \mathbf{a} + U$, there exists some $\mathbf{u}_1 \in U$ such that $\mathbf{x} = \mathbf{a} + \mathbf{u}_1$; on the other hand, since $\mathbf{x} \in \mathbf{b} + U$, there exists some $\mathbf{u}_2 \in U$ such that $\mathbf{x} = \mathbf{b} + \mathbf{u}_2$. So, $\mathbf{a} + \mathbf{u}_1 = \mathbf{b} + \mathbf{u}_2$, and it follows that $\mathbf{b} = \mathbf{a} + (\mathbf{u}_1 - \mathbf{u}_2)$. Since $\mathbf{u}_1, \mathbf{u}_2 \in U$, and since U is a linear subspace of V, we have that $\mathbf{u}_1 - \mathbf{u}_2 \in U$; consequently, $\mathbf{b} = \mathbf{a} + (\mathbf{u}_1 - \mathbf{u}_2) \in \mathbf{a} + U = M$, contrary to the fact that $\mathbf{b} \in V \setminus M$. □

Given a vector space V over a field $\mathbb{F}$, we define the *dimension* of an affine subspace $M = \mathbf{a} + U$ of V (where $\mathbf{a}$ is a vector and U a linear subspace of V) to be

$$\dim(M) \quad := \quad \dim(U).$$

By Theorem 5.1.1(c), this is well defined.

Corollary 5.1.2. *Let V be a vector space over a field $\mathbb{F}$. Then linear subspaces of V are precisely those affine spaces of V that contain $\mathbf{0}$. In other words, for all $U \subseteq V$, the following are equivalent:*

(i) U is a linear subspace of V;

(ii) U is an affine subspace of V and $\mathbf{0} \in U$.

[4] This is "obvious," but here is a formal proof. By symmetry, it suffices to show that $U' \subseteq U$. Fix $\mathbf{u}' \in U'$. Then $\mathbf{a}' + \mathbf{u}' \in \mathbf{a}' + U' = \mathbf{a}' + U$, and it follows that there exists some $\mathbf{u} \in U$ such that $\mathbf{a}' + \mathbf{u}' = \mathbf{a}' + \mathbf{u}$. By subtracting $\mathbf{a}'$ from both sides, we get $\mathbf{u}' = \mathbf{u}$; since $\mathbf{u} \in U$, we deduce that $\mathbf{u}' \in U$. So, $U' \subseteq U$.

Proof. Fix $U \subseteq V$. Suppose first that (i) holds. Then $\mathbf{0} \in U$ (by Theorem 3.1.7), and moreover, $U = \mathbf{0} + U$. So, (ii) holds.

Suppose now that (ii) holds. Since U is an affine subspace of V, we know that there exists a vector $\mathbf{a} \in V$ and a linear subspace U' of V such that $U = \mathbf{a} + U'$. Moreover, by (ii), we have that $\mathbf{0} \in U$, and so by Theorem 5.1.1(b), we have that $U = \mathbf{0} + U'$. So, $U = U'$. Since U' is a linear subspace of V, we see that (i) holds. □

Recall that the intersection of two linear subspaces is a linear subspace (see subsection 3.1.3). In the case of affine subspaces, we have the following corollary.

Corollary 5.1.3. *Let V be a vector space over a field $\mathbb{F}$, and let M_1 and M_2 be affine subspaces of V. Then either $M_1 \cap M_2 = \emptyset$, or $M_1 \cap M_2$ is an affine subspace of V.*

Proof. We may assume that $M_1 \cap M_2 \neq \emptyset$, for otherwise we are done. Fix any $\mathbf{a} \in M_1 \cap M_2$. By Theorem 5.1.1, M_1 and M_2 can be written as $M_1 = \mathbf{a} + U_1$ and $M_2 = \mathbf{a} + U_2$, for some linear subspaces U_1 and U_2 of V. Then $U := U_1 \cap U_2$ is a linear subspace of V (see subsection 3.1.3). Moreover, it is clear that $M_1 \cap M_2 = \mathbf{a} + U$,[5] and so $M_1 \cap M_2$ is an affine subspace. □

5.2 Affine functions

Suppose that V_1 and V_2 are vector spaces over a field $\mathbb{F}$. A function $f : V_1 \to V_2$ is called an *affine function* if there exists a linear function $g : V_1 \to V_2$ and a vector $\mathbf{b} \in V_2$ such that for all $\mathbf{x} \in V_1$, we have that $f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b}$.

Obviously, every linear function f is affine (we simply take $g := f$ and $\mathbf{b} := \mathbf{0}$). Moreover, we have the following proposition.

Proposition 5.2.1. *Let V_1 and V_2 be vector spaces over a field $\mathbb{F}$, and let $f : V_1 \to V_2$ be an affine function. Then f is linear if and only if $f(\mathbf{0}) = \mathbf{0}$.*

Proof. If f is linear, then Proposition 4.1.6 guarantees that $f(\mathbf{0}) = \mathbf{0}$. For the reverse implication, we assume that $f(\mathbf{0}) = \mathbf{0}$, and we show that f is linear. Since f is an affine function, we know that there exists a linear function $g : V_1 \to V_2$ and a vector $\mathbf{b} \in V_2$ such that for all $\mathbf{x} \in V_1$, we have that $f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b}$. But now

$$\mathbf{0} \;=\; f(\mathbf{0}) \;=\; g(\mathbf{0}) + \mathbf{b} \;\overset{(*)}{=}\; \mathbf{0} + \mathbf{b} \;=\; \mathbf{b}$$

where (*) follows from the fact that g is linear, and so $g(\mathbf{0}) = \mathbf{0}$ (by Proposition 4.1.6). So, $f(\mathbf{x}) = g(\mathbf{x})$ for all $\mathbf{x} \in V_1$, that is, $f = g$. Since g is linear, so is f. □

[5] This is "obvious," but here is a full proof. It is clear that $\mathbf{a} + U \subseteq M_1 \cap M_2$. For the reverse inclusion, we fix some $\mathbf{x} \in M_1 \cap M_2$, and we show that $\mathbf{x} \in \mathbf{a} + U$. Since $\mathbf{x} \in M_1 = \mathbf{a} + U_1$, we know that there exists some $\mathbf{u}_1 \in U_1$ such that $\mathbf{x} = \mathbf{a} + \mathbf{u}_1$. Similarly, since $\mathbf{x} \in M_2 = \mathbf{a} + U_2$, there exists some $\mathbf{u}_2 \in U_2$ such that $\mathbf{x} = \mathbf{a} + \mathbf{u}_2$. So, $\mathbf{a} + \mathbf{u}_1 = \mathbf{a} + \mathbf{u}_2$, and consequently, $\mathbf{u}_1 = \mathbf{u}_2$. Since $\mathbf{u}_1 \in U_1$ and $\mathbf{u}_2 \in U_2$, we deduce that $\mathbf{u}_1 = \mathbf{u}_2$ belongs to $U_1 \cap U_2 = U$. But now $\mathbf{x} = \mathbf{a} + \mathbf{u}_1 \in \mathbf{a} + U$, and we are done.

5.2.1 Making new affine functions out of old ones

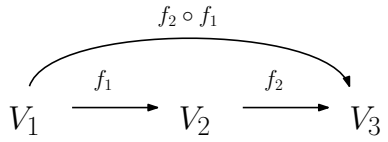
Theorem 5.2.2 (below) is an analog of Theorem 4.1.7 for affine functions.

Theorem 5.2.2. *Let V_1, V_2, V_3 be vector spaces over a field $\mathbb{F}$. Then all the following hold:*

(a) for all affine functions $f_1, f_2 : V_1 \to V_2$, we have that $f_1 + f_2$ is an affine function;

(b) for all affine functions $f : V_1 \to V_2$ and scalars α, we have that αf is an affine function;

(c) for all affine functions $f_1 : V_1 \to V_2$ and $f_2 : V_2 \to V_3$, we have that $f_2 \circ f_1$ is an affine function.



Proof. We prove (c). The proofs of (a) and (b) are left as an exercise. Fix affine functions $f_1 : V_1 \to V_2$ and $f_2 : V_2 \to V_3$. Since $f_1 : V_1 \to V_2$ is an affine function, there exists a linear function $g_1 : V_1 \to V_2$ and a vector $\mathbf{b}_2 \in V_2$ such that for all $\mathbf{x} \in V_1$, we have that $f_1(\mathbf{x}) = g_1(\mathbf{x}) + \mathbf{b}_2$. Similarly, since $f_2 : V_2 \to V_3$ is an affine function, there exists a linear function $g_2 : V_2 \to V_3$ and a vector $\mathbf{b}_3 \in V_3$ such that for all $\mathbf{x} \in V_2$, we have that $f_2(\mathbf{x}) = g_2(\mathbf{x}) + \mathbf{b}_3$. But now for all $\mathbf{x} \in V_1$, we have that

$$\begin{aligned}
(f_2 \circ f_1)(\mathbf{x}) &= f_2\big(f_1(\mathbf{x})\big) \\
&= f_2\big(g_1(\mathbf{x}) + \mathbf{b}_2\big) \\
&= g_2\big(g_1(\mathbf{x}) + \mathbf{b}_2\big) + \mathbf{b}_3 \\
&= g_2\big(g_1(\mathbf{x})\big) + g_2(\mathbf{b}_2) + \mathbf{b}_3 \qquad \text{because } g_2 \text{ is linear} \\
&= (g_2 \circ g_1)(\mathbf{x}) + \big(g_2(\mathbf{b}_2) + \mathbf{b}_3\big).
\end{aligned}$$

Since g_1 and g_2 are linear, Proposition 4.1.7(c) guarantees that $g_2 \circ g_1$ is linear. On the other hand, $g_2(\mathbf{b}_2) + \mathbf{b}_3$ is a vector in V_3. So, $f_2 \circ f_1$ is an affine function. □

5.2.2 The image of an affine subspace under an affine function

By Theorem 4.2.3(a), for a linear function $f : U \to V$ (where U and V are vector spaces over a field $\mathbb{F}$), and for a linear subspace U' of U, we have that $f[U']$ is a subspace of V, and moreover, by Corollary 4.2.9, we have that $\dim\big(f[U']\big) \leq \min\big\{\dim(U'), \dim(V)\big\}$. In the case of affine functions and affine subspaces, we have the following theorem.

Theorem 5.2.3. *Let V_1 and V_2 be vector spaces over a field $\mathbb{F}$, let $f : V_1 \to V_2$ be an affine function given by*

$$f(\mathbf{x}) \;=\; g(\mathbf{x}) + \mathbf{b} \qquad \textit{for all } \mathbf{x} \in V_1,$$

where $g : V_1 \to V_2$ is a linear function and $\mathbf{b}$ is a fixed vector in V_2, and let let $M_1 = \mathbf{a}_1 + U_1$ be an affine subspace of V_1 (where $\mathbf{a}_1$ is a vector and U_1 a linear subspace of V_1). Then

$$f[M_1] \;=\; \big(g(\mathbf{a}_1) + \mathbf{b}\big) + g[U_1],$$

and consequently, $f[M_1]$ is an affine subspace of V_2. Moreover,

$$dim\big(f[M_1]\big) \;\leq\; \min\big\{dim(M_1), dim(V)\big\}.$$

Proof. First, we have the following:

$$\begin{aligned}
f[M_1] &= \big\{f(\mathbf{x}) \mid \mathbf{x} \in M_1\big\} \\
&= \big\{f(\mathbf{a}_1 + \mathbf{u}) \mid \mathbf{u} \in U_1\big\} && \text{becuase } M_1 = \mathbf{a}_1 + U_1 \\
&= \big\{g(\mathbf{a}_1 + \mathbf{u}) + \mathbf{b} \mid \mathbf{u} \in U_1\big\} && \text{because } f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b} \text{ for all } \mathbf{x} \in V_1 \\
&= \big\{g(\mathbf{a}_1) + g(\mathbf{u}) + \mathbf{b} \mid \mathbf{u} \in U_1\big\} && \text{because } g \text{ is linear} \\
&= \big(g(\mathbf{a}_1) + \mathbf{b}\big) + \big\{g(\mathbf{u}) \mid \mathbf{u} \in U_1\big\} \\
&= \big(g(\mathbf{a}_1) + \mathbf{b}\big) + g[U_1].
\end{aligned}$$

Since $g : V_1 \to V_2$ is a linear function and U_1 is a linear subspace of V_1, Theorem 4.2.3(a) guarantees that $g[U_1]$ is a linear subspace of V_2. So, $f[M_1] = \big(g(\mathbf{a}_1) + \mathbf{b}\big) + g[U_1]$ is an affine subspace of V_2.

It remains to show that $\dim\big(f[M_1]\big) \leq \dim(M_1)$. For this, we observe the following:

$$\begin{aligned}
\dim\big(f[M_1]\big) &= \dim\Big(\big(g(\mathbf{a}_1) + \mathbf{b}\big) + g[U_1]\Big) \\
&= \dim\big(g[U_1]\big) && \text{by definition, since } g[U_1] \text{ is a linear subspace of } V_2 \\
&\leq \min\big\{\dim(U_1), \dim(V)\big\} && \text{by Corollary 4.2.9} \\
&= \min\big\{\dim(M_1), \dim(V)\big\} && \text{by definition, since } M_1 = \mathbf{a}_1 + U_1.
\end{aligned}$$

This completes the argument. □

Corollary 5.2.4. *Let V_1 and V_2 be vector spaces over a field $\mathbb{F}$, and let $f : V_1 \to V_2$ be an affine function given by*

$$f(\mathbf{x}) \;=\; g(\mathbf{x}) + \mathbf{b} \qquad \textit{for all } \mathbf{x} \in V_1,$$

where $g : V_1 \to V_2$ is a linear function and $\mathbf{b}$ is a fixed vector in V_2. Then

$$\mathit{Im}(f) \;=\; \big(g(\mathbf{a}_1) + \mathbf{b}\big) + \mathit{Im}(g),$$

and consequently, $\mathit{Im}(f)$ is an affine subspace of V_2. Moreover,

$$\mathit{dim}\big(\mathit{Im}(f)\big) \;=\; \mathit{rank}(g) \;\leq\; \min\big\{\mathit{dim}(V_1), \mathit{dim}(V_2)\big\}.$$

Proof. Since V_1 is a linear subspace of itself, we have the following:

$$\begin{aligned}
\mathrm{Im}(f) &= f[V_1] \\
&= f[\mathbf{0} + V_1] \\
&= \big(g(\mathbf{0}) + \mathbf{b}\big) + g[V_1] && \text{by Theorem 5.2.3} \\
&= \mathbf{b} + g[V_1] && \text{because } g \text{ is linear and therefore (by Proposition 4.1.6) satisfies } g(\mathbf{0}) = \mathbf{0} \\
&= \mathbf{b} + \mathrm{Im}(g).
\end{aligned}$$

Since g is linear, Theorem 4.2.3(b) guarantees that $\mathrm{Im}(g)$ is a linear subspace of V_2, and it follows that $\mathrm{Im}(f)$ is an affine subspace of V_2. Finally, we have the following:

$$\begin{aligned}
\dim\big(\mathrm{Im}(f)\big) &= \dim\big(\mathbf{b} + \mathrm{Im}(g)\big) \\
&= \dim\big(\mathrm{Im}(g)\big) && \text{by definition, since } \mathrm{Im}(g) \text{ is a linear subspace of } V_2 \\
&= \mathrm{rank}(g) && \text{by definition, since } g \text{ is linear} \\
&\leq \min\big\{\dim(V_1), \dim(V_2)\big\} && \text{by Corollary 4.2.8.}
\end{aligned}$$

This completes the argument. □

Geometric considerations. Suppose that $f : \mathbb{R}^m \to \mathbb{R}^n$ is an affine (possibly linear) function. As we know, affine subspaces of $\mathbb{R}^m$ are points (technically, sets that contain exactly one point), lines, planes, and higher dimensional generalizations. Theorem 5.2.3 guarantees that f maps every affine subspace M of $\mathbb{R}^m$ onto an affine subspace of $\mathbb{R}^n$, and moreover, $\dim\big(f[M]\big) \leq \dim(M)$. So, f maps lines onto lines or points, and it maps planes onto planes, lines, or points. Obvious higher-dimensional generalizations apply. (Compare these remarks to the discussion in subsection 1.10.2.)

5.2.3 The preimage of an affine subspace under an affine function

By Theorem 4.2.3(c), for any linear function $f : U \to V$ (where U and V are vector spaces over a field $\mathbb{F}$), and any linear subspace V' of V, we have that $f^{-1}[V']$ is a linear subspace of the domain U. Theorem 5.2.5 (below) is an analog of Theorem 4.2.3(c) for affine functions and affine subspaces. Note that it is not quite true that the preimage of an affine subspace under an affine function must be an affine subspace. This is because such a preimage may possibly be empty.[6] (Note that this never happens with preimages of linear subspaces under linear functions: such preimages always contain at least the zero vector.) However, as Theorem 5.2.5 states, if the preimage of an affine subspace of the codomain under an affine function is not empty, then that preimage is indeed an affine subspace of the domain, as we would expect.

Theorem 5.2.5. *Let V_1 and V_2 be vector spaces over a field $\mathbb{F}$, and let $f : V_1 \to V_2$ be an affine function given by*

$$f(\mathbf{x}) \;=\; g(\mathbf{x}) + \mathbf{b} \qquad \text{for all } \mathbf{x} \in V_1,$$

where $g : V_1 \to V_2$ is a linear function and $\mathbf{b}$ is a fixed vector in V_2. Further, let $M_2 = \mathbf{a}_2 + U_2$ be an affine subspace of V_2 (where $\mathbf{a}_2$ is a vector and U_2 a linear subspace of V_2). Then both the following hold:

(a) for all $\mathbf{a}_1 \in f^{-1}[M_2]$, we have that $f^{-1}[M_2] = \mathbf{a}_1 + g^{-1}[U_2]$;

(b) $f^{-1}[M_2]$ is either empty or an affine subspace of V_1.

Remark: If $f^{-1}[M_2] = \emptyset$ (which is possible), then (a) is vacuously true (since in this case, there are no vectors $\mathbf{a}_1$ in $f^{-1}[M_2]$).

[6] For a simple example, consider any two vector spaces U and V over the same field $\mathbb{F}$, and assume that V is non-trivial. Consider the function $f : U \to V$ that maps all elements of U to the zero vector (i.e. $f(\mathbf{u}) = \mathbf{0}$ for all $\mathbf{u} \in U$). Then f is a linear (and therefore affine) function. Now choose any vector $\mathbf{b} \in V \setminus \{\mathbf{0}\}$. Then $\{\mathbf{b}\} = \mathbf{b} + \{\mathbf{0}\}$ is an affine subspace of the codomain V, and we have that $f^{-1}[\{\mathbf{b}\}] = \emptyset$.

Proof. We first prove (a). Fix $\mathbf{a}_1 \in f^{-1}[M_2]$. Then $f(\mathbf{a}_1) \in M_2$, and so by Theorem 5.1.1, we have that $M_2 = f(\mathbf{a}_1) + U_2$ for some linear subspace U_2 of V_2. We now have the following:

$$\begin{aligned}
f^{-1}[M_2] &= \{\mathbf{x} \in V_1 \mid f(\mathbf{x}) \in M_2\} \\
&= \{\mathbf{x} \in V_1 \mid f(\mathbf{x}) \in f(\mathbf{a}_1) + U_2\} \\
&= \{\mathbf{x} \in V_1 \mid g(\mathbf{x}) + \mathbf{b} \in (g(\mathbf{a}_1) + \mathbf{b}) + U_2\} \\
&= \{\mathbf{x} \in V_1 \mid g(\mathbf{x}) \in g(\mathbf{a}_1) + U_2\} \\
&= \{\mathbf{x} \in V_1 \mid g(\mathbf{x}) - g(\mathbf{a}_1) \in U_2\} \\
&\overset{(*)}{=} \{\mathbf{x} \in V_1 \mid g(\mathbf{x} - \mathbf{a}_1) \in U_2\} \\
&\overset{(**)}{=} \mathbf{a}_1 + \{\mathbf{y} \in V_1 \mid g(\mathbf{y}) \in U_2\} \\
&= \mathbf{a}_1 + g^{-1}[U_2],
\end{aligned}$$

where (*) follows from the fact that g is linear, and in (**) we set $\mathbf{y} = \mathbf{x} - \mathbf{a}_1$.[7] This proves (a).

It remains to prove (b). We may assume that $f^{-1}[M_2] \neq \emptyset$, for otherwise we are done. Fix any $\mathbf{a}_1 \in f^{-1}[M_2]$. Then by (a), we have that $f^{-1}[M_2] = \mathbf{a}_1 + g^{-1}[U_2]$. Since $g : V_1 \to V_2$ is a linear function and U_2 is a linear subspace of V_2, Theorem 4.2.3(c) guarantees that $g^{-1}[U_2]$ is a linear subspace of V_1. So, $f^{-1}[M_2] = \mathbf{a}_1 + g^{-1}[U_2]$ is an affine subspace of V_1. This proves (b). □

Corollary 5.2.6. *Let V_1 and V_2 be vector spaces over a field $\mathbb{F}$, and let $f : V_1 \to V_2$ be an affine function given by*

$$f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b} \qquad \text{for all } \mathbf{x} \in V_1,$$

where $g : V_1 \to V_2$ is a linear function and $\mathbf{b}$ is a fixed vector in V_2. Further, let $\mathbf{c}$ be any vector in V_2. Then both the following hold:

(a) if $\mathbf{a} \in V_1$ is any solution of the equation $f(\mathbf{x}) = \mathbf{c}$,[8] then the solution set of the equation $f(\mathbf{x}) = \mathbf{c}$ is $\mathbf{a} + Ker(g)$;

(b) the solution set of the equation $f(\mathbf{x}) = \mathbf{c}$ is either empty or an affine subspace of V_1.

[7] We are also using the fact that $V_1 = \{\mathbf{x} - \mathbf{a}_1 \mid \mathbf{x} \in V_1\}$. (Proof?)

[8] This simply means that $f(\mathbf{a}) = \mathbf{c}$.

Remark: If the equation $f(\mathbf{x}) = \mathbf{c}$ has no solutions (which is possible), then (a) is vacuously true.

Proof. We first prove (a). Suppose that $\mathbf{a} \in V_1$ is any solution of the equation $f(\mathbf{x}) = \mathbf{c}$; we must show that the solution set of the equation $f(\mathbf{x}) = \mathbf{c}$ is precisely the set $\mathbf{a} + \mathrm{Ker}(g)$. Now, note that the solution set of the equation $f(\mathbf{x}) = \mathbf{c}$ is precisely the set $f^{-1}[\{\mathbf{c}\}]$, and in particular, we have that $\mathbf{a} \in f^{-1}[\{\mathbf{c}\}]$. Moreover, since $\{\mathbf{c}\} = \mathbf{c} + \{\mathbf{0}\}$, and $\{\mathbf{0}\}$ is a linear subspace of V_2, we have that $\{\mathbf{c}\}$ is an affine subspace of V_2. But now

We can now apply Theorem 5.2.5, as follows:

$$f^{-1}[\{\mathbf{c}\}] \stackrel{(*)}{=} \mathbf{a} + g^{-1}[\{\mathbf{0}\}] \stackrel{(**)}{=} \mathbf{a} + \mathrm{Ker}(g)$$

where (*) follows from Theorem 5.2.5, and (**) follows from the definition of $\mathrm{Ker}(g)$. This proves (a).

It remains to prove (b). We may assume that the equation $f(\mathbf{x}) = \mathbf{c}$ is consistent, for otherwise, its solution set is empty, and we are done. Fix any solution $\mathbf{a}$ of the equation $f(\mathbf{x}) = \mathbf{c}$. Then by (a), the solution set of the equation $f(\mathbf{x}) = \mathbf{c}$ is $\mathbf{a} + \mathrm{Ker}(g)$. But by Theorem 4.2.3(d), $\mathrm{Ker}(g)$ is a linear subspace of V_1. So, $\mathbf{a} + \mathrm{Ker}(g)$ is an affine subspace of V_1. This proves (b). □

Corollary 5.2.7. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times m}$ and $\mathbf{b} \in \mathbb{F}^n$. Then both the following hold:*

(a) if $\mathbf{a}$ is any solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, then the solution set of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{a} + Nul(A)$;

(b) if the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent, then its solution set is an affine subspace of $\mathbb{F}^m$.

Proof. Let $f_A : \mathbb{F}^m \to \mathbb{F}^n$ be given by $f_A(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^m$. By Proposition 1.10.4, f_A is a linear (and therefore affine) function. If $\mathbf{a}$ is any solution of $A\mathbf{x} = \mathbf{b}$, then it is also a solution of $f_A(\mathbf{x}) = \mathbf{b}$, and by Corollary 5.2.6, the latter is precisely equal to

$$\mathbf{a} + \mathrm{Ker}(f_A) \stackrel{(*)}{=} \mathbf{a} + \mathrm{Nul}(A),$$

where (*) follows from Proposition 4.2.1(b). This proves (a). Part (b) follows from (a) and from the fact that $\mathrm{Nul}(A)$ is a subspace of $\mathbb{F}^m$ (by Proposition 3.3.23). □

Let us take a look at a simple example illustrating Corollary 5.2.7. Consider the the following matrix and vector, with entries understood to be in $\mathbb{Z}_3$:

$$A := \begin{bmatrix} 1 & 2 & 2 & 2 & 1 \\ 2 & 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & 2 & 0 \\ 1 & 1 & 2 & 1 & 1 \end{bmatrix}, \qquad \mathbf{b} := \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \end{bmatrix}.$$

Let us solve the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. We form the augmented matrix

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{ccccc|c} 1 & 2 & 2 & 2 & 1 & 2 \\ 2 & 2 & 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & 2 & 0 & 1 \\ 1 & 1 & 2 & 1 & 1 & 0 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}\Big(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\Big) = \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 2 & 2 & 2 \\ 0 & 1 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We see that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent, and that the general solution of this equation is

$$\mathbf{x} = \begin{bmatrix} s+t+2 \\ 2s+2 \\ s+2t+1 \\ s \\ t \end{bmatrix}, \qquad \text{where } s,t \in \mathbb{Z}_3.$$

By separating parameters, we obtain

$$\mathbf{x} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \qquad \text{where } s,t \in \mathbb{Z}_3.$$

So, the solution set of the equation $A\mathbf{x} = \mathbf{b}$ is

$$\Big\{\begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \mid s,t \in \mathbb{Z}_3\Big\} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathrm{Span}\Big(\begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}\Big).$$

But note that $\mathbf{a} := \begin{bmatrix} 2 & 2 & 1 & 0 & 0 \end{bmatrix}^T$ is one solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, and that the null space of A is precisely

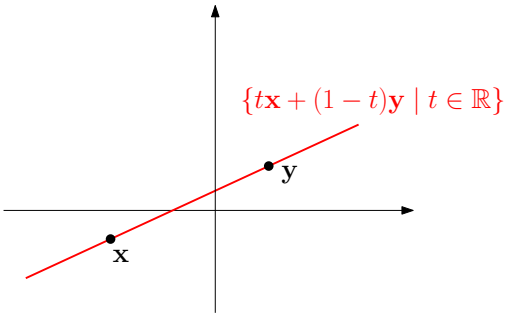
$$\mathrm{Nul}(A) = \mathrm{Span}\Big(\begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}\Big).$$

So, the solution set of $A\mathbf{x} = \mathbf{b}$ is precisely $\mathbf{a} + \mathrm{Nul}(A)$, which is consistent with Corollary 5.2.7.

Geometric considerations. Suppose that we are given a matrix $A \in \mathbb{R}^{n\times m}$ and a vector $\mathbf{b} \in \mathbb{R}^n$. By Corollary 5.2.7(b), the solution set of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is either empty or an affine subspace of $\mathbb{R}^m$, i.e. a point (technically, a set that contains exactly one point), a line, a plane, or a higher-dimensional generalization in $\mathbb{R}^m$.

5.3 Affine combinations and affine hulls

Recall from analytic geometry that if $\mathbf{x}$ and $\mathbf{y}$ are distinct points (vectors) in $\mathbb{R}^2$, then the line in $\mathbb{R}^2$ that passes through $\mathbf{x}$ and $\mathbf{y}$ is $\{t\mathbf{x} + (1-t)\mathbf{y} \mid t \in \mathbb{R}\}$. This in fact holds for all distinct points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ (not just $\mathbb{R}^2$). Affine combinations are a generalization of this concept.



Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ ($n \geq 1$) are vectors in a vector space V over a field $\mathbb{F}$. An *affine combination* of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is any sum of the form $\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n$, where $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ satisfy $\alpha_1 + \cdots + \alpha_n = 1$. The set of all affine combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, denoted $\mathrm{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, is called the *affine hull* (or *affine span*) of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. So, we have that

$$\mathrm{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \quad := \quad \Big\{ \sum_{i=1}^{n} \alpha_i \mathbf{x}_i \mid \alpha_1, \ldots, \alpha_n \in \mathbb{F}, \ \sum_{i=1}^{n} \alpha_i = 1 \Big\}.$$

Since $\mathbf{x}_i = 0\mathbf{x}_1 + \cdots + 0\mathbf{x}_{i-1} + 1\mathbf{x}_i + 0\mathbf{x}_{i+1} + \cdots + 0\mathbf{x}_n$ for all $i \in \{1, \ldots, n\}$, we see that $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathrm{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. As Theorem 5.3.1 (below) shows, affine subspaces of V are precisely those non-empty subsets of V that are closed under affine combinations. As a corollary (see Corollary 5.3.2), we deduce that all affine hulls are affine subspaces of V. We note that Theorem 5.3.1 are the affine subspace analogs of Theorems 3.1.7 and 3.1.11(b), respectively.

Theorem 5.3.1. *Let V be a vector space over a field $\mathbb{F}$, and let $M \subseteq V$. Then the following are equivalent:*

(i) M is an affine subspace of V;

*(ii) M is **non-empty** and closed under affine combinations, that is, for all vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in M$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\alpha_1 + \cdots + \alpha_n = 1$, we have that $\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n \in M$.*

Proof. Assume first that (i) holds. Let us prove (ii). Set $M = \mathbf{a} + U$, where $\mathbf{a}$ is a vector and U a linear subspace of V, as in the definition of an affine subspace. By Theorem 5.1.1(a), we have that $\mathbf{a} \in M$, and in particular, $M \neq \emptyset$. It remains to show that M is closed under affine combinations. Fix $\mathbf{x}_1, \ldots, \mathbf{x}_n \in M$, and fix $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\alpha_1 + \cdots + \alpha_n = 1$; we must show that $\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n$ belongs to M. Since $\mathbf{x}_1, \ldots, \mathbf{x}_n \in M = \mathbf{a} + U$, there exist vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n \in U$ such that $\mathbf{x}_1 = \mathbf{a} + \mathbf{u}_1, \ldots, \mathbf{x}_n = \mathbf{a} + \mathbf{u}_n$. We now have that

$$\begin{aligned} \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n &= \alpha_1(\mathbf{a} + \mathbf{u}_1) + \cdots + \alpha_n(\mathbf{a} + \mathbf{u}_1) \\ &= \underbrace{(\alpha_1 + \cdots + \alpha_n)}_{=1}\mathbf{a} + (\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n) \\ &= \mathbf{a} + \underbrace{(\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n)}_{:=\mathbf{u}}. \end{aligned}$$

Since $\mathbf{u}_1, \ldots, \mathbf{u}_n \in U$, and U is a linear subspace of V, we have that $\mathbf{u} \in U$. So, $\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n = \mathbf{a} + \mathbf{u} \in \mathbf{a} + U = M$. This proves (ii).

Conversely, suppose that (ii) holds. We must prove (i). Using the fact that $M \neq \emptyset$, we fix some $\mathbf{a} \in M$. Set $U := \{\mathbf{x} - \mathbf{a} \mid \mathbf{x} \in M\}$. Clearly, $M = \mathbf{a} + U$. It remains to show that U is a linear subspace of V. By Theorem 3.1.7, it suffices to show that $\mathbf{0} \in U$, and that U is closed under vector addition and scalar multiplication.

First, since $\mathbf{a} \in M$, we have that $\mathbf{0} = \mathbf{a} - \mathbf{a} \in U$.

Next, fix $\mathbf{u}_1, \mathbf{u}_2 \in U$. We must show that $\mathbf{u}_1 + \mathbf{u}_2 \in U$. Since $\mathbf{u}_1, \mathbf{u}_2 \in U$, there exist $\mathbf{x}_1, \mathbf{x}_2 \in M$ such that $\mathbf{u}_1 = \mathbf{x}_1 - \mathbf{a}$ and $\mathbf{u}_2 = \mathbf{x}_2 - \mathbf{a}$. Then

$$\mathbf{u}_1 + \mathbf{u}_2 = (\mathbf{x}_1 - \mathbf{a}) + (\mathbf{x}_2 - \mathbf{a}) = \Big(\underbrace{1\mathbf{x}_1 + 1\mathbf{x}_2 + (-1)\mathbf{a}}_{:=\mathbf{y}}\Big) - \mathbf{a}.$$

Since $\mathbf{x}_1, \mathbf{x}_2, \mathbf{a} \in M$ and $1 + 1 + (-1) = 1$, and since (ii) holds, we see that $\mathbf{y} \in M$. But now $\mathbf{u}_1 + \mathbf{u}_2 = \mathbf{y} - \mathbf{a} \in U$.

Finally, fix $\mathbf{u} \in U$ and $\alpha \in \mathbb{F}$; we must show that $\alpha\mathbf{u} \in U$. Since $\mathbf{u} \in U$, we know that there exists some $\mathbf{x} \in M$ such that $\mathbf{u} = \mathbf{x} - \mathbf{a}$. But now

$$\alpha\mathbf{u} = \alpha(\mathbf{x} - \mathbf{a}) = \Big(\underbrace{\alpha\mathbf{x} + (1 - \alpha)\mathbf{a}}_{:=\mathbf{y}}\Big) - \mathbf{a}.$$

Since $\mathbf{x}, \mathbf{a} \in M$, and since (ii) holds, we have that $\mathbf{y} = \alpha\mathbf{x} + (1 - \alpha)\mathbf{a} \in M$. But now $\alpha\mathbf{u} = \mathbf{y} - \mathbf{a} \in U$.

We have now shown that U is a linear subspace of V, and it follows that (i) holds. □

Corollary 5.3.2. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ $(n \geq 1)$ be vectors in a vector space V over a field $\mathbb{F}$. Then $M := \mathit{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is an affine subspace of V.*

Proof. Since $\mathbf{x}_1, \ldots, \mathbf{x}_n \in M$, we see that $M \neq \emptyset$. In view of Theorem 5.3.1, it now suffices to show that M is closed under affine combinations. Fix $\mathbf{y}_1, \ldots, \mathbf{y}_m \in M$ and $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that $\alpha_1 + \cdots + \alpha_m = 1$. We must show that $\mathbf{y} := \alpha_1\mathbf{y}_1 + \cdots + \alpha_m\mathbf{y}_m$ belongs to M. Since $\mathbf{y}_1, \ldots, \mathbf{y}_m \in M$, we see that for all $i \in \{1, \ldots, m\}$, $\mathbf{y}_i$ is an affine combination of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, that is, there exist scalars $\beta_{i,1}, \ldots, \beta_{i,n} \in \mathbb{F}$ such that $\mathbf{y}_i = \sum_{j=1}^n \beta_{i,j}\mathbf{x}_j$ and $\sum_{j=1}^n \beta_{i,j} = 1$. But now

$$\mathbf{y} \;=\; \sum_{i=1}^m \alpha_i \mathbf{y}_i \;=\; \sum_{i=1}^m \alpha_i \Big(\sum_{j=1}^n \beta_{i,j}\mathbf{x}_j \Big) \;=\; \sum_{j=1}^n \sum_{i=1}^m \alpha_i\beta_{i,j}\mathbf{x}_j.$$

For each $j \in \{1, \ldots, n\}$, we set $\gamma_j := \sum_{i=1}^m \alpha_i\beta_{i,j}$. Then $\mathbf{y} = \sum_{j=1}^n \gamma_n\mathbf{x}_j$. It now remains to show that $\sum_{j=1}^n \gamma_j = 1$, for this will imply that $\mathbf{y}$ is an affine combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, that is, that $\mathbf{y} \in M$, which is what we need to show. We compute:

$$\sum_{j=1}^n \gamma_j \;=\; \sum_{j=1}^n \sum_{i=1}^m \alpha_i\beta_{i,j} \;=\; \sum_{i=1}^m \alpha_i \Big(\sum_{j=1}^n \beta_{i,j} \Big) \;\overset{(*)}{=}\; \sum_{i=1}^m \alpha_i \;=\; 1,$$

where (*) follows from the fact that $\sum_{j=1}^n \beta_{i,j} = 1$. This completes the argument. □

Corollary 5.3.3. *Let V be a vector space over a field $\mathbb{F}$, let M be an affine subspace of V, and let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ $(n \geq 1)$ be vectors in V. Then the following are equivalent:*

(i) $M = \mathit{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$;

(ii) $\mathbf{x}_1, \ldots, \mathbf{x}_n \in M$, and every vector in M is an affine combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Proof. Obviously, (i) implies (ii). For the reverse implication, we assume that (ii) holds, and we prove (i). Since every vector in M is an affine combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, we have that $M \subseteq \mathrm{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Let us prove the reverse inclusion. Fix $\mathbf{x} \in \mathrm{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. By (ii), we have that $\mathbf{x}_1, \ldots, \mathbf{x}_n \in M$, and by Theorem 5.3.1, we know that M is closed under affine combinations. Since $\mathbf{x}$ is an affine combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, we deduce that $\mathbf{x} \in M$. This proves that $\mathrm{Aff}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \subseteq M$. Thus, (i) holds. □

5.4 Affine frames and affine bases

We have extensively studied bases of (finite-dimensional) vector spaces. For affine subspaces, we have two analogues of bases: "affine frames" and "affine bases."

5.4.1 Affine frames

Let n be a non-negative integer, and let M be an n-dimensional affine subspace of a vector space V over a field $\mathbb{F}$. An *affine frame* of M is an ordered $(n+1)$-tuple $(\mathbf{a}, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ of vectors of V such that M can be written in the form $M = \mathbf{a} + U$, where U is a linear subspace of V,[9] and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of U.

Remark: Suppose that M is an affine subspace of a vector space V over a field $\mathbb{F}$, and that $\mathbf{a}, \mathbf{u}_1, \ldots, \mathbf{u}_n \in V$. It then follows from the definition that $(\mathbf{a}, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ is an affine frame of M if and only if vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent and $M = \mathbf{a} + \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$.[10]

Remark: Infinite-dimensional affine subspaces do not have affine frames.

By Theorem 3.2.7, if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of a vector space V over a field $\mathbb{F}$, then every vector in V can be written as a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in a unique way. Our next theorem is an analogue of this result for affine subspaces and affine frames.

Theorem 5.4.1. *Let M be an affine subspace of a vector space V over a field $\mathbb{F}$, and let $(\mathbf{a}, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ be an affine frame of M. Then for all $\mathbf{x} \in M$, there exist unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{x} = \mathbf{a} + \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$.*

Proof. Set $U := \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, so that $M = \mathbf{a} + U$. Fix $\mathbf{x} \in M$. We must show that there exist unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{x} = \mathbf{a} + \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$.

We first prove existence. Since $\mathbf{x} \in M = \mathbf{a} + U$, there exists some $\mathbf{u} \in U$ such that $\mathbf{x} = \mathbf{a} + \mathbf{u}$. Since $\mathbf{u} \in U = \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, we know that there exist scalars $\alpha_1, \ldots, \alpha_n$ such that $\mathbf{u} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$. So, $\mathbf{x} = \mathbf{a} + \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$. This proves existence.

Let us prove uniqueness. Fix scalars $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{F}$ such that $\mathbf{x} = \mathbf{a} + \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$ and $\mathbf{x} = \mathbf{a} + \beta_1\mathbf{u}_1 + \cdots + \beta_n\mathbf{u}_n$. Then $\mathbf{a} + \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n = \mathbf{a} + \beta_1\mathbf{u}_1 + \cdots + \beta_n\mathbf{u}_n$, and consequently, $(\alpha_1 - \beta_1)\mathbf{u}_1 + \cdots + (\alpha_n - \beta_n)\mathbf{u}_n = \mathbf{0}$. Since the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is linearly independent,[11] we have that $\alpha_1 - \beta_1 = \cdots = \alpha_n - \beta_n = 0$. It follows that $\alpha_i = \beta_i$ for all $i \in \{1, \ldots, n\}$. This proves uniqueness. □

[9] By Theorem 5.1.1(c), the linear subspace U is unique, i.e. it depends only on M, and not on the choice of $\mathbf{a}$.

[10] This is "obvious," but here is a formal proof. Suppose first that vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent and $M = \mathbf{a} + \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of $U := \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, and we deduce that $(\mathbf{a}, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ is an affine frame of $M = \mathbf{a} + U$. Suppose conversely that $(\mathbf{a}, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ is an affine frame of M. Then there exists some linear subspace U of V such that $M = \mathbf{a} + U$ and such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of U. But then vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent and $U = \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, and consequently, $M = \mathbf{a} + U = \mathbf{a} + \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$.

[11] This follows from the fact that $(\mathbf{a}, \mathbf{u}_1, \ldots, \mathbf{u}_n)$ is an affine frame of $M = \mathbf{a} + U$, and consequently, $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of U.

5.4.2 Affine independence

Given vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in a vector space V over a field $\mathbb{F}$, we say that vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in V$ are *affinely independent*, or that the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is *affinely independent*, if for all $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n = \mathbf{0} \qquad \text{and} \qquad \alpha_1 + \cdots + \alpha_n = 0,$$

we have that $\alpha_1 = \cdots = \alpha_n = 0$.

Proposition 5.4.2. *Let V be a vector space over a field $\mathbb{F}$, and let $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ ($n \geq 0$) be vectors in V. Then the following are equivalent:*

(i) $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ *are affinely independent;*

(ii) there exists some $i \in \{0, 1, \ldots, n\}$ *such that vectors*

$$\mathbf{x}_0 - \mathbf{x}_i, \ldots, \mathbf{x}_{i-1} - \mathbf{x}_i, \mathbf{x}_{i+1} - \mathbf{x}_i, \ldots, \mathbf{x}_n - \mathbf{x}_i$$

are linearly independent;

(iii) for all $i \in \{0, 1, \ldots, n\}$*, vectors*

$$\mathbf{x}_0 - \mathbf{x}_i, \ldots, \mathbf{x}_{i-1} - \mathbf{x}_i, \mathbf{x}_{i+1} - \mathbf{x}_i, \ldots, \mathbf{x}_n - \mathbf{x}_i$$

are linearly independent.

Proof. Obviously, (iii) implies (ii). We will show that (ii) implies (i), and that (i) implies (iii).

Suppose that (ii) holds. Let us prove (i). By (ii) and by symmetry, we may assume that $\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0$ are linearly independent. Now, fix scalars $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\alpha_0\mathbf{x}_0 + \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n = \mathbf{0}$ and $\alpha_0 + \alpha_1 + \cdots + \alpha_n = 0$. We must show that $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$. Since $\alpha_0 + \alpha_1 + \cdots + \alpha_n = 0$, we have that $\alpha_0 = -\alpha_1 - \cdots - \alpha_n$, and so

$$\begin{aligned} \mathbf{0} &= \alpha_0\mathbf{x}_0 + \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n \\ &= (-\alpha_1 - \cdots - \alpha_n)\mathbf{x}_0 + \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n \\ &= \alpha_1(\mathbf{x}_1 - \mathbf{x}_0) + \cdots + \alpha_n(\mathbf{x}_n - \mathbf{x}_0) \end{aligned}$$

Since vectors $\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0$ are linearly independent, we see that $\alpha_1 = \cdots = \alpha_n = 0$. Since $\alpha_0 = -\alpha_1 - \cdots - \alpha_n$, it follows that $\alpha_0 = 0$. This proves (i).

Suppose now that (i) holds. Let us prove (iii). By symmetry, it suffices to show that $\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0$ are linearly independent. Fix scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\alpha_1(\mathbf{x}_1 - \mathbf{x}_0) + \cdots + \alpha_n(\mathbf{x}_n - \mathbf{x}_0) = \mathbf{0}$. Then

$$\underbrace{(-\alpha_1 - \cdots - \alpha_n)}_{:=\alpha_0}\mathbf{x}_0 + \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n = \mathbf{0}.$$

Since $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are affinely independent, we now get that $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$, and we deduce that (iii) holds. □

5.4.3 Affine bases

Let M be an affine subspace of a vector space V over a field $\mathbb{F}$. An *affine basis* (also called a *barycentric frame*) of M is a non-empty ordered set $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of vectors in M such that

- vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are affinely independent;
- $M = \mathrm{Aff}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n)$.

As we shall see (see Theorem 5.4.4 and the remark following it), for any non-negative integer n, every affine basis of an n-dimensional affine subspace contains exactly $n+1$ vectors.

Remark: Suppose that M is an affine subspace of a vector space V over a field $\mathbb{F}$, and let $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ be vectors in V. In view of Corollary 5.3.3, we have that $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is an affine basis of M if and only if all the following hold:

1. $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n \in M$;
2. vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are affinely independent;
3. every vector in M can be expressed as an affine combination of $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$.

Recall that by Theorem 3.2.7, if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of a vector space V over a field $\mathbb{F}$, then every vector in V can be written as a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in a unique way. Theorem 5.4.1 was an analogue of this result for affine frames. Theorem 5.4.3 (below) is an analogue of that same result for affine bases.

Theorem 5.4.3. *Let M be an affine subspace of a vector space V over a field $\mathbb{F}$, and let $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be an affine basis of M. Then for all $\mathbf{x} \in M$, there exist unique scalars $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{F}$, called the* barycentric coordinates *of $\mathbf{x}$ with respect to the affine basis $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$, such that $\mathbf{x} = \sum_{i=0}^{n} \alpha_i \mathbf{x}_i$ and $\sum_{i=0}^{n} \alpha_i = 1$.*

Proof. Fix $\mathbf{x} \in M$. The existence of scalars $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{x} = \sum_{i=0}^{n} \alpha_i \mathbf{x}_i$ and $\sum_{i=0}^{n} \alpha_i = 1$ follows from the fact that $M = \mathrm{Aff}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n)$.

It remains to prove uniqueness. So, fix $\alpha_0, \alpha_1, \ldots, \alpha_n, \beta_0, \beta_1, \ldots, \beta_n \in \mathbb{F}$ such that

- $\mathbf{x} = \sum_{i=0}^{n} \alpha_i \mathbf{x}_i$ and $\sum_{i=0}^{n} \alpha_i = 1$;
- $\mathbf{x} = \sum_{i=0}^{n} \beta_i \mathbf{x}_i$ and $\sum_{i=0}^{n} \beta_i = 1$.

Then $\sum_{i=0}^{n} \alpha_i \mathbf{x}_i = \sum_{i=0}^{n} \beta_i \mathbf{x}_i$, and we deduce that $\sum_{i=0}^{n} (\alpha_i - \beta_i)\mathbf{x}_i = \mathbf{0}$. On the other hand,

$$\sum_{i=0}^{n} (\alpha_i - \beta_i) \;=\; \Big(\sum_{i=0}^{n} \alpha_i\Big) - \Big(\sum_{i=0}^{n} \beta_i\Big) \;=\; 1 - 1 \;=\; 0.$$

Since vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are affinely independent, we now deduce that $\alpha_0 - \beta_0 = \alpha_1 - \beta_1 = \cdots = \alpha_n - \beta_n = 0$. Therefore, $\alpha_i = \beta_i$ for all $i \in \{0, 1, \ldots, n\}$. This proves uniqueness. □

5.4.4 A relationship between affine bases and affine frames

Theorem 5.4.4. *Let V be a vector space over a field $\mathbb{F}$, let M be an affine subspace of V, and let $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n \in V$. Then the following are equivalent:*

(i) $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ *is an affine basis of M;*

(ii) $(\mathbf{x}_0, \mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0)$ *is an affine frame of M.*

Remark: Since every affine frame of an n-dimensional affine subspace contains $n+1$ vectors, Theorem 5.4.4 implies that every affine basis of an n-dimensional affine subspace contains exactly $n+1$ vectors.

Proof. First, we know that (i) and (ii) are, respectively, equivalent to (1) and (2) below:

(1) vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are **affinely** independent and

$$M \;=\; \mathrm{Aff}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n);$$

(2) vectors $\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0$ are **linearly** independent and

$$M \;=\; \mathbf{x}_0 + \mathrm{Span}(\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0).$$

So, it suffices to show that (1) and (2) are equivalent. By Proposition 5.4.2, vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ are affinely independent if and only if vectors $\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0$ are linearly independent. It now remains to show that $\mathrm{Aff}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathbf{x}_0 + \mathrm{Span}(\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0)$. For this, we compute:

$$\begin{aligned}
&\mathrm{Aff}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n) \\
=\;& \big\{\alpha_0\mathbf{x}_0 + \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n \mid \alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{F},\ \alpha_0 + \alpha_1 + \cdots + \alpha_n = 1\big\} \\
=\;& \big\{(1 - \alpha_1 - \cdots - \alpha_n)\mathbf{x}_0 + \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n \mid \alpha_1, \ldots, \alpha_n \in \mathbb{F}\big\} \\
=\;& \big\{\mathbf{x}_0 + \alpha_1(\mathbf{x}_1 - \mathbf{x}_0) + \cdots + \alpha_n(\mathbf{x}_n - \mathbf{x}_0) \mid \alpha_1, \ldots, \alpha_n \in \mathbb{F}\big\} \\
=\;& \mathbf{x}_0 + \mathrm{Span}(\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_n - \mathbf{x}_0).
\end{aligned}$$

This completes the argument. □

Remark: If $M = \{\mathbf{a}\}$ is a one-element affine subspace of a vector space V over a field $\mathbb{F}$, then $(\mathbf{a})$ is the (unique) affine frame and $\{\mathbf{a}\}$ the (unique) affine basis of M.

Remark: Suppose we are given a matrix $A \in \mathbb{F}^{n\times m}$ and a vector $\mathbf{b} \in \mathbb{F}^n$ (where $\mathbb{F}$ is a field). By Corollary 5.2.7, the solution set of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is either empty or an affine subspace of $\mathbb{F}^m$. Moreover, we have the following:

- if the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is **inconsistent**, then its solution set is **empty**, and consequently, it is **not** and affine subspace of $\mathbb{F}^m$ and therefore does **not** have an affine frame or an affine basis;
- if the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution, say $\mathbf{x}_0$, then $\{\mathbf{x}_0\}$ is the solution set of $A\mathbf{x} = \mathbf{b}$, and we see that $(\mathbf{x}_0)$ is the (unique) affine frame and $\{\mathbf{x}_0\}$ the (unique) affine basis of the solution set of $A\mathbf{x} = \mathbf{b}$;
- if the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has more than one solution, then an affine frame and an affine basis of the solution set of $A\mathbf{x} = \mathbf{b}$ can be computed by following the procedure from the solution of Example 5.4.5 (below).

Example 5.4.5. *Consider the following matrix and vector, both with entries in $\mathbb{Z}_2$:*

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Show that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent, and consequently (by Corollary 5.2.7(b)), an affine subspace of $\mathbb{Z}_2^6$. Find an affine frame and an affine basis of the solution set of $A\mathbf{x} = \mathbf{b}$.

Solution. We form the augmented matrix

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{array}\right],$$

and by row reducing, we obtain

$$\mathrm{RREF}\left(\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]\right) = \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

So, the general solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x} = \begin{bmatrix} r+s+t+1 \\ s+t+1 \\ r+t+1 \\ r \\ s \\ t \end{bmatrix} \qquad \text{where } r, s, t \in \mathbb{Z}_2.$$

In particular, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent, and so by Corollary 5.2.7(b), the solution set of this equation is an affine subspace of $\mathbb{Z}_2^6$.

Now, the solution set of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is the following (color coding is for emphasis):

$$S := \left\{ \begin{bmatrix} r+s+t+1 \\ s+t+1 \\ r+t+1 \\ r \\ s \\ t \end{bmatrix} \mid r,s,t \in \mathbb{Z}_2 \right\}$$

$$= \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + r \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \mid r,s,t \in \mathbb{Z}_2 \right\}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \mathrm{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

We now see that

$$\left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right)$$

is an affine frame of the solution set S of $A\mathbf{x} = \mathbf{b}$, whereas (by Theorem 5.4.4)

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

$$= \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is an affine basis of S. □

Chapter 6

Scalar (inner) products, norms, and orthogonality

6.1 The scalar product

So far, we have worked with vector spaces over arbitrary fields $\mathbb{F}$. In this chapter, we impose some additional structure on vector spaces, namely the "scalar product" (also called "inner product") and the "norm." A scalar product is a way of multiplying two vectors and obtaining a scalar. A norm is a way of measuring the distance of a vector from the origin, or alternatively, measuring the length of a vector. As a trade-off for imposing this additional structure, we restrict ourselves to vector spaces over only two fields: $\mathbb{R}$ and $\mathbb{C}$. The theory that we develop in this chapter would not work for vector spaces over general fields $\mathbb{F}$.

Terminology: Vector spaces over $\mathbb{R}$ are called *real vector spaces*, and vector spaces over $\mathbb{C}$ are called *complex vector spaces*.

6.1.1 The scalar product in real vector spaces

A *scalar product* (also called *inner product*) in a real vector space V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that satisfies the following four axioms:

r.1. for all $\mathbf{x} \in V$, $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, and equality holds if and only if $\mathbf{x} = \mathbf{0}$;

r.2. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$;

r.3. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{R}$, $\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$;

r.4. for all $\mathbf{x}, \mathbf{y} \in V$, $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.

The name "scalar product" comes from the fact that we multiply two vectors and obtain a scalar as a result.

Note that axioms r.2 and r.3 from the definition above guarantee that the scalar product in a real vector space V is linear in the first variable (when we keep the second variable fixed). But in fact, axioms r.2, r.3, and r.4 guarantee that it is linear in the second variable as well (when we keep the first variable fixed). More precisely, we have the following:

r.2'. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$;

r.3'. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{R}$, $\langle \mathbf{x}, \alpha\mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$.

To see that r.2' holds, note that for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, we have the following:

$$\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle \overset{\text{r.4}}{=} \langle \mathbf{y} + \mathbf{z}, \mathbf{x} \rangle \overset{\text{r.2}}{=} \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{z}, \mathbf{x} \rangle \overset{\text{r.4}}{=} \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle.$$

On the other hand, to see that r.3' holds, note that for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{R}$, we have the following:

$$\langle \mathbf{x}, \alpha\mathbf{y} \rangle \overset{\text{r.4}}{=} \langle \alpha\mathbf{y}, \mathbf{x} \rangle \overset{\text{r.3}}{=} \alpha \langle \mathbf{y}, \mathbf{x} \rangle \overset{\text{r.4}}{=} \alpha \langle \mathbf{x}, \mathbf{y} \rangle.$$

The standard scalar product in $\mathbb{R}^n$. Perhaps the best known example of a scalar product is the "standard scalar product" (sometimes also called the "dot product") in $\mathbb{R}^n$. The *standard scalar product* of vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$ is given by

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^{n} x_i y_i.$$

(By Proposition 6.1.1, this really is a scalar product in $\mathbb{R}^n$.) For example, for vectors $\begin{bmatrix} 1 & -2 & 5 \end{bmatrix}^T$ and $\begin{bmatrix} -3 & 2 & 1 \end{bmatrix}^T$ in $\mathbb{R}^3$, we compute:

$$\begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} = 1 \cdot (-3) + (-2) \cdot 2 + 5 \cdot 1 = -2.$$

Note that for vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \end{bmatrix} = \begin{bmatrix} \mathbf{x} \cdot \mathbf{y} \end{bmatrix}.$$

So, if we identify 1×1 matrices with scalars, then we simply get that

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proposition 6.1.1. *The standard scalar product in $\mathbb{R}^n$ is a scalar product.*

Proof. We need to check that the standard scalar product $\cdot$ in $\mathbb{R}^n$ satisfies the four axioms from the definition of a scalar product in a real vector space.

r.1. For a vector $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x} \cdot \mathbf{x} \;=\; \sum_{i=1}^{n} x_i^2 \;\overset{(*)}{\geq}\; 0,$$

and (*) is an equality if and only if $x_1 = \dots = x_n = 0$, i.e. if and only if $\mathbf{x} = \mathbf{0}$.

r.2. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$, $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$, and $\mathbf{z} = \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\begin{aligned} (\mathbf{x}+\mathbf{y}) \cdot \mathbf{z} &= \sum_{i=1}^{n} (x_i + y_i) z_i \\ &= \Big(\sum_{i=1}^{n} x_i z_i \Big) + \Big(\sum_{i=1}^{n} y_i z_i \Big) \\ &= \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}. \end{aligned}$$

r.3. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$ and a scalar $\alpha \in \mathbb{R}$, we have that

$$(\alpha \mathbf{x}) \cdot \mathbf{y} \;=\; \sum_{i=1}^{n} (\alpha x_i) y_i \;=\; \alpha \sum_{i=1}^{n} x_i y_i \;=\; \alpha (\mathbf{x} \cdot \mathbf{y}).$$

r.4. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x} \cdot \mathbf{y} \;=\; \sum_{i=1}^{n} x_i y_i \;=\; \sum_{i=1}^{n} y_i x_i \;=\; \mathbf{y} \cdot \mathbf{x}.$$

This proves that the standard scalar product in $\mathbb{R}^n$ really is a scalar product. □

We note that a similar type of scalar product can be defined for matrices. Indeed, for matrices $A = \begin{bmatrix} a_{i,j} \end{bmatrix}_{n \times m}$ and $B = \begin{bmatrix} b_{i,j} \end{bmatrix}_{n \times m}$ in $\mathbb{R}^{n \times m}$, we can define $\langle A, B \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij}$. It is easy to verify that this really is a scalar product in $\mathbb{R}^{n \times m}$ (the proof is similar to that of Proposition 6.1.1).

Remark: The standard scalar product is only one of many possible scalar products in $\mathbb{R}^n$. A full characterization of all possible scalar products in $\mathbb{R}^n$ (and more generally, in all non-trivial, finite-dimensional real vector spaces) is given by Theorem 10.5.1.

An example with integrals. For readers who have studied calculus, Proposition 6.1.2 (below) gives an example of a scalar product involving integrals.

Proposition 6.1.2. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $\mathcal{C}_{[a,b]}$ be the (real) vector space of all continuous functions from the closed interval $[a, b]$ to $\mathbb{R}$.*[1] *Then the function $\langle \cdot, \cdot \rangle : \mathcal{C}_{[a,b]} \times \mathcal{C}_{[a,b]} \to \mathbb{R}$ defined by*

$$\langle f, g \rangle \quad := \quad \int_a^b f(x)g(x)dx$$

for all $f, g \in \mathcal{C}_{[a,b]}$ is a scalar product.

Proof. We must verify that the four axioms from the definition of a scalar product are satisfied. We first prove that axioms r.2, r.3, and r.4 are satisfied, and then we prove that axiom r.1 is satisfied (our proof of r.1 relies on r.3 and r.4, which is why we prove r.1 last).

r.2. For $f_1, f_2, f_3 \in \mathcal{C}_{[a,b]}$, we have that

$$\begin{aligned}
\langle f_1 + f_2, f_3 \rangle &= \int_a^b \Big(f_1(x) + f_2(x) \Big) f_3(x)dx \\
&= \int_a^b \Big(f_1(x)f_3(x) + f_2(x)f_3(x) \Big) dx \\
&= \int_a^b f_1(x)f_3(x)dx + \int_a^b f_2(x)f_3(x)dx \\
&= \langle f_1, f_3 \rangle + \langle f_2, f_3 \rangle.
\end{aligned}$$

r.3. For $f_1, f_2 \in \mathcal{C}_{[a,b]}$ and $\alpha \in \mathbb{R}$, we have that

$$\begin{aligned}
\langle \alpha f_1, f_2 \rangle &= \int_a^b \Big(\alpha f_1(x) \Big) f_2(x)dx \\
&= \alpha \int_a^b f_1(x)f_2(x)dx \\
&= \alpha \langle f_1, f_2 \rangle.
\end{aligned}$$

r.4. For $f_1, f_2 \in \mathcal{C}_{[a,b]}$, we have that

$$\langle f_1, f_2 \rangle \quad = \quad \int_a^b f_1(x)f_2(x)dx \quad = \quad \int_a^b f_2(x)f_1(x)dx \quad = \quad \langle f_2, f_1 \rangle.$$

[1] Recall from calculus that all such functions are integrable.

r.1. Let $f \in \mathcal{C}_{[a,b]}$. Then

$$\langle f, f \rangle \;=\; \int_a^b f(x)^2 dx \;\overset{(*)}{\geq}\; 0,$$

where (*) follows from the fact that $f(x)^2 \geq 0$ for all $x \in [a, b]$. If $f(x) = 0$ for all $x \in [a, b]$, then obviously, $\langle f, f \rangle = 0$. Suppose now that there exists some $x_0 \in [a, b]$ such that $f(x_0) \neq 0$. We must show that $\langle f, f \rangle > 0$.

Suppose first that $f(x_0) > 0$. Set $m = \frac{f(x_0)}{2}$. (Clearly, $m > 0$.) Then since f is continuous on $[a, b]$, there exist $a_0, b_0 \in \mathbb{R}$ such that $a \leq a_0 \leq x_0 \leq b_0 \leq b$ and $a_0 < b_0$, and such that for all $x \in [a_0, b_0]$, we have that $f(x) \geq m$.[2] We now compute:

$$\begin{aligned}
\langle f, f \rangle \;&=\; \int_a^b f(x)^2 dx \\
&=\; \int_a^{a_0} f(x)^2 dx + \int_{a_0}^{b_0} f(x)^2 dx + \int_{b_0}^b f(x)^2 dx \\
&\overset{(*)}{\geq}\; \int_{a_0}^{b_0} f(x)^2 dx \\
&\overset{(**)}{\geq}\; m^2(b_0 - a_0) \\
&>\; 0,
\end{aligned}$$

where (*) follows from the fact that $f(x)^2 \geq 0$ for all $x \in [a, a_0] \cup [b_0, b]$, and (**) follows from the fact that $f(x)^2 \geq m^2$ for all $x \in [a_0, b_0]$.

Suppose now that $f(x_0) < 0$. Then $-f(x_0) > 0$. So, by an argument completely analogous to the above (applied to $-f$ instead of f), we obtain $\langle -f, -f \rangle > 0$. We now use axioms r.3 and r.4 (which we have already verified) to obtain the following:

$$\langle -f, -f \rangle \;\overset{\text{r.3}}{=}\; -\langle f, -f \rangle \;\overset{\text{r.4}}{=}\; -\langle -f, f \rangle \;\overset{\text{r.3}}{=}\; \langle f, f \rangle.$$

Since $\langle -f, -f \rangle > 0$, it follows that $\langle f, f \rangle > 0$. □

[2] This is essentially because, by the continuity of f, we have that "$x \approx x_0 \Longrightarrow f(x) \approx f(x_0)$." So, since $f(x_0) > m$, there exists some (sufficiently small) subinterval $[a_0, b_0]$ of $[a, b]$ such that $x_0 \in [a_0, b_0]$ and such that for all $x \in [a_0, b_0]$, we have that $f(x) \geq m$. Here is a formal proof, for those readers who would like one. Set $\varepsilon := m = \frac{f(x_0)}{2}$. By the continuity of f, there exists some $\delta > 0$ such that for all $x \in [a, b]$, if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$. Now, set $a_0 := \max\left\{a, x_0 - \frac{\delta}{2}\right\}$ and $b_0 := \min\left\{b, x_0 + \frac{\delta}{2}\right\}$. Then $a \leq a_0 \leq x_0 \leq b_0 \leq b$ and $a_0 < b_0$. Moreover, by construction, for all $x \in [a_0, b_0]$, we have that $|x - x_0| \leq \frac{\delta}{2} < \delta$, and consequently, $|f(x) - f(x_0)| < \varepsilon$, i.e. $f(x_0) - \varepsilon < f(x) < f(x_0) + \varepsilon$. Since $f(x_0) = 2m$ and $\varepsilon = m$, we deduce that for all $x \in [a_0, b_0]$, we have that $m < f(x) < 3m$, and in particular, $f(x) \geq m$.

6.1.2 The scalar product in complex vector spaces

Recall from section 0.3 that the *complex conjugate* of a complex number $z = a + bi$ (where $a, b \in \mathbb{R}$) is defined to be the number $\overline{z} := a - bi$.

A *scalar product* (also called *inner product*) in a complex vector space V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ that satisfies the following four axioms:

c.1. for all $\mathbf{x} \in V$, $\langle \mathbf{x}, \mathbf{x} \rangle$ is a real number, $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, and equality holds if and only if $\mathbf{x} = \mathbf{0}$;

c.2. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$;

c.3. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$, $\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$;

c.4. for all $\mathbf{x}, \mathbf{y} \in V$, $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$.

Note that axioms c.2 and c.3 from the definition above guarantee that the scalar product in a complex vector space V is linear in the first variable (when we keep the second variable fixed). Unlike in the real case, it is **not** linear in the second variable (when we keep the first variable fixed). We do, however, have the following:

c.2'. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$;

c.3'. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$, $\langle \mathbf{x}, \alpha \mathbf{y} \rangle = \overline{\alpha} \langle \mathbf{x}, \mathbf{y} \rangle$.

To see that c.2' holds, note that for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, we have the following:

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle &\overset{\text{c.4}}{=} \overline{\langle \mathbf{y} + \mathbf{z}, \mathbf{x} \rangle} \\ &\overset{\text{c.2}}{=} \overline{\langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{z}, \mathbf{x} \rangle} \\ &= \overline{\langle \mathbf{y}, \mathbf{x} \rangle} + \overline{\langle \mathbf{z}, \mathbf{x} \rangle} \\ &\overset{\text{c.4}}{=} \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle. \end{aligned}$$

On the other hand, to see that c.3' holds, note that for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$, we have the following:

$$\langle \mathbf{x}, \alpha \mathbf{y} \rangle \overset{\text{c.4}}{=} \overline{\langle \alpha \mathbf{y}, \mathbf{x} \rangle} \overset{\text{c.3}}{=} \overline{\alpha \langle \mathbf{y}, \mathbf{x} \rangle} = \overline{\alpha}\, \overline{\langle \mathbf{y}, \mathbf{x} \rangle} \overset{\text{c.4}}{=} \overline{\alpha} \langle \mathbf{x}, \mathbf{y} \rangle.$$

The standard scalar product in $\mathbb{C}^n$. The *standard scalar product* of vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{C}^n$ is given by

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^{n} x_i \overline{y_i}.$$

(By Proposition 6.1.3, this really is a scalar product in $\mathbb{C}^n$.) For example, for vectors $\begin{bmatrix} 1-2i & -2+i \end{bmatrix}^T$ and $\begin{bmatrix} 2+i & 1+3i \end{bmatrix}^T$ in $\mathbb{C}^2$, we compute:

$$\begin{aligned} \begin{bmatrix} 1-2i \\ -2+i \end{bmatrix} \cdot \begin{bmatrix} 2+i \\ 1+3i \end{bmatrix} &= (1-2i)\overline{(2+i)} + (-2+i)\overline{(1+3i)} \\ &= (1-2i)(2-i) + (-2+i)(1-3i) \\ &= 1+2i. \end{aligned}$$

Proposition 6.1.3. *The standard scalar product in $\mathbb{C}^n$ is a scalar product.*

Proof. We need to check that the standard scalar product $\cdot$ in $\mathbb{C}$ satisfies the four axioms from the definition of a scalar product in a complex vector space.

c.1. For a vector $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$ in $\mathbb{C}^n$, we have that

$$\mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^{n} x_i \overline{x_i} \overset{(*)}{=} \sum_{i=1}^{n} |x_i|^2 \overset{(**)}{\geq} 0,$$

where (*) follows from Proposition 0.3.2. Moreover, note that the inequality (**) is an equality if and only if $x_1 = \dots = x_n = 0$, i.e. if and only if $\mathbf{x} = \mathbf{0}$.

c.2. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$, $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$, and $\mathbf{z} = \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}^T$ in $\mathbb{C}^n$, we have that

$$\begin{aligned} (\mathbf{x}+\mathbf{y}) \cdot \mathbf{z} &= \sum_{i=1}^{n} (x_i + y_i)\overline{z_i} \\ &= \Big(\sum_{i=1}^{n} x_i \overline{z_i} \Big) + \Big(\sum_{i=1}^{n} y_i \overline{z_i} \Big) \\ &= \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}. \end{aligned}$$

c.3. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{C}^n$ and a scalar $\alpha \in \mathbb{C}$, we have that

$$(\alpha \mathbf{x}) \cdot \mathbf{y} = \sum_{i=1}^{n} (\alpha x_i)\overline{y_i} = \alpha \sum_{i=1}^{n} x_i \overline{y_i} = \alpha(\mathbf{x} \cdot \mathbf{y}).$$

c.4. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{C}^n$, we have that

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i \overline{y_i} = \sum_{i=1}^{n} \overline{\overline{x_i} y_i} = \overline{\sum_{i=1}^{n} \overline{x_i} y_i} = \overline{\sum_{i=1}^{n} y_i \overline{x_i}} = \overline{\mathbf{y} \cdot \mathbf{x}}.$$

This proves that the standard scalar product in $\mathbb{C}^n$ really is a scalar product. $\square$

6.1.3 Orthogonality

Given a real or complex vector space V, equipped with a scalar product $\langle \cdot, \cdot \rangle$, we say that vectors $\mathbf{x}$ and $\mathbf{y}$ in V are *orthogonal*, and we write $\mathbf{x} \perp \mathbf{y}$, if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. When our scalar product is the **standard** scalar product in $\mathbb{R}^n$, this corresponds to the usual geometric interpretation (a detailed explanation is given in subsection 6.2.1). However, for general scalar products, this is how we **define** orthogonality.[3]

Proposition 6.1.4. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. Then all the following hold:*

(a) for all vectors $\mathbf{x}, \mathbf{y} \in V$, we have that $\mathbf{x} \perp \mathbf{y}$ if and only if $\mathbf{y} \perp \mathbf{x}$;

(b) for all vectors $\mathbf{x}, \mathbf{y} \in V$ and scalars α, β,[4] if $\mathbf{x} \perp \mathbf{y}$ then $(\alpha \mathbf{x}) \perp (\beta \mathbf{y})$;

(c) for all vectors $\mathbf{x} \in V$, we have that $\mathbf{x} \perp \mathbf{0}$ and $\mathbf{0} \perp \mathbf{x}$.

Proof. We prove the proposition for the case when V is a complex vector space. The real case is similar but slightly easier (because we do not have to deal with complex conjugates).

(a) For vectors $\mathbf{x}, \mathbf{y} \in V$, we have the following sequence of equivalences:

$$\begin{aligned}
\mathbf{x} \perp \mathbf{y} \quad &\iff \quad \langle \mathbf{x}, \mathbf{y} \rangle = 0 && \text{by definition} \\
&\iff \quad \overline{\langle \mathbf{y}, \mathbf{x} \rangle} = 0 && \text{by c.4} \\
&\iff \quad \langle \mathbf{y}, \mathbf{x} \rangle = 0 && \\
&\iff \quad \mathbf{y} \perp \mathbf{x} && \text{by definition.}
\end{aligned}$$

(b) Fix vectors $\mathbf{x}, \mathbf{y} \in V$ and scalars $\alpha, \beta \in \mathbb{C}$, and assume that $\mathbf{x} \perp \mathbf{y}$. Then we compute:

$$\begin{aligned}
\langle \alpha \mathbf{x}, \beta \mathbf{y} \rangle &= \alpha \langle \mathbf{x}, \beta \mathbf{y} \rangle && \text{by c.3} \\
&= \alpha \overline{\beta} \langle \mathbf{x}, \mathbf{y} \rangle && \text{by c.3'} \\
&= \alpha \overline{\beta} 0 && \text{beause } \mathbf{x} \perp \mathbf{y} \\
&= 0.
\end{aligned}$$

[3] For example, for the scalar product defined on $\mathcal{C}_{[-\pi,\pi]}$ in Proposition 6.1.2, we have that $\sin x \perp \cos x$, since $\langle \sin x, \cos x \rangle = \int_{-\pi}^{\pi} \sin x \cos x dx = 0$.

[4] Here, α and β are real or complex numbers, depending on whether V is a real or complex vector space.

So, $(\alpha\mathbf{x}) \perp (\beta\mathbf{y})$.

(c) Fix any vector $\mathbf{x} \in V$. We then have that

$$\langle \mathbf{0}, \mathbf{x} \rangle \;=\; \langle 0\mathbf{0}, \mathbf{x} \rangle \;\overset{\text{c.3}}{=}\; 0\langle \mathbf{0}, \mathbf{x} \rangle \;=\; 0,$$

and so $\mathbf{0} \perp \mathbf{x}$. The fact that $\mathbf{x} \perp \mathbf{0}$ now follows from (a). □

Suppose that V is a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. For a vector $\mathbf{v} \in V$ and a set of vectors $A \subseteq V$,[5] we say that $\mathbf{v}$ is *orthogonal* to A, and we write $\mathbf{v} \perp A$, provided that $\mathbf{v}$ is orthogonal to all vectors in A.[6] For sets of vectors $A, B \subseteq V$,[7] we say that A is *orthogonal* to B, and we write $A \perp B$, if every vector in A is orthogonal to every vector in B.

Proposition 6.1.5. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. Let $\mathbf{a}_1, \ldots, \mathbf{a}_p, \mathbf{b}_1, \ldots, \mathbf{b}_q \in V$, and assume that $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \perp \{\mathbf{b}_1, \ldots, \mathbf{b}_q\}$. Then Span$(\mathbf{a}_1, \ldots, \mathbf{a}_p) \perp$ Span$(\mathbf{b}_1, \ldots, \mathbf{b}_q)$.*

Proof. Fix $\mathbf{a} \in \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_p)$ and $\mathbf{b} \in \mathrm{Span}(\mathbf{b}_1, \ldots, \mathbf{b}_q)$. Then there exist scalars $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ such that $\mathbf{a} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_p\mathbf{a}_p$ and $\mathbf{b} = \beta_1\mathbf{b}_1 + \cdots + \beta_q\mathbf{b}_q$. We now compute:

$$\begin{aligned}
\langle \mathbf{a}, \mathbf{b} \rangle &= \Big\langle \sum_{i=1}^{p} \alpha_i \mathbf{a}_i, \sum_{j=1}^{q} \beta_j \mathbf{b}_j \Big\rangle \\
&= \sum_{i=1}^{p} \Big\langle \alpha_i \mathbf{a}_i, \sum_{j=1}^{q} \beta_j \mathbf{b}_j \Big\rangle \qquad \text{by r.2 or c.2} \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \underbrace{\langle \alpha_i \mathbf{a}_i, \beta_j \mathbf{b}_j \rangle}_{\overset{(*)}{=} 0} \qquad \text{by r.2' or c.2'} \\
&= 0,
\end{aligned}$$

where (*) follows from Proposition 6.1.4(b) and from the fact that $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\} \perp \{\mathbf{b}_1, \ldots, \mathbf{b}_q\}$. This proves that $\mathbf{a} \perp \mathbf{b}$, and the result follows. □

6.2 The norm

In this section, we introduce the notion of a "norm" $|| \cdot ||$ in a real or complex vector space V. The idea is that for a vector $\mathbf{x} \in V$, $||\mathbf{x}||$ is the distance from $\mathbf{x}$ to

[5] A may, but need not be, a subspace of V.

[6] By definition, this means that for all $\mathbf{a} \in A$, we have that $\langle \mathbf{v}, \mathbf{a} \rangle = 0$.

[7] Again, A and B may or may not be subspaces of V.

the origin, or alternatively, the length of the vector $\mathbf{x}$; $||\mathbf{x}||$ is always supposed to be a non-negative real number (even if V is a complex vector space). For vectors $\mathbf{x}, \mathbf{y} \in V$, $||\mathbf{x} - \mathbf{y}||$ is supposed to be the distance between $\mathbf{x}$ and $\mathbf{y}$. Distance can be defined in a variety of ways. We first study norms induced by a scalar product (see subsections 6.2.1 and 6.2.2). The definition of a norm in general is given in subsection 6.2.3, and some additional examples of norms are given in subsection 6.2.4.

6.2.1 The norm induced by a scalar product

Given a scalar product $\langle \cdot, \cdot \rangle$ in a real or complex vector space V, we define the *norm in V induced by $\langle \cdot, \cdot \rangle$* to be the function $|| \cdot || : V \to \mathbb{R}$ given by

$$||\mathbf{x}|| \quad := \quad \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

for all $\mathbf{x} \in V$. In view of r.1 and c.1, for all $\mathbf{x} \in V$, we have that $||\mathbf{x}||$ is a non-negative **real** number,[8] and moreover, $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

Proposition 6.2.1. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Then for all vectors $\mathbf{x} \in V$ and scalars α,*[9] *we have that*

$$||\alpha \mathbf{x}|| \quad = \quad |\alpha| \, ||\mathbf{x}||.$$

Proof. If the vector space V is real, then for all vectors $\mathbf{x} \in V$ and scalars $\alpha \in \mathbb{R}$, we have that

$$||\alpha \mathbf{x}|| \quad = \quad \sqrt{\langle \alpha \mathbf{x}, \alpha \mathbf{x} \rangle} \quad \stackrel{(*)}{=} \quad \sqrt{\alpha^2 \langle \mathbf{x}, \mathbf{x} \rangle} \quad = \quad |\alpha| \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \quad = \quad |\alpha| \, ||\mathbf{x}||,$$

where (*) follows from r.3 and r.3'.

On the other hand, if the vector space V is complex, then for all vectors $\mathbf{x} \in V$ and scalars $\alpha \in \mathbb{C}$, we have that

$$\begin{aligned} ||\alpha \mathbf{x}|| &= \sqrt{\langle \alpha \mathbf{x}, \alpha \mathbf{x} \rangle} \\ &= \sqrt{\alpha \overline{\alpha} \langle \mathbf{x}, \mathbf{x} \rangle} \qquad \text{by c.3 and c.3'} \\ &= \sqrt{|\alpha|^2 \langle \mathbf{x}, \mathbf{x} \rangle} \qquad \text{by Proposition 0.3.2} \\ &= |\alpha| \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \\ &= |\alpha| \, ||\mathbf{x}||. \end{aligned}$$

This completes the argument. □

[8] This happens even if V is a complex vector space.

[9] So, α is a real or complex number, depending on whether the vector space V is real or complex.

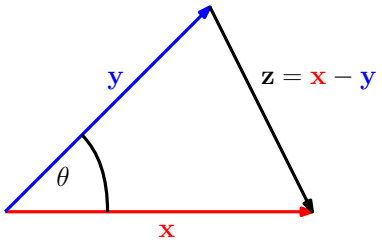
Note that if $||\cdot||$ is the norm induced by the **standard** scalar product in $\mathbb{R}^n$, then for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$||\mathbf{x}|| \quad = \quad \sqrt{\mathbf{x} \cdot \mathbf{x}} \quad = \quad \sqrt{\sum_{i=1}^{n} x_i^2}.$$

So, we simply get the standard Euclidean length in $\mathbb{R}^n$. We note that if $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ are non-zero vectors in $\mathbb{R}^n$, then we have that

$$\mathbf{x} \cdot \mathbf{y} \quad = \quad ||\mathbf{x}|| \, ||\mathbf{y}|| \, \cos\theta,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. To see this, consider the triangle formed by $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z} := \mathbf{x} - \mathbf{y}$,[10] and let θ be the angle between $\mathbf{x}$ and $\mathbf{y}$ in this triangle.



We then compute

$$\begin{aligned} ||\mathbf{z}||^2 \quad &= \quad \mathbf{z} \cdot \mathbf{z} \\ &= \quad (\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) \\ &= \quad \underbrace{\mathbf{x} \cdot \mathbf{x}}_{=||\mathbf{x}||^2} - \mathbf{x} \cdot \mathbf{y} - \mathbf{y} \cdot \mathbf{x} + \underbrace{\mathbf{y} \cdot \mathbf{y}}_{=||\mathbf{y}||^2} \\ &= \quad ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2\mathbf{x} \cdot \mathbf{y} \end{aligned}$$

On the other hand, the Law of Cosines (for triangles) tells us that

$$||\mathbf{z}||^2 \quad = \quad ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2||\mathbf{x}|| \, ||\mathbf{y}|| \, \cos\theta.$$

So, $||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2\mathbf{x} \cdot \mathbf{y} = ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2||\mathbf{x}|| \, ||\mathbf{y}|| \, \cos\theta$, and consequently,

$$\mathbf{x} \cdot \mathbf{y} \quad = \quad ||\mathbf{x}|| \, ||\mathbf{y}|| \, \cos\theta,$$

as we had claimed. Note that this means that non-zero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are orthogonal in the usual geometric sense (i.e. the angle between them is $90°$) if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.[11]

[10] This triangle may possibly be "degenerate" (i.e. one-dimensional). This happens if $\mathbf{x}$ and $\mathbf{y}$ are scalar multiples of each other.

[11] This is because for an angle θ, with $0° \leq \theta \leq 180°$, we have that $\cos\theta = 0$ if and only if $\theta = 90°$.

Warning: The formula $\mathbf{x} \cdot \mathbf{y} = ||\mathbf{x}||\,||\mathbf{y}||\cos\theta$ that we obtained above only works for the **standard** scalar product in $\mathbb{R}^n$ and the norm induced by it. Do not attempt to use it for general scalar products!

6.2.2 The Pythagorean theorem, the Cauchy–Schwarz inequality, and the triangle inequality

The Pythagorean theorem. *Let V be a real or complex vector space, equipped with a scalar product $\langle\cdot,\cdot\rangle$ and the induced norm $||\cdot||$. Then for all $\mathbf{x},\mathbf{y} \in V$ such that $\mathbf{x} \perp \mathbf{y}$, we have that*

$$||\mathbf{x}+\mathbf{y}||^2 \;=\; ||\mathbf{x}||^2 + ||\mathbf{y}||^2.$$

Proof. Fix $\mathbf{x},\mathbf{y} \in V$ such that $\mathbf{x} \perp \mathbf{y}$. Then $\langle \mathbf{x},\mathbf{y}\rangle = 0$ and $\langle \mathbf{y},\mathbf{x}\rangle = 0$. So,

$$\begin{aligned} ||\mathbf{x}+\mathbf{y}||^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle \\ &= \underbrace{\langle \mathbf{x},\mathbf{x}\rangle}_{=||\mathbf{x}||^2} + \underbrace{\langle \mathbf{x},\mathbf{y}\rangle}_{=0} + \underbrace{\langle \mathbf{y},\mathbf{x}\rangle}_{=0} + \underbrace{\langle \mathbf{y},\mathbf{y}\rangle}_{=||\mathbf{y}||^2} \\ &= ||\mathbf{x}||^2 + ||\mathbf{y}||^2, \end{aligned}$$

which is what we needed to show. □

The Cauchy–Schwarz inequality. *Let V be a real or complex vector space, equipped with a scalar product $\langle\cdot,\cdot\rangle$ and the induced norm $||\cdot||$. Then*

$$|\langle \mathbf{x},\mathbf{y}\rangle| \;\leq\; ||\mathbf{x}||\,||\mathbf{y}||$$

for all $\mathbf{x},\mathbf{y} \in V$.

Proof. Fix $\mathbf{x},\mathbf{y} \in V$. We may assume that $\langle \mathbf{x},\mathbf{y}\rangle \neq 0$, for otherwise, the result is immediate. Note that this implies that $\mathbf{x},\mathbf{y} \neq \mathbf{0}$, and consequently, $||\mathbf{x}||, ||\mathbf{y}|| \neq 0$. We set

$$\mathbf{z} \;:=\; \tfrac{\langle \mathbf{y},\mathbf{y}\rangle}{\langle \mathbf{x},\mathbf{y}\rangle}\mathbf{x} - \mathbf{y},$$

and we compute

$$\langle \mathbf{z},\mathbf{y}\rangle \;=\; \langle \tfrac{\langle \mathbf{y},\mathbf{y}\rangle}{\langle \mathbf{x},\mathbf{y}\rangle}\mathbf{x} - \mathbf{y}, \mathbf{y}\rangle \;\overset{(*)}{=}\; \tfrac{\langle \mathbf{y},\mathbf{y}\rangle}{\langle \mathbf{x},\mathbf{y}\rangle}\langle \mathbf{x},\mathbf{y}\rangle - \langle \mathbf{y},\mathbf{y}\rangle \;=\; 0,$$

where (*) follows from r.2 and r.3 if V is a real vector space, or from c.2 and c.3 if V is a complex vector space. We have now shown that $\mathbf{z} \perp \mathbf{y}$, and so by the Pythagorean theorem, we have that

$$||\mathbf{z}+\mathbf{y}||^2 \;=\; ||\mathbf{z}||^2 + ||\mathbf{y}||^2.$$

But by construction, $\mathbf{z}+\mathbf{y} = \frac{\langle \mathbf{y},\mathbf{y}\rangle}{\langle \mathbf{x},\mathbf{y}\rangle}\mathbf{x}$, and consequently:

$$||\mathbf{z}+\mathbf{y}|| \;=\; ||\tfrac{\langle \mathbf{y},\mathbf{y}\rangle}{\langle \mathbf{x},\mathbf{y}\rangle}\mathbf{x}|| \;\overset{(*)}{=}\; |\tfrac{\langle \mathbf{y},\mathbf{y}\rangle}{\langle \mathbf{x},\mathbf{y}\rangle}|\,||\mathbf{x}|| \;=\; \tfrac{|\langle \mathbf{y},\mathbf{y}\rangle|}{|\langle \mathbf{x},\mathbf{y}\rangle|}\,||\mathbf{x}|| \;=\; \tfrac{||\mathbf{y}||^2}{|\langle \mathbf{x},\mathbf{y}\rangle|}\,||\mathbf{x}||,$$

where (*) follows from Proposition 6.2.1. So,

$$\tfrac{||\mathbf{y}||^4}{|\langle \mathbf{x},\mathbf{y}\rangle|^2}\,||\mathbf{x}||^2 \;=\; ||\mathbf{z}+\mathbf{y}||^2 \;=\; ||\mathbf{z}||^2+||\mathbf{y}||^2 \;\geq\; ||\mathbf{y}||^2,$$

which yields

$$\tfrac{||\mathbf{y}||^4}{|\langle \mathbf{x},\mathbf{y}\rangle|^2}\,||\mathbf{x}||^2 \;\geq\; ||\mathbf{y}||^2.$$

Since $\langle \mathbf{x},\mathbf{y}\rangle$ and $||\mathbf{y}||$ are both non-zero, we have that $\frac{|\langle \mathbf{x},\mathbf{y}\rangle|^2}{||\mathbf{y}||^2}$ is defined and positive. So, we may multiply both sides of the inequality above by $\frac{|\langle \mathbf{x},\mathbf{y}\rangle|^2}{||\mathbf{y}||^2}$ to obtain

$$||\mathbf{x}||^2||\mathbf{y}||^2 \;\geq\; |\langle \mathbf{x},\mathbf{y}\rangle|^2.$$

By taking the square root of both sides, we get

$$||\mathbf{x}||\,||\mathbf{y}|| \;\geq\; |\langle \mathbf{x},\mathbf{y}\rangle|,$$

which is what we needed to show. □

Corollary 6.2.2. *For all $x_1,\ldots,x_n,y_1,\ldots,y_n\in\mathbb{R}$, we have that*

$$\Big(\sum_{i=1}^n x_iy_i\Big)^2 \;\leq\; \Big(\sum_{i=1}^n x_i^2\Big)\Big(\sum_{i=1}^n y_i^2\Big).$$

Proof. If we consider the standard scalar product in $\mathbb{R}^n$, the Cauchy-Schwarz inequality yields

$$\Big|\sum_{i=1}^n x_iy_i\Big| \;\leq\; \sqrt{\sum_{i=1}^n x_i^2}\sqrt{\sum_{i=1}^n y_i^2}.$$

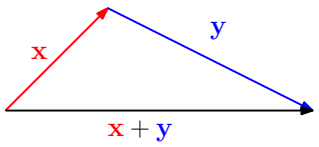
for all $x_1,\ldots,x_n,y_1,\ldots,y_n\in\mathbb{R}$. By squaring both sides, we obtain the desired inequality. □

As a further corollary of the Cauchy-Schwarz inequality, we obtain the following.

The triangle inequality. *Let V be a real or complex vector space, equipped with a scalar product $\langle\cdot,\cdot\rangle$ and the induced norm $||\cdot||$. Then*

$$||\mathbf{x}+\mathbf{y}|| \;\leq\; ||\mathbf{x}||+||\mathbf{y}||$$

for all $\mathbf{x},\mathbf{y}\in V$.



Proof. We prove the result for the case when V is a complex vector space. The real case is similar but slightly easier (because we do not have to deal with complex conjugates). We first remark that for all complex numbers $z = a+ib$ (where $a, b \in \mathbb{R}$), we have that

- $z + \overline{z} = 2a = 2\mathrm{Re}(z)$;
- $\mathrm{Re}(z) = a \leq |a| \leq \sqrt{a^2 + b^2} = |z|$.

Now, fix $\mathbf{x}, \mathbf{y} \in V$. Then we have the following:

$$\begin{aligned}
||\mathbf{x}+\mathbf{y}||^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle \\
&\overset{(*)}{=} \underbrace{\langle \mathbf{x},\mathbf{x}\rangle}_{=||\mathbf{x}||^2} + \langle \mathbf{x},\mathbf{y}\rangle + \langle \mathbf{y},\mathbf{x}\rangle + \underbrace{\langle \mathbf{y},\mathbf{y}\rangle}_{=||\mathbf{y}||^2} \\
&= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + \langle \mathbf{x},\mathbf{y}\rangle + \langle \mathbf{y},\mathbf{x}\rangle \\
&\overset{(**)}{=} ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + \langle \mathbf{x},\mathbf{y}\rangle + \overline{\langle \mathbf{x},\mathbf{y}\rangle} \\
&= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2\mathrm{Re}\big(\langle \mathbf{x},\mathbf{y}\rangle\big) \\
&\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2|\langle \mathbf{x},\mathbf{y}\rangle| \\
&\overset{(***)}{\leq} ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2||\mathbf{x}||\,||\mathbf{y}|| \\
&= (||\mathbf{x}|| + ||\mathbf{y}||)^2,
\end{aligned}$$

where (*) follows from c.2 and c.2', (**) follows from c.4, and (***) follows from the Cauchy-Schwarz inequality. By taking the square root of both sides, we obtain

$$||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||,$$

which is what we needed to show. □

6.2.3 The norm in general

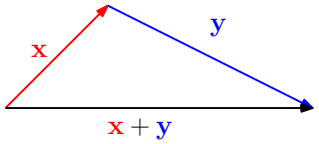
A *norm* in a real or complex vector space V is a function $||\cdot|| : V \to \mathbb{R}$ that satisfies the following three axioms:

n.1. for all vectors $\mathbf{x} \in V$, we have that $||\mathbf{x}|| \geq 0$, and equality holds if and only if $\mathbf{x} = \mathbf{0}$;

n.2. for all vectors $\mathbf{x} \in V$ and scalars α,[12] we have that $||\alpha\mathbf{x}|| = |\alpha| \, ||\mathbf{x}||$;

n.3. for all vectors $\mathbf{x}, \mathbf{y} \in V$, we have that $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

As stated at the beginning of section 6.2, a norm in a real or complex vector space V gives a way of measuring the distance of a vector from the origin, or equivalently, measuring the length of a vector. The norm of a vector is always a non-negative real number (regardless of whether the vector space is real or complex). We note that n.3 is referred to as the "triangle inequality." The idea is that vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x}+\mathbf{y}$ form a triangle (see the picture below), and the length of the third side cannot be greater than the sum of lengths of the other two sides.



It follows from the results of subsection 6.2.1 that any norm induced by a scalar product in a real or complex vector space V really is a norm, i.e. it is a function from V to $\mathbb{R}$ that satisfies axioms n.1, n.2, and n.3 above. The fact that axiom n.1 is satisfied is immediate from the construction of a norm induced by a scalar product, the fact that n.2 is satisfied follows from Proposition 6.2.1, and the fact that n.3 is satisfied follows from the triangle inequality proven in subsection 6.2.2.

Unit vectors. Suppose that V is a real or complex vector space, equipped with a norm $||\cdot||$. A vector $\mathbf{v} \in V$ is called a *unit vector* if $||\mathbf{v}|| = 1$. (In view of n.1, any unit vector is, in particular, a non-zero vector.) For notational convenience, given a vector $\mathbf{v}$ and a scalar $\alpha \neq 0$, we often write $\frac{\mathbf{v}}{\alpha}$ instead of $\alpha^{-1}\mathbf{v}$ or $\frac{1}{\alpha}\mathbf{v}$. In particular, for a non-zero vector $\mathbf{v} \in V$, we may write $\frac{\mathbf{v}}{||\mathbf{v}||}$ (as in Proposition 6.2.3 below).

Proposition 6.2.3. *Let V be a real or complex vector space, equipped with a norm $||\cdot||$. Then for all non-zero vectors $\mathbf{v} \in V$, we have that $||\mathbf{v}|| > 0$ and that $||\frac{\mathbf{v}}{||\mathbf{v}||}|| = 1$, and in particular, $\frac{\mathbf{v}}{||\mathbf{v}||}$ is a unit vector.*

Proof. Fix a non-zero vector $\mathbf{v} \in V$. By n.1, we have that $||\mathbf{v}|| > 0$. We further have that

$$||\tfrac{\mathbf{v}}{||\mathbf{v}||}|| \overset{\text{n.2}}{=} |\tfrac{1}{||\mathbf{v}||}| \, ||\mathbf{v}|| \overset{(*)}{=} \tfrac{1}{||\mathbf{v}||} \, ||\mathbf{v}|| = 1,$$

where (*) follows from the fact that $||\mathbf{v}|| > 0$. This completes the argument. □

[12] So, α is a real or complex number, depending on whether the vector space V is real or complex.

Terminology/Remark: Suppose that V is a real or complex vector space, equipped with a norm $||\cdot||$. To *normalize* a non-zero vector $\mathbf{v}$ in V means to multiply that vector by $\frac{1}{||\mathbf{v}||}$ ("divide by its length"). By Proposition 6.2.3, when we normalize a non-zero vector, we produce a unit vector.

6.2.4 Other examples of norms

For a positive integer p, we define the *p-norm* in $\mathbb{R}^n$, denoted by $||\cdot||_p$, by setting

$$||\mathbf{x}||_p \quad := \quad \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{\frac{1}{p}}$$

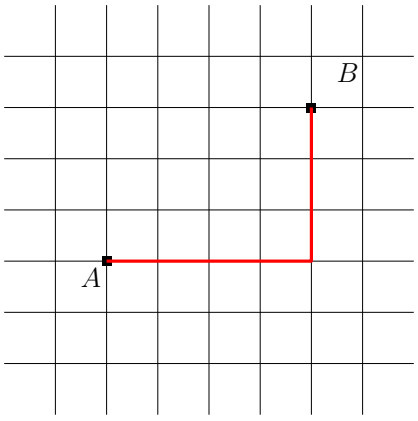
for all $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$. We omit the proof of the fact that this really is a norm in $\mathbb{R}^n$. We do note, however, that for $p = 2$, we get

$$||\mathbf{x}||_2 \quad = \quad \sqrt{\sum_{i=1}^{n} x_i^2},$$

which is precisely the norm induced by the standard scalar product in $\mathbb{R}^n$, i.e. the standard Euclidean norm in $\mathbb{R}^n$. For $p = 1$, we get

$$||\mathbf{x}||_1 \quad = \quad \sum_{i=1}^{n} |x_i|.$$

We note that the $||\cdot||_1$ norm is sometimes called the "Manhattan norm." This is because streets and avenues in Manhattan form a perfect grid (more or less), and so $||\cdot||_1$ gives the actual walking distance between two places in Manhattan (see the picture below).



Another norm of interest is the so called "Chebyshev distance" in $\mathbb{R}^n$, denoted by $||\cdot||_\infty$. It is defined by

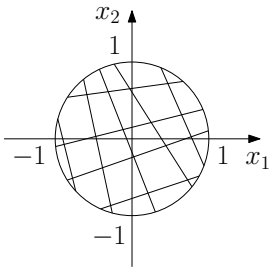
$$||\mathbf{x}||_\infty \quad := \quad \max\big\{|x_1|, \dots, |x_n|\big\}$$

for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$.[13]

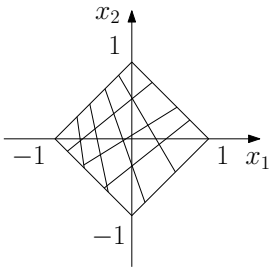
The *unit disk* in a real or complex vector space V, equipped with a norm $||\cdot||$, is the set

$$\{\mathbf{x} \in V \mid ||\mathbf{x}|| \leq 1\}.$$

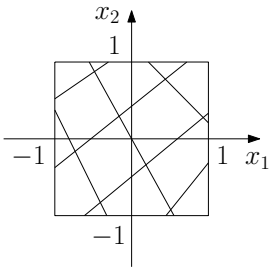
The unit disks in $\mathbb{R}^2$ with respect to the norms $||\cdot||_2$, $||\cdot||_1$, and $||\cdot||_\infty$ are represented in the picture below.



unit disk in $\mathbb{R}^2$
w.r.t. $||\cdot||_2$



unit disk in $\mathbb{R}^2$
w.r.t. $||\cdot||_1$



unit disk in $\mathbb{R}^2$
w.r.t. $||\cdot||_\infty$

Finally, if you have studied calculus, recall that for $a, b \in \mathbb{R}$ such that $a < b$, $\mathcal{C}_{[a,b]}$ is the (real) vector space of all continuous functions from $[a, b]$ to $\mathbb{R}$. For a real number $p \geq 1$, we have the norm $||\cdot||_p$ on $\mathcal{C}_{[a,b]}$ given by

$$||f||_p \quad = \quad \left(\int_a^b |f(x)|^p \right)^{\frac{1}{p}}$$

for all $f \in \mathcal{C}_{[a,b]}$, and we also have the norm $||\cdot||_\infty$ on $\mathcal{C}_{[a,b]}$ given by

$$||f||_\infty \quad = \quad \max_{x \in [a,b]} |f(x)|$$

for all $f \in \mathcal{C}_{[a,b]}$. Once again, we omit the proof of the fact that $||\cdot||_p$ (for a real number $p \geq 1$) and $||\cdot||_\infty$ really are norms in $\mathcal{C}_{[a,b]}$.

6.3 Orthogonal and orthonormal bases. Gram-Schmidt orthogonalization

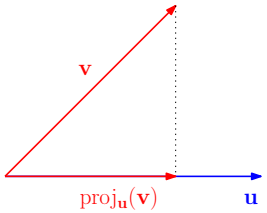
6.3.1 Vector projection

Suppose we are given a real or complex vector space V, equipped with a scalar product $\langle \cdot, \cdot \rangle$. For a **non-zero** vector $\mathbf{u} \in V$ and any vector $\mathbf{v} \in V$, the *orthogonal projection* of $\mathbf{v}$ onto $\mathbf{u}$ is the vector

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) \quad := \quad \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}.$$

[13] It is not hard to check that $||\cdot||_\infty$ really is a norm in $\mathbb{R}^n$, i.e. that it satisfies axioms n.1, n.2, and n.3. The details are left as an exercise.

Since $\mathbf{u} \neq \mathbf{0}$, r.1 or c.1 guarantees that $\langle \mathbf{u}, \mathbf{u} \rangle > 0$, and so the expression above is well defined (that is, we are not dividing by zero). As we can see, $\text{proj}_{\mathbf{u}}(\mathbf{v})$ is a scalar multiple of $\mathbf{u}$.

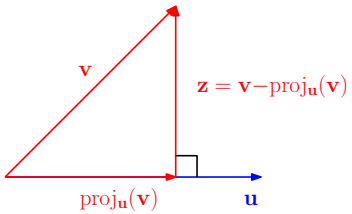


As the picture above suggests, for any scalar $\alpha \neq 0$, the projection of $\mathbf{v}$ onto $\alpha\mathbf{u}$ is the same as the projection of $\mathbf{v}$ onto $\mathbf{u}$. Indeed, if V is a complex vector space, then we have that

$$\begin{aligned}
\text{proj}_{\alpha\mathbf{u}}(\mathbf{v}) &= \tfrac{\langle \mathbf{v}, \alpha\mathbf{u} \rangle}{\langle \alpha\mathbf{u}, \alpha\mathbf{u} \rangle}(\alpha\mathbf{u}) && \text{by definition} \\
&= \tfrac{\langle \mathbf{v}, \alpha\mathbf{u} \rangle}{\alpha\langle \mathbf{u}, \alpha\mathbf{u} \rangle}(\alpha\mathbf{u}) && \text{by c.3} \\
&= \tfrac{\overline{\alpha}\langle \mathbf{v}, \mathbf{u} \rangle}{\alpha\overline{\alpha}\langle \mathbf{u}, \mathbf{u} \rangle}(\alpha\mathbf{u}) && \text{by c.3'} \\
&= \tfrac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u} \\
&= \text{proj}_{\mathbf{u}}(\mathbf{v}) && \text{by definition.}
\end{aligned}$$

If V is a real vector space, then the proof is similar to the above, except that we use r.3 and r.3' instead of c.3 and c.3', respectively, and instead of $\overline{\alpha}$, we simply have α. Moreover, we have the following proposition.

Proposition 6.3.1. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. Let $\mathbf{u}$ be a non-zero vector in V, let $\mathbf{v}$ be any vector in V, and set $\mathbf{z} := \mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$. Then $\mathbf{z} \perp \mathbf{u}$.*



Proof. We compute

$$\begin{aligned}
\langle \mathbf{z}, \mathbf{u} \rangle &= \big\langle \mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v}), \mathbf{u} \big\rangle \\
&= \big\langle \mathbf{v} - \tfrac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u}, \mathbf{u} \big\rangle
\end{aligned}$$

$$\begin{aligned} &\overset{(*)}{=} \langle \mathbf{v}, \mathbf{u} \rangle - \tfrac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \langle \mathbf{u}, \mathbf{u} \rangle \\ &= \langle \mathbf{v}, \mathbf{u} \rangle - \langle \mathbf{v}, \mathbf{u} \rangle \\ &= 0, \end{aligned}$$

where (*) follows from r.2 and r.3 (in the real case) or from c.2 and c.3 (in the complex case). This proves that $\mathbf{z} \perp \mathbf{u}$, which is what we needed to show. □

6.3.2 Orthogonal and orthonormal sets. Orthogonal and orthonormal bases

Suppose we are given a real or complex vector space V, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. An *orthogonal set of vectors* in V is a set of pairwise orthogonal vectors in V. An *orthonormal set of vectors* is an orthogonal set of unit vectors (i.e. vectors of length 1). An *orthogonal basis* (resp. *orthonormal basis*) of V is an orthogonal (resp. orthonormal) set in V that is also a basis of V.

Proposition 6.3.2. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Then both the following hold:*

*(a) any **orthogonal** set of **non-zero** vectors in V is linearly independent;*

*(b) any **orthonormal** set of vectors in V is linearly independent.*

Proof. Any orthonormal set of vectors is an orthogonal set of non-zero vectors (because $\mathbf{0}$ is not a unit vector). So, (a) immediately implies (b).

It remains to prove (a). Fix an orthogonal set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of non-zero vectors in V. We must show that this set is linearly independent. Fix scalars $\alpha_1, \ldots, \alpha_k$ such that

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k = \mathbf{0}.$$

We must show that $\alpha_1 = \cdots = \alpha_k = 0$. Fix any $i \in \{1, \ldots, k\}$. Then

$$\langle \underbrace{\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k}_{=\mathbf{0}}, \mathbf{u}_i \rangle = \langle \mathbf{0}, \mathbf{u}_i \rangle \overset{(*)}{=} 0,$$

where (*) follows from Proposition 6.1.4(c). On the other hand, note that

$$\langle \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k, \mathbf{u}_i \rangle \overset{(*)}{=} \alpha_1 \langle \mathbf{u}_1, \mathbf{u}_i \rangle + \cdots + \alpha_k \langle \mathbf{u}_k, \mathbf{u}_i \rangle \overset{(**)}{=} \alpha_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle,$$

where (*) follows from r.2 and r.3 (in the real case) or from c.2 and c.3 (in the complex case), and (**) follows from the fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal

set.[14] So,

$$\alpha_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle \quad = \quad 0.$$

Since $\mathbf{u}_i \neq \mathbf{0}$, r.1 or c.1 guarantees that $\langle \mathbf{u}_i, \mathbf{u}_i \rangle \neq 0$; consequently, $\alpha_i = 0$. Since $i \in \{1, \ldots, k\}$ was chosen arbitrarily, it follows that $\alpha_1 = \cdots = \alpha_k = 0$, and we are done. □

Proposition 6.3.3. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an **orthogonal** set of vectors in V. Then all the following hold:*

*(a) for all scalars $\alpha_1, \ldots, \alpha_k$, we have that $\{\alpha_1 \mathbf{u}_1, \ldots, \alpha_k \mathbf{u}_k\}$ is an **orthogonal** set of vectors;*

*(b) if vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are all non-zero, then $\left\{ \frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right\}$ is an **orthonormal** set of vectors, and consequently, an orthonormal basis of Span$(\mathbf{u}_1, \ldots, \mathbf{u}_k)$;*

(c) if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of V, then $\left\{ \frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right\}$ is an orthonormal basis of V.

Proof. Part (a) follows immediately from Proposition 6.1.4(b), and part (c) is a special case of part (b).[15] It remains to prove (b). Assume that vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are all non-zero. By (a), $\left\{ \frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right\}$ is an orthogonal set. On the other hand, by Proposition 6.2.3, vectors $\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||}$ are all unit vectors. So, by definition, $\left\{ \frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right\}$ is an orthonormal set. In particular, by Proposition 6.3.2(b), the set $\left\{ \frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right\}$ is linearly independent. Moreover, by Proposition 3.1.12, we have that Span$(\mathbf{u}_1, \ldots, \mathbf{u}_k) =$ Span$\left(\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right)$. So, $\left\{ \frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right\}$ is an orthonormal basis of Span$(\mathbf{u}_1, \ldots, \mathbf{u}_k)$. This proves (b). □

Proposition 6.3.4. *Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Set $n := dim(V)$. Then both the following hold:*

(a) any orthogonal set of n non-zero vectors in V is an orthogonal basis of V;

(b) any orthonormal set of n vectors in V is an orthonormal basis of V.

Proof. By Proposition 6.3.2(a), any orthogonal set of non-zero vectors is linearly independent, and by Corollary 3.2.20(a), any linearly independent set of size n in an n-dimensional vector space is a basis of that vector space. This proves (a). Part (b) follows from (a), since any orthonormal set of vectors is, in particular, an orthogonal set of non-zero vectors (because $\mathbf{0}$ is not a unit vector). □

[14] In particular, we have that $\langle \mathbf{u}_j, \mathbf{u}_i \rangle = 0$ for all $j \in \{1, \ldots, k\} \setminus \{i\}$.

[15] Indeed, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of V, then vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are all non-zero and $V =$ Span$(\mathbf{u}_1, \ldots, \mathbf{u}_k)$, and so by (b), $\left\{ \frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} \right\}$ is an orthonormal basis of V.

6.3.3 Coordinate vectors with respect to orthogonal and orthonormal bases. Fourier coefficients

If we have an orthogonal basis of a real or complex vector space (equipped with a scalar product and the norm induced by it), then every vector in that vector space can be expressed as a linear combination of those basis vectors in a particularly nice way, that is, we have a convenient formula for the coefficients in front of the basis vectors (see Theorem 6.3.5). If our basis is orthonormal, then we get an even nicer formula for the coefficients (see Corollary 6.3.6).

Theorem 6.3.5. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an **orthogonal** basis of V. Then for all $\mathbf{v} \in V$, we have that*

$$\mathbf{v} \;=\; \sum_{i=1}^{n} proj_{\mathbf{u}_i}(\mathbf{v}) \;=\; \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i,$$

and consequently,

$$\left[\ \mathbf{v}\ \right]_{\mathcal{B}} \;=\; \left[\ \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \quad \ldots \quad \frac{\langle \mathbf{v}, \mathbf{u}_n \rangle}{\langle \mathbf{u}_n, \mathbf{u}_n \rangle}\ \right]^T.$$

Proof. The second statement follows from the first and from the definition of a coordinate vector. It remains to prove the first statement. Fix a vector $\mathbf{v} \in V$. By definition, for all $i \in \{1, \ldots, n\}$, we have that $\mathrm{proj}_{\mathbf{u}_i}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$. So, it suffices to show that

$$\mathbf{v} \;=\; \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Since $\mathbf{v} \in V$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of V, there exist scalars $\alpha_1, \ldots, \alpha_n$ such that

$$\mathbf{v} \;=\; \sum_{i=1}^{n} \alpha_i \mathbf{u}_i.$$

Now, fix any index $j \in \{1, \ldots, n\}$. We then have that

$$\langle \mathbf{v}, \mathbf{u}_j \rangle \;=\; \left\langle \sum_{i=1}^{n} \alpha_i \mathbf{u}_i, \mathbf{u}_j \right\rangle \;\overset{(*)}{=}\; \sum_{i=1}^{n} \alpha_i \langle \mathbf{u}_i, \mathbf{u}_j \rangle \;\overset{(**)}{=}\; \alpha_j \langle \mathbf{u}_j, \mathbf{u}_j \rangle,$$

where (*) follows from r.2 and r.3 (in the real case) or from c.2 and c.3 (in the complex case), and (**) follows from the fact that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are pairwise orthogonal. Since $\mathbf{u}_j \neq \mathbf{0}$ (because $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of V), r.1 or c.1 guarantees that $\langle \mathbf{u}_j, \mathbf{u}_j \rangle \neq 0$, and we deduce that

$$\alpha_j \;=\; \frac{\langle \mathbf{v}, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle}.$$

Since $j \in \{1, \ldots, n\}$ was chosen arbitrarily, we now deduce that

$$\mathbf{v} \;=\; \sum_{i=1}^{n} \alpha_i \mathbf{u}_i \;=\; \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i,$$

and we are done. □

Corollary 6.3.6. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an* ***orthonormal*** *basis of V. Then for all $\mathbf{v} \in V$, we have that*

$$\mathbf{v} \quad = \quad \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{u}_i \rangle \ \mathbf{u}_i,$$

and consequently,

$$\left[\ \mathbf{v} \ \right]_{\mathcal{B}} \quad = \quad \left[\ \langle \mathbf{v}, \mathbf{u}_1 \rangle \quad \ldots \quad \langle \mathbf{v}, \mathbf{u}_n \rangle \ \right]^T.$$

Terminology: Coefficients $\langle \mathbf{v}, \mathbf{u}_i \rangle$ from Corollary 6.3.6 are called *Fourier coefficients*.

Proof. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of V, we know that $||\mathbf{u}_1|| = \cdots = ||\mathbf{u}_n|| = 1$, and consequently (by the construction of $|| \cdot ||$), we have that $\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \cdots = \langle \mathbf{u}_n, \mathbf{u}_n \rangle = 1$. The result now follows immediately from Theorem 6.3.6. □

6.3.4 Gram-Schmidt orthogonalization

In this subsection, we describe the "Gram-Schmidt orthogonalization process," which gives a recipe for transforming an arbitrary basis of a real or complex vector space (equipped with a scalar product and the norm induced by it) into an orthogonal (and even orthonormal) basis. But first, we need a technical proposition.

Proposition 6.3.7. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal set of non-zero vectors in V. Let $\mathbf{v} \in V$, and set $\mathbf{y} := \sum_{i=1}^{k} proj_{\mathbf{u}_i}(\mathbf{v}) = \sum_{i=1}^{k} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$ and $\mathbf{z} := \mathbf{v} - \mathbf{y}$. Then all the following hold:*

(a) $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{z}\}$ is an orthogonal set of vectors;

(b) $\mathbf{z} = \mathbf{0}$ if and only if $\mathbf{v} \in Span(\mathbf{u}_1, \ldots, \mathbf{u}_k)$;

(c) $Span(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}) = Span(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{z})$.

Proof. First of all, Proposition 6.3.2 guarantees that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a linearly independent set, and we deduce that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$.

Let us first prove (a). By hypothesis, vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are pairwise orthogonal. On the other hand, for each $j \in \{1, \ldots, k\}$, we have the following:

$$\begin{aligned} \langle \mathbf{z}, \mathbf{u}_j \rangle \quad &= \quad \langle \mathbf{v} - \sum_{i=1}^{k} \tfrac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i, \mathbf{u}_j \rangle \\ &\overset{(*)}{=} \quad \langle \mathbf{v}, \mathbf{u}_j \rangle - \sum_{i=1}^{k} \tfrac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \langle \mathbf{u}_i, \mathbf{u}_j \rangle \end{aligned}$$

$$\begin{aligned} &\overset{(**)}{=} \quad \langle \mathbf{v}, \mathbf{u}_j \rangle - \tfrac{\langle \mathbf{v}, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \langle \mathbf{u}_j, \mathbf{u}_j \rangle \\ &= \quad \langle \mathbf{v}, \mathbf{u}_j \rangle - \langle \mathbf{v}, \mathbf{u}_j \rangle \\ &= \quad 0, \end{aligned}$$

where (*) follows from r.2 and r.3 (in the real case) or from c.2 and c.3 (in the complex case), and (**) follows from the fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set. Thus, $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{z}\}$ is an orthogonal set of vectors. This proves (a).

Next, we prove (b). Clearly, $\mathbf{z} = \mathbf{0}$ if and only if $\mathbf{v} = \sum_{i=1}^{k} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$. If $\mathbf{v} = \sum_{i=1}^{k} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$, then $\mathbf{v}$ is a linear combination of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$, and consequently, $\mathbf{v} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$. On the other hand, if $\mathbf{v} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$, then Theorem 6.3.5 guarantees $\mathbf{v} = \sum_{i=1}^{k} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$.[16] This proves (b).

Finally, we prove (c). Fix any vector $\mathbf{x} \in V$. We must show that $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v})$ if and only if $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{z})$. We prove both directions (as we shall see, they are very similar).

Suppose first that $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v})$. Then there exist scalars $\alpha_1, \ldots, \alpha_k, \beta$ such that $\mathbf{x} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \beta \mathbf{v}$. But now

$$\begin{aligned} \mathbf{x} &= \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \beta \mathbf{v} \\ &= \Big(\sum_{i=1}^{k} \alpha_i \mathbf{u}_i \Big) + \beta(\mathbf{y} + \mathbf{z}) \\ &= \Big(\sum_{i=1}^{k} \alpha_i \mathbf{u}_i \Big) + \beta\Big(\big(\sum_{i=1}^{k} \tfrac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \big) + \mathbf{z} \Big) \\ &= \Big(\sum_{i=1}^{k} \big(\alpha_i + \beta \tfrac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \big) \mathbf{u}_i \Big) + \beta \mathbf{z}, \end{aligned}$$

and we deduce that $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{z})$.

Suppose, conversely, that $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{z})$. Then there exist scalars $\alpha_1, \ldots, \alpha_k, \beta$ such that $\mathbf{x} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \beta \mathbf{z}$. But now

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \beta \mathbf{z}$$

[16] This is because $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$.

$$= \Big(\sum_{i=1}^{k} \alpha_i \mathbf{u}_i\Big) + \beta(\mathbf{v}-\mathbf{y})$$

$$= \Big(\sum_{i=1}^{k} \alpha_i \mathbf{u}_i\Big) + \beta\Big(\mathbf{v} - \big(\sum_{i=1}^{k} \tfrac{\langle \mathbf{v},\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i\big)\Big)$$

$$= \Big(\sum_{i=1}^{k} \big(\alpha_i - \beta\tfrac{\langle \mathbf{v},\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\big)\mathbf{u}_i\Big) + \beta\mathbf{v},$$

and we deduce that $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_k,\mathbf{v})$. This proves (c). □

The Gram-Schmidt orthogonalization process (version 1). *Let V be a real or complex vector space, equipped with a scalar product $\langle\cdot,\cdot\rangle$ and the induced norm $||\cdot||$, and let $\mathbf{v}_1,\ldots,\mathbf{v}_k$ be linearly independent vectors in V. For all $\ell \in \{1,\ldots,k\}$, set*

$$\mathbf{u}_\ell \;:=\; \mathbf{v}_\ell - \sum_{i=1}^{\ell-1} proj_{\mathbf{u}_i}(\mathbf{v}_\ell) \;=\; \mathbf{v}_\ell - \sum_{i=1}^{\ell-1} \tfrac{\langle \mathbf{v}_\ell,\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i.$$

Then $\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}$ is an orthogonal basis of Span$(\mathbf{v}_1,\ldots,\mathbf{v}_k)$, and $\Big\{\frac{\mathbf{u}_1}{||\mathbf{u}_1||},\ldots,\frac{\mathbf{u}_k}{||\mathbf{u}_k||}\Big\}$ is an orthonormal basis of Span$(\mathbf{v}_1,\ldots,\mathbf{v}_k)$.

Remark: Before proceeding with the proof, it may be helpful to note that the sequence $\mathbf{u}_1,\ldots,\mathbf{u}_k$ is obtained (recursively) as follows:

- $\mathbf{u}_1 := \mathbf{v}_1$;
- $\mathbf{u}_2 := \mathbf{v}_2 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_2)$;
- $\mathbf{u}_3 := \mathbf{v}_3 - \Big(\mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_3) + \mathrm{proj}_{\mathbf{u}_2}(\mathbf{v}_3)\Big)$;

  $\vdots$

- $\mathbf{u}_k := \mathbf{v}_k - \Big(\mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_k) + \mathrm{proj}_{\mathbf{u}_2}(\mathbf{v}_k) + \cdots + \mathrm{proj}_{\mathbf{u}_{k-1}}(\mathbf{v}_k)\Big)$.

This describes precisely the sequence $\mathbf{u}_1,\ldots,\mathbf{u}_k$ from the statement of the Gram-Schmidt orthogonalization process (version 1), only in a less compact (but perhaps more readable) form. We also note that it may be helpful to read Example 6.3.8 before reading the proof below, since Example 6.3.8 illustrates the Gram-Schmidt orthogonalization process on a concrete numerical example.

Proof. We first prove that $\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}$ is an orthogonal basis of Span$(\mathbf{v}_1,\ldots,\mathbf{v}_k)$. For each $\ell \in \{1,\ldots,k\}$, we set $U_\ell := \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_\ell)$, and we prove (inductively) that $\{\mathbf{u}_1,\ldots,\mathbf{u}_\ell\}$ is an orthogonal basis of U_ℓ. Obviously, this is enough, because for $k=\ell$, we get that $\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}$ is an orthogonal basis of $U_k = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k)$.

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent, we see that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are all non-zero, and in particular, $\{\mathbf{v}_1\}$ is linearly independent. Since $U_1 = \text{Span}(\mathbf{v}_1)$ and $\mathbf{u}_1 = \mathbf{v}_1$, we deduce that $\{\mathbf{u}_1\}$ is a basis of U_1, and this basis is obviously orthogonal (since it contains only one vector).

Now, fix $\ell \in \{1, \ldots, k-1\}$, and assume inductively that $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ is an orthogonal basis of U_ℓ. We must show that $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{u}_{\ell+1}\}$ is an orthogonal basis of $U_{\ell+1}$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ are two bases of U_ℓ,[17] it is clear that $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{v}_{\ell+1}) = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_\ell, \mathbf{v}_{\ell+1}) = U_{\ell+1}$.[18] On the other hand, by the construction of $\mathbf{u}_{\ell+1}$ and by Proposition 6.3.7(c), we have that $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{v}_{\ell+1}) = \text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{u}_{\ell+1})$. So, $\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{u}_{\ell+1}) = U_{\ell+1}$. Since $\dim(U_{\ell+1}) = \ell + 1$,[19] the fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{u}_{\ell+1}\}$ spans $U_{\ell+1}$ implies that $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{u}_{\ell+1}\}$ is in fact a basis of $U_{\ell+1}$ (this follows from Corollary 3.2.20). By the induction hypothesis, vectors $\mathbf{u}_1, \ldots, \mathbf{u}_\ell$ are pairwise orthogonal non-zero vectors,[20] and so by the construction of $\mathbf{u}_{\ell+1}$ and by Proposition 6.3.7(a), we have that $\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{u}_{\ell+1}$ are pairwise orthogonal. So, $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{u}_{\ell+1}\}$ is an orthogonal basis of $U_{\ell+1}$. This completes the induction.

We have now shown that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. By Proposition 6.3.3(b), this implies that $\left\{\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||}\right\}$ is an orthonormal basis of $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. This completes the argument. □

Example 6.3.8. *Consider the following linearly independent vectors in $\mathbb{R}^4$:*

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -5 \\ 10 \\ 2 \\ 11 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix}.$$

Set $U := \text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Using the Gram-Schmidt orthogonalization process (version 1):

(a) compute an orthogonal basis of U (with respect to the standard scalar product $\cdot$ in $\mathbb{R}^4$).

(b) compute an orthonormal basis of U (with respect to the standard scalar product $\cdot$ in $\mathbb{R}^4$ and the norm $||\cdot||$ induced by it).

[17] The fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ is a basis of U_ℓ follows from the induction hypothesis. The fact that $\{\mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ is a basis of U_ℓ follows from the fact that $\{\mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ is linearly independent (because $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent) and $U_\ell = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_\ell)$ (by construction).

[18] Details?

[19] This is because $\{\mathbf{v}_1, \ldots, \mathbf{v}_{\ell+1}\}$ is a basis of $U_{\ell+1}$.

[20] The fact that $\mathbf{u}_1, \ldots, \mathbf{u}_\ell$ are all non-zero follows from the fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ is a basis of U_ℓ (by the induction hypothesis).

Remark: To see that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ really are linearly independent, we compute

$$\mathrm{RREF}\Big(\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \Big) \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and we deduce that $\mathrm{rank}\Big(\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \Big) = 3$, i.e. $\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}$ has full column rank. So, by Theorem 3.3.12(a), vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent.

Solution. (a) First, we set

$$\mathbf{u}_1 \;:=\; \mathbf{v}_1 \;=\; \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}.$$

Next, we compute:

$$\mathbf{u}_2 \;:=\; \mathbf{v}_2 - \tfrac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 \;=\; \begin{bmatrix} -5 \\ 10 \\ 2 \\ 11 \end{bmatrix} - \frac{\begin{bmatrix} -5 \\ 10 \\ 2 \\ 11 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}}{\begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} -5 \\ 10 \\ 2 \\ 11 \end{bmatrix} - \tfrac{(-5)\cdot 3 + 10\cdot 4 + 2\cdot(-4) + 11\cdot 3}{3\cdot 3 + 4\cdot 4 + (-4)\cdot(-4) + 3\cdot 3} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix}.$$

Finally, we compute:

$$\mathbf{u}_3 \;:=\; \mathbf{v}_3 - \Big(\tfrac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \tfrac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 \Big)$$

$$= \begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix} - \left(\frac{\begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}}{\begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix} + \frac{\begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix}}{\begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix} \cdot \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix}} \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix} - \left(\frac{8 \cdot 3 + 19 \cdot 4 + 11 \cdot (-4) + (-2) \cdot 3}{3 \cdot 3 + 4 \cdot 4 + (-4) \cdot (-4) + 3 \cdot 3} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix} + \frac{8 \cdot (-8) + 19 \cdot 6 + 11 \cdot 6 + (-2) \cdot 8}{(-8) \cdot (-8) + 6 \cdot 6 + 6 \cdot 6 + 8 \cdot 8} \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 9 \\ 12 \\ 12 \\ -9 \end{bmatrix}.$$

So,

$$\mathcal{B} := \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \left\{ \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}, \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix}, \begin{bmatrix} 9 \\ 12 \\ 12 \\ -9 \end{bmatrix} \right\}$$

is an orthogonal basis of U.

(b) To obtain an orthonormal basis of U, we normalize the vectors of the orthogonal basis $\mathcal{B}$ of U that we obtained in part (a). First, we compute:

$$||\mathbf{u}_1|| = \sqrt{\mathbf{u}_1 \cdot \mathbf{u}_1} = \sqrt{\begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}}$$

$$= \sqrt{3 \cdot 3 + 4 \cdot 4 + (-4) \cdot (-4) + 3 \cdot 3}$$

$$= 5\sqrt{2}$$

$$||\mathbf{u}_2|| = \sqrt{\mathbf{u}_2 \cdot \mathbf{u}_2} = \sqrt{\begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix} \cdot \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix}}$$

$$= \sqrt{(-8)\cdot(-8)+6\cdot 6+6\cdot 6+8\cdot 8}$$
$$= 10\sqrt{2}$$

$$||\mathbf{u}_3|| = \sqrt{\mathbf{u}_3\cdot\mathbf{u}_3} = \sqrt{\begin{bmatrix} 9\\12\\12\\-9\end{bmatrix}\cdot\begin{bmatrix} 9\\12\\12\\-9\end{bmatrix}}$$
$$= \sqrt{9\cdot 9+12\cdot 12+12\cdot 12+(-9)\cdot(-9)}$$
$$= 15\sqrt{2}.$$

We now see that

$$\mathcal{C} := \left\{\tfrac{\mathbf{u}_1}{||\mathbf{u}_1||},\tfrac{\mathbf{u}_2}{||\mathbf{u}_2||},\tfrac{\mathbf{u}_3}{||\mathbf{u}_3||}\right\} = \left\{\tfrac{1}{5\sqrt{2}}\begin{bmatrix} 3\\4\\-4\\3\end{bmatrix},\tfrac{1}{5\sqrt{2}}\begin{bmatrix} -4\\3\\3\\4\end{bmatrix},\tfrac{1}{5\sqrt{2}}\begin{bmatrix} 3\\4\\4\\-3\end{bmatrix}\right\}$$

is an orthonormal basis of U. □

Remark: Suppose that V is a real or complex vector space, equipped with a scalar product $\langle\cdot,\cdot\rangle$ and the induced norm $||\cdot||$. Suppose, furthermore, that we are given a list $\mathbf{v}_1,\ldots,\mathbf{v}_k$ of vectors in V, which may possibly be linearly dependent. How would we find an orthogonal (or orthonormal) basis of $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k)$? In this case, we would first need to find a basis of $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k)$, and then apply the Gram-Schmidt process to that basis. If $V=\mathbb{R}^n$, then Theorem 3.3.4 guarantees that such a basis is formed by the pivot columns of the matrix $\begin{bmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_k\end{bmatrix}$. A numerical example is given below.

Example 6.3.9. *Consider the following vectors in* $\mathbb{R}^3$*:*

$$\mathbf{v}_1=\begin{bmatrix}1\\2\\1\end{bmatrix},\quad \mathbf{v}_2=\begin{bmatrix}-2\\-4\\-2\end{bmatrix},\quad \mathbf{v}_3=\begin{bmatrix}3\\5\\-1\end{bmatrix},\quad \mathbf{v}_4=\begin{bmatrix}5\\9\\1\end{bmatrix}.$$

Set $U := \mathit{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4)$*. Compute an orthogonal basis of* U *(with respect to the standard scalar product* $\cdot$ *in* $\mathbb{R}^3$*).*

Solution. First, we form the matrix

$$A := \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4\end{bmatrix} = \begin{bmatrix}1 & -2 & 3 & 5\\2 & -4 & 5 & 9\\1 & -2 & -1 & 1\end{bmatrix},$$

and by row reducing, we obtain

$$\mathrm{RREF}(A) \;=\; \begin{bmatrix} 1 & -2 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

So, the pivot columns of A are its first and third column, and therefore, $\{\mathbf{v}_1, \mathbf{v}_3\}$ is a basis of $\mathrm{Col}(A) = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$. We now apply the Gram-Schmidt orthogonalization process to vectors $\mathbf{v}_1, \mathbf{v}_3$. First, we set:

$$\mathbf{u}_1 \;:=\; \mathbf{v}_1 \;=\; \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

Further, we compute:

$$\mathbf{u}_2 \;:=\; \mathbf{v}_2 - \tfrac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 \;=\; \begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix} - \frac{\begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix} - \tfrac{3\cdot 1 + 5 \cdot 2 + (-1)\cdot 1}{1 \cdot 1 + 2 \cdot 2 + 1 \cdot 1} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}.$$

We now see that

$$\mathcal{B} \;:=\; \{\mathbf{u}_1, \mathbf{u}_2\} \;=\; \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} \right\}$$

is an orthogonal basis of U.

Remark: We could obtain an orthonormal basis of U (with respect to the standard scalar product $\cdot$ and the induced norm $||\cdot||$) by normalizing the vectors in the orthogonal basis $\mathcal{B}$ of U. We simply compute $||\mathbf{u}_1|| = \sqrt{6}$ and $||\mathbf{u}_2|| = \sqrt{11}$, and we deduce that

$$\mathcal{C} \;:=\; \left\{ \tfrac{\mathbf{u}_1}{||\mathbf{u}_1||}, \tfrac{\mathbf{u}_2}{||\mathbf{u}_2||} \right\} \;=\; \left\{ \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{11} \\ 1/\sqrt{11} \\ -3/\sqrt{11} \end{bmatrix} \right\}$$

is an orthonormal basis of U. □

The Gram-Schmidt orthogonalization process (version 2). *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be linearly independent vectors in V. For all $\ell \in \{1, \ldots, k\}$, set*

$$\begin{aligned} \mathbf{u}_\ell &= \mathbf{v}_\ell - \sum_{i=1}^{\ell-1} proj_{\mathbf{z}_i}(\mathbf{v}_\ell) = \mathbf{v}_\ell - \sum_{i=1}^{\ell-1} \langle \mathbf{v}_\ell, \mathbf{z}_i \rangle \, \mathbf{z}_i; \\ \mathbf{z}_\ell &= \frac{\mathbf{u}_\ell}{||\mathbf{u}_\ell||}. \end{aligned}$$

Then $\{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$ is an orthonormal basis of Span$(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.

The proof of correctness of the Gram-Schmidt orthogonalization process (version 2) is similar to that of version 1, and we omit it. Let us, however, explain the main difference. The Gram-Schmidt orthogonalization process (version 2) recursively constructs two sequences of vectors, namely, $\mathbf{u}_1, \ldots, \mathbf{u}_k$ and $\mathbf{z}_1, \ldots, \mathbf{z}_k$, as follows:

- $\mathbf{u}_1 = \mathbf{v}_1$;
- $\mathbf{z}_1 = \frac{\mathbf{u}_1}{||\mathbf{u}_1||}$;
- $\mathbf{u}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{z}_1}(\mathbf{v}_2)$;
- $\mathbf{z}_2 = \frac{\mathbf{u}_2}{||\mathbf{u}_2||}$;
- $\mathbf{u}_3 = \mathbf{v}_3 - \Big(\text{proj}_{\mathbf{z}_1}(\mathbf{v}_3) + \text{proj}_{\mathbf{z}_2}(\mathbf{v}_3)\Big)$;
- $\mathbf{z}_3 = \frac{\mathbf{u}_3}{||\mathbf{u}_3||}$;

  $\vdots$
- $\mathbf{u}_k = \mathbf{v}_k - \Big(\text{proj}_{\mathbf{z}_1}(\mathbf{v}_k) + \text{proj}_{\mathbf{z}_2}(\mathbf{v}_k) + \cdots + \text{proj}_{\mathbf{z}_{k-1}}(\mathbf{v}_k)\Big)$;
- $\mathbf{z}_k = \frac{\mathbf{u}_k}{||\mathbf{u}_k||}$.

So, at each step, we obtain a vector $\mathbf{u}_\ell$ that is orthogonal to the previously constructed vectors $\mathbf{z}_1, \ldots, \mathbf{z}_{\ell-1}$, and then we normalize $\mathbf{u}_\ell$ to obtain the unit vector $\mathbf{z}_\ell$ that points in the same direction as $\mathbf{u}_\ell$. (In version 1, we skip this normalization process during our recursive construction. At the very end, we may optionally normalize all the vectors in the orthonormal basis that we obtain and thus create an orthonormal basis.) Note that, for all $\ell \in \{1, \ldots, k\}$ and $i \in \{1, \ldots, \ell - 1\}$, we have that $\text{proj}_{\mathbf{z}_i}(\mathbf{v}_\ell) = \langle \mathbf{v}_\ell, \mathbf{z}_i \rangle \, \mathbf{z}_i$. This is because $||\mathbf{z}_i|| = 1$, and so $\langle \mathbf{z}_i, \mathbf{z}_i \rangle = 1$, and therefore, $\text{proj}_{\mathbf{z}_i}(\mathbf{v}_\ell) = \frac{\langle \mathbf{v}_\ell, \mathbf{z}_i \rangle}{\langle \mathbf{z}_i, \mathbf{z}_i \rangle} \mathbf{z}_i = \langle \mathbf{v}_\ell, \mathbf{z}_i \rangle \, \mathbf{z}_i$.

We now return to Example 6.3.8, and we compute an orthonormal basis using the Gram-Schmidt process (version 2).

Example 6.3.10. *Consider the following linearly independent vectors in $\mathbb{R}^4$:*

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -5 \\ 10 \\ 2 \\ 11 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix}.$$

Set $U := Span(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Using the Gram-Schmidt orthogonalization process (version 2), compute an orthonormal basis of U (with respect to the standard scalar product $\cdot$ in $\mathbb{R}^4$ and the norm $||\cdot||$ induced by it).

Proof. We set

$$\mathbf{u}_1 := \mathbf{v}_1 = \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix},$$

we compute $||\mathbf{u}_1|| = \sqrt{\mathbf{u}_1 \cdot \mathbf{u}_1} = 5\sqrt{2}$, and we set

$$\mathbf{z}_1 := \frac{\mathbf{u}_1}{||\mathbf{u}_1||} = \frac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}.$$

Next, we set

$$\begin{aligned} \mathbf{u}_2 &:= \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{z}_1)\, \mathbf{z}_1 \\ &= \begin{bmatrix} -5 \\ 10 \\ 2 \\ 11 \end{bmatrix} - \Big(\begin{bmatrix} -5 \\ 10 \\ 2 \\ 11 \end{bmatrix} \cdot \big(\tfrac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}\big)\Big) \big(\tfrac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}\big) \\ &= \begin{bmatrix} -8 \\ 6 \\ 6 \\ 8 \end{bmatrix}. \end{aligned}$$

We compute $||\mathbf{u}_2|| = \sqrt{\mathbf{u}_2 \cdot \mathbf{u}_2} = 10\sqrt{2}$, and we set

$$\mathbf{z}_2 := \frac{\mathbf{u}_2}{||\mathbf{u}_2||} = \frac{1}{5\sqrt{2}} \begin{bmatrix} -4 \\ 3 \\ 3 \\ 4 \end{bmatrix}.$$

Next, we set

$$\begin{aligned}
\mathbf{u}_3 &:= \mathbf{v}_3 - \Big((\mathbf{v}_3 \cdot \mathbf{z}_1)\, \mathbf{z}_1 + (\mathbf{v}_3 \cdot \mathbf{z}_2)\, \mathbf{z}_2 \Big) \\
&= \begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix} - \left(\begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix} \cdot \big(\tfrac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}\big) \big(\tfrac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}\big) + \right. \\
&\qquad \left. + \left(\begin{bmatrix} 8 \\ 19 \\ 11 \\ -2 \end{bmatrix} \cdot \big(\tfrac{1}{5\sqrt{2}} \begin{bmatrix} -4 \\ 3 \\ 3 \\ 4 \end{bmatrix}\big) \big(\tfrac{1}{5\sqrt{2}} \begin{bmatrix} -4 \\ 3 \\ 3 \\ 4 \end{bmatrix}\big) \right) \\
&= \begin{bmatrix} 9 \\ 12 \\ 12 \\ -9 \end{bmatrix}.
\end{aligned}$$

We compute $||\mathbf{u}_3|| = \sqrt{\mathbf{u}_3 \cdot \mathbf{u}_3} = 15\sqrt{2}$, and we set

$$\mathbf{z}_3 := \tfrac{\mathbf{u}_3}{||\mathbf{u}_3||} = \tfrac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ 4 \\ -3 \end{bmatrix}.$$

We now have that

$$\mathcal{C} := \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\} = \left\{ \tfrac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ -4 \\ 3 \end{bmatrix}, \tfrac{1}{5\sqrt{2}} \begin{bmatrix} -4 \\ 3 \\ 3 \\ 4 \end{bmatrix}, \tfrac{1}{5\sqrt{2}} \begin{bmatrix} 3 \\ 4 \\ 4 \\ -3 \end{bmatrix} \right\}$$

is an orthonormal basis of U. □

We complete this subsection with the following important corollary of the Gram-Schmidt orthogonalization process.

Corollary 6.3.11. *Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let U be a subspace of V. Then all the following hold:*

(a) U has an orthogonal basis;

(b) any orthogonal basis of U can be extended to an orthogonal basis of V;[21]

[21] This means that for any orthogonal basis $\mathcal{B}$ of U, there exists an orthogonal basis $\mathcal{C}$ of V such that $\mathcal{B} \subseteq \mathcal{C}$.

(c) U has an orthonormal basis;

(d) any orthonormal basis of U can be extended to an orthonormal basis of V.[22]

Proof. Since V is finite-dimensional, Theorem 3.2.21 guarantees that the subspace U of V is also finite-dimensional. Consider any basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of U. Then the Gram-Schmidt orthogonalization process (version 1) applied to the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ yields a sequence of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal and $\left\{\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||}\right\}$ an orthonormal basis of $U = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. This proves (a) and (c).

For (b), consider any orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of U, and using Theorem 3.2.19, extend it to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ of V. We apply the Gram-Schmidt orthogonalization process (version 1) to the sequence $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$, and we obtain a sequence $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$ such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of V. However, since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ were pairwise orthogonal to begin with, we see from the description of the Gram-Schmidt orthogonalization process that $\mathbf{u}_1 = \mathbf{v}_1, \ldots, \mathbf{u}_k = \mathbf{v}_k$. So, the orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ of V extends the orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of U. This proves (b).

For (d), consider any orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of U. In particular, the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of U is orthogonal, and so by (b), it can be extended to an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ of V. Then by Proposition 6.3.3(c), $\left\{\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_n}{||\mathbf{u}_n||}\right\}$ is an orthonormal basis of V. But since the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of U is orthonormal, we know that $||\mathbf{u}_1|| = \cdots = ||\mathbf{u}_k|| = 1$, and it follows that $\frac{\mathbf{u}_1}{||\mathbf{u}_1||} = \mathbf{u}_1, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||} = \mathbf{u}_k$. So, our orthonormal basis $\left\{\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_n}{||\mathbf{u}_n||}\right\}$ of V in fact extends the orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of U. This proves (d). □

6.4 The orthogonal complement of a subspace

Suppose we are given a real or complex vector space V, equipped with a scalar product $\langle \cdot, \cdot \rangle$. For a set $A \subseteq V$,[23] the *orthogonal complement* of A, denoted by $A^\perp$, is the set of all vectors in V that are orthogonal to A, that is,

$$\begin{aligned} A^\perp &= \{\mathbf{v} \in V \mid \mathbf{v} \perp A\} \\ &= \{\mathbf{v} \in V \mid \mathbf{v} \perp \mathbf{a} \ \ \forall \mathbf{a} \in A\} \\ &= \{\mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{a} \rangle = 0 \ \ \forall \mathbf{a} \in A\}. \end{aligned}$$

[22] This means that for any orthonormal basis $\mathcal{B}$ of U, there exists an orthonormal basis $\mathcal{C}$ of V such that $\mathcal{B} \subseteq \mathcal{C}$.

[23] Here, A may or may not be a subspace of V.

Proposition 6.4.1. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. Let $A, B \subseteq V$. Then*

(a) $A^\perp$ is a subspace of V;[24]

(b) if $A \subseteq B$, then $A^\perp \supseteq B^\perp$.

Proof. (a) We use Theorem 3.1.7.

First, by Proposition 6.1.4(c), we have that $\mathbf{0} \perp \mathbf{v}$ for all $\mathbf{v} \in V$. In particular, $\mathbf{0} \perp \mathbf{a}$ for all $\mathbf{a} \in A$. So, $\mathbf{0} \in A^\perp$.

Next, fix vectors $\mathbf{x}_1, \mathbf{x}_2 \in A^\perp$. Then for all $\mathbf{a} \in A$, we have that $\langle \mathbf{x}_1, \mathbf{a} \rangle = \langle \mathbf{x}_2, \mathbf{a} \rangle = 0$, and consequently,

$$\langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{a} \rangle \quad = \quad \langle \mathbf{x}_1, \mathbf{a} \rangle + \langle \mathbf{x}_2, \mathbf{a} \rangle \quad = \quad 0 + 0 \quad = 0,$$

i.e. $(\mathbf{x}_1 + \mathbf{x}_2) \perp \mathbf{a}$. So, $\mathbf{x}_1 + \mathbf{x}_2 \in A^\perp$.

Finally, fix a vector $\mathbf{x} \in A^\perp$ and a scalar α. Then for all $\mathbf{a} \in A$, we have that $\langle \mathbf{x}, \mathbf{a} \rangle = 0$, and consequently,

$$\langle \alpha \mathbf{x}, \mathbf{a} \rangle \quad = \quad \alpha \langle \mathbf{x}, \mathbf{a} \rangle \quad = \quad \alpha 0 \quad = \quad 0,$$

i.e. $\alpha \mathbf{x} \perp \mathbf{a}$. So, $\alpha \mathbf{x} \in A^\perp$.

By Theorem 3.1.7, it now follows that $A^\perp$ is a subspace of V.

(b) Suppose that $A \subseteq B$. Then any vector that is orthogonal to all vectors in B is, in particular, orthogonal to all vectors in A. So, $A^\perp \supseteq B^\perp$. □

Proposition 6.4.2. *Let V be a real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}^\perp = \mathit{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)^\perp$.*

Proof. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subseteq \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$, Proposition 6.4.1(b) guarantees that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}^\perp \supseteq \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)^\perp$. Let us prove the reverse inclusion. Fix $\mathbf{x} \in \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}^\perp$. We must show that $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)^\perp$. Fix $\mathbf{u} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$. Then there exist scalars $\alpha_1, \ldots, \alpha_k$ such that $\mathbf{u} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k$. But now

$$\begin{aligned} \langle \mathbf{u}, \mathbf{x} \rangle &= \langle \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k, \mathbf{x} \rangle \\ &= \alpha_1 \langle \mathbf{u}_1, \mathbf{x} \rangle + \cdots + \alpha_k \langle \mathbf{u}_k, \mathbf{x} \rangle \\ &\overset{(*)}{=} \alpha_1 0 + \cdots + \alpha_k 0 \\ &= 0, \end{aligned}$$

[24] Note that it is possible that $A = \emptyset$. In this case, we simply get that $A^\perp = V$. This is because every vector in V is (vacuously) orthogonal to every vector in the empty set.

where (*) follows from the fact that $\mathbf{x} \in \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}^\perp$. This proves that $\mathbf{x} \perp \mathbf{u}$, and consequently, $\mathbf{x} \in \text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)^\perp$. □

Recall from subsection 3.1.3 that if V is a vector space over a field $\mathbb{F}$, and U and W are subspaces of V, then

$$U + W \quad := \quad \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U,\ \mathbf{w} \in W\}$$

is a subspace of V. Moreover, recall from subsection 3.2.6, that if $V = U + W$ and $U \cap W = \{\mathbf{0}\}$, then we say that V is the *direct sum* of U and W, and we write $V = U \oplus W$.

Theorem 6.4.3. *Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let U be a subspace of V.*[25] *Then $U^\perp$ is a subspace of V, and all the following hold:*

(a) if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, and $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an extension of that basis to an orthogonal basis of V,[26] *then $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$;*

(b) if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal basis of U, and $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an extension of that basis to an orthonormal basis of V,[27] *then $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of $U^\perp$;*

(c) $(U^\perp)^\perp = U$;

(d) $V = U \oplus U^\perp$, that is, $V = U + U^\perp$ and $U \cap U^\perp = \{\mathbf{0}\}$;

(e) $dim(V) = dim(U) + dim(U^\perp)$.

Proof. By Proposition 6.4.1(a), $U^\perp$ is a subspace of V. It remains to prove (a)-(e).

We first prove (a). Assume that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, and that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an extension of that basis to an orthogonal basis of V. We must show that $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$. Clearly, $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal set of vectors, and so it suffices to show that $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is in fact a basis of $U^\perp$. We already know that $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is linearly independent (because it is a subset of the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ of V), and so we need only show that $\text{Span}(\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n) = U^\perp$.

Let us first prove that $\text{Span}(\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n) \supseteq U^\perp$. Fix $\mathbf{x} \in U^\perp$. Then $\mathbf{x} \in V$, and so since $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of V, Theorem 6.3.5 guarantees that

$$\mathbf{x} \quad = \quad \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

[25] By Theorem 3.2.21, the fact that V is finite-dimensional implies that U is also finite-dimensional.
[26] The existence of $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ follows from Corollary 6.3.11(a,b).
[27] The existence of $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ follows from Corollary 6.3.11(c,d).

Since $\mathbf{x} \in U^\perp$, and since $\mathbf{u}_1, \ldots, \mathbf{u}_k \in U$, we know that $\langle \mathbf{x}, \mathbf{u}_i \rangle = 0$ for all $i \in \{1, \ldots, k\}$. Consequently,

$$\mathbf{x} \quad = \quad \sum_{i=k+1}^{n} \tfrac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Thus, $\mathbf{x}$ is a linear combination of the vectors $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$, and we deduce that $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n)$. This proves that $\mathrm{Span}(\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n) \supseteq U^\perp$.

For the reverse inclusion, we fix an arbitrary $\mathbf{x} \in \mathrm{Span}(\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n)$, and we show that $\mathbf{x} \in U^\perp$. Fix scalars $\alpha_{k+1}, \ldots, \alpha_n$ such that

$$\mathbf{x} \quad = \quad \alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_n \mathbf{u}_n.$$

Fix any $\mathbf{u} \in U$; we must show that $\mathbf{x} \perp \mathbf{u}$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of U, we know that there exist scalars $\alpha_1, \ldots, \alpha_k$ such that

$$\mathbf{u} \quad = \quad \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k.$$

Now, if V is a real vector space, then we have that

$$\begin{aligned} \langle \mathbf{x}, \mathbf{u} \rangle \quad &= \quad \langle \alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_n \mathbf{u}_n, \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k \rangle \\ &= \quad \sum_{i=k+1}^{n} \sum_{j=1}^{k} \alpha_i \alpha_j \langle \mathbf{u}_i, \mathbf{u}_j \rangle \quad \overset{(*)}{=} \quad 0, \end{aligned}$$

where (*) follows from the fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal set. Similarly, if V is a complex vector space, then we have that

$$\begin{aligned} \langle \mathbf{x}, \mathbf{u} \rangle \quad &= \quad \langle \alpha_{k+1}\mathbf{u}_{k+1} + \cdots + \alpha_n \mathbf{u}_n, \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k \rangle \\ &= \quad \sum_{i=k+1}^{n} \sum_{j=1}^{k} \alpha_i \overline{\alpha_j} \langle \mathbf{u}_i, \mathbf{u}_j \rangle \quad \overset{(*)}{=} \quad 0, \end{aligned}$$

where (*) follows from the fact that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal set. In either case, we get that $\mathbf{x} \perp \mathbf{u}$, and consequently, $\mathbf{x} \in U^\perp$. It follows that $\mathrm{Span}(\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n) \subseteq U^\perp$. This proves (a). Part (b) follows immediately from part (a).[28]

It remains to prove (c), (d), and (e). First, since V is finite-dimensional, so is U. So, by Corollary 6.3.11(a), U has an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. By Corollary 6.3.11(b), the orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of U can be extended to

[28] This is "obvious," but here are the details. Assume that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal basis of U, and that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an extension of that basis to an orthonormal basis of V. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is, in particular, an orthogonal basis of U, and $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an extension of that basis to an orthogonal basis of V. So, by (a), $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$. But all vectors in $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ are unit vectors (because $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of V). So, $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of $U^\perp$. This proves (b).

an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ of V. By (a), $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$. But then $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n, \mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of V that extends $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$,[29] and so by (a) applied to the vector space $U^\perp$, we have that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of $(U^\perp)^\perp$. But now $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of both U and $(U^\perp)^\perp$, and it follows that $U = (U^\perp)^\perp$, i.e. (c) holds. Further, we have the following:

- $\dim(U) = k$, since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of U;
- $\dim(U^\perp) = n - k$, since $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is a basis of $U^\perp$;
- $\dim(V) = n$, since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is a basis of V.

It now immediately follows that $\dim(V) = \dim(U) + \dim(U^\perp)$, i.e. (e) holds.

Finally, we prove (d). Let us first show that $U \cap U^\perp = \{\mathbf{0}\}$. Since U and $U^\perp$ are both subspaces of V, they both contain $\mathbf{0}$, and consequently, $\mathbf{0} \in U \cap U^\perp$. Now, fix any $\mathbf{u} \in U \cap U^\perp$; we must show that $\mathbf{u} = \mathbf{0}$. Since $\mathbf{u} \in U$ and $\mathbf{u} \in U^\perp$, we have that $\mathbf{u} \perp \mathbf{u}$, i.e. $\langle \mathbf{u}, \mathbf{u} \rangle = 0$. But then by the definition of a scalar product, we have that $\mathbf{u} = \mathbf{0}$. This proves that $U \cap U^\perp = \{\mathbf{0}\}$. It remains to show that $V = U + U^\perp$. It is clear that $U + U^\perp \subseteq V$, and so we need only show that $V \subseteq U + U^\perp$. Fix any $\mathbf{v} \in V$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is a basis of V, we know that there exist scalars $\alpha_1, \ldots, \alpha_n$ such that $\mathbf{v} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n$. Set $\mathbf{v}_1 := \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k$ and $\mathbf{v}_2 := \alpha_{k+1} \mathbf{u}_{k+1} + \cdots + \alpha_n \mathbf{u}_n$. Then $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of U, we see that $\mathbf{v}_1 \in U$, and since $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is a basis of $U^\perp$, we see that $\mathbf{v}_2 \in U^\perp$. So, $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ belongs to $U + U^\perp$, and it follows that $V \subseteq U + U^\perp$. This proves (d), and we are done. □

As a corollary of Theorem 6.4.3(a-b), we obtain the following computationally useful proposition.

Proposition 6.4.4. *Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be any linearly independent set of vectors V, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ be an extension of that linearly independent set to a basis of V.*[30] *Set $U := Span(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.*

(a) If the Gram-Schmidt orthogonalization process (version 1) is applied to input vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ to produce output vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$, then both the following hold:

- *$\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, and $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$;*

[29] Indeed, since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of V, so is $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n, \mathbf{u}_1, \ldots, \mathbf{u}_k\}$ (we simply reordered vectors).

[30] By Theorem 3.2.19, any linearly independent set of vectors in a finite-dimensional vector space can be extended to a basis of that vector space.

- $\left\{\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||}\right\}$ *is an orthonormal basis of* U*, and* $\left\{\frac{\mathbf{u}_{k+1}}{||\mathbf{u}_{k+1}||}, \ldots, \frac{\mathbf{u}_n}{||\mathbf{u}_n||}\right\}$ *is an orthonormal basis of* $U^\perp$*.*

(b) If the Gram-Schmidt orthogonalization process (version 2) is applied to input vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ *to produce output vectors* $\mathbf{z}_1, \ldots, \mathbf{z}_k, \mathbf{z}_{k+1}, \ldots, \mathbf{z}_n$*, then* $\{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$ *is an orthonormal basis of* U*, and* $\{\mathbf{z}_{k+1}, \ldots, \mathbf{z}_n\}$ *is an orthonormal basis of* $U^\perp$*.*

Proof. Let us prove (a). The Gram-Schmidt orthogonalization process (version 1) applied to $\mathbf{v}_1, \ldots, \mathbf{v}_k$ produces vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$, which form an orthogonal basis of U. If we now continue the Gram-Schmidt orthogonalization process (version 1) with vectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$, then we obtain vectors $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$ such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of V. But now by Theorem 6.4.3(a), we know that $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$. If we then normalize our vecotrs $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$, then Proposition 6.3.3(b) guarantees that we obtain an orthonormal basis $\left\{\frac{\mathbf{u}_1}{||\mathbf{u}_1||}, \ldots, \frac{\mathbf{u}_k}{||\mathbf{u}_k||}\right\}$ of U and an orthonormal basis $\left\{\frac{\mathbf{u}_{k+1}}{||\mathbf{u}_{k+1}||}, \ldots, \frac{\mathbf{u}_n}{||\mathbf{u}_n||}\right\}$ of $U^\perp$. This proves (a).

The proof of (b) is similar to that of (a), except that we apply Theorem 6.4.3(b) instead of Theorem 6.4.3(a). □

Example 6.4.5. *Consider the following vectors in* $\mathbb{R}^4$*:*

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ 3 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 2 \\ 4 \\ 4 \\ 2 \end{bmatrix}.$$

Compute an orthonormal basis of $U := Span(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$ *and an orthonormal basis of* $U^\perp$*.*

Solution. First, we need to find a basis of U and extend it to a basis of $\mathbb{R}^4$. For this, we use Proposition 3.3.21. We consider the standard basis $\mathcal{E}_4 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ of $\mathbb{R}^4$, and we form the matrix

$$\begin{aligned} C &:= \left[\begin{array}{cccc|cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{array}\right] \\ &= \left[\begin{array}{cccc|cccc} 1 & 2 & 0 & 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & 0 & 1 & 0 & 0 \\ 1 & 2 & 3 & 4 & 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 2 & 0 & 0 & 0 & 1 \end{array}\right]. \end{aligned}$$

By row reducing, we obtain

$$\text{RREF}(C) = \left[\begin{array}{cccc|cccc} 1 & 2 & 0 & 2 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 2/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{array}\right].$$

As we can see, the pivot columns of C are its first, third, fifth, and sixth column. So, by Proposition 3.3.21, $\{\mathbf{a}_1, \mathbf{a}_3\}$ is a basis of U, and $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{e}_1, \mathbf{e}_2\}$ is a basis of $\mathbb{R}^4$ that extends $\{\mathbf{a}_1, \mathbf{a}_3\}$. By applying the Gram-Schmidt orthogonalization process (version 2) to the vectors $\mathbf{a}_1, \mathbf{a}_3, \mathbf{e}_1, \mathbf{e}_2$, we obtain the following vectors:

$$\mathbf{z}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \\ 0 \end{bmatrix}, \qquad \mathbf{z}_2 = \begin{bmatrix} -2/\sqrt{15} \\ 1/\sqrt{15} \\ 1/\sqrt{15} \\ 3/\sqrt{15} \end{bmatrix},$$

$$\mathbf{z}_3 = \begin{bmatrix} 2/\sqrt{10} \\ -1/\sqrt{10} \\ -1/\sqrt{10} \\ 2/\sqrt{10} \end{bmatrix}, \qquad \mathbf{z}_4 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}.$$

By Proposition 6.4.4(b), $\{\mathbf{z}_1, \mathbf{z}_2\}$ is an orthonormal basis of U, whereas $\{\mathbf{z}_3, \mathbf{z}_4\}$ is an orthonormal basis of $U^\perp$.

Remark: We could also have applied the Gram-Schmidt orthogonalization process (version 1) to the vectors $\mathbf{a}_1, \mathbf{a}_3, \mathbf{e}_1, \mathbf{e}_2$, and then normalized the output vectors. We would have gotten the same vectors $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4$ as above. Proposition 6.4.4(a) would then imply that $\{\mathbf{z}_1, \mathbf{z}_2\}$ is an orthonormal basis of U, whereas $\{\mathbf{z}_3, \mathbf{z}_4\}$ is an orthonormal basis of $U^\perp$. □

6.5 Orthogonal projection onto a subspace

Theorem 6.5.1. *Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $||\cdot||$. Let U be a subspace of V, and let $\mathbf{x} \in V$. Then there exists a unique vector $\mathbf{x}_U \in U$ that has the property that*

$$||\mathbf{x} - \mathbf{x}_U|| = \min_{\mathbf{u} \in U} ||\mathbf{x} - \mathbf{u}||.$$

Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, then this vector $\mathbf{x}_U$ is given by the formula

$$\mathbf{x}_U = \sum_{i=1}^{k} proj_{\mathbf{u}_i}(\mathbf{x}) = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Remark: Note that if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ happens to be an **orthonormal** basis of U, then we get that $\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \cdots = \langle \mathbf{u}_k, \mathbf{u}_k \rangle = 1$, and so the formula for $\mathbf{x}_U$ from Theorem 6.5.1 turns into

$$\mathbf{x}_U = \sum_{i=1}^{k} \mathrm{proj}_{\mathbf{u}_i}(\mathbf{x}) = \sum_{i=1}^{k} \langle \mathbf{x}, \mathbf{u}_i \rangle \, \mathbf{u}_i.$$

Moreover, we note that if $\mathbf{x} \in U$, then $\mathbf{x}_U = \mathbf{x}$, since in this case, the expression $||\mathbf{x} - \mathbf{u}||$ (for $\mathbf{u} \in U$) is minimized for $\mathbf{u} = \mathbf{x}$.

Proof of Theorem 6.5.1. Using Corollary 6.3.11, we fix an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of U, and we extend it to an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ of V. By Theorem 6.4.3(a), $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$. Set

$$\mathbf{u}^* := \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

(So, $\mathbf{u}^*$ is defined via the formula from the statement of the theorem. The reason we call it $\mathbf{u}^*$ rather than $\mathbf{x}_U$ is because we have not proven the existence and uniqueness of $\mathbf{x}_U$ yet. However, this is just a minor stylistic matter!) Since $\mathbf{u}^*$ is a linear combination of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$, which form a basis of U, we see that $\mathbf{u}^* \in U$. Now, fix any $\mathbf{u} \in U$. We must show that $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$, and that equality holds if and only if $\mathbf{u}^* = \mathbf{u}$. Clearly, this is sufficient to prove the theorem.

Let us first prove that $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$. Since $\mathbf{u}^*, \mathbf{u} \in U$, and since U is a subspace of V, it is clear that $\mathbf{u}^* - \mathbf{u} \in U$. So, it suffices to show that $\mathbf{x} - \mathbf{u}^* \in U^\perp$. By Theorem 6.3.5, we have that

$$\mathbf{x} = \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i,$$

and it follows that

$$\mathbf{x} - \mathbf{u}^* = \sum_{i=k+1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

So, $\mathbf{x} - \mathbf{u}^*$ is a linear combination of the vectors $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$; since those $n-k$ vectors form a basis of $U^\perp$, it follows that $\mathbf{x} - \mathbf{u}^* \in U^\perp$. This proves that $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$.

Now that we have shown that vectors $\mathbf{u}^* - \mathbf{u}$ and $\mathbf{x} - \mathbf{u}^*$ are orthogonal to each other, we can apply the Pythagorean theorem (see subsection 6.2.2) to them, as follows:

$$\begin{aligned} ||\mathbf{x} - \mathbf{u}||^2 &= ||(\mathbf{x} - \mathbf{u}^*) + (\mathbf{u}^* - \mathbf{u})||^2 \\ &\overset{(*)}{=} ||\mathbf{x} - \mathbf{u}^*||^2 + ||\mathbf{u}^* - \mathbf{u}||^2 \\ &\geq ||\mathbf{x} - \mathbf{u}^*||^2, \end{aligned}$$

where (*) follows from the Pythagorean theorem. Consequently, we have that $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$. Moreover, the inequality above is an equality if and only if $||\mathbf{u}^* - \mathbf{u}|| = 0$, i.e. if and only if $\mathbf{u}^* = \mathbf{u}$. This completes the argument. $\square$

Terminology/Notation: The vector $\mathbf{x}_U$ from Theorem 6.5.1 is called the *orthogonal projection* of $\mathbf{x}$ onto U.

Corollary 6.5.2. *Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let $\mathbf{u}$ be any non-zero vector in V, and set $U := Span(\mathbf{u})$.*[31] *Then for every $\mathbf{x} \in V$, we have that*

$$\mathbf{x}_U \quad = \quad proj_{\mathbf{u}}(\mathbf{x}) \quad = \quad \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}.$$

Proof. Clearly, $\{\mathbf{u}\}$ is an orthogonal basis of U. So, the result follows immediately from Theorem 6.5.1. □

Corollary 6.5.3. *Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let U be a subspace of V, and let $\mathbf{x} \in V$. Then*

$$\mathbf{x} \quad = \quad \mathbf{x}_U + \mathbf{x}_{U^\perp}.$$

Moreover, this is the unique way of expressing $\mathbf{x}$ as a sum of a vector in U and a vector in $U^\perp$.[32]

Proof. By Corollary 6.3.11, U has an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, and moreover, this basis can be extended to an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ of V. By Theorem 6.4.3(a), we have that $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$. Now, by Theorem 6.5.1, we have that

$$\mathbf{x}_U \quad = \quad \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \qquad \text{and} \qquad \mathbf{x}_{U^\perp} \quad = \quad \sum_{i=k+1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

On the other hand, by Theorem 6.3.5, we have that

$$\mathbf{x} \quad = \quad \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Consequently,

$$\mathbf{x} \quad = \quad \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \quad = \quad \left(\sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \right) + \left(\sum_{i=k+1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \right) \quad = \quad \mathbf{x}_U + \mathbf{x}_{U^\perp}.$$

It remains to prove the uniqueness part of the corollary. So, suppose that $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$ are such that $\mathbf{x} = \mathbf{y} + \mathbf{z}$. We must prove that $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$. We have that

$$\mathbf{x}_U + \mathbf{x}_{U^\perp} \quad = \quad \mathbf{x} \quad = \quad \mathbf{y} + \mathbf{z},$$

[31] So, U is a one-dimensional subspace of V.

[32] This means that for all $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$, if $\mathbf{x} = \mathbf{y} + \mathbf{z}$, then $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$.

and consequently,

$$\mathbf{x}_U - \mathbf{y} \quad = \quad \mathbf{z} - \mathbf{x}_{U^\perp}.$$

But $\mathbf{x}_U - \mathbf{y} \in U$ and $\mathbf{z} - \mathbf{x}_{U^\perp} \in U^\perp$. Since $U \cap U^\perp = \{\mathbf{0}\}$ (by Theorem 6.4.3(d)), it follows that $\mathbf{x}_U - \mathbf{y} = \mathbf{z} - \mathbf{x}_{U^\perp} = \mathbf{0}$, and consequently, $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$. This completes the argument. □

Remark: We note that the uniqueness part of Corollary 6.5.3 could also have been obtained as an immediate consequence of Theorems 6.4.3(d) and 3.2.24. However, note that the proof of Theorem 3.2.24 is actually quite similar to the proof of the uniqueness part of Corollary 6.5.3 that we gave above.

The linearity of orthogonal projection onto a subspace. Suppose that V is a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and suppose that U is a subspace of V. We can then define the function $\text{proj}_U : V \to V$ by setting $\text{proj}_U(\mathbf{x}) = \mathbf{x}_U$ for all $\mathbf{x} \in V$ (where $\mathbf{x}_U$ is the orthogonal projection of $\mathbf{x}$ onto U, as in Theorem 6.5.1). Clearly, $\text{proj}_U(\mathbf{u}) = \mathbf{u}$ for all $\mathbf{u} \in U$. Moreover, we have that $\text{Im}\big(\text{proj}_U\big) = U$ and $\text{proj}_U[U] = U$. Using the formula from Theorem 6.5.1, we can easily see that the function proj_U is linear. Indeed, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is any orthogonal basis of U (such a basis exists by Corollary 6.3.11), then the following hold:

- for all $\mathbf{x}, \mathbf{y} \in V$, we have that

$$\begin{aligned}
\text{proj}_U(\mathbf{x}+\mathbf{y}) &\overset{(*)}{=} \sum_{i=1}^{k} \tfrac{\langle \mathbf{x}+\mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \\
&\overset{(**)}{=} \sum_{i=1}^{k} \tfrac{\langle \mathbf{x}, \mathbf{u}_i \rangle + \langle \mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \\
&= \Big(\sum_{i=1}^{k} \tfrac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \Big) + \Big(\sum_{i=1}^{k} \tfrac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \Big) \\
&\overset{(*)}{=} \text{proj}_U(\mathbf{x}) + \text{proj}_U(\mathbf{y}),
\end{aligned}$$

  where both instances of (*) follow from Theorem 6.5.1, and (**) follows from r.2 or c.2;

- for all $\mathbf{x} \in V$ and scalars α, we have that

$$\begin{aligned}
\text{proj}_U(\alpha \mathbf{x}) &\overset{(*)}{=} \sum_{i=1}^{k} \tfrac{\langle \alpha \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \\
&\overset{(**)}{=} \sum_{i=1}^{k} \tfrac{\alpha \langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i
\end{aligned}$$

$$= \alpha \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$$

$$\overset{(*)}{=} \alpha \text{proj}_U(\mathbf{x}),$$

where both instances of (*) follow from Theorem 6.5.1, and (**) follows from r.3 or c.3.

6.6 Orthogonal projection onto subspaces of $\mathbb{R}^n$

In this section, we assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$. Recall that if we identify 1×1 matrices with scalars, then we have that $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Now, suppose that U is a subspace of $\mathbb{R}^n$. As we saw above (see the comment following the proof of Corollary 6.5.3), $\text{proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (and its image is U). Since proj_U is linear, it has a standard matrix (note that this matrix belongs to $\mathbb{R}^{n \times n}$). In this section, we give formulas for the standard matrices of orthogonal projections onto various subspaces of $\mathbb{R}^n$.

In section 3.3, we defined the row space of a matrix A to be the span of the rows of A, and Proposition 3.3.1(b) states that $\text{Row}(A) = \{\mathbf{u}^T \mid \mathbf{u} \in \text{Col}(A^T)\}$. In this section, it will be convenient to slightly modify the definition of the row space, as follows:

$$\text{Row}(A) \quad := \quad \text{Col}(A^T).$$

So, we (re)defined the row space of a matrix to be the span of the transposes of its rows. For example, for the matrix

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 3 & 2 & 3 \\ 3 & 4 & 3 & 4 \end{bmatrix},$$

we have that

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix},$$

and consequently,

$$\text{Row}(A) = \text{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 3 \\ 4 \end{bmatrix} \right).$$

(If this change of definition bothers you, then every time you see Row($\square$), mentally replace it with Col($\square^T$).)

Theorem 6.6.1. *Let $A \in \mathbb{R}^{n\times m}$. Then Row$(A)^\perp$ = Nul(A) and Row(A) = Nul$(A)^\perp$.*

Proof. In view of Theorem 6.4.3(c), it suffices to show that $\text{Row}(A)^\perp = \text{Nul}(A)$.[33] Set

$$A = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix},$$

so that $\text{Row}(A) = \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.[34] Now, for all vectors $\mathbf{x} \in \mathbb{R}^m$, we have the following sequence of equivalences:

$$\begin{aligned}
\mathbf{x} \in \text{Nul}(A) &\iff A\mathbf{x} = \mathbf{0} \\
&\iff \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} \mathbf{x} = \mathbf{0} \\
&\iff \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{a}_n \cdot \mathbf{x} \end{bmatrix} = \mathbf{0} \\
&\iff \mathbf{a}_i \cdot \mathbf{x} = 0 \quad \forall i \in \{1, \ldots, n\} \\
&\iff \mathbf{a}_i \perp \mathbf{x} \quad \forall i \in \{1, \ldots, n\} \\
&\iff \mathbf{x} \in \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}^\perp \\
&\overset{(*)}{\iff} \mathbf{x} \in \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)^\perp \\
&\iff \mathbf{x} \in \text{Row}(A)^\perp,
\end{aligned}$$

where (*) follows from the fact that $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}^\perp = \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_m)^\perp$ (by Proposition 6.4.2). This proves that $\text{Nul}(A) = \text{Row}(A)^\perp$, and we are done. $\square$

[33] Indeed, by Theorem 6.4.3(c), we have that $(\text{Row}(A)^\perp)^\perp = \text{Row}(A)$. So, if $\text{Row}(A)^\perp = \text{Nul}(A)$, then $\text{Nul}(A)^\perp = (\text{Row}(A)^\perp)^\perp = \text{Row}(A)$.

[34] Indeed, we have the following:

$$\text{Row}(A) = \text{Col}(A^T) = \text{Col}\Big(\begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix} \Big) = \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n).$$

Corollary 6.6.2. *Let $A \in \mathbb{R}^{n\times m}$. Then all the following hold:*

(a) Nul(A^TA) = Nul(A);

(b) Row(A^TA) = Row(A);

(c) rank(A^TA) = rank(A).

Proof. We first prove (a). Note that $A^TA \in \mathbb{R}^{m\times m}$, and that both Nul$(A)$ and Nul(A^TA) are subspaces of $\mathbb{R}^m$. Now, fix any $\mathbf{x} \in \mathbb{R}^m$. We must show that $\mathbf{x} \in \text{Nul}(A^TA)$ if and only if $\mathbf{x} \in \text{Nul}(A)$.

Suppose first that $\mathbf{x} \in \text{Nul}(A)$. Then $A\mathbf{x} = \mathbf{0}$, and consequently, $A^TA\mathbf{x} = \mathbf{0}$. So, $\mathbf{x} \in \text{Nul}(A^TA)$.

Suppose, conversely, that $\mathbf{x} \in \text{Nul}(A^TA)$. Then $A^TA\mathbf{x} = \mathbf{0}$, and it follows that $\mathbf{x}^TA^TA\mathbf{x} = \mathbf{0}$. But note that $\mathbf{x}^TA^TA\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = (A\mathbf{x}) \cdot (A\mathbf{x}) = ||A\mathbf{x}||^2$; consequently, $||A\mathbf{x}||^2 = 0$. It follows that $||A\mathbf{x}|| = 0$, and therefore, $A\mathbf{x} = \mathbf{0}$, i.e. $\mathbf{x} \in \text{Nul}(A)$. This proves (a).

For (b), we observe that

$$\begin{aligned} \text{Row}(A^TA) &= \text{Nul}(A^TA)^\perp && \text{by Theorem 6.6.1} \\ &= \text{Nul}(A)^\perp && \text{by (a)} \\ &= \text{Row}(A) && \text{by Theorem 6.6.1.} \end{aligned}$$

Finally, for (c), we have the following:

$$\begin{aligned} \text{rank}(A^TA) &= \dim\big(\text{Row}(A^TA)\big) && \text{by Theorem 3.3.9} \\ &= \dim\big(\text{Row}(A)\big) && \text{by (b)} \\ &= \text{rank}(A) && \text{by Theorem 3.3.9.} \end{aligned}$$

This completes the argument. □

Theorem 6.6.3. *Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto Col(A), that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C :=$ Col(A) is given by*

$$\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof. Fix $\mathbf{x} \in \mathbb{R}^n$. We must first check that the expression $A(A^TA)^{-1}A^T\mathbf{x}$ is defined and belongs to $C = \text{Col}(A)$. First, note that $A^TA \in \mathbb{R}^{m\times m}$, and that by Corollary 6.6.2(a), we have that $\text{rank}(A^TA) = \text{rank}(A) = m$. So, by the Invertible

Matrix Theorem,[35] A^TA is invertible, and we see that $(A^TA)^{-1}$ is defined and belongs to $\mathbb{R}^{m\times m}$. Since $A \in \mathbb{R}^{n\times m}$, $(A^TA)^{-1} \in \mathbb{R}^{m\times m}$, and $A^T \in \mathbb{R}^{m\times n}$, we see that $A(A^TA)^{-1}A^T \in \mathbb{R}^{n\times n}$; since $\mathbf{x} \in \mathbb{R}^n$, we see that $A(A^TA)^{-1}A^T\mathbf{x}$ is defined and belongs to $\mathbb{R}^n$. Meanwhile, $(A^TA)^{-1}A^T\mathbf{x}$ is a vector in $\mathbb{R}^m$, and so

$$A(A^TA)^{-1}A^T\mathbf{x} \;=\; \underbrace{A}_{\in\mathbb{R}^{n\times m}}\Big(\underbrace{(A^TA)^{-1}A^T\mathbf{x}}_{\in\mathbb{R}^m}\Big)$$

is a linear combination of the columns of A. By definition, this means that $A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Col}(A) = C$.

In view of Corollary 6.5.3, it is now enough to prove that

$$\big(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}\big) \in C^\perp,$$

for it will then follow that $\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}$,[36] which is what we need to show. But note that

$$C^\perp \;=\; \mathrm{Col}(A)^\perp \;=\; \mathrm{Row}(A^T)^\perp \;\overset{(*)}{=}\; \mathrm{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, it in fact suffices to show that the vector $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}$ belongs to $\mathrm{Nul}(A^T)$. For this, we compute:

$$A^T\Big(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}\Big) \;=\; A^T\mathbf{x} - \underbrace{A^TA(A^TA)^{-1}}_{=I_m}A^T\mathbf{x} \;=\; \mathbf{0}.$$

This proves that $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Nul}(A^T)$, and we are done. □

Remark: Suppose that we are given a matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{R}^{n\times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$. If $\mathrm{rank}(A) = m$ (i.e. A has full column rank), then the matrix that we need is $A(A^TA)^{-1}A^T$, as per Theorem 6.6.3. But what if $\mathrm{rank}(A) < m$? In that case, we let B be the matrix obtained from A by deleting all the non-pivot columns of A. By Theorem 3.3.4, the columns of B form a basis of $\mathrm{Col}(A)$, and we see that $\mathrm{Col}(A) = \mathrm{Col}(B)$. Moreover, all the columns of B are pivot columns, and so B has full column rank. But now the matrix B satisfies the hypotheses of Theorem 6.6.3. So, the standard matrix of orthogonal projection onto $\mathrm{Col}(A) = \mathrm{Col}(B)$ is $B(B^TB)^{-1}B^T$.

As a special case of Theorem 6.6.3, we get the following.

[35] See subsection 1.11.7 or 3.3.6.

[36] Indeed, if we can show that $\big(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}\big) \in C^\perp$, then we get that

$$\mathbf{x} \;=\; \underbrace{A(A^TA)^{-1}A^T\mathbf{x}}_{\in C} + \big(\underbrace{\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}}_{\in C^\perp}\big),$$

which (by Corollary 6.5.3) implies that $\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}$ and $\mathbf{x}_{C^\perp} = \mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}$.

Corollary 6.6.4. *Let* $\mathbf{a}$ *be a non-zero vector in* $\mathbb{R}^n$*. Then the standard matrix of orthogonal projection onto the line* $L := Span(\mathbf{a})$ *is the matrix*

$$\mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T \;=\; \mathbf{a}(\mathbf{a}\cdot\mathbf{a})^{-1}\mathbf{a}^T \;=\; \tfrac{1}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T.$$

Consequently, for every vector $\mathbf{x}\in\mathbb{R}^n$*, we have that*

$$\mathbf{x}_L \;=\; proj_L(\mathbf{x}) \;=\; \tfrac{1}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T\mathbf{x}.$$

Proof. We have that $L = \mathrm{Span}(\mathbf{a}) = \mathrm{Col}(\mathbf{a})$, where we think of the vector $\mathbf{a}$ simply as a one-column matrix. Moreover, since $\mathbf{a}\neq 0$, we know that $\mathrm{rank}(\mathbf{a}) = 1$, i.e. the one-column matrix $\mathbf{a}$ has full column rank. The result now follows immediately from Theorem 6.6.3. □

Theorem 6.6.5. *Let* U *be a subspace of* $\mathbb{R}^n$*, and let* $P\in\mathbb{R}^{n\times n}$ *be the standard matrix of* $proj_U$*. Then* $I_n - P$ *is the standard matrix of* $proj_{U^\perp}$*, that is, for all* $\mathbf{x}\in\mathbb{R}^n$*, the orthogonal projection of* $\mathbf{x}$ *onto* $U^\perp$ *is given by* $\mathbf{x}_{U^\perp} = (I_n - P)\mathbf{x}$*.*

Proof. We observe that for all $\mathbf{x}\in\mathbb{R}^n$, we have that

$$(I_n - P)\mathbf{x} \;=\; I_n\mathbf{x} - P\mathbf{x} \;\overset{(*)}{=}\; \mathbf{x} - \mathbf{x}_U \;\overset{(**)}{=}\; \mathbf{x}_{U^\perp},$$

where (*) follows from the fact that P is the standard matrix of proj_U,[37] and (**) follows from Corollary 6.5.3. So, $I_n - P$ is indeed the standard matrix of $\mathrm{proj}_{U^\perp}$. □

Corollary 6.6.6. *Let* $A\in\mathbb{R}^{n\times m}$ *be a matrix of rank* n *(i.e.* A *is a matrix of full row rank). Then the matrix* $I_m - A^T(AA^T)^{-1}A$ *is the standard matrix of orthogonal projection onto* $N := Nul(A)$*, that is, for all* $\mathbf{x}\in\mathbb{R}^m$*, the orthogonal projection of* $\mathbf{x}$ *onto* N *is given by* $\mathbf{x}_N = \big(I_m - A^T(AA^T)^{-1}A\big)\mathbf{x}$*.*

Proof. First, note that

$$\mathrm{Nul}(A) \;\overset{(*)}{=}\; \mathrm{Row}(A)^\perp \;=\; \mathrm{Col}(A^T)^\perp.$$

where (*) follows from Theorem 6.6.1. Note further that $A^T\in\mathbb{R}^{m\times n}$ and that (by Corollary 3.3.11) $\mathrm{rank}(A^T) = \mathrm{rank}(A) = n$, i.e. A^T has full column rank. So, by Theorem 6.6.3, the standard matrix of orthogonal projection onto $\mathrm{Col}(A^T)$ is $A^T(AA^T)^{-1}A$. Finally, by Theorem 6.6.5, the standard matrix of orthogonal projection onto $\mathrm{Col}(A^T)^\perp = \mathrm{Nul}(A)$ is $I_m - A^T(AA^T)^{-1}A$. This completes the argument. □

[37] Technically, the fact that P is the standard matrix of proj_U guarantees that $P\mathbf{x} = \mathbf{x}_U$. The fact that $I_n\mathbf{x} = \mathbf{x}$ follows from Proposition 1.4.5.

Remark: Suppose that we are given a matrix $A \in \mathbb{R}^{n\times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\text{Nul}(A)$. If $\text{rank}(A) = n$ (i.e. A has full row rank), then the matrix that we need is $I_m - A^T(AA^T)^{-1}A$, as per Corollary 6.6.6. But what if $\text{rank}(A) < n$? If $A = O_{n\times m}$ (i.e. A is a zero matrix), then $\text{Nul}(A) = \mathbb{R}^m$, and so the standard matrix of orthogonal projection onto $\text{Nul}(A)$ is the identity matrix I_m. Assume now that $A \neq O_{n\times m}$. In this case, we let B be the matrix obtained from $\text{RREF}(A)$ by deleting any zero rows that $\text{RREF}(A)$ may have. By Proposition 3.3.26, $\text{Nul}(A) = \text{Nul}(B)$. But the matrix B has full row rank, which means that we can apply Corollary 6.6.6 to it. So, the standard matrix of orthogonal projection onto $\text{Nul}(A) = \text{Nul}(B)$ is $I_m - B^T(BB^T)^{-1}B$.

6.7 The method of least squares

In some real-world applications, we may be interested in finding the best **approximate** solution to a (possibly inconsistent) matrix-vector equation $A\mathbf{x} = \mathbf{b}$. More formally, suppose we are given a norm $||\cdot||$ on $\mathbb{R}^n$, a matrix $A \in \mathbb{R}^{n\times m}$, and a vector $\mathbf{b} \in \mathbb{R}^n$. We would then like to find a vector $\mathbf{x}$ for which

$$||A\mathbf{x} - \mathbf{b}||$$

is as small as possible. If $A\mathbf{x} = \mathbf{b}$ is consistent, then any solution of that equation will minimize $||A\mathbf{x} - \mathbf{b}||$. However, what if the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent? Then the answer will obviously depend on which norm that we are using. In the remainder of this section, we will work only with the **norm induced by the standard scalar product** in $\mathbb{R}^n$, i.e. the standard Euclidean norm. Recall that this is the norm $||\cdot||$ given by

$$||\mathbf{x}|| \quad = \quad \sqrt{\mathbf{x}\cdot\mathbf{x}} \quad = \quad \sqrt{x_1^2 + \cdots + x_n^2}$$

for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$.

Theorem 6.7.1. *Let $A \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation*

$$A^TA\mathbf{x} \quad = \quad A^T\mathbf{b}$$

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

Proof. We are looking for vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize the expression $||A\mathbf{x} - \mathbf{b}||$. Our goal is to show is that the vectors we are looking for are precisely those that satisfy $A^TA\mathbf{x} = A^T\mathbf{b}$.

By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the

definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized. Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$. So, we are looking for those $\mathbf{x}$'s for which $\mathbf{b} - A\mathbf{x} \in C^\perp$. But note that

$$C^\perp \quad = \quad \text{Col}(A)^\perp \quad = \quad \text{Row}(A^T)^\perp \quad \stackrel{(*)}{=} \quad \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, we in fact looking for vectors $\mathbf{x}$ for which $\mathbf{b} - A\mathbf{x} \in \text{Nul}(A^T)$, i.e. those that satisfy $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, which is obviously equivalent to $A^TA\mathbf{x} = A^T\mathbf{b}$.

It remains to show that the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ is consistent. By our argument above, a vector $\mathbf{x} \in \mathbb{R}^m$ satisfies $A^TA\mathbf{x} = A^T\mathbf{b}$ if and only if it satisfies the equation $A\mathbf{x} = \mathbf{b}_C$. Since the latter equation is consistent (this follows from the definition of C and the existence of $\mathbf{b}_C$), so is the former. □

Terminology: Suppose we are given a matrix $A \in \mathbb{R}^{n\times m}$ and a vector $\mathbf{b} \in \mathbb{R}^n$. Vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize the expression $||A\mathbf{x} - \mathbf{b}||$ are called the *least-squares solutions* of the equation $A\mathbf{x} = \mathbf{b}$ (such solutions are often denoted by $\hat{\mathbf{x}}$), whereas the number

$$\min_{\mathbf{x}\in\mathbb{R}^m} ||A\mathbf{x} - \mathbf{b}||$$

is called the *least-squares error* for the equation $A\mathbf{x} = \mathbf{b}$. By Theorem 6.7.1, the equation $A\mathbf{x} = \mathbf{b}$ has at least one least-squares solution $\hat{\mathbf{x}}$, and consequently, the least-squares error is defined and is equal to $||A\hat{\mathbf{x}} - \mathbf{b}||$.

Remark: Obviously, if $A\mathbf{x} = \mathbf{b}$ is consistent, then the least-squares solutions of $A\mathbf{x} = \mathbf{b}$ are precisely the solutions of the equation $A\mathbf{x} = \mathbf{b}$ itself. This is because if $A\mathbf{x} = \mathbf{b}$ is consistent, then the solutions of that equation minimize the expression $||A\mathbf{x} - \mathbf{b}||$ (indeed, $||A\mathbf{x} - \mathbf{b}|| = 0$ if and only if $A\mathbf{x} = \mathbf{b}$). Moreover, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent if and only if the least-squares error of this equation is zero.

Example 6.7.2. *Let*

$$A \quad = \quad \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} \qquad \textit{and} \qquad \mathbf{b} \quad = \quad \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Find all least-squares solutions $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$, as well as the least-squares error. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent?

Solution. We apply Theorem 6.7.1. So, we need to find the solutions of the equation $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. We first compute

$$A^TA \quad = \quad \begin{bmatrix} 6 & 6 \\ 6 & 42 \end{bmatrix} \qquad \text{and} \qquad A^T\mathbf{b} \quad = \quad \begin{bmatrix} 6 \\ -6 \end{bmatrix},$$

and then we compute

$$\text{RREF}\Big(\left[\begin{array}{c|c} A^TA & A^T\mathbf{b} \end{array}\right] \Big) \;=\; \left[\begin{array}{cc|c} 1 & 0 & 4/3 \\ 0 & 1 & -1/3 \end{array}\right].$$

It follows that

$$\hat{\mathbf{x}} \;=\; \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix}$$

is the unique solution of the matrix-vector equation $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$, and consequently, the unique least-squares solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$.

The least-squares error of $A\mathbf{x} = \mathbf{b}$ is

$$||A\hat{\mathbf{x}} - \mathbf{b}|| \;=\; || \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix} ||$$

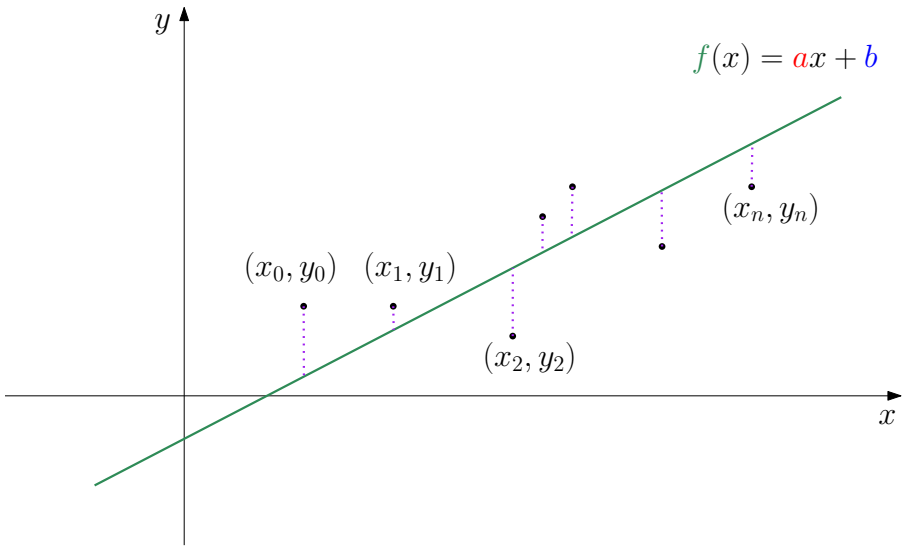
$$= \; || \begin{bmatrix} -1 \\ -3 \\ 3 \\ -1 \end{bmatrix} || \;=\; 2\sqrt{5}.$$

Since the least-squares error of the equation $A\mathbf{x} = \mathbf{b}$ is strictly positive, we see that the equation is inconsistent. □

Remark: In the example above, the equation $A\mathbf{x} = \mathbf{b}$ had a unique least-squares solution $\hat{\mathbf{x}}$, and we obtained the least-squares error of $A\mathbf{x} = \mathbf{b}$ by computing $||A\hat{\mathbf{x}} - \mathbf{b}||$. But what if $A\mathbf{x} = \mathbf{b}$ had more than one least-squares solution? In that case, we would choose one least-squares solution $\hat{\mathbf{x}}$ (any one will do), and we would compute $||A\hat{\mathbf{x}} - \mathbf{b}||$. By the definition of a least-squares solution, the value of $||A\hat{\mathbf{x}} - \mathbf{b}||$ is the same regardless of which least-squares solution $\hat{\mathbf{x}}$ we choose.

6.7.1 Data fitting

Suppose we are given a collection of two or more data points, and we wish to find a line that best fits them. How would we do this? First, let us be a bit more precise. We will be plotting our data points, say $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$, in $\mathbb{R}^2$. Most commonly, the x-axis is time (measured in whatever time units happen to be convenient for the problem that we are studying), whereas the y-axis is the quantity that we are measuring, such as population size, the average global temperature, the number of products of a certain type produced or consumed in a given region, etc. We are looking for a line $f(x) = ax + b$ that best fits our data points (see the graph below).



So, we set up a system of linear equations shown below.

$$\begin{array}{rcl} ax_0 + b & = & y_0 \\ ax_1 + b & = & y_1 \\ & \vdots & \\ ax_n + b & = & y_n \end{array}$$

This linear system can be rewritten as the matrix-vector equation below, where the vector $\begin{bmatrix} a \\ b \end{bmatrix}$ is the unknown.

$$\begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Except in rare cases, the system above will be inconsistent. For this reason, we will look for the least-squares solution(s) $\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$ of the system, which yields the line $\hat{f}(x) = \hat{a}x + \hat{b}$. This (approximate) solution minimizes the following quantity:

$$\left\| \begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \right\| = \left\| \begin{bmatrix} ax_0 + b - y_0 \\ ax_1 + b - y_1 \\ \vdots \\ ax_n + b - y_n \end{bmatrix} \right\|$$

$$= \quad || \begin{bmatrix} f(x_0) - y_0 \\ f(x_1) - y_1 \\ \vdots \\ f(x_n) - y_n \end{bmatrix} ||$$

$$= \quad \sqrt{\sum_{i=0}^{n} \big(f(x_i) - y_i\big)^2}.$$

So, we are effectively minimizing the sum of squares of the vertical distances between our data points and the line, i.e. the sum of squares of the lengths of the purple dotted line segments shown in the graph above.

Example 6.7.3. *Using the method of least squares, find the line that best fits the data points* $(1, 2)$*,* $(2, 3)$*,* $(3, 3)$*,* $(5, 6)$*.*

Solution. We are looking for the function $f(x) = ax + b$ that best fits these four data points. We get the linear system below.

$$\begin{array}{rcl} 1a + b & = & 2 \\ 2a + b & = & 3 \\ 3a + b & = & 3 \\ 5a + b & = & 6 \end{array}$$

At a glance, we can see that this system is inconsistent; so, we will not be able to find an exact solution and will instead have to settle for an approximate one. This system can be rewritten as a matrix-vector equation below, where $\begin{bmatrix} a \\ b \end{bmatrix}$ is the unknown.

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad = \quad \begin{bmatrix} 2 \\ 3 \\ 3 \\ 6 \end{bmatrix}$$

We multiply both sides by the transpose of the matrix on the left, and we get the following (where a and b became $\hat{a}$ and $\hat{b}$, respectively, because we are now approximating):

$$\begin{bmatrix} 1 & 2 & 3 & 5 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & 2 & 3 & 5 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \\ 6 \end{bmatrix}.$$

After performing matrix multiplication, we obtain

$$\begin{bmatrix} 39 & 11 \\ 11 & 5 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \quad = \quad \begin{bmatrix} 47 \\ 14 \end{bmatrix}.$$

We now form the augmented matrix of the matrix-vector equation above, and we row reduce to obtain:

$$\mathrm{RREF}\Big(\left[\begin{array}{cc|c} 39 & 11 & 47 \\ 11 & 5 & 14 \end{array}\right]\Big) \;=\; \left[\begin{array}{cc|c} 1 & 0 & 81/74 \\ 0 & 1 & 29/74 \end{array}\right].$$

This yields the least-squares solution

$$\left[\begin{array}{c} \hat{a} \\ \hat{b} \end{array}\right] \;=\; \left[\begin{array}{c} 81/74 \\ 29/74 \end{array}\right].$$

So, the line that best fits our data points is

$$\hat{f}(x) \;=\; \tfrac{81}{74}x + \tfrac{29}{74}.$$

□

6.8 Orthogonal matrices

Throughout this section, we assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.

6.8.1 Orthogonal matrices: definition and characterization

A matrix $Q \in \mathbb{R}^{n\times n}$ is *orthogonal* if it satisfies $Q^TQ = I_n$. Obviously, matrices I_n and $-I_n$ are orthogonal. Moreover, by Theorem 2.3.14, permutation matrices are orthogonal (as long as we consider the 0's and 1's in those matrices as being real numbers). The matrices mentioned so far all have entries only $-1, 0, 1$. However, there are many other orthogonal matrices: a couple of important examples will be given in subsection 6.8.3. In this subsection, we prove a theorem (Theorem 6.8.1 below) that gives several equivalent characterizations of orthogonal matrices.

Theorem 6.8.1. *Let $Q \in \mathbb{R}^{n\times n}$. Then the following are equivalent:*

(a) Q is orthogonal (i.e. satisfies $Q^TQ = I_n$);

(b) Q is invertible and satisfies $Q^{-1} = Q^T$;

(c) $QQ^T = I_n$;

(d) Q^T is orthogonal;

(e) Q is invertible and Q^{-1} is orthogonal;

(f) the columns of Q form an orthonormal basis of $\mathbb{R}^n$;

(g) the columns of Q^T form an orthonormal basis of $\mathbb{R}^n$.

Proof. By Corollary 3.3.18, we have that (a), (b), and (c) are equivalent. Moreover, since $(Q^T)^T = Q$, we have that (c) and (d) are equivalent. This proves that (a), (b), (c), and (d) are equivalent.

Next, (b) and (d) together imply (e). Suppose now that (e) holds. Then by applying "(a) $\Longrightarrow$ (b)" to the matrix Q^{-1}, we see that Q^{-1} is invertible and satisfies $(Q^{-1})^{-1} = (Q^{-1})^T$. Consequently, $Q^{-1} = Q^T$, and it follows that (b) holds.

So far, we have established that (a), (b), (c), (d), and (e) are equivalent.

Let us now show that (a) and (f) are equivalent. Set $Q = \begin{bmatrix} \mathbf{q}_1 & \dots & \mathbf{q}_n \end{bmatrix}$. Then

$$\begin{aligned} Q^T Q &= \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \dots & \mathbf{q}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{q}_1 \cdot \mathbf{q}_1 & \mathbf{q}_1 \cdot \mathbf{q}_2 & \dots & \mathbf{q}_1 \cdot \mathbf{q}_n \\ \mathbf{q}_2 \cdot \mathbf{q}_1 & \mathbf{q}_2 \cdot \mathbf{q}_2 & \dots & \mathbf{q}_2 \cdot \mathbf{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_n \cdot \mathbf{q}_1 & \mathbf{q}_n \cdot \mathbf{q}_2 & \dots & \mathbf{q}_n \cdot \mathbf{q}_n \end{bmatrix}. \end{aligned}$$

So, $Q^T Q = I_n$ if and only if $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ is an orthonormal set. But by Proposition 6.3.4(b), any orthonormal set of n vectors in $\mathbb{R}^n$ is in fact an orthonormal basis of $\mathbb{R}^n$. It now follows that (a) and (f) are equivalent. Analogously, (d) and (g) are equivalent. This completes the argument. □

6.8.2 Making new orthogonal matrices out of old ones

Proposition 6.8.2. *Let*

$$Q = \begin{bmatrix} \mathbf{q}_1 & \dots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^T \\ \vdots \\ \mathbf{r}_n^T \end{bmatrix}$$

be an orthogonal matrix in $\mathbb{R}^n$. Then all the following hold:

(a) for all $\alpha_1, \dots, \alpha_n \in \{-1, 1\}$, the matrix $\begin{bmatrix} \alpha_1 \mathbf{q}_1 & \dots & \alpha_n \mathbf{q}_n \end{bmatrix}$ is orthogonal;

(b) for all $\alpha_1, \dots, \alpha_n \in \{-1, 1\}$, the matrix $\begin{bmatrix} \alpha_1 \mathbf{r}_1^T \\ \vdots \\ \alpha_n \mathbf{r}_n^T \end{bmatrix}$ is orthogonal;

(c) the matrix $-Q$ is orthogonal.

Remark: Proposition 6.8.2 guarantees that if we multiply one row or one column of an orthogonal matrix by -1, then the resulting matrix is again orthogonal. Obviously, we can iterate the process and obtain a sequence of orthogonal matrices.

Proof. We first prove (a). Fix $\alpha_1, \ldots, \alpha_n \in \{-1, 1\}$. Since $Q = \begin{bmatrix} \mathbf{q}_1 & \ldots & \mathbf{q}_n \end{bmatrix}$ is orthogonal, Theorem 6.8.1 guarantees that $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ is an orthonormal basis of $\mathbb{R}^n$. But then $\{\alpha_1\mathbf{q}_1, \ldots, \alpha_n\mathbf{q}_n\}$ is also an orthonormal basis of $\mathbb{R}^n$,[38] and so once again by Theorem 6.8.1, the matrix $\begin{bmatrix} \alpha_1\mathbf{q}_1 & \ldots & \alpha_n\mathbf{q}_n \end{bmatrix}$ is orthogonal. This proves (a).

Next, we prove (b). Fix $\alpha_1, \ldots, \alpha_n \in \{-1, 1\}$. Since Q is orthogonal, Theorem 6.8.1 guarantees that $Q^T = \begin{bmatrix} \mathbf{r}_1, \ldots, \mathbf{r}_n \end{bmatrix}$ is also orthogonal. By (a) applied to the orthogonal matrix Q^T, we get that $\begin{bmatrix} \alpha_1\mathbf{r}_1, \ldots, \alpha_n\mathbf{r}_n \end{bmatrix}$ is orthogonal. We now apply Theorem 6.8.1 to the matrix $\begin{bmatrix} \alpha_1\mathbf{r}_1, \ldots, \alpha_n\mathbf{r}_n \end{bmatrix}$, and we deduce that its transpose is orthogonal. This proves (b).

Finally, part (c) is simply a special case of (a) for $\alpha_1 = \cdots = \alpha_n = -1$. □

Proposition 6.8.3. *If $Q_1, Q_2 \in \mathbb{R}^{n\times n}$ are orthogonal, then so is their product Q_1Q_2.*

Proof. Assume $Q_1, Q_2 \in \mathbb{R}^{n\times n}$ are orthogonal. Then $Q_1^TQ_1 = I_n$ and $Q_2^TQ_2 = I_n$, and consequently,

$$(Q_1Q_2)^T(Q_1Q_2) \quad = \quad Q_2^T \underbrace{Q_1^TQ_1}_{=I_n} Q_2 \quad = \quad Q_2^TQ_2 \quad = \quad I_n.$$

So, Q_1Q_2 is indeed orthogonal. □

Proposition 6.8.4. *Let $Q_1 \in \mathbb{R}^{m\times m}$ and $Q_2 \in \mathbb{R}^{n\times n}$ be orthogonal matrices. Then the $(m+n)\times(m+n)$ matrix*

$$Q \quad = \quad \left[\begin{array}{c|c} Q_1 & O_{m\times n} \\ \hline O_{n\times m} & Q_2 \end{array}\right]$$

is an orthogonal matrix in $\mathbb{R}^{(m+n)\times(m+n)}$.

Proof. By hypothesis, we have that $Q_1^TQ_1 = I_m$ and $Q_2^TQ_2 = I_n$. We now compute:

[38] This is "obvious," but here are the details. First of all, since $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ is an orthonormal basis of $\mathbb{R}^n$, it is, in particular, an orthogonal set of unit vectors in $\mathbb{R}^n$. By Proposition 6.3.3(a), the fact that $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ is an orthogonal set implies that $\{\alpha_1\mathbf{q}_1, \ldots, \alpha_n\mathbf{q}_n\}$ is also an orthogonal set. On the other hand, for all $i \in \{1, \ldots, n\}$, we have that

$$||\alpha_i\mathbf{q}_i|| \overset{(*)}{=} |\alpha_i|\,||\mathbf{q}_i|| \overset{(**)}{=} ||\mathbf{q}_i|| \overset{(***)}{=} 1,$$

where (*) follows from Proposition 6.2.1, (**) follows from the fact that $\alpha_i \in \{-1, 1\}$, and (***) follows from the fact that $\mathbf{q}_i$ is a unit vector. But now $\{\alpha_1\mathbf{q}_1, \ldots, \alpha_n\mathbf{q}_n\}$ is an orthonormal set of n vectors in $\mathbb{R}^n$, and so Proposition 6.3.4(b) guarantees that $\{\alpha_1\mathbf{q}_1, \ldots, \alpha_n\mathbf{q}_n\}$ is in fact an orthonormal basis of $\mathbb{R}^n$.

$$\begin{aligned}
Q^TQ &= \left[\begin{array}{c|c} Q_1^T & O_{m\times n} \\ \hline O_{n\times m} & Q_2^T \end{array}\right]\left[\begin{array}{c|c} Q_1 & O_{m\times n} \\ \hline O_{n\times m} & Q_2 \end{array}\right] \\
&= \left[\begin{array}{c|c} Q_1^TQ_1 + O_{m\times n}O_{n\times m} & Q_1^TO_{m\times n} + O_{m\times n}Q_2 \\ \hline O_{n\times m}Q_1 + Q_2^TO_{n\times m} & O_{n\times m}O_{m\times n} + Q_2^TQ_2 \end{array}\right] \\
&= \left[\begin{array}{c|c} Q_1^TQ_1 & O_{m\times n} \\ \hline O_{n\times m} & Q_2^TQ_2 \end{array}\right] \\
&= \left[\begin{array}{c|c} I_m & O_{m\times n} \\ \hline O_{n\times m} & I_n \end{array}\right] \\
&= I_{m+n}.
\end{aligned}$$

So, Q is indeed an orthogonal matrix. □

6.8.3 The Householder matrix

For a non-zero vector $\mathbf{a}$ in $\mathbb{R}^n$, the *Householder matrix* is the $n \times n$ matrix

$$H(\mathbf{a}) := I_n - \tfrac{2}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T = I_n - \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T.$$

To see that $H(\mathbf{a})$ really is an orthogonal matrix, we perform the following simple calculation:

$$\begin{aligned}
H(\mathbf{a})^T\, H(\mathbf{a}) &= (I_n - \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T)^T\,(I_n - \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T) \\
&= (I_n^T - \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}(\mathbf{a}\mathbf{a}^T)^T)\,(I_n - \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T) \\
&= (I_n - \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T)\,(I_n - \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T) \\
&= I_n - \tfrac{4}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T + \tfrac{4}{(\mathbf{a}\cdot\mathbf{a})^2}\mathbf{a}\underbrace{\mathbf{a}^T\mathbf{a}}_{=\mathbf{a}\cdot\mathbf{a}}\mathbf{a}^T \\
&= I_n - \tfrac{4}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T + \tfrac{4}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T \\
&= I_n.
\end{aligned}$$

Let us now discuss the geometric meaning of this matrix. By Corollary 6.6.4, the standard matrix of orthogonal projection onto the line Span($\mathbf{a}$) is

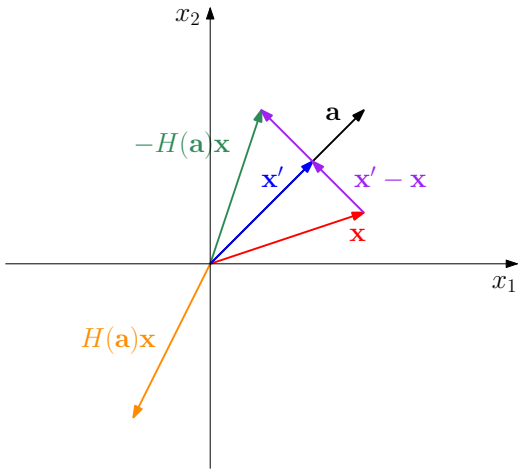
$$\mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T = \mathbf{a}(\mathbf{a}\cdot\mathbf{a})^{-1}\mathbf{a}^T = \tfrac{1}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T.$$

Now, if $\mathbf{x}$ is any vector in $\mathbb{R}^n$, and $\mathbf{x}'$ represents the orthogonal projection of $\mathbf{x}$ onto

$\text{Span}(\mathbf{a}) = \text{Col}(\begin{bmatrix} \mathbf{a} \end{bmatrix})$,[39] then the reflection of $\mathbf{x}$ about the line $\text{Span}(\mathbf{a}) = \text{Col}(\begin{bmatrix} \mathbf{a} \end{bmatrix})$ is given by

$$\begin{aligned} \mathbf{x} + 2(\mathbf{x}' - \mathbf{x}) &= 2\mathbf{x}' - \mathbf{x} \\ &= \tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T\mathbf{x} - I_n\mathbf{x} \\ &= (\tfrac{2}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T - I_n)\mathbf{x} \\ &= -H(\mathbf{a})\mathbf{x}. \end{aligned}$$

Thus, $-H(\mathbf{a})$ is the standard matrix of reflection about the $\text{Span}(\mathbf{a})$ line. The Householder matrix $H(\mathbf{a})$ itself is the standard matrix of the linear operation that first reflects about the $\text{Span}(\mathbf{a})$ line and then reflects about the origin. In the case of $\mathbb{R}^2$, this is illustrated in the picture below.



Remark: Suppose that $\mathbf{a}$ is a non-zero vector in $\mathbb{R}^n$. Then the standard matrix of reflection about the line $L := \text{Span}(\mathbf{a})$ in $\mathbb{R}^n$ is an orthogonal matrix. Indeed, as we saw above, the Householder matrix $H(\mathbf{a})$ is an orthogonal matrix. By Proposition 6.8.2(c), it follows that $-H(\mathbf{a})$ is also an orthogonal matrix, and as we saw above, $-H(\mathbf{a})$ is the standard matrix of reflection about the line $L = \text{Span}(\mathbf{a})$ in $\mathbb{R}^n$.

6.8.4 The Givens matrix

Given an integer $n \geq 2$, indices $i, j \in \{1, \ldots, n\}$ such that $i < j$, and real numbers c and s such that $c^2 + s^2 = 1$, we define the *Givens matrix* $G_{i,j}(c,s) = \begin{bmatrix} g_{i,j} \end{bmatrix}_{n\times n}$ as follows:

- $g_{i,i} = g_{j,j} = c$;

[39] So, $\mathbf{x}' = \frac{1}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T\mathbf{x}$, since $\frac{1}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T$ is the standard matrix of orthogonal projection onto $\text{Span}(\mathbf{a})$.

- $g_{k,k} = 1$ for all $k \in \{1, \ldots, n\} \setminus \{i, j\}$;
- $g_{i,j} = -s$;
- $g_{j,i} = s$;
- all other entries of the matrix are zero.

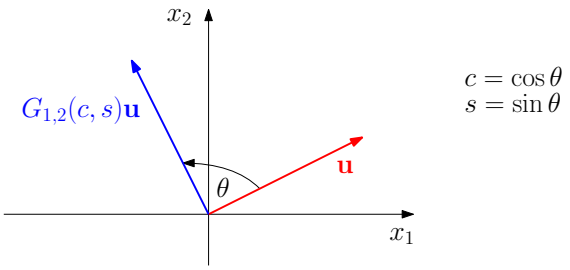
So, the matrix $G_{i,j}(c, s)$ can be represented as follows:

$$G_{i,j}(c,s) = \begin{array}{c} \\ \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccccccccc} & & & & i & & & & j & & & \end{array} \\ \left[\begin{array}{cccccccccccc} 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & c & 0 & \ldots & 0 & -s & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & s & 0 & \ldots & 0 & c & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & & 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & & 0 & 0 & 0 & \ldots & 1 \end{array}\right] \end{array}$$

It is not hard to check that the columns of $G_{i,j}(c, s)$ form an orthonormal set of vectors in $\mathbb{R}^n$, and therefore (by Proposition 6.3.4) an orthonormal basis of $\mathbb{R}^n$. So, by Theorem 6.8.1, the Givens matrix $G_{i,j}(c, s)$ really is orthogonal. Let us now give a geometric interpretation of this matrix. Since $c^2 + s^2 = 1$, we see that there exists a real number (angle in radians) θ such that $c = \cos\theta$ and $s = \sin\theta$. With this set-up, we see that $G_{i,j}(c, s)$ represents rotation about the origin by angle θ in the $x_i x_j$-plane. This is particularly easy to see in the case when $n = 2$. In that case, we have that

$$G_{1,2}(c,s) = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

which is precisely the standard matrix of counterclockwise rotation about the origin by angle θ.



6.8.5 Orthogonal matrices, length, and angles

Theorem 6.8.5. *Let $Q = \left[\ q_{i,j}\ \right]_{n\times n}$ be an orthogonal matrix in $\mathbb{R}^{n\times n}$. Then all the following hold:*

(a) for all $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, $(Q\mathbf{x})\cdot(Q\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$;

(b) for all $\mathbf{x}\in\mathbb{R}^n$, $||Q\mathbf{x}|| = ||\mathbf{x}||$;

(c) for all $i,j\in\{1,\ldots,n\}$, $|q_{i,j}|\leq 1$.

Proof. (a) For $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, we have the following:

$$(Q\mathbf{x})\cdot(Q\mathbf{y}) \;=\; (Q\mathbf{x})^T(Q\mathbf{x}) \;=\; \mathbf{x}^T\underbrace{Q^TQ}_{=I_n}\mathbf{y} \;=\; \mathbf{x}^T\mathbf{y} \;=\; \mathbf{x}\cdot\mathbf{y}.$$

(b) For $\mathbf{x}\in\mathbb{R}^n$, we have the following:

$$||Q\mathbf{x}|| \;=\; \sqrt{(Q\mathbf{x})\cdot(Q\mathbf{x})} \;\overset{(a)}{=}\; \sqrt{\mathbf{x}\cdot\mathbf{x}} \;=\; ||\mathbf{x}||.$$

(c) By Theorem 6.8.1, the columns of Q form an orthonormal basis. In particular, all columns of Q are unit vectors, and it follows that all entries of Q have absolute value at most 1. □

Remark: By Theorem 6.8.5(b), multiplication by an orthogonal matrix (on the left) preserves vector length. On the other hand, recall that for non-zero vectors $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, we have that $\mathbf{x}\cdot\mathbf{y} = ||\mathbf{x}||\,||\mathbf{y}||\cos\theta$, where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. So, Theorem 6.8.5(a-b) implies that multiplication (on the left) by an orthogonal matrix preserves angles between non-zero vectors.

6.9 Scalar product, coordinate vectors, and matrices of linear functions

Proposition 6.9.1. *Let V be a real or complex vector space, equipped with the scalar product $\langle\cdot,\cdot\rangle$ and the induced norm $||\cdot||$, and let $\mathcal{B} = \{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ be an **orthonormal***

basis of V. Let $\cdot$ be the standard scalar product in $\mathbb{R}^n$ or $\mathbb{C}^n$ (depending on whether the vector space V is real or complex). Then for all $\mathbf{x}, \mathbf{y} \in V$, we have that

$$\langle \mathbf{x}, \mathbf{y} \rangle \;=\; \left[\ \mathbf{x}\ \right]_{\mathcal{B}} \cdot \left[\ \mathbf{y}\ \right]_{\mathcal{B}}.$$

Proof. We prove the result for the case when V is a complex vector space.[40] The proof for the real case is similar but slightly easier (because we do not have to deal with complex conjugates). Fix $\mathbf{x}, \mathbf{y} \in V$. Since $\mathcal{B}$ is an orthonormal basis of V, Corollary 6.3.6 guarantees that

$$\mathbf{x} \;=\; \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{u}_i \rangle\ \mathbf{u}_i \qquad \text{and} \qquad \mathbf{y} \;=\; \sum_{i=1}^{n} \langle \mathbf{y}, \mathbf{u}_i \rangle\ \mathbf{u}_i,$$

and consequently,

$$\left[\ \mathbf{x}\ \right]_{\mathcal{B}} \;=\; \begin{bmatrix} \langle \mathbf{x}, \mathbf{u}_1 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{u}_n \rangle \end{bmatrix} \qquad \text{and} \qquad \left[\ \mathbf{y}\ \right]_{\mathcal{B}} \;=\; \begin{bmatrix} \langle \mathbf{y}, \mathbf{u}_1 \rangle \\ \vdots \\ \langle \mathbf{y}, \mathbf{u}_n \rangle \end{bmatrix}.$$

We now compute:

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{y} \rangle &= \left\langle \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{u}_i \rangle\ \mathbf{u}_i, \sum_{i=1}^{n} \langle \mathbf{y}, \mathbf{u}_i \rangle\ \mathbf{u}_i \right\rangle \\
&= \left\langle \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{u}_i \rangle\ \mathbf{u}_i, \sum_{j=1}^{n} \langle \mathbf{y}, \mathbf{u}_j \rangle\ \mathbf{u}_j \right\rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle \langle \mathbf{x}, \mathbf{u}_i \rangle\ \mathbf{u}_i, \langle \mathbf{y}, \mathbf{u}_j \rangle\ \mathbf{u}_j \right\rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \mathbf{x}, \mathbf{u}_i \rangle \overline{\langle \mathbf{y}, \mathbf{u}_j \rangle}\ \langle \mathbf{u}_i, \mathbf{u}_j \rangle \\
&\overset{(*)}{=} \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{u}_i \rangle \overline{\langle \mathbf{y}, \mathbf{u}_i \rangle} \\
&= \begin{bmatrix} \langle \mathbf{x}, \mathbf{u}_1 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{u}_n \rangle \end{bmatrix} \cdot \begin{bmatrix} \langle \mathbf{y}, \mathbf{u}_1 \rangle \\ \vdots \\ \langle \mathbf{y}, \mathbf{u}_n \rangle \end{bmatrix} \\
&= \left[\ \mathbf{x}\ \right]_{\mathcal{B}} \cdot \left[\ \mathbf{y}\ \right]_{\mathcal{B}},
\end{aligned}$$

where (*) follows from the fact that $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal set. □

[40] In this case, $\cdot$ is the standard scalar product in $\mathbb{C}^n$.

Theorem 6.9.2. *Let U and V be non-trivial, finite-dimensional* ***real*** *vector spaces. Assume that U is equipped with a scalar product $\langle \cdot, \cdot \rangle_U$ and the induced norm $|| \cdot ||_U$, and that V is equipped with a scalar product $\langle \cdot, \cdot \rangle_V$ and the induced norm $|| \cdot ||_V$. Let $\mathcal{B}_U = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and $\mathcal{B}_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be* ***orthonormal*** *bases of U and V, respectively, and let $f : U \to V$ be a linear function. Then the following two statements are equivalent:*

(i) the columns of the $n \times m$ matrix ${}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}$ form an orthonormal set of vectors in $\mathbb{R}^n$ (with respect to the standard scalar product $\cdot$ and the induced norm $|| \cdot ||$);[41]

(ii) f preserves the scalar product, that is, for all vectors $\mathbf{x}, \mathbf{y} \in U$, we have that $\big\langle f(\mathbf{x}), f(\mathbf{y}) \big\rangle_V = \langle \mathbf{x}, \mathbf{y} \rangle_U$.

Proof. Set ${}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U} = \begin{bmatrix} \mathbf{c}_1 & \ldots & \mathbf{c}_m \end{bmatrix}$. We observe that

$$({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U})^T\ {}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U} = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_m^T \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_m \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{c}_1 \cdot \mathbf{c}_1 & \mathbf{c}_1 \cdot \mathbf{c}_2 & \ldots & \mathbf{c}_1 \cdot \mathbf{c}_m \\ \mathbf{c}_2 \cdot \mathbf{c}_1 & \mathbf{c}_2 \cdot \mathbf{c}_2 & \ldots & \mathbf{c}_2 \cdot \mathbf{c}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_m \cdot \mathbf{c}_1 & \mathbf{c}_m \cdot \mathbf{c}_2 & \ldots & \mathbf{c}_m \cdot \mathbf{c}_m \end{bmatrix}.$$

So, we see that (i) holds if and only if $({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U})^T\ {}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U} = I_m$.

Next, by Proposition 6.9.1, the following hold for all $\mathbf{x}, \mathbf{y} \in U$:

(1) $\langle \mathbf{x}, \mathbf{y} \rangle_U = [\ \mathbf{x}\ ]_{\mathcal{B}_U} \cdot [\ \mathbf{y}\ ]_{\mathcal{B}_U}$;

(2) $\big\langle f(\mathbf{x}), f(\mathbf{y}) \big\rangle_V = [\ f(\mathbf{x})\ ]_{\mathcal{B}_V} \cdot [\ f(\mathbf{y})\ ]_{\mathcal{B}_V}$.

Now, for all $\mathbf{x}, \mathbf{y} \in U$, we have that

$$\big\langle f(\mathbf{x}), f(\mathbf{y}) \big\rangle_V \overset{(2)}{=} [\ f(\mathbf{x})\ ]_{\mathcal{B}_V} \cdot [\ f(\mathbf{y})\ ]_{\mathcal{B}_V}$$

$$= \big([\ f(\mathbf{x})\ ]_{\mathcal{B}_V}\big)^T\ [\ f(\mathbf{y})\ ]_{\mathcal{B}_V}$$

[41] However, despite Theorem 6.8.1, this does not necessarily mean that the matrix ${}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}$ is orthogonal. This is because ${}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}$ is an $n \times m$ matrix, and it is possible that $m \neq n$, in which case ${}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}$ is not a square matrix. Only square matrices can be orthogonal!

$$
\begin{aligned}
&= \Big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}\ \ [\ \mathbf{x}\ ]_{\mathcal{B}_U} \Big)^T \Big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}\ \ [\ \mathbf{y}\ ]_{\mathcal{B}_U} \Big) \\
&= \big([\ \mathbf{x}\ ]_{\mathcal{B}_U} \big)^T \big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U} \big)^T\ {}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}\ \ [\ \mathbf{y}\ ]_{\mathcal{B}_U}.
\end{aligned}
$$

Suppose first that (i) holds. Then $({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U})^T\ {}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U} = I_m$, and consequently, for all $\mathbf{x}, \mathbf{y} \in U$, we have that

$$
\begin{aligned}
\big\langle f(\mathbf{x}), f(\mathbf{y}) \big\rangle_V &= \big([\ \mathbf{x}\ ]_{\mathcal{B}_U} \big)^T \underbrace{\big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U} \big)^T\ {}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}}_{=I_m}\ [\ \mathbf{y}\ ]_{\mathcal{B}_U} \\
&= \big([\ \mathbf{x}\ ]_{\mathcal{B}_U} \big)^T\ [\ \mathbf{y}\ ]_{\mathcal{B}_U} \\
&= [\ \mathbf{x}\ ]_{\mathcal{B}_U} \cdot [\ \mathbf{y}\ ]_{\mathcal{B}_U} \\
&\overset{(1)}{=} \langle \mathbf{x}, \mathbf{y} \rangle_U.
\end{aligned}
$$

Thus, (ii) holds.

Suppose now that (ii) holds. Then for all $i, j \in \{1, \ldots, m\}$, we have that

$$
\begin{aligned}
\mathbf{e}_i^m \cdot \mathbf{e}_j^m &= [\ \mathbf{u}_i\ ]_{\mathcal{B}_U} \cdot [\ \mathbf{u}_j\ ]_{\mathcal{B}_U} \\
&\overset{(1)}{=} \langle \mathbf{u}_i, \mathbf{u}_j \rangle_U \\
&\overset{\text{(ii)}}{=} \big\langle f(\mathbf{u}_i), f(\mathbf{u}_j) \big\rangle_V \\
&\overset{(2)}{=} [\ f(\mathbf{u}_i)\ ]_{\mathcal{B}_V} \cdot [\ f(\mathbf{u}_j)\ ]_{\mathcal{B}_V} \\
&= \big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}\ \ [\ \mathbf{u}_i\ ]_{\mathcal{B}_U} \big) \cdot \big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}\ \ [\ \mathbf{u}_j\ ]_{\mathcal{B}_U} \big) \\
&= \big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}\ \mathbf{e}_i^m \big) \cdot \big({}_{\mathcal{B}_V}[\ f\ ]_{\mathcal{B}_U}\ \mathbf{e}_j^m \big) \\
&= \mathbf{c}_i \cdot \mathbf{c}_j.
\end{aligned}
$$

So, for all distinct $i, j \in \{1, \ldots, n\}$, we have that $\mathbf{c}_i \cdot \mathbf{c}_j = \mathbf{e}_i^m \cdot \mathbf{e}_j^m = 0$, and it follows that $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ is an orthogonal set of vectors in $\mathbb{R}^n$. On the other hand, for all $i \in \{1, \ldots, m\}$, we have that $||\mathbf{c}_i|| = \sqrt{\mathbf{c}_i \cdot \mathbf{c}_i} = \sqrt{\mathbf{e}_i^m \cdot \mathbf{e}_i^m} = ||\mathbf{e}_i^m|| = 1$. Thus, $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ is an orthonormal set of vectors in $\mathbb{R}^n$, that is, (i) holds. □

Chapter 7

Determinants

7.1 Determinants: definition, examples, and basic properties

The *determinant* of a matrix $A = \left[\ a_{i,j}\ \right]_{n\times n}$ with entries in some field $\mathbb{F}$, denoted by $\det(A)$ or $|A|$, is defined by

$$\begin{aligned}\det(A) &:= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \dots a_{n,\sigma(n)}.\end{aligned}$$

Let us try to explain this definition. Each permutation $\sigma \in S_n$ gives us one way of selecting one entry of A out of each row and each column: we select entries $a_{1,\sigma(1)}, \dots, a_{n,\sigma(n)}$, multiply them together, and then multiply that product by $\operatorname{sgn}(\sigma)$, which yields the product $\operatorname{sgn}(\sigma) a_{1,\sigma(1)} \dots a_{n,\sigma(n)}$.[1] We then sum up all products of this type (there are $|S_n| = n!$ many of them), and we obtain the determinant of our matrix.

Note that if the entries of our square matrix belong to a field of characteristic 2 (i.e. a field in which $1 + 1 = 0$, such as the field $\mathbb{Z}_2$), then $1 = -1$, and so $\operatorname{sgn}(\sigma)$ can be ignored (because it is always equal to 1). However, if our field is of characteristic

[1] For example, for $n = 4$ and $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix} = (134)(2)$, we select the boxed entries below,

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix},$$

and we obtain the product $\operatorname{sgn}(\sigma) a_{1,3} a_{2,2} a_{3,4} a_{4,1} = a_{1,3} a_{2,2} a_{3,4} a_{4,1}$, since $\operatorname{sgn}(\sigma) = 1$.

other than 2 (i.e. if $1+1 \neq 0$ in our field, and consequently, $1 \neq -1$), then we must keep track of $\mathrm{sgn}(\sigma)$ in each summand from the definition of a determinant. (For a more detailed discussion of the characteristic of a field, see subsection 2.4.4.)

Notation: We typically write

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix}$$

instead of

$$\det\Big(\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix} \Big).$$

Remark: Only **square** matrices have determinants. Moreover, the determinant of a matrix in $\mathbb{F}^{n\times n}$ is always a scalar in $\mathbb{F}$.

The following proposition easily follows from the definition of a determinant. We note that permutation matrices were discussed in subsection 2.3.7.

Proposition 7.1.1. *Let n be a positive integer, and let $\pi \in S_n$, and consider the matrix P_π of the permutation π (where the 0's and 1's in P_π can be considered as belonging to an arbitrary field $\mathbb{F}$). Then*

$$det(P_\pi) \quad = \quad sgn(\pi).$$

Proof. Set $P_\pi = \begin{bmatrix} p_{i,j} \end{bmatrix}_{n\times n}$, so that

$$p_{i,j} \quad = \quad \begin{cases} 1 & \text{if} \quad j = \pi(i) \\ 0 & \text{if} \quad j \neq \pi(i) \end{cases}$$

for all $i, j \in \{1, \dots, n\}$. By definition,

$$\det(P_\pi) \quad = \quad \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) p_{1,\sigma(1)} p_{2,\sigma(2)} \cdots p_{n,\sigma(n)}.$$

The only permutation $\sigma \in S_n$ for which none of $p_{1,\sigma(1)}, p_{2,\sigma(2)}, \dots, p_{n,\sigma(n)}$ is 0 is the permutation $\sigma = \pi$. So,

$$\det(P_\pi) \quad = \quad \mathrm{sgn}(\pi) p_{1,\pi(1)} p_{2,\pi(2)} \cdots p_{n,\pi(n)} \quad \overset{(*)}{=} \quad \mathrm{sgn}(\pi),$$

where (*) follows from the fact that $p_{i,\pi(i)} = 1$ for all $i \in \{1, \dots, n\}$. □

Remark: Note that the identity matrix I_n is the matrix of the identity permutation 1 in S_n. Since $\mathrm{sgn}(1) = 1$, Proposition 7.1.1 guarantees that $\det(I_n) = 1$.

Proposition 7.1.2. *We have the following formulas for the determinants of* 1×1, 2×2, *and* 3×3 *matrices (with entries in some field* $\mathbb{F}$*):*

(a) $\left| \; a_{1,1} \; \right| = a_{1,1}$;[2]

(b) $\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$;

(c) $\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = \begin{cases} a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} \\ -a_{1,3}a_{2,2}a_{3,1} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3}. \end{cases}$

Proof. (a) S_1 has just one element, namely $\sigma_1 = (1)$, with $\mathrm{sgn}(\sigma_1) = 1$. So, we have that

$$\left| \; a_{1,1} \; \right| \;=\; \mathrm{sgn}(\sigma_1)a_{1,\sigma_1(1)} \;=\; a_{1,1}.$$

(b) S_2 has two elements, listed below, along with their signs.

- $\sigma_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} = (1)(2)$, with $\mathrm{sgn}(\sigma_1) = 1$;
- $\sigma_2 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = (12)$, with $\mathrm{sgn}(\sigma_2) = -1$.

So, we have that

$$\begin{aligned} \begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} &= \mathrm{sgn}(\sigma_1)a_{1,\sigma_1(1)}a_{2,\sigma_1(2)} + \mathrm{sgn}(\sigma_2)a_{1,\sigma_2(1)}a_{2,\sigma_2(2)} \\ &= a_{1,1}a_{2,2} - a_{1,2}a_{2,1}. \end{aligned}$$

(c) S_3 has six elements, listed below, along with their signs.

- $\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = (1)(2)(3)$, with $\mathrm{sgn}(\sigma_1) = 1$;
- $\sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = (123)$, with $\mathrm{sgn}(\sigma_2) = 1$;
- $\sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = (132)$, with $\mathrm{sgn}(\sigma_3) = 1$;

[2] Be careful not to confuse this with the absolute value! (The notation is admittedly somewhat unfortunate/ambiguous.) If there is any danger of confusion, it is always possible to write $\det(\begin{bmatrix} a_{1,1} \end{bmatrix})$ instead of $\left| \; a_{1,1} \; \right|$.

- $\sigma_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = (13)(2)$, with $\operatorname{sgn}(\sigma_4) = -1$;
- $\sigma_5 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = (1)(23)$, with $\operatorname{sgn}(\sigma_5) = -1$;
- $\sigma_6 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = (12)(3)$, with $\operatorname{sgn}(\sigma_6) = -1$.

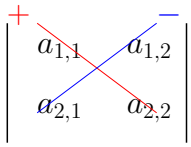
So, we have that

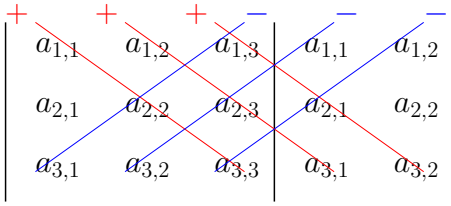
$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = \begin{cases} \operatorname{sgn}(\sigma_1)a_{1,\sigma_1(1)}a_{2,\sigma_1(2)}a_{3,\sigma_1(3)} \\ +\operatorname{sgn}(\sigma_2)a_{1,\sigma_2(1)}a_{2,\sigma_2(2)}a_{3,\sigma_2(3)} \\ +\operatorname{sgn}(\sigma_3)a_{1,\sigma_3(1)}a_{2,\sigma_3(2)}a_{3,\sigma_3(3)} \\ +\operatorname{sgn}(\sigma_4)a_{1,\sigma_4(1)}a_{2,\sigma_4(2)}a_{3,\sigma_4(3)} \\ +\operatorname{sgn}(\sigma_5)a_{1,\sigma_5(1)}a_{2,\sigma_5(2)}a_{3,\sigma_5(3)} \\ +\operatorname{sgn}(\sigma_6)a_{1,\sigma_6(1)}a_{2,\sigma_6(2)}a_{3,\sigma_6(3)} \end{cases}$$

$$= \begin{cases} a_{1,1}a_{2,2}a_{3,3} \\ +a_{1,2}a_{2,3}a_{3,1} \\ +a_{1,3}a_{2,1}a_{3,2} \\ -a_{1,3}a_{2,2}a_{3,1} \\ -a_{1,1}a_{2,3}a_{3,2} \\ -a_{1,2}a_{2,1}a_{3,3}. \end{cases}$$

□

Determinants of 2×2 and 3×3 matrices can be represented schematically, as shown below.

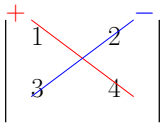




We multiply the entries along each of the red lines and add them up, and then we multiply the entries along each of the blue lines and subtract them. In each case, the result we get is precisely the formula from Proposition 7.1.2. For example, we can compute the determinant of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

in $\mathbb{R}^{2\times 2}$ by forming the diagram



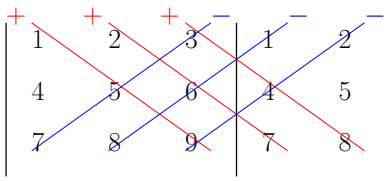
and the computing

$$\det(A) \quad = \quad \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} \quad = \quad 1 \cdot 4 - 2 \cdot 3 \quad = \quad -2.$$

Similarly, we can compute the determinant of the matrix

$$B \quad = \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

in $\mathbb{R}^{3\times 3}$ by forming the diagram



and then computing

$$\begin{aligned} \det(B) \quad &= \quad \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} \\ &= \quad 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 - 3 \cdot 5 \cdot 7 - 1 \cdot 6 \cdot 8 - 2 \cdot 4 \cdot 9 \\ &= \quad 0. \end{aligned}$$

Warning: Do not try this with matrices of larger size!

Theorem 7.1.3. *Let $\mathbb{F}$ be a field. For all $A \in \mathbb{F}^{n\times n}$, we have that*

$$det(A^T) \quad = \quad det(A).$$

Proof. We set $A = \left[\ a_{i,j}\ \right]_{n\times n}$ and $A^T = \left[\ a^T_{i,j}\ \right]_{n\times n}$. So, for all $i, j \in \{1, \ldots, n\}$, we have $a_{i,j} = a^T_{j,i}$. Now, we compute:

$$\det(A^T) \quad = \quad \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a^T_{i,\sigma(i)}$$

$$\begin{aligned}
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i} \\
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{j,\sigma^{-1}(j)} \\
&\overset{(*)}{=} \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma^{-1}) \prod_{j=1}^{n} a_{j,\sigma^{-1}(j)} \\
&= \sum_{\pi\in S_n} \operatorname{sgn}(\pi) \prod_{j=1}^{n} a_{j,\pi(j)} \\
&= \det(A),
\end{aligned}$$

where (*) follows from Proposition 2.3.2. □

7.1.1 Some matrices whose determinants are zero

Proposition 7.1.4. *Let $\mathbb{F}$ be a field, and let $A = \left[\ a_{i,j}\ \right]_{n\times n}$ be a matrix in $\mathbb{F}^{n\times n}$. If A has a zero row or a zero column,[3] then $\det(A) = 0$.*

Proof. In view of Theorem 7.1.3, it suffices to consider the case when A has a zero row.[4] Suppose that that the p-th row of A is a zero row. Then for all $\sigma \in S_n$, we have that $a_{p,\sigma(p)} = 0$. Consequently,

$$\det(A) \;=\; \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \dots a_{n,\sigma(n)} \;=\; 0,$$

which is what we needed to show. □

Proposition 7.1.5. *Let $\mathbb{F}$ be a field, and let $A = \left[\ a_{i,j}\ \right]_{n\times n}$ be a matrix in $\mathbb{F}^{n\times n}$. If A has two identical rows or two identical columns, then $\det(A) = 0$.*

Proof. In view of Theorem 7.1.3, it suffices to consider the case when A has two identical rows.[5] So, suppose that for some distinct $p, q \in \{1, \dots, n\}$, the p-th and

[3] A *zero row* is a row in which all entries are zero. Similarly, a *zero column* is a column in which all entries are zero.

[4] Let us explain this in more detail. Suppose we have shown that if a matrix in $\mathbb{F}^{n\times n}$ has a zero row, then its determinant is zero. Suppose now that B is a matrix in $\mathbb{F}^{n\times n}$ that has a zero column. Then B^T has a zero row, and we see that $\det(B) \overset{(*)}{=} \det(B^T) \overset{(**)}{=} 0$, where (*) follows from Theorem 7.1.3, and (**) follows from the fact that B^T has a zero row.

[5] Let us explain this in more detail. Suppose we have shown that if a matrix in $\mathbb{F}^{n\times n}$ has two identical rows, then its determinant is zero. Let us prove this for matrices with two identical columns. Suppose B is a matrix in $\mathbb{F}^{n\times n}$ with two identical columns. Then B^T has two identical rows, and we see that $\det(B) \overset{(*)}{=} \det(B^T) \overset{(**)}{=} 0$, where (*) follows from Theorem 7.1.3, and (**) follows from the fact that B^T has two identical rows.

q-th row of A are the same. (In particular, $n \geq 2$.) Now, let A_n be the alternating group of degree n, i.e. the group of all even permutations in S_n, and let O_n be the set of all odd permutations in S_n.[6] Obviously,

$$S_n \quad = \quad A_n \cup O_n \qquad \text{and} \qquad A_n \cap O_n \quad = \quad \emptyset.$$

Next, consider the transposition $\tau = (pq)$. By Proposition 2.3.2, for all $\sigma \in S_n$, we have that $\text{sgn}(\sigma \circ \tau) = -\text{sgn}(\sigma)$; it then readily follows that $O_n = \{\sigma \circ \tau \mid \sigma \in A_n\}$,[7] and obviously, for all distinct $\sigma_1, \sigma_2 \in A_n$, we have that $\sigma_1 \circ \tau \neq \sigma_2 \circ \tau$.[8]

Claim. For all $\sigma \in S_n$, we have that $\prod_{i=1}^n a_{i,\sigma(i)} = \prod_{i=1}^n a_{i,\sigma\circ\tau(i)}$.

Proof of the Claim. Fix $\sigma \in S_n$. First, note that

- $a_{p,\sigma(p)} = a_{p,\sigma\circ\tau(q)} \overset{(*)}{=} a_{q,\sigma\circ\tau(q)}$,
- $a_{q,\sigma(q)} = a_{q,\sigma\circ\tau(p)} \overset{(*)}{=} a_{p,\sigma\circ\tau(p)}$,

where both instances of (*) follow from the fact that the p-th and q-th row of A are the same. So, $a_{p,\sigma(p)}a_{q,\sigma(q)} = a_{p,\sigma\circ\tau(p)}a_{q,\sigma\circ\tau(q)}$. On the other hand, it is clear that for all $i \in \{1, \ldots, n\} \setminus \{p, q\}$, we have that $a_{i,\sigma(i)} = a_{i,\sigma\circ\tau(i)}$. It follows that $\prod_{i=1}^n a_{i,\sigma(i)} = \prod_{i=1}^n a_{i,\sigma\circ\tau(i)}$, which is what we needed to show. ♦

We now compute:

$$\begin{aligned} \det(A) \quad &= \quad \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1,\sigma(1)} \ldots a_{n,\sigma(n)} \\ &= \quad \sum_{\sigma \in A_n} \underbrace{\text{sgn}(\sigma)}_{=1} a_{1,\sigma(1)} \ldots a_{n,\sigma(n)} + \sum_{\pi \in O_n} \underbrace{\text{sgn}(\pi)}_{=-1} a_{1,\pi(1)} \ldots a_{n,\pi(n)} \end{aligned}$$

[6] Unlike A_n, O_n is not a group.

[7] Let us check this. First, suppose that $\sigma \in A_n$. Then

$$\text{sgn}(\sigma \circ \tau) \quad \overset{(*)}{=} \quad -\text{sgn}(\sigma) \quad \overset{(**)}{=} \quad -1,$$

where (*) follows from Proposition 2.3.2, and (**) follows from the fact that σ is even. So, $\sigma \circ \tau \in O_n$.

Conversely, suppose that $\pi \in O_n$. Set $\sigma := \pi \circ \tau$. Then

$$\text{sgn}(\sigma) \quad = \quad \text{sgn}(\pi \circ \tau) \quad \overset{(*)}{=} \quad -\text{sgn}(\pi) \quad \overset{(**)}{=} \quad 1,$$

where (*) follows from Proposition 2.3.2, and (**) follows from the fact that π is odd. So, $\sigma \in A_n$. But τ is a transposition, and consequently, $\tau^{-1} = \tau$. So, $\pi = \pi \circ \tau \circ \tau = \sigma \circ \tau$.

[8] This follows from the fact that τ is a bijection. So, if we had that $\sigma_1 \circ \tau = \sigma_2 \circ \tau$, then we would have that

$$\sigma_1 \quad = \quad \sigma_1 \circ \tau \circ \tau^{-1} \quad = \quad \sigma_2 \circ \tau \circ \tau^{-1} \quad = \quad \sigma_2.$$

$$
\begin{aligned}
&= \sum_{\sigma\in A_n} a_{1,\sigma(1)}\cdots a_{n,\sigma(n)} - \sum_{\pi\in O_n} a_{1,\pi(1)}\cdots a_{n,\pi(n)} \\
&= \sum_{\sigma\in A_n} a_{1,\sigma(1)}\cdots a_{n,\sigma(n)} - \sum_{\sigma\in A_n} a_{1,\sigma\circ\tau(1)}\cdots a_{n,\sigma\circ\tau(n)} \\
&\overset{(*)}{=} 0,
\end{aligned}
$$

where (*) follows from the Claim. □

7.2 The linearity of determinants in one row or one column

In general, for matrices $A, B \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is some field) and a scalar $\alpha \in \mathbb{F}$, we have that

$$\det(A+B) \not= \det(A)+\det(B) \qquad \text{and} \qquad \det(\alpha A) \not= \alpha\det(A).$$

We do, however, have the following proposition.

Proposition 7.2.1. *Let $\mathbb{F}$ be a field, and let $\mathbf{a}_1,\ldots,\mathbf{a}_{p-1},\mathbf{a}_{p+1},\ldots,\mathbf{a}_n \in \mathbb{F}^n$. Then both the following hold:*

(a) the function $f_{C_p}:\mathbb{F}^n\to\mathbb{F}$ given by

$$f_{C_p}(\mathbf{x}) = \det\Big(\begin{bmatrix}\mathbf{a}_1 & \ldots & \mathbf{a}_{p-1} & \mathbf{x} & \mathbf{a}_{p+1} & \ldots & \mathbf{a}_n\end{bmatrix}\Big)$$

for all $\mathbf{x}\in\mathbb{F}^n$ is linear;

(b) the function $f_{R_p}:\mathbb{F}^n\to\mathbb{F}$ given by

$$f_{R_p}(\mathbf{x}) = \det\Big(\begin{bmatrix}\mathbf{a}_1^T\\ \vdots \\ \mathbf{a}_{p-1}^T \\ \mathbf{x}^T \\ \mathbf{a}_{p+1}^T \\ \vdots \\ \mathbf{a}_n^T\end{bmatrix}\Big)$$

for all $\mathbf{x}\in\mathbb{F}^n$ is linear.

Remark: Before reading the proof, the reader might want to take a look at Example 7.2.2 (below), since it illustrates how Proposition 7.2.1 can be used in practice.

Proof. Clearly, (b) and Theorem 7.1.3 imply (a).[9] So, it suffices to prove (b).

We first set up some notation. For each index $i \in \{1, \ldots, n\} \setminus \{p\}$, we set $\mathbf{a}_i = \begin{bmatrix} a_{i,1} & \ldots & a_{i,n} \end{bmatrix}^T$, so that $\mathbf{a}_i^T = \begin{bmatrix} a_{i,1} & \ldots & a_{i,n} \end{bmatrix}$. Now, let us prove that f_{R_p} is linear.

1. Fix $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, and set $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \ldots & y_n \end{bmatrix}^T$. We compute:

$$f_{R_p}(\mathbf{x}+\mathbf{y}) = \det\left(\begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{p-1}^T \\ (\mathbf{x}+\mathbf{y})^T \\ \mathbf{a}_{p+1}^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}\right) = \begin{vmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \ldots & a_{p-1,n} \\ x_1+y_1 & \ldots & x_n+y_n \\ a_{p+1,1} & \ldots & a_{p+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \ldots & a_{n,n} \end{vmatrix}$$

$$= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{p-1,\sigma(p-1)} \Big(x_{\sigma(p)} + y_{\sigma(p)}\Big) a_{p+1,\sigma(p+1)} \cdots a_{n,\sigma(n)}$$

$$= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{p-1,\sigma(p-1)} x_{\sigma(p)} a_{p+1,\sigma(p+1)} \cdots a_{n,\sigma(n)} + \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{p-1,\sigma(p-1)} y_{\sigma(p)} a_{p+1,\sigma(p+1)} \cdots a_{n,\sigma(n)}$$

$$= \begin{vmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \ldots & a_{p-1,n} \\ x_1 & \ldots & x_n \\ a_{p+1,1} & \ldots & a_{p+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \ldots & a_{n,n} \end{vmatrix} + \begin{vmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \ldots & a_{p-1,n} \\ y_1 & \ldots & y_n \\ a_{p+1,1} & \ldots & a_{p+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \ldots & a_{n,n} \end{vmatrix}$$

$$= \det\left(\begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{p-1}^T \\ \mathbf{x}^T \\ \mathbf{a}_{p+1}^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}\right) + \det\left(\begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{p-1}^T \\ \mathbf{y}^T \\ \mathbf{a}_{p+1}^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}\right) = f_{R_p}(\mathbf{x}) + f_{R_p}(\mathbf{y}).$$

[9] Details?

2. Fix $\mathbf{x} \in \mathbb{F}^n$ and $\alpha \in \mathbb{F}$, and set $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$. We compute:

$$f_{R_p}(\alpha\mathbf{x}) = \det\left(\begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{p-1}^T \\ \alpha\mathbf{x}^T \\ \mathbf{a}_{p+1}^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}\right) = \begin{vmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \dots & a_{p-1,n} \\ \alpha x_1 & \dots & \alpha x_n \\ a_{p+1,1} & \dots & a_{p+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{vmatrix}$$

$$= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \dots a_{p-1,\sigma(p-1)} \left(\alpha x_{\sigma(p)}\right) a_{p+1,\sigma(p+1)} \dots a_{n,\sigma(n)}$$

$$= \alpha \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \dots a_{p-1,\sigma(p-1)} x_{\sigma(p)} a_{p+1,\sigma(p+1)} \dots a_{n,\sigma(n)}$$

$$= \alpha \begin{vmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \dots & a_{p-1,n} \\ x_1 & \dots & x_n \\ a_{p+1,1} & \dots & a_{p+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{vmatrix} = \alpha \det\left(\begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{p-1}^T \\ \mathbf{x}^T \\ \mathbf{a}_{p+1}^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}\right) = \alpha f_{R_p}(\mathbf{x}).$$

By 1. and 2., f_{R_p} is linear, i.e. (b) holds. □

Example 7.2.2. *By Proposition 7.2.1, we have the following (entries are understood to be in $\mathbb{R}$, and the row/column being manipulated is in red to facilitate reading):*

- $\begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 0 & 1 & 5 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & 4 \\ 0 & -2 & 5 \end{vmatrix} + \begin{vmatrix} 1 & 1 & 1 \\ 2 & 1 & 4 \\ 0 & 3 & 5 \end{vmatrix};$
- $\begin{vmatrix} 3 & 2 & 4 \\ 6 & -1 & 0 \\ -3 & 0 & 5 \end{vmatrix} = 3 \begin{vmatrix} 1 & 2 & 4 \\ 2 & -1 & 0 \\ -1 & 0 & 5 \end{vmatrix};$
- $\begin{vmatrix} 1 & 2 & 3 \\ 2 & 2 & 3 \\ 7 & 3 & -2 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 2 & 3 \\ 4 & 4 & -2 \end{vmatrix} + \begin{vmatrix} 1 & 2 & 3 \\ 2 & 2 & 3 \\ 3 & -1 & 0 \end{vmatrix};$

- $\begin{vmatrix} 2 & -2 & 4 \\ 1 & 0 & -2 \\ 2 & 1 & 4 \end{vmatrix} = 2 \begin{vmatrix} 1 & -1 & 2 \\ 1 & 0 & -2 \\ 2 & 1 & 4 \end{vmatrix}$.

As an easy corollary of Proposition 7.2.1, we obtain the following.

Proposition 7.2.3. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$, and let $\alpha \in \mathbb{F}$. Then*

$$det(\alpha A) \quad = \quad \alpha^n \; det(A).$$

Proof. We apply Proposition 7.2.1 n times, once to each row (or alternatively, once to each column) of αA, and the result follows.[10] □

7.3 Computing determinants via elementary row and column operations

In this section, we examine how performing elementary row and column operations affects the value of the determinant, and how we can use these operations to compute the determinant of a square matrix. We studied elementary row operations in chapter 1. Elementary column operations are defined completely analogously, only for columns instead of rows. Elementary column operations should **not** be used for solving linear systems. However, it turns out that both elementary row operations and elementary column operations behave well with respect to determinants, i.e. they change the value of the determinant in a way that we can describe precisely, as we shall see.

7.3.1 Computing the determinant of a triangular matrix

Given a square matrix $A = \left[\; a_{i,j} \; \right]_{n\times n}$ in $\mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is some field), we say that

- A is *upper triangular* if all entries of A below the main diagonal are zero, i.e. if for all $i, j \in \{1, \ldots, n\}$ such that $i > j$, we have that $a_{i,j} = 0$;
- A is *lower triangular* if all entries of A above the main diagonal are zero, i.e. if for all $i, j \in \{1, \ldots, n\}$ such that $i < j$, we have that $a_{i,j} = 0$;
- A is *triangular* if it is upper triangular or lower triangular.

A schematic representation of an upper triangular and a lower triangular matrix is given below (*'s represent arbitrary elements of the field $\mathbb{F}$, and the main diagonal is in red in both cases).

[10] Details?

$$\begin{bmatrix} * & * & * & \dots & * & * \\ 0 & * & * & \dots & * & * \\ 0 & 0 & * & \dots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & * & * \\ 0 & 0 & 0 & \dots & 0 & * \end{bmatrix} \qquad \begin{bmatrix} * & 0 & 0 & \dots & 0 & 0 \\ * & * & 0 & \dots & 0 & 0 \\ * & * & * & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & * & \dots & * & 0 \\ * & * & * & \dots & * & * \end{bmatrix}$$

upper triangular matrix lower triangular matrix

Note that any square matrix in row echelon form is in fact an upper triangular matrix. (However, not all upper triangular matrices are in row echelon form.) So, the row reduction algorithm performed on a square matrix will, in particular, yield an upper triangular matrix.

It turns out that the determinant of any triangular matrix is particularly easy to compute, as we now show.

Proposition 7.3.1. *Let $\mathbb{F}$ be a field, and let $A = \left[\ a_{i,j}\ \right]_{n\times n}$ be a triangular matrix in $\mathbb{F}^{n\times n}$. Then*

$$det(A) \quad = \quad \prod_{i=1}^{n} a_{i,i} \quad = \quad a_{1,1}a_{2,2}\dots a_{n,n},$$

that is, $det(A)$ is equal to the product of entries on the main diagonal of A.

Remark: For example, we can compute the determinants of the following matrices in $\mathbb{R}^{3\times 3}$ as follows:

- $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{vmatrix} = 1\cdot 4\cdot 6 = 24;$
- $\begin{vmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{vmatrix} = 1\cdot 3\cdot 6 = 18.$

Proof. Note that the transpose of an upper triangular matrix is a lower triangular matrix, and moreover, the main diagonal remains unchanged when we take the transpose of a square matrix. So, in view of Theorem 7.1.3, it suffices to prove the result for the case when A is lower triangular. Now, note that for all $\sigma \in S_n \setminus \{1\}$,[11] there exists some index $i \in \{1,\dots,n\}$ such that $i < \sigma(i)$,[12] and consequently, $a_{i,\sigma(i)} = 0$ (since A is lower triangular). It follows that for all $\sigma \in S_n \setminus \{1\}$, we have that $a_{1,\sigma(1)}a_{2,\sigma(2)}\dots a_{n,\sigma(n)} = 0$, and consequently,

[11] Recall that 1 is the identity permutation in S_n.

[12] This is "obvious," but here is a formal proof. Fix a permutation $\sigma \in S_n \setminus \{1\}$, and let $i \in \{1,\dots,n\}$ be minimal with the property that $\sigma(i) \neq i$. Set $j := \sigma(i)$. Then

- $\sigma(1) = 1, \dots, \sigma(i-1) = i-1$,
- $\sigma(i) = j \neq i$.

If $j < i$, then $\sigma(j) = j = \sigma(i)$, contrary to the fact that σ is a permutation (and in particular, one-to-one). So, $j > i$, i.e. $\sigma(i) > i$.

$$\begin{aligned}\det(A) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \dots a_{n,\sigma(n)} \\ &= \operatorname{sgn}(1) a_{1,1} a_{2,2} \dots a_{n,n} \\ &= a_{1,1} a_{2,2} \dots a_{n,n}.\end{aligned}$$

This completes the argument. □

7.3.2 Determinants and elementary row and column operations

Theorem 7.3.2. *Let $\mathbb{F}$ be a field, and let $A = \left[\ a_{i,j}\ \right]_{n\times n}$ be a matrix in $\mathbb{F}^{n\times n}$. Then all the following hold:*

(a) if a matrix B is obtained by swapping two rows or swapping two columns of A, then

$$det(B) = -det(A);$$

(b) if a matrix B is obtained by multiplying some row or some column of A by a scalar $\alpha \in \mathbb{F}\setminus\{0\}$, then

$$det(B) = \alpha det(A) \quad and \quad det(A) = \alpha^{-1} det(B);$$

(c) if a matrix B is obtained from A by adding a scalar multiple of one row (resp. column) of A to another row (resp. column) of A, then

$$det(B) = det(A).$$

Proof. In view of Theorem 7.1.3, it suffices to prove the result for row operations only.

(a) Fix distinct indices $p, q \in \{1, \dots, n\}$, and suppose that B is obtained by swapping rows p and q of A ("$R_p \leftrightarrow R_q$"). Set $B = \left[\ b_{i,j}\ \right]_{n\times n}$, so that

- for all $j \in \{1, \dots, n\}$, we have that $b_{p,j} = a_{q,j}$ and $b_{q,j} = a_{p,j}$;
- for all $i \in \{1, \dots, n\} \setminus \{p, q\}$ and $j \in \{1, \dots, n\}$, we have that $b_{i,j} = a_{i,j}$.

Next, consider the transposition $\tau = (pq)$ in S_n.

Claim. For all $\sigma \in S_n$, we have that $\prod_{i=1}^{n} b_{i,\sigma(i)} = \prod_{i=1}^{n} a_{i,\sigma\circ\tau(i)}$.

Proof of the Claim. First, we note that

- $b_{p,\sigma(p)} = a_{q,\sigma(p)} = a_{q,\sigma\circ\tau(q)}$;

- $b_{q,\sigma(q)} = a_{p,\sigma(q)} = a_{p,\sigma\circ\tau(p)}$.

So, $b_{p,\sigma(p)}b_{q,\sigma(q)} = a_{p,\sigma\circ\tau(p)}a_{q,\sigma\circ\tau(q)}$. On the other hand, for all $i \in \{1,\ldots,n\} \setminus \{p,q\}$, we have that $b_{i,\sigma(i)} = a_{i,\sigma\circ\tau(i)}$. It follows that $\prod_{i=1}^{n} b_{i,\sigma(i)} = \prod_{i=1}^{n} a_{i,\sigma\circ\tau(i)}$, which is what we needed to show. ♦

We now compute:

$$\begin{aligned}
\det(B) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \textstyle\prod_{i=1}^{n} b_{i,\sigma(i)} \\
&\overset{(*)}{=} \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \textstyle\prod_{i=1}^{n} a_{i,\sigma\circ\tau(i)} \\
&\overset{(**)}{=} \sum_{\sigma\in S_n} \Big(-\operatorname{sgn}(\sigma\circ\tau)\Big) \textstyle\prod_{i=1}^{n} a_{i,\sigma\circ\tau(i)} \\
&= -\sum_{\sigma\in S_n} \operatorname{sgn}(\sigma\circ\tau) \textstyle\prod_{i=1}^{n} a_{i,\sigma\circ\tau(i)} \\
&= -\sum_{\pi\in S_n} \operatorname{sgn}(\pi) \textstyle\prod_{i=1}^{n} a_{i,\pi(i)} \\
&= -\det(A),
\end{aligned}$$

where (*) follows from the Claim, and (**) follows from Proposition 2.3.2. This proves (a).

(b) Fix an index $p \in \{1,\ldots,n\}$ and a scalar $\alpha \in \mathbb{F} \setminus \{0\}$, and suppose that B is obtained by multiplying the p-th row of A by α ("$R_p \to \alpha R_p$"). Set $B = \left[\ b_{i,j}\ \right]_{n\times n}$, so that

- for all $j \in \{1,\ldots,n\}$, we have that $b_{p,j} = \alpha a_{p,j}$;
- for all $i \in \{1,\ldots,n\} \setminus \{p\}$ and $j \in \{1,\ldots,n\}$, we have that $b_{i,j} = a_{i,j}$.

We now compute:

$$\begin{aligned}
\det(B) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) b_{1,\sigma(1)} \ldots b_{n,\sigma(n)} \\
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \ldots a_{p-1,\sigma(p-1)} \Big(\alpha a_{p,\sigma(p)}\Big) a_{p+1,\sigma(p+1)} \ldots a_{n,\sigma(n)} \\
&= \alpha \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \ldots a_{n,\sigma(n)} \\
&= \alpha\det(A).
\end{aligned}$$

Since $\alpha \neq 0$, we deduce that $\det(A) = \alpha^{-1}\det(B)$. This proves (b).

(c) Fix distinct indices $p, q \in \{1, \ldots, n\}$ and a scalar $\alpha \in \mathbb{F}$, and suppose that B is obtained by adding α times row p to row q ("$R_q \to R_q + \alpha R_p$"). Set $B = \left[\ b_{i,j}\ \right]_{n\times n}$, so that

- for all $j \in \{1, \ldots, n\}$, we have that $b_{q,j} = a_{q,j} + \alpha a_{p,j}$;
- for all $i \in \{1, \ldots, n\} \setminus \{q\}$ and $j \in \{1, \ldots, n\}$, we have that $b_{i,j} = a_{i,j}$.

We now compute (the q-th row is in red for emphasis):

$$\begin{aligned}
\det(B) &= \begin{vmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{q-1,1} & \ldots & a_{q-1,n} \\ \color{red}{a_{q,1} + \alpha a_{p,1}} & \color{red}{\ldots} & \color{red}{a_{q,n} + \alpha a_{p,n}} \\ a_{q+1,1} & \ldots & a_{q+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \ldots & a_{n,n} \end{vmatrix} \\
&\overset{(*)}{=} \underbrace{\begin{vmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{q-1,1} & \ldots & a_{q-1,n} \\ \color{red}{a_{q,1}} & \color{red}{\ldots} & \color{red}{a_{q,n}} \\ a_{q+1,1} & \ldots & a_{q+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \ldots & a_{n,n} \end{vmatrix}}_{=\det(A)} + \alpha \underbrace{\begin{vmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{q-1,1} & \ldots & a_{q-1,n} \\ \color{red}{a_{p,1}} & \color{red}{\ldots} & \color{red}{a_{p,n}} \\ a_{q+1,1} & \ldots & a_{q+1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \ldots & a_{n,n} \end{vmatrix}}_{\overset{(**)}{=}0} \\
&= \det(A),
\end{aligned}$$

where (*) follows from the fact that the determinant is linear in the q-th row (by Proposition 7.2.1), and (**) follows from the fact that a matrix with two identical rows (in this case, the p-th and q-th row) has determinant zero (by Proposition 7.1.5). □

Example 7.3.3. *Compute the determinant of the matrix below (with entries understood to be in $\mathbb{R}$).*

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 2 & 4 & 4 \\ 3 & 3 & 7 \end{bmatrix}$$

Solution. We perform elementary row operations on A (keeping track of the way that this changes the value of the determinant, as per Theorem 7.3.2) until we

transform A into a matrix in row echelon form. Square matrices in row echelon form are upper triangular, and so by Proposition 7.3.1, we can obtain their determinant by multiplying the entries on the main diagonal. We now compute:

$$\begin{aligned}
\det(A) &= \begin{vmatrix} 2 & 4 & 6 \\ 2 & 4 & 4 \\ 3 & 3 & 7 \end{vmatrix} \\
&\overset{R_2 \to R_2 - R_1}{=} \begin{vmatrix} 2 & 4 & 6 \\ 0 & 0 & -2 \\ 3 & 3 & 7 \end{vmatrix} \\
&\overset{R_2 \leftrightarrow R_3}{=} - \begin{vmatrix} 2 & 4 & 6 \\ 3 & 3 & 7 \\ 0 & 0 & -2 \end{vmatrix} \\
&\overset{R_1 \to \frac{1}{2} R_1}{=} -2 \begin{vmatrix} 1 & 2 & 3 \\ 3 & 3 & 7 \\ 0 & 0 & -2 \end{vmatrix} \\
&\overset{R_2 \to R_1 - 3R_1}{=} -2 \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -2 \\ 0 & 0 & -2 \end{vmatrix} \\
&\overset{(*)}{=} (-2)1(-3)(-2) \\
&= -12,
\end{aligned}$$

where (*) follows by taking the determinant of an upper triangular matrix. □

Example 7.3.4. *Compute the determinant of the matrix below (with entries understood to be in $\mathbb{Z}_3$).*

$$A = \begin{bmatrix} 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 0 & 2 & 1 \\ 2 & 0 & 1 & 1 & 2 \\ 2 & 2 & 0 & 0 & 1 \\ 1 & 0 & 2 & 1 & 2 \end{bmatrix}$$

Solution. Here, we just notice that the second column is the sum of the first and third. This allows us to turn the second column into a zero column via two elementary column operations, which implies that $\det(A) = 0$. The detailed computation is as follows:

$$\begin{aligned}
\det(A) &= \begin{vmatrix} 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 0 & 2 & 1 \\ 2 & 0 & 1 & 1 & 2 \\ 2 & 2 & 0 & 0 & 1 \\ 1 & 0 & 2 & 1 & 2 \end{vmatrix} \\
&\overset{C_2 \to C_2 - C_1}{=} \begin{vmatrix} 1 & 1 & 1 & 1 & 2 \\ 1 & 0 & 0 & 2 & 1 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 0 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 2 \end{vmatrix} \\
&\overset{C_2 \to C_2 - C_3}{=} \begin{vmatrix} 1 & 0 & 1 & 1 & 2 \\ 1 & 0 & 0 & 2 & 1 \\ 2 & 0 & 1 & 1 & 2 \\ 2 & 0 & 0 & 0 & 1 \\ 1 & 0 & 2 & 1 & 2 \end{vmatrix} \\
&\overset{(*)}{=} 0,
\end{aligned}$$

where (*) follows from the fact that a matrix with a zero column has determinant zero (by Proposition 7.1.4). We could also have noticed that the matrix in the second line of the computation above has two identical columns, and so by Proposition 7.1.5, its determinant is zero. Finally, we note that our two elementary column operations could also have been written as "$C_2 \to C_2 + 2C_1$" and "$C_2 \to C_2 + 2C_3$," since in $\mathbb{Z}_3$, we have that $-1 = 2$. □

7.4 Determinants and matrix invertibility

Theorem 7.4.1. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$. Then A is invertible if and only if $\det(A) \neq 0$.*

Proof. We can transform A into a matrix in reduced row echelon form via a sequence of elementary row operations. By Theorem 7.3.2, each elementary row operation has the effect of multiplying the value of the determinant by some non-zero scalar. So, there exists some scalar $\alpha \in \mathbb{F} \setminus \{0\}$ such that $\det(A) = \alpha \det\big(\mathrm{RREF}(A)\big)$. Therefore, $\det(A) = 0$ if and only if $\det\big(\mathrm{RREF}(A)\big) = 0$. Moreover, $\mathrm{RREF}(A)$ is an upper triangular matrix, and so (by Proposition 7.3.1) its determinant is zero if and only if at least one entry on its main diagonal is zero. We now have the following sequence of equivalent statements:

$$\det(A) = 0 \iff \det\big(\mathrm{RREF}(A)\big) = 0$$

$$\iff \quad \text{at least one entry on the main diagonal of RREF}(A) \text{ is } 0$$

$$\overset{(*)}{\iff} \quad \text{RREF}(A) \neq I_n$$

$$\overset{(**)}{\iff} \quad A \text{ is not invertible,}$$

where (*) follows from the fact that RREF(A) is a square matrix in reduced row echelon form, and (**) follows from the Invertible Matrix Theorem (see subsection 1.11.7 or 3.3.6). It now obviously follows that A is invertible if and only if $\det(A) \neq 0$, which is what we needed to show. □

7.4.1 The Invertible Matrix Theorem (version 3)

In this subsection, we expand the previous version of the Invertible Matrix Theorem (see subsection 3.3.6) to include Theorem 7.4.1.

The Invertible Matrix Theorem (version 3). *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a **square** matrix. Further, let $f : \mathbb{F}^n \to \mathbb{F}^n$ be given by $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$.*[13] *Then the following are equivalent:*

(a) A is invertible (i.e. A has an inverse);

(b) A^T is invertible;

(c) RREF$(A) = I_n$;

(d) RREF$\left(\left[\begin{array}{c:c} A & I_n \end{array}\right]\right) = \left[\begin{array}{c:c} I_n & B \end{array}\right]$ for some matrix $B \in \mathbb{F}^{n\times n}$;

(e) rank$(A) = n$;

(f) rank$(A^T) = n$;

(g) A is a product of elementary matrices;

(h) the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$);

(i) there exists some vector $\mathbf{b} \in \mathbb{F}^n$ such that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;

(j) for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;

[13] Since f is a matrix transformation, Proposition 1.10.4 guarantees that f is linear. Moreover, A is the standard matrix of f.

(k) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *has at most one solution;*

(l) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *is consistent;*

(m) f *is one-to-one;*

(n) f *is onto;*

(o) f *is an isomorphism;*

(p) there exists a matrix $B \in \mathbb{F}^{n\times n}$ *such that* $BA = I_n$ *(i.e.* A *has a left inverse);*

(q) there exists a matrix $C \in \mathbb{F}^{n\times n}$ *such that* $AC = I_n$ *(i.e.* A *has a right inverse);*

(r) the columns of A *are linearly independent;*

(s) the columns of A *span* $\mathbb{F}^n$ *(i.e.* $Col(A) = \mathbb{F}^n$*);*

(t) the columns of A *form a basis of* $\mathbb{F}^n$*;*

(u) the rows of A *are linearly independent;*

(v) the rows of A *span* $\mathbb{F}^{1\times n}$ *(i.e.* $Row(A) = \mathbb{F}^{1\times n}$*);*

(w) the rows of A *form a basis of* $\mathbb{F}^{1\times n}$*;*

(x) $Nul(A) = \{\mathbf{0}\}$ *(i.e.* $dim\big(Nul(A)\big) = 0$*);*

(y) $det(A) \neq 0$*.*

Proof. Items (a)-(x) are the same as those from the Invertible Matrix Theorem (version 2) from subsection 3.3.6. The equivalence of (a) and (y) follows from Theorem 7.4.1. □

7.5 The multiplicative property of determinants

Suppose that $\mathbb{F}$ is some field. In general, for matrices $A, B \in \mathbb{F}^{n\times n}$ and a scalar $\alpha \in \mathbb{F}$, we have that

$$\det(A+B) \not= \det(A) + \det(B) \qquad \text{and} \qquad \det(\alpha A) \not= \alpha\det(A).$$

However, as we shall see (see Theorem 7.5.2 below), we do have that

$$\det(AB) = \det(A)\det(B).$$

We begin with a technical proposition. (Recall from subsection 1.11.5 that an *elementary matrix* is any matrix obtained by performing one elementary row operation on an identity matrix I_n.[14])

[14]Here, it is possible that $E = I_n$. In this case, we can take R to be the multiplication of the first row by the scalar 1.

Proposition 7.5.1. *Let $\mathbb{F}$ be a field, let $A, E \in \mathbb{F}^{n\times n}$, and assume that E is an elementary matrix. Then $det(EA) = det(E)det(A)$.*

Proof. Let R be an elementary row operation that corresponds to the elementary matrix E, i.e. E is the matrix obtained by performing R on I_n. By Proposition 1.11.11(a), EA is the matrix obtained by performing R on A. Now, by Theorem 7.3.2, there exists some scalar $\alpha \in \mathbb{F} \setminus \{0\}$ such that for any matrix $M \in \mathbb{F}^{n\times n}$, the determinant of the matrix obtained by performing the elementary row operation R on M is $\alpha\det(M)$. So,

- $\det(E) = \alpha\det(I_n) = \alpha$;
- $\det(EA) = \alpha\det(A)$.

It follows that

$$\det(EA) \quad = \quad \alpha\det(A) \quad = \quad \det(E)\det(A),$$

which is what we needed to show. □

We are now ready to prove the multiplicative property of determinants.

Theorem 7.5.2. *Let $\mathbb{F}$ be a field, and let $A, B \in \mathbb{F}^{n\times n}$. Then*

$$det(AB) \quad = \quad det(A)det(B).$$

Proof. Suppose first that at least one of A, B is non-invertible. Then by Corollary 3.3.16, AB is also non-invertible. But by Theorem 7.4.1, non-invertible matrices have determinant zero, and so $\det(AB) = 0 = \det(A)\det(B)$.[15]

From now on, we assume that A and B are both invertible. Therefore, by the Invertible Matrix Theorem (see subsection 1.11.7, 3.3.6, or 7.4.1), each of them can be written as a product of elementary matrices, say $A = E_1^A \dots E_p^A$ and $B = E_1^B \dots E_q^B$, where $E_1^A, \dots, E_p^A, E_1^B, \dots, E_q^B$ are elementary matrices. So, $AB = E_1^A \dots E_p^A E_1^B \dots E_q^B$. By repeatedly applying Proposition 7.5.1, we get that

- $\det(A) = \det(E_1^A) \dots \det(E_p^A)$;
- $\det(B) = \det(E_1^B) \dots \det(E_q^B)$;
- $\det(AB) = \det(E_1^A) \dots \det(E_p^A)\det(E_1^B) \dots \det(E_q^B)$.

But now

$$\det(AB) \quad = \quad \underbrace{\det(E_1^A) \dots \det(E_p^A)}_{=\det(A)} \underbrace{\det(E_1^B) \dots \det(E_q^B)}_{=\det(B)} \quad = \quad \det(A)\det(B),$$

which is what we needed to show. □

[15] If A is non-invertible, then $\det(A) = 0$, and if B is non-invertible, then $\det(B) = 0$. In either case, $\det(A)\det(B) = 0$.

Corollary 7.5.3. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be an invertible matrix. Then*

$$det(A^{-1}) \;=\; \tfrac{1}{det(A)}.$$

Proof. Since $AA^{-1} = I_n$, we see that

$$\det(A)\det(A^{-1}) \;\overset{(*)}{=}\; \det(AA^{-1}) \;=\; \det(I_n) \;=\; 1,$$

where (*) follows from Theorem 7.5.2. We now see that $\det(A^{-1}) = \frac{1}{\det(A)}$, which is what we needed to show. □

Similar matrices and determinants. Recall from subsection 4.5.2 that matrices $A, B \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) are said to be *similar* if there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$. Using Theorem 7.5.2 and Corollary 7.5.3, we can easily show that similar matrices have the same determinant.

Corollary 7.5.4. *Let $\mathbb{F}$ be a field, and let A and B be similar matrices in $\mathbb{F}^{n\times n}$. Then $det(A) = det(B)$.*

Proof. Since A and B are similar, there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$. We then have that

$$\begin{aligned}
\det(B) &= \det(P^{-1}AP) \\
&= \det(P^{-1})\det(A)\det(P) && \text{by Theorem 7.5.2} \\
&= \tfrac{1}{\det(P)}\det(A)\det(P) && \text{by Corollary 7.5.3} \\
&= \det(A),
\end{aligned}$$

which is what we needed to show. □

Determinants of (some) linear functions. Suppose that V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and that $f : V \to V$ is a linear function. Then we define the *determinant* of f to be

$$\det(f) \;:=\; \det\Big({}_{\mathcal{B}}\big[\; f \;\big]_{\mathcal{B}}\Big),$$

where $\mathcal{B}$ is any basis of V. Let us explain why this is well defined, that is, why the value of $\det(f)$ that we get depends only on f, and not on the particular choice of the basis $\mathcal{B}$. Suppose that $\mathcal{C}$ is any basis of V. Then by Theorem 4.5.16, matrices ${}_{\mathcal{B}}\big[\; f \;\big]_{\mathcal{B}}$ and ${}_{\mathcal{C}}\big[\; f \;\big]_{\mathcal{C}}$ are similar, and consequently (by Corollary 7.5.4), they have the same determinant. So, $\det(f)$ is well defined.

Remark: Note that we defined determinants only for linear functions whose domain and codomain are one and the same, and moreover, are finite-dimensional and non-null.

Determinants of orthogonal matrices. As another corollary of Theorem 7.5.2, we obtain the following.

Corollary 7.5.5. *Let A be an orthogonal matrix in $\mathbb{R}^{n\times n}$. Then $det(A) = \pm 1$ (i.e. $det(A)$ is either $+1$ or -1).*

Proof. Since A is orthogonal, it satisfies $A^T A = I_n$ (by definition). Therefore,

$$1 \;=\; \det(I_n) \;=\; \det(A^T A) \;\overset{(*)}{=}\; \det(A^T)\det(A) \;\overset{(**)}{=}\; \det(A)^2,$$

where (*) follows from Theorem 7.5.2, and (**) follows from Theorem 7.1.3. But now we see that $\det(A) = \pm 1$, which is what we needed to show. □

Warning: The converse of Corollary 7.5.5 is false, i.e. matrices whose determinant is ± 1 need not be orthogonal. For example, the matrix $A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}$ satisfies $\det(A) = 1$, but A is not orthogonal. More generally, suppose that A is any invertible matrix in $\mathbb{R}^{n\times n}$. Then by Theorem 7.4.1, we have that $\det(A) \neq 0$. We now form the matrix B by multiplying one row or one column of A by $\frac{1}{\det(A)}$, and we see that $\det(B) = 1$. However, B need not be orthogonal.

7.6 Laplace expansion

Our goal in this section is to prove a formula ("Laplace explansion") for the determinant of a square matrix in terms of determinants of smaller matrices. We begin with a technical proposition.

Proposition 7.6.1. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{(n-1)\times(n-1)}$ (where $n \geq 2$) and $\mathbf{a} \in \mathbb{F}^{n-1}$. Then*

$$det\Big(\left[\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{a}^T & 1 \end{array}\right]_{n\times n} \Big) \;=\; det(A).$$

Proof. First, set $\left[\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{a}^T & 1 \end{array}\right]_{n\times n} = \left[\; a_{i,j} \;\right]_{n\times n}$, so that all the following hold:

- $A = \left[\; a_{i,j} \;\right]_{(n-1)\times(n-1)}$;
- $a_{n,n} = 1$;
- for all $i \in \{1, \ldots, n-1\}$, $a_{i,n} = 0$;

- for all $j \in \{1, \ldots, n-1\}$, $a_{n,j}$ is the j-th entry of the vector $\mathbf{a}$.

Next, for all $\sigma \in S_{n-1}$, let $\sigma^* \in S_n$ be given by $\sigma^*(i) = \sigma(i)$ for all $i \in \{1, \ldots, n-1\}$ and $\sigma^*(n) = n$. So, for any $\sigma \in S_{n-1}$, the disjoint cycle decomposition of σ^* is obtained by adding the one-element cycle (n) to the disjoint cycle decomposition of σ, and consequently, $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\sigma^*)$.[16] Set $S_n^* := \{\sigma^* \mid \sigma \in S_{n-1}\} = \{\pi \in S_n \mid \pi(n) = n\}$. We then have the following:

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in S_{n-1}} \mathrm{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n-1,\sigma(n-1)} \\
&= \sum_{\sigma \in S_{n-1}} \mathrm{sgn}(\sigma^*) a_{1,\sigma^*(1)} \cdots a_{n-1,\sigma^*(n-1)} \underbrace{a_{n,\sigma^*(n)}}_{=1} \\
&= \sum_{\pi \in S_n^*} \mathrm{sgn}(\pi) a_{1,\pi(1)} \cdots a_{n-1,\pi(n-1)} a_{n,\pi(n)} \\
&\overset{(*)}{=} \sum_{\pi \in S_n} \mathrm{sgn}(\pi) a_{1,\pi(1)} \cdots a_{n-1,\pi(n-1)} a_{n,\pi(n)} \\
&= \det\Big(\left[\begin{array}{c:c} A & \mathbf{0} \\ \hdashline \mathbf{a}^T & 1 \end{array}\right]_{n \times n} \Big),
\end{aligned}$$

where (*) follows from the fact that for all $\pi \in S_n \setminus S_n^*$, we have that $a_{n,\pi(n)} = 0$. $\square$

We now introduce some terminology and notation. For a matrix $A = \left[\ a_{i,j}\ \right]_{n \times n}$ (where $n \geq 2$) with entries in some field $\mathbb{F}$, and for indices $p, q \in \{1, \ldots, n\}$, $A_{p,q}$ is the $(n-1) \times (n-1)$ matrix obtained from A by deleting the p-th row and q-th column (see Example 7.6.2 below). The determinants

$$\det(A_{i,j}), \qquad \text{with } i, j \in \{1, \ldots, n\}$$

are referred to as the *first minors* of A, whereas numbers

$$C_{i,j} := (-1)^{i+j} \det(A_{i,j}) \qquad \text{with } i, j \in \{1, \ldots, n\}$$

are referred to as the *cofactors* of A.

Example 7.6.2. *For the matrix*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix},$$

we have the following:

[16] Indeed, fix any $\sigma \in S_{n-1}$, and suppose that the disjoint cycle decomposition of σ contains exactly k cycles (when cycles of length one are included). Then the disjoint cycle decomposition of σ^* contains exactly $k+1$ cycles (when cycles of length one are included). But now $\mathrm{sgn}(\sigma) = (-1)^{(n-1)-k} = (-1)^{n-k-1} = (-1)^{n-(k+1)} = \mathrm{sgn}(\sigma^*)$.

- $A_{1,1} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}$;
- $A_{1,2} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}$;
- $A_{1,3} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}$;
- $A_{2,1} = \begin{bmatrix} 2 & 3 \\ 8 & 9 \end{bmatrix}$;
- $A_{2,2} = \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix}$;
- $A_{2,3} = \begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}$;
- $A_{3,1} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}$;
- $A_{3,2} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$;
- $A_{3,3} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$.

We now prove a recursive formula for computing determinants in terms of minors or cofactors. It allows us to compute the determinant of a square matrix in terms of determinants of smaller square matrices. This formula is called "Laplace expansion" or "cofactor expansion."

Laplace expansion. *Let $\mathbb{F}$ be a field, and let $A = \left[\ a_{i,j}\ \right]_{n\times n}$ (where $n \geq 2$) be a matrix in $\mathbb{F}^{n\times n}$. Then both the following hold:*

(a) ***[expansion along the i-th row]*** *for all $i \in \{1, \ldots, n\}$, we have that*

$$det(A) \quad = \quad \sum_{j=1}^{n} (-1)^{i+j}\ a_{i,j}\ det(A_{i,j});$$

(b) ***[expansion along the j-th column]*** *for all $j \in \{1, \ldots, n\}$, we have that*

$$det(A) \quad = \quad \sum_{i=1}^{n} (-1)^{i+j}\ a_{i,j}\ det(A_{i,j}).$$

Remark: If we write $C_{i,j} := (-1)^{i+j}\det(A_{i,j})$ for all $i, j \in \{1, \ldots, n\}$ (so, the $C_{i,j}$'s are the cofactors of A), then the formula from (a) becomes $\det(A) = \sum_{j=1}^{n} a_{i,j}C_{i,j}$, and the formula from (b) becomes $\det(A) = \sum_{i=1}^{n} a_{i,j}C_{i,j}$. This is why Laplace expansion is also referred to as "cofactor expansion."

Proof. In view of Theorem 7.1.3, it is enough to prove (b). Fix $j \in \{1, \ldots, n\}$. We must show that

$$\det(A) \quad = \quad \sum_{i=1}^{n} (-1)^{i+j}\ a_{i,j}\ \det(A_{i,j}).$$

First, set $A = \left[\ \mathbf{a}_1 \quad \ldots \quad \mathbf{a}_n\ \right]$. Then $\mathbf{a}_j = \sum_{i=1}^{n} a_{i,j}\mathbf{e}_i$, and so

$$\det(A) \quad = \quad \det\Big(\left[\ \mathbf{a}_1 \quad \ldots \quad \mathbf{a}_{j-1} \quad \mathbf{a}_j \quad \mathbf{a}_{j+1} \quad \ldots \quad \mathbf{a}_n\ \right]\Big)$$

$$= \det\Big(\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \sum_{i=1}^{n} a_{i,j}\mathbf{e}_i & \mathbf{a}_{j+1} & \dots & \mathbf{a}_n \end{bmatrix} \Big)$$

$$\overset{(*)}{=} \sum_{i=1}^{n} a_{i,j} \det\Big(\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \mathbf{e}_i & \mathbf{a}_{j+1} & \dots & \mathbf{a}_n \end{bmatrix} \Big),$$

where (*) follows from Proposition 7.2.1(a). Fix an arbitrary index $i \in \{1, \dots, n\}$. To complete the proof, it now suffices to show that

$$\det\Big(\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \mathbf{e}_i & \mathbf{a}_{j+1} & \dots & \mathbf{a}_n \end{bmatrix} \Big) = (-1)^{i+j} \det(A_{i,j}).$$

By iteratively performing $n - j$ column swaps on the matrix

$$B_i := \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \mathbf{e}_i & \mathbf{a}_{j+1} & \dots & \mathbf{a}_n \end{bmatrix},$$

we can obtain the matrix

$$C_i := \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \mathbf{a}_{j+1} & \dots & \mathbf{a}_n & \mathbf{e}_i \end{bmatrix}.$$

By iteratively performing $n - i$ row swaps on the matrix C_i, we can obtain the matrix

$$\left[\begin{array}{c|c} A_{i,j} & \mathbf{0} \\ \hline \mathbf{a}^T & 1 \end{array}\right],$$

where $\mathbf{a}^T$ is the row vector of length $n-1$ obtained from the i-th row of A by deleting its j-th entry.[17] Since swapping two rows or two columns has the effect of changing the sign of the determinant, we see that

$$\begin{aligned} \det(B_i) &= (-1)^{n-j} \det(C_i) \\ &= (-1)^{n-j}(-1)^{n-i} \det\Big(\left[\begin{array}{c|c} A_{i,j} & \mathbf{0} \\ \hline \mathbf{a}^T & 1 \end{array}\right] \Big) \\ &\overset{(*)}{=} (-1)^{2n-i-j} \det(A_{i,j}) \\ &= (-1)^{i+j} \det(A_{i,j}), \end{aligned}$$

where (*) follows from Proposition 7.6.1. This completes the argument. □

[17] So, $\mathbf{a}^T = \begin{bmatrix} a_{i,1} & \dots & a_{i,j-1} & a_{i,j+1} & \dots & a_{i,n} \end{bmatrix}^T$, and in particular $\mathbf{a}^T \in \mathbb{F}^{n-1}$.

Example 7.6.3. *Consider the matrix*

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 3 & 4 & 5 \\ 7 & 0 & 8 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$*. Compute* $det(A)$ *in two ways:*

(a) via Laplace expansion along the third row;

(b) via Laplace expansion along the second column.

Solution. (a) We compute:

$$\begin{aligned} \det(A) &= \begin{vmatrix} 2 & 0 & 1 \\ 3 & 4 & 5 \\ 7 & 0 & 8 \end{vmatrix} \\ &= (-1)^{3+1}\, 7 \begin{vmatrix} 0 & 1 \\ 4 & 5 \end{vmatrix} + (-1)^{3+2}\, 0 \begin{vmatrix} 2 & 1 \\ 3 & 5 \end{vmatrix} + (-1)^{3+3}\, 8 \begin{vmatrix} 2 & 0 \\ 3 & 4 \end{vmatrix} \\ &= 7 \underbrace{\begin{vmatrix} 0 & 1 \\ 4 & 5 \end{vmatrix}}_{=-4} + 8 \underbrace{\begin{vmatrix} 2 & 0 \\ 3 & 4 \end{vmatrix}}_{=8} = 36. \end{aligned}$$

(b) We compute:

$$\begin{aligned} \det(A) &= \begin{vmatrix} 2 & 0 & 1 \\ 3 & 4 & 5 \\ 7 & 0 & 8 \end{vmatrix} \\ &= (-1)^{1+2}\, 0 \begin{vmatrix} 3 & 5 \\ 7 & 8 \end{vmatrix} + (-1)^{2+2}\, 4 \begin{vmatrix} 2 & 1 \\ 7 & 8 \end{vmatrix} + (-1)^{3+2}\, 0 \begin{vmatrix} 2 & 1 \\ 3 & 5 \end{vmatrix} \\ &= 4 \underbrace{\begin{vmatrix} 2 & 1 \\ 7 & 8 \end{vmatrix}}_{=9} = 36. \end{aligned}$$

□

Example 7.6.4. *Compute the determinant of the matrix*

$$A = \begin{bmatrix} 1 & 2 & 0 & -1 & -2 \\ 3 & -4 & 0 & -2 & -1 \\ 1 & 2 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 & 2 \\ 2 & -1 & 0 & 1 & 3 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$.

Solution. As a general rule, it is best to expand along a row or column that has a lot of zeros (if such a row or column exists), since that reduces the amount of calculation that we need to perform. In the calculation below, the row or column along which we are about to expand is in red (to facilitate reading).

$$\det(A) = \begin{vmatrix} 1 & 2 & 0 & -1 & -2 \\ 3 & -4 & 0 & -2 & -1 \\ 1 & 2 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 & 2 \\ 2 & -1 & 0 & 1 & 3 \end{vmatrix} = (-1)^{3+3}\, 2 \begin{vmatrix} 1 & 2 & -1 & -2 \\ 3 & -4 & -2 & -1 \\ 1 & 0 & 0 & 2 \\ 2 & -1 & 1 & 3 \end{vmatrix}$$

$$= 2 \begin{vmatrix} 1 & 2 & -1 & -2 \\ 3 & -4 & -2 & -1 \\ 1 & 0 & 0 & 2 \\ 2 & -1 & 1 & 3 \end{vmatrix}$$

$$= 2\left((-1)^{3+1}\, 1 \begin{vmatrix} 2 & -1 & -2 \\ -4 & -2 & -1 \\ -1 & 1 & 3 \end{vmatrix} + (-1)^{3+4}\, 2 \begin{vmatrix} 1 & 2 & -1 \\ 3 & -4 & -2 \\ 2 & -1 & 1 \end{vmatrix} \right)$$

$$= 2\Big(\underbrace{\begin{vmatrix} 2 & -1 & -2 \\ -4 & -2 & -1 \\ -1 & 1 & 3 \end{vmatrix}}_{=-11} - 2 \underbrace{\begin{vmatrix} 1 & 2 & -1 \\ 3 & -4 & -2 \\ 2 & -1 & 1 \end{vmatrix}}_{-25} \Big) = 78,$$

where the determinants of the two 3×3 matrices from the last line can be obtained in various ways: Laplace expansion, elementary row/column operations, or our diagram for computing determinants of 3×3 matrices (described section 7.1). □

Example 7.6.5. *Compute the determinant of the matrix*

$$A = \begin{bmatrix} 1 & -1 & 2 \\ -2 & 4 & 1 \\ 3 & -3 & 5 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$.

Solution. We combine various methods for computing determinants, as follows:

$$\begin{aligned}
\det(A) &= \begin{vmatrix} 1 & -1 & 2 \\ -2 & 4 & 1 \\ 3 & -3 & 5 \end{vmatrix} \\
&\overset{C_2 \to C_2 + C_1}{=} \begin{vmatrix} 1 & 0 & 2 \\ -2 & 2 & 1 \\ 3 & 0 & 5 \end{vmatrix} \\
&= (-1)^{2+2}\, 2 \underbrace{\begin{vmatrix} 1 & 2 \\ 3 & 5 \end{vmatrix}}_{=-1} && \text{Laplace expansion along 2nd column} \\
&= -2.
\end{aligned}$$

□

Using Laplace expansion, one can easily show (see Theorem 7.6.6 and Corollary 7.6.7 below) that the determinant of a matrix obtained by arranging several square matrices along the main diagonal and placing zeros everywhere else is equal to the product of the determinants of those square matrices along the main diagonal.

Theorem 7.6.6. *Let* $\mathbb{F}$ *be a field, and let* $A \in \mathbb{F}^{n \times n}$ *and* $B \in \mathbb{F}^{m \times m}$ *be square matrices. Then*

$$det\Big(\begin{bmatrix} A & O_{n \times m} \\ O_{m \times n} & B \end{bmatrix} \Big) = det(A)\ det(B).$$

Proof (outilne). This can be proven (for example) by induction on n, via Laplace expansion along the leftmost column. The details are left as an exercise. □

Corollary 7.6.7. *Let* $\mathbb{F}$ *be a field, and let* $A_1 \in \mathbb{F}^{n_1 \times n_1}, A_2 \in \mathbb{F}^{n_2 \times n_2}, \ldots, A_k \in \mathbb{F}^{n_k \times n_k}$ *be square matrices. Then*

$$det\Big(\begin{bmatrix} A_1 & O_{n_1 \times n_2} & \cdots & O_{n_1 \times n_k} \\ O_{n_2 \times n_1} & A_2 & \cdots & O_{n_2 \times n_k} \\ \vdots & \vdots & \ddots & \vdots \\ O_{n_k \times n_1} & O_{n_k \times n_2} & \cdots & A_k \end{bmatrix} \Big) = \prod_{i=1}^{k} det(A_i).$$

Proof. This follows from Theorem 7.6.6 via an easy induction on k. □

7.7 Cramer's rule

Before stating Cramer's rule, we set up some notation. For a matrix $A \in \mathbb{F}^{n\times n}$, a vector $\mathbf{b} \in \mathbb{F}^n$, and an index $j \in \{1, \ldots, n\}$, we denote by $A_j(\mathbf{b})$ the matrix obtained from A by replacing the j-th column of A with $\mathbf{b}$. For example, for

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix},$$

we have that

$$A_1(\mathbf{b}) = \begin{bmatrix} 4 & 1 & 1 \\ 5 & 2 & 2 \\ 6 & 0 & 3 \end{bmatrix}, \quad A_2(\mathbf{b}) = \begin{bmatrix} 1 & 4 & 1 \\ 0 & 5 & 2 \\ 0 & 6 & 3 \end{bmatrix}, \quad A_3(\mathbf{b}) = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2 & 5 \\ 0 & 0 & 6 \end{bmatrix}.$$

Further, in the remainder of this section, it will be convenient to use the fraction notation discussed in subsection 2.4.3.

Cramer's rule. *Let $\mathbb{F}$ be a field, and let A be an invertible matrix in $\mathbb{F}^{n\times n}$, and let $\mathbf{b} \in \mathbb{F}^n$. Then the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution, namely*

$$\mathbf{x} = \begin{bmatrix} \frac{det\big(A_1(\mathbf{b})\big)}{det(A)} & \frac{det\big(A_2(\mathbf{b})\big)}{det(A)} & \cdots & \frac{det\big(A_n(\mathbf{b})\big)}{det(A)} \end{bmatrix}^T.$$

Proof. Since A is invertible, we know that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution, namely, $\mathbf{x} = A^{-1}\mathbf{b}$. Now, for this solution $\mathbf{x}$, we set $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$. Our goal is to show that

$$\mathbf{x} = \begin{bmatrix} \frac{\det\big(A_1(\mathbf{b})\big)}{\det(A)} & \frac{\det\big(A_2(\mathbf{b})\big)}{\det(A)} & \cdots & \frac{\det\big(A_n(\mathbf{b})\big)}{\det(A)} \end{bmatrix}^T.$$

Fix an index $j \in \{1, \ldots, n\}$. We must show that

$$x_j = \frac{\det\big(A_j(\mathbf{b})\big)}{\det(A)}.$$

Set $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$. Using the fact that $A\mathbf{x} = \mathbf{b}$ and the fact that $A\mathbf{x} = \sum_{i=1}^n x_i\mathbf{a}_i$ (by the definition of matrix-vector multiplication), we compute:

$$\begin{aligned} \det(A_j(\mathbf{b})) &= \det\Big(\begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_{j-1} & \mathbf{b} & \mathbf{a}_{j+1} & \ldots \mathbf{a}_n \end{bmatrix} \Big) \\ &= \det\Big(\begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_{j-1} & A\mathbf{x} & \mathbf{a}_{j+1} & \ldots \mathbf{a}_n \end{bmatrix} \Big) \\ &= \det\Big(\begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_{j-1} & \sum_{i=1}^n x_i\mathbf{a}_i & \mathbf{a}_{j+1} & \ldots \mathbf{a}_n \end{bmatrix} \Big) \end{aligned}$$

$$\begin{aligned} &\overset{(*)}{=} \quad \sum_{i=1}^{n} x_i \det\Big(\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \mathbf{a}_i & \mathbf{a}_{j+1} & \dots \mathbf{a}_n \end{bmatrix} \Big) \\ &\overset{(**)}{=} \quad x_j \det\Big(\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \mathbf{a}_j & \mathbf{a}_{j+1} & \dots \mathbf{a}_n \end{bmatrix} \Big) \\ &= \quad x_j \det(A), \end{aligned}$$

where (*) follows from Proposition 7.2.1(a), and (**) follows from the fact that for all $i \in \{1, \dots, n\} \setminus \{j\}$, the matrix

$$\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{j-1} & \mathbf{a}_i & \mathbf{a}_{j+1} & \dots \mathbf{a}_n \end{bmatrix}$$

has two identical columns and therefore (by Proposition 7.1.5) has determinant zero. We have now shown that

$$\det\big(A_j(\mathbf{b})\big) \quad = \quad x_j \det(A).$$

Since A is invertible, Theorem 7.4.1 guarantees that $\det(A) \neq 0$. So, we can divide both sides of the equality above by $\det(A)$ to obtain

$$x_j \quad = \quad \frac{\det\big(A_j(\mathbf{b})\big)}{\det(A)}.$$

This completes the argument. □

Example 7.7.1. *Let*

$$A \quad = \quad \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \qquad \textit{and} \qquad \mathbf{b} \quad = \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

with entries understood to be in $\mathbb{Z}_3$. *Solve the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$.

Solution. Note that $\det(A) = 2$, and in particular, A is invertible (by Theorem 7.4.1). So, Cramer's rule applies. We compute:

- $\det\big(A_1(\mathbf{b})\big) = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 2 & 2 \\ 0 & 1 & 1 \end{vmatrix} = 2;$
- $\det\big(A_2(\mathbf{b})\big) = \begin{vmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{vmatrix} = 1;$
- $\det\big(A_3(\mathbf{b})\big) = \begin{vmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 0.$

By Cramer's rule, $A\mathbf{x} = \mathbf{b}$ has a unique solution, namely

$$\begin{aligned}\mathbf{x} &= \left[\begin{array}{ccc} \frac{\det\left(A_1(\mathbf{b})\right)}{\det(A)} & \frac{\det\left(A_2(\mathbf{b})\right)}{\det(A)} & \frac{\det\left(A_3(\mathbf{b})\right)}{\det(A)} \end{array}\right]^T \\ &= \left[\begin{array}{ccc} \frac{2}{2} & \frac{1}{2} & \frac{0}{2} \end{array}\right]^T \\ &= \left[\begin{array}{ccc} 1 & 2 & 0 \end{array}\right]^T.\end{aligned}$$

□

7.8 The adjugate matrix

Given a field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n\times n}$ ($n \geq 2$), with cofactors $C_{i,j} = (-1)^{i+j}\det(A_{i,j})$ (for $i, j \in \{1, \ldots, n\}$), the *cofactor matrix* of A is the matrix $\left[\begin{array}{c} C_{i,j} \end{array}\right]_{n\times n}$. The *adjugate matrix* (also called the *classical adjoint*) of A, denoted by $\mathrm{adj}(A)$, is the transponse of the cofactor matrix of A, i.e.

$$\mathrm{adj}(A) := \left(\left[\begin{array}{c} C_{i,j} \end{array}\right]_{n\times n}\right)^T.$$

So, the i, j-th entry of $\mathrm{adj}(A)$ is the cofactor $C_{j,i}$ (note the swapping of the indices).

Example 7.8.1. *Consider the matrix*

$$A = \left[\begin{array}{ccc} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{array}\right],$$

with entries understood to be in $\mathbb{R}$. *Compute the cofactor and adjugate matrices of the matrix* A.

Solution. For all $i, j \in \{1, 2, 3\}$, we let $C_{i,j} = (-1)^{i+j}\det(A_{i,j})$. We compute:

- $C_{1,1} = (-1)^{1+1}\left|\begin{array}{cc} 2 & 2 \\ 0 & 3 \end{array}\right| = 6$;
- $C_{1,2} = (-1)^{1+2}\left|\begin{array}{cc} 0 & 2 \\ 0 & 3 \end{array}\right| = 0$;
- $C_{1,3} = (-1)^{1+3}\left|\begin{array}{cc} 0 & 2 \\ 0 & 0 \end{array}\right| = 0$;
- $C_{2,1} = (-1)^{2+1}\left|\begin{array}{cc} 1 & 1 \\ 0 & 3 \end{array}\right| = -3$;

- $C_{2,2} = (-1)^{2+2} \begin{vmatrix} 1 & 1 \\ 0 & 3 \end{vmatrix} = 3;$
- $C_{2,3} = (-1)^{2+3} \begin{vmatrix} 1 & 1 \\ 0 & 0 \end{vmatrix} = 0;$
- $C_{3,1} = (-1)^{3+1} \begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} = 0;$
- $C_{3,2} = (-1)^{3+2} \begin{vmatrix} 1 & 1 \\ 0 & 2 \end{vmatrix} = -2;$
- $C_{3,3} = (-1)^{3+3} \begin{vmatrix} 1 & 1 \\ 0 & 2 \end{vmatrix} = 2.$

So, the cofactor matrix of A is

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,1} & C_{3,2} & C_{3,3} \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ -3 & 3 & 0 \\ 0 & -2 & 2 \end{bmatrix}.$$

The adjugate matrix of A is the transpose of the cofactor matrix, i.e.

$$\text{adj}(A) = \begin{bmatrix} 6 & -3 & 0 \\ 0 & 3 & -2 \\ 0 & 0 & 2 \end{bmatrix}.$$

□

Theorem 7.8.2. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$ ($n \geq 2$). Then*

$$adj(A)\ A = A\ adj(A) = det(A) I_n.$$

Consequently, if A is invertible, then $A^{-1} = \frac{1}{det(A)} adj(A)$.

Proof. Let us first show that the first statement implies the second. Indeed, if A is invertible, then $\det(A) \neq 0$, and so if the first statement holds, then we get that

$$\left(\tfrac{1}{\det(A)}\text{adj}(A)\right) A = A \left(\tfrac{1}{\det(A)}\text{adj}(A)\right) = I_n,$$

and consequently, $A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$.

It remains to prove the first statement, i.e. that $\text{adj}(A)\ A = A\ \text{adj}(A) = \det(A) I_n$. We will do this by proving that the matrices $\text{adj}(A)\ A$, $A\ \text{adj}(A)$, and $\det(A) I_n$ have the same corresponding entries. Fix indices $i, j \in \{1, \ldots, n\}$. The i,j-th entry of the matrix $\det(A) I_n$ is $\det(A)$ if $i = j$, and is zero if $i \neq j$. We must show this holds for the i,j-th entry of the matrices $\text{adj}(A)\ A$ and $A\ \text{adj}(A)$ as well.

We first consider the matrix $\text{adj}(A)\ A$. The i-th row of $\text{adj}(A)$ is $\begin{bmatrix} C_{1,i} & \dots & C_{n,i} \end{bmatrix}$, and the j-th column of A is $\begin{bmatrix} a_{1,j} & \dots & a_{n,j} \end{bmatrix}^T$. So, the i,j-th entry of $\text{adj}(A)\ A$ is $\sum_{k=1}^{n} a_{k,j}C_{k,i}$. Now, let B_1 be the matrix obtained by replacing the i-th column of A by the j-th column of A. Then $\det(B_1) = \sum_{k=1}^{n} a_{k,j}C_{k,i}$ (via Laplace expansion along the i-th column of B_1). But if $i = j$, then $\det(B_1) = \det(A)$ (because $B_1 = A$), and if $i \neq j$, then $\det(B_1) = 0$ (because B_1 has two identical columns, namely, the i-th and j-th column).[18]

We now consider the matrix $A\ \text{adj}(A)$. The i-th row of A is $\begin{bmatrix} a_{i,1} & \dots & a_{i,n} \end{bmatrix}$, and the j-th column of $\text{adj}(A)$ is $\begin{bmatrix} C_{j,1} & \dots & C_{j,n} \end{bmatrix}^T$. So, the i,j-th entry of $A\ \text{adj}(A)$ is $\sum_{k=1}^{n} a_{i,k}C_{j,k}$. Now, let B_2 be the matrix obtained by replacing the j-th row of A by the i-th row of A. Then $\det(B_2) = \sum_{k=1}^{n} a_{i,k}C_{j,k}$ (via Laplace expansion along the j-th row of B_2). But if $i = j$, then $\det(B_2) = \det(A)$ (because $B_2 = A$), and if $i \neq j$, then $\det(B_2) = 0$ (because B_2 has two identical rows, namely, the i-th and j-th row).[19] □

Example 7.8.3. *Show that the matrix*

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix},$$

(with entries understood to be in $\mathbb{R}$) is invertible, and using Theorem 7.8.5, find its inverse A^{-1}.

Solution. The matrix A is upper triangular, and so its determinant can be computed by multiplying the entries along the main diagonal. So, $\det(A) = 1 \cdot 2 \cdot 3 = 6$. Since $\det(A) \neq 0$, Theorem 7.4.1 guarantees that A is invertible. In Example 7.8.1, we compute the adjugate matrix of A:

$$\text{adj}(A) = \begin{bmatrix} 6 & -3 & 0 \\ 0 & 3 & -2 \\ 0 & 0 & 2 \end{bmatrix}.$$

So, by Theorem 7.8.5, we have that

$$A^{-1} = \frac{1}{\det(A)}\text{adj}(A) = \frac{1}{6}\begin{bmatrix} 6 & -3 & 0 \\ 0 & 3 & -2 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 1/2 & -1/3 \\ 0 & 0 & 1/3 \end{bmatrix}.$$

□

[18] By Proposition 7.1.5, the determinant of a square matrix with two identical columns is zero.
[19] By Proposition 7.1.5, the determinant of a square matrix with two identical rows is zero.

Corollary 7.8.4. *Let $\mathbb{F}$ be a field, and let $a, b, c, d \in \mathbb{F}$. Then the matrix*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is invertible if and only if $ad \neq bc$, and in this case, the inverse of A is given by the formula

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Proof. By Theorem 7.4.1, we know that A is invertible if and only if $\det(A) \neq 0$. Clearly, $\det(A) = ad - bc$, and it follows that A is invertible if and only if $ad - bc \neq 0$, i.e. if and only if $ad \neq bc$.

Now, assume that A is invertible, so that $ad \neq bc$. We first compute the cofactors $C_{i,j}$ of A:

- $C_{1,1} = (-1)^{1+1}\det(A_{1,1}) = d$;
- $C_{1,2} = (-1)^{1+2}\det(A_{1,2}) = -c$;
- $C_{2,1} = (-1)^{2+1}\det(A_{2,1}) = -b$;
- $C_{2,2} = (-1)^{2+2}\det(A_{2,2}) = a$.

The cofactor matrix of A is

$$\begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{bmatrix} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}.$$

The adjugate matrix of A is the transpose of the cofactor matrix, i.e.

$$\mathrm{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

By Theorem 7.8.5, we now have that

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A) = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

which is what we needed to show. □

We complete this section with another proof of a slightly weaker version of Theorem 7.8.2, one that only applies to invertible matrices A. We give this proof in order to illustrate a nice application of Cramer's rule.

Theorem 7.8.5. *Let $\mathbb{F}$ be a field, and let A be an **invertible** matrix in $\mathbb{F}^{n \times n}$. Then*

$$A^{-1} = \frac{1}{det(A)}\, adj(A).$$

Proof. Since A is invertible, Theorem 7.4.1 guarantees that $\det(A) \neq 0$, and in particular, the expression $\frac{1}{\det(A)}\text{adj}(A)$ is defined. We will prove the theorem by showing that matrices A^{-1} and $\frac{1}{\det(A)}\text{adj}(A)$ have the same corresponding entries. Fix indices $i, j \in \{1, \ldots, n\}$. By the definition of $\text{adj}(A)$, we see that the i,j-th entry of the matrix $\frac{1}{\det(A)}\text{adj}(A)$ is

$$\frac{(-1)^{j+i}\det(A_{j,i})}{\det(A)}.$$

We will use Cramer's rule to show that this is also the i,j-th entry of the matrix A^{-1}.

Set $A^{-1} = \left[\begin{array}{ccc} \mathbf{a}_1^* & \ldots & \mathbf{a}_n^* \end{array} \right]$. Since $AA^{-1} = I_n$, we have that

$$A \left[\begin{array}{ccc} \mathbf{a}_1^* & \ldots & \mathbf{a}_n^* \end{array} \right] = \left[\begin{array}{ccc} \mathbf{e}_1 & \ldots & \mathbf{e}_n \end{array} \right],$$

and consequently (by the definition of matrix-vector multiplication), that

$$\left[\begin{array}{ccc} A\mathbf{a}_1^* & \ldots & A\mathbf{a}_n^* \end{array} \right] = \left[\begin{array}{ccc} \mathbf{e}_1 & \ldots & \mathbf{e}_n \end{array} \right].$$

In particular, the two matrices above have the same j-th column, and so $A\mathbf{a}_j^* = \mathbf{e}_j$, i.e. $\mathbf{a}_j^*$ is the solution of the equation $A\mathbf{x} = \mathbf{e}_j$ (this solution is unique because A is invertible). So, by Cramer's rule, we have that

$$\mathbf{a}_j^* = \left[\begin{array}{ccc} \frac{\det\left(A_1(\mathbf{e}_j)\right)}{\det(A)} & \ldots & \frac{\det\left(A_n(\mathbf{e}_j)\right)}{\det(A)} \end{array} \right]^T.$$

The i-th entry of $\mathbf{a}_j^*$ is

$$\frac{\det\left(A_i(\mathbf{e}_j)\right)}{\det(A)}.$$

By Laplace explansion along the i-th column, we get that

$$\det\left(A_i(\mathbf{e}_j)\right) = (-1)^{j+i}\det(A_{j,i}).$$

So, the i-th entry of $\mathbf{a}_j^*$ (which is precisely the i,j-th entry of A^{-1}) is

$$\frac{(-1)^{j+i}\det(A_{j,i})}{\det(A)},$$

which is what we needed to show. □

7.9 The Vandermonde matrix

For a positive integer n and real numbers $a_0, a_1, \ldots, a_n$, we define the matrix

$$V(a_0, a_1, \ldots, a_n) := \left[\begin{array}{cccc} 1 & 1 & \ldots & 1 \\ a_0 & a_1 & \ldots & a_n \\ a_0^2 & a_1^2 & \ldots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_0^n & a_1^n & \ldots & a_n^n \end{array} \right]_{(n+1)\times(n+1)},$$

called the *Vandermonde matrix*.

Proposition 7.9.1. *For all positive integers n and real numbers $a_0, a_1, \ldots, a_n$, we have that*

$$det\Big(V(a_0, a_1, \ldots, a_n)\Big) \;=\; \prod_{i>j}(a_i - a_j) \;=\; \prod_{i=1}^{n}\prod_{j=0}^{i-1}(a_i - a_j).$$

and consequently, the matrix $V(a_0, a_1, \ldots, a_n)$ is invertible if and only if $a_0, a_1, \ldots, a_n$ are pairwise distinct.

Proof. By Theorem 7.4.1, we know that a square matrix is invertible if and only if its determinant is non-zero. So, the second statement of the proposition follows immediately from the first, i.e. from the formula for the determinant of the Vandermonde matrix.

We prove the formula for the determinant of the Vandermonde by induction on n. For $n = 1$, we note that for any $a_0, a_1 \in \mathbb{R}$, we have that

$$\det\Big(V(a_0, a_1)\Big) \;=\; \begin{vmatrix} 1 & 1 \\ a_0 & a_1 \end{vmatrix} \;=\; (a_1 - a_0).$$

Now, fix a positive integer n, and assume inductively that our formula is correct for n, i.e. that for all real numbers $a_0, a_1, \ldots, a_n$, we have that $\det\Big(V(a_0, a_1, \ldots, a_n)\Big) = \prod_{i=1}^{n}\prod_{j=0}^{i-1}(a_i - a_j)$. We must show that the formula is correct for $n + 1$. Fix $a_0, a_1, \ldots, a_n, a_{n+1} \in \mathbb{R}$; we will show that

$$\det\Big(V(a_0, a_1, \ldots, a_n, a_{n+1})\Big) \;=\; \prod_{i=1}^{n+1}\prod_{j=0}^{i-1}(a_i - a_j).$$

If some two of the numbers $a_0, a_1, \ldots, a_n, a_{n+1}$ are the same, then this is obvious. Indeed, in this case, the matrix $V(a_0, a_1, \ldots, a_n, a_{n+1})$ has two identical columns and therefore (by Proposition 7.1.5) has determinant zero, and on the other hand, $\prod_{i=1}^{n+1}\prod_{j=0}^{i}(a_i - a_j) = 0$ (because one of the factors is zero). So, from now on, we may assume that $a_0, a_1, \ldots, a_n, a_{n+1}$ are pairwise distinct.

Set

$$f(t) \;:=\; \begin{vmatrix} 1 & 1 & \ldots & 1 & 1 \\ a_0 & a_1 & \ldots & a_n & t \\ a_0^2 & a_1^2 & \ldots & a_n^2 & t^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_0^n & a_1^n & \ldots & a_n^n & t^n \\ a_0^{n+1} & a_1^{n+1} & \ldots & a_n^{n+1} & t^{n+1} \end{vmatrix},$$

so that $f(a_{n+1}) = V(a_0, a_1, \ldots, a_n, a_{n+1})$. By performing Laplace expansion along the rightmost column, we see that $f(t)$ is a polynomial of degree $n + 1$, and that its

leading coefficient (i.e. coefficient in front of t^{n+1}) is

$$k \;:=\; \begin{vmatrix} 1 & 1 & \dots & 1 \\ a_0 & a_1 & \dots & a_n \\ a_0^2 & a_1^2 & \dots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_0^n & a_1^n & \dots & a_n^n \end{vmatrix} \overset{(*)}{=} \prod_{i=1}^{n} \prod_{j=0}^{i-1} (a_i - a_j),$$

where (*) follows from the induction hypothesis. Moreover, for each $i \in \{0, 1, \dots, n\}$, $f(a_i)$ is the determinant of a square matrix that has two identical columns and is therefore (by Proposition 7.1.5) equal to zero. Thus, $a_0, a_1, \dots, a_n$ are all roots of the $(n+1)$-th degree polynomial $f(t)$. So, $f(t)$ can be factored as

$$f(t) \;=\; k \prod_{j=0}^{n} (t - a_j) \;=\; \Big(\prod_{i=1}^{n} \prod_{j=0}^{i-1} (a_i - a_j) \Big) \Big(\prod_{j=0}^{n} (t - a_j) \Big).$$

Consequently,

$$\det\Big(V(a_0, a_1, \dots, a_n, a_{n+1}) \Big) \;=\; f(a_{n+1}) \;=\; \prod_{i=1}^{n+1} \prod_{j=0}^{i-1} (a_i - a_j).$$

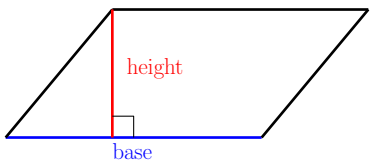
This completes the induction. □

7.10 Determinants and volume

Throughout this section, we assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.

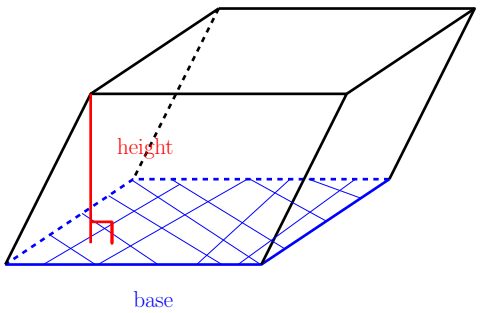
For a parallelogram, we have the familiar formula

$$\begin{pmatrix} \text{area of} \\ \text{parallelogram} \end{pmatrix} \;=\; (\text{length of base}) \;\times\; (\text{height}).$$



We have a similar formula for the volume of a parallelepiped:

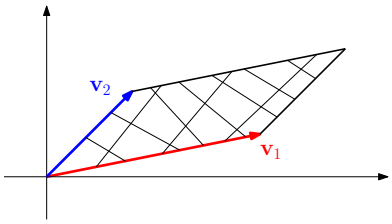
$$\begin{pmatrix} \text{volume of} \\ \text{parallelepiped} \end{pmatrix} \;=\; (\text{area of base}) \;\times\; (\text{height}).$$



We would now like to generalize this to arbitrary dimensions. Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$, the *m-parallelepiped* determined by vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is the set

$$\Big\{ c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m \;\Big|\; c_1, \ldots, c_m \in \mathbb{R},\ 0 \leq c_1, \ldots, c_m \leq 1 \Big\}.$$

For instance, given two vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$, neither of which is a scalar multiple of the other, the 2-parallelepiped determined by $\mathbf{v}_1, \mathbf{v}_2$ is just the usual parallelogram determined by these two vectors (see the picture below).

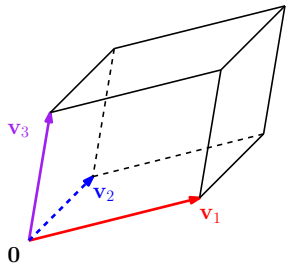


For vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, neither of which is a scalar multiple of each other, the 2-parallelepiped determined by $\mathbf{v}_1, \mathbf{v}_2$ is still a parallelogram, but this parallelogram lies in the plane (2-dimensional subspace) $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$ of $\mathbb{R}^n$. What happens if one of $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ is a scalar multiple of the other, say $\mathbf{v}_2 = \alpha\mathbf{v}_1$ for some scalar $\alpha \in \mathbb{R}$? Then the 2-parallelepiped determined by $\mathbf{v}_1$ and $\mathbf{v}_2$ is just set

$$\begin{aligned}
& \Big\{ c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \,\Big|\, c_1, c_2 \in \mathbb{R},\ 0 \leq c_1, c_2 \leq 1 \Big\} \\
= \; & \Big\{ c_1\mathbf{v}_1 + c_2\alpha\mathbf{v}_1 \;\Big|\; c_1, c_2 \in \mathbb{R},\ 0 \leq c_1, c_2 \leq 1 \Big\} \\
= \; & \Big\{ (c_1 + c_2\alpha)\mathbf{v}_1 \;\Big|\; c_1, c_2 \in \mathbb{R},\ 0 \leq c_1, c_2 \leq 1 \Big\} \\
= \; & \Big\{ c(1+\alpha)\mathbf{v}_1 \;\Big|\; c \in \mathbb{R},\ 0 \leq c \leq 1 \Big\},
\end{aligned}$$

which is 1-dimensional (a line segment) if $\mathbf{v}_1 \neq \mathbf{0}$, and is 0-dimensional (containing only the zero vector) if $\mathbf{v}_1 = \mathbf{0}$. We can think of these as "degenerate parallelograms."

Similarly, for three linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^n$, the 3-parallelepiped defined by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is just the usual parallelepiped whose edges are determined by these three vectors (see the picture below).



If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is not linearly independent, then the 3-parallelepiped determined by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is either a parallelogram, or a line segment, or $\{\mathbf{0}\}$, depending on the dimension of $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Once again, we can think of these as "degenerate parallelepipeds."

For more than three vectors, we get higher-dimensional generalizations.

We would now like to define the "volume" (more precisely, the "m-volume") of an m-parallelepiped in $\mathbb{R}^n$. We do this recursively, as follows.

- The *1-volume* of the 1-parallelepiped determined by the vector $\mathbf{v}_1 \in \mathbb{R}^n$ is defined to be
$$V_1(\mathbf{v}_1) \quad := \quad ||\mathbf{v}_1||.$$
- For a positive integer m, the $(m+1)$-volume of the $(m+1)$-parallelepiped determined by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1} \in \mathbb{R}^n$ is defined to be
$$V_{m+1}(\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}) \quad := \quad V_m(\mathbf{v}_1, \ldots, \mathbf{v}_m)\, ||\mathbf{v}_{m+1}^{\perp}||,$$
where $\mathbf{v}_{m+1}^{\perp} = \text{proj}_{\text{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m)^{\perp}}(\mathbf{v}_{m+1})$.[20]

In this recursive formula, the m-parallelepiped determined by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is our "base" and $||\mathbf{v}_{m+1}^{\perp}||$ is our "height." So, we get the formula

$$\left(\begin{array}{c}(m+1)\text{-volume of} \\ (m+1)\text{-parallelepiped}\end{array}\right) \quad = \quad (m\text{-volume of base}) \quad \times \quad (\text{height}).$$

Note that 1-volume represents (1-dimensional) length, 2-volume represents (2-dimensional) area, and 3-volume represents (3-dimensional) volume. For $m \geq 4$, m-volume is an m-dimensional generalization of these concepts.

Obviously, we would like volume to be non-negative and invariant under vector reordering (i.e. the m-volume of an m-parallelepiped should not change if we merely reorder the vectors determining this m-parallelepiped). The former readily follows

[20] Equivalently (by Corollary 6.5.3): $\mathbf{v}_{m+1}^{\perp} = \mathbf{v}_{m+1} - \text{proj}_{\text{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m)}(\mathbf{v}_{m+1})$.

from the definition of volume (see Proposition 7.10.1 below). However, the latter is not entirely obvious. We prove this in Corollary 7.10.4, which in fact follows from the formula for volume as the determinant of a certain square matrix (see Theorem 7.10.2).

Proposition 7.10.1. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$. Then $V_m(\mathbf{v}_1, \ldots, \mathbf{v}_m) \geq 0$, and equality holds if and only if $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly dependent set.*

Proof. For each $i \in \{1, \ldots, m-1\}$, set $\mathbf{v}_{i+1}^{\perp} := \text{proj}_{\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_i)^{\perp}}(\mathbf{v}_{i+1})$. It then follows from our recursive definition of $V(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ that

$$V_m(\mathbf{v}_1, \ldots, \mathbf{v}_m) \quad = \quad ||\mathbf{v}_1|| \, ||\mathbf{v}_2^{\perp}|| \, \ldots \, ||\mathbf{v}_m^{\perp}||.$$

Since the length of any vector is non-negative, we see that $V_m(\mathbf{v}_1, \ldots, \mathbf{v}_m) \geq 0$. Moreover, equality holds if and only if at least one of the vectors $\mathbf{v}_1, \mathbf{v}_2^{\perp}, \ldots, \mathbf{v}_m^{\perp}$ is $\mathbf{0}$. We will show that the latter happens if and only if the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent.

Suppose first that at least one of $\mathbf{v}_1, \mathbf{v}_2^{\perp}, \ldots, \mathbf{v}_m^{\perp}$ is $\mathbf{0}$. If $\mathbf{v}_1 = \mathbf{0}$, then obviously, $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent. Suppose now that $\mathbf{v}_i^{\perp} = \mathbf{0}$ for some index $i \in \{2, \ldots, m\}$. Then $\text{proj}_{\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})}(\mathbf{v}_i) = \mathbf{v}_i - \mathbf{v}_i^{\perp} = \mathbf{v}_i$, and so $\mathbf{v}_i \in \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})$, i.e. $\mathbf{v}_i$ is a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$. So, $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent.

Suppose now that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent. Then by Proposition 3.2.12, there exists some $i \in \{1, \ldots, m\}$ such that $\mathbf{v}_i$ is a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$. If $i = 1$, then $\mathbf{v}_1 = \mathbf{0}$. On the other hand, if $i \geq 2$, then $\mathbf{v}_i \in \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})$, and so $\mathbf{v}_i^{\perp} = \mathbf{0}$.[21] □

Theorem 7.10.2. *Let $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$, and set $A := \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix}$. Then*

$$V_m(\mathbf{a}_1, \ldots, \mathbf{a}_m) \quad = \quad \sqrt{det(A^T A)}.$$

Proof. For each $i \in \{1, \ldots, m\}$, set $A_i := \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_i \end{bmatrix}$. We will prove inductively that for all $i \in \{1, \ldots, m\}$, we have that $V_i(\mathbf{a}_1, \ldots, \mathbf{a}_i) = \sqrt{\det(A_i^T A_i)}$. Obviously, this is enough, since $A_m = A$.

For $i = 1$, we observe that

$$A_1^T A_1 \quad = \quad \begin{bmatrix} \mathbf{a}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{a}_1 \end{bmatrix} \quad = \quad \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{a}_1 \end{bmatrix},$$

and consequently,

$$\sqrt{\det(A_1^T A_1)} \quad = \quad \sqrt{\mathbf{a}_1 \cdot \mathbf{a}_1} \quad = \quad ||\mathbf{a}_1|| \quad = \quad V_1(\mathbf{a}_1).$$

[21] Indeed, if $\mathbf{v}_i \in \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})$, then $\text{proj}_{\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})}(\mathbf{v}_i) = \mathbf{v}_i$, and consequently, $\mathbf{v}_i^{\perp} = \mathbf{v}_i - \text{proj}_{\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})}(\mathbf{v}_i) = \mathbf{0}$.

We may now assume that $m \geq 2$, for otherwise we are done by what we just showed. Fix $i \in \{1, \ldots, m-1\}$, and assume inductively that $V_i(\mathbf{a}_1, \ldots, \mathbf{a}_i) = \sqrt{\det(A_i^T A_i)}$. We must show that $V_{i+1}(\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{a}_{i+1}) = \sqrt{\det(A_{i+1}^T A_{i+1})}$. Set

- $\mathbf{a}_{i+1}^{||} := \text{proj}_{\text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_i)}(\mathbf{a}_{i+1})$;
- $\mathbf{a}_{i+1}^{\perp} := \text{proj}_{\text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_i)^{\perp}}(\mathbf{a}_{i+1})$.

By Corollary 6.5.3, we have that $\mathbf{a}_{i+1} = \mathbf{a}_{i+1}^{||} + \mathbf{a}_{i+1}^{\perp}$. Since $\mathbf{a}_{i+1}^{||} \in \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_i)$, there exist scalars $c_1, \ldots, c_i \in \mathbb{R}$ such that $\mathbf{a}_{i+1}^{||} = c_1\mathbf{a}_1 + \cdots + c_i\mathbf{a}_i$, and consequently,

$$\mathbf{a}_{i+1}^{\perp} \quad = \quad \mathbf{a}_{i+1} - \mathbf{a}_i^{||} \quad = \quad \mathbf{a}_{i+1} - c_1\mathbf{a}_1 - \cdots - c_i\mathbf{a}_i.$$

Now, let B_{i+1} be the matrix obtained from A_{i+1} by replacing the rightmost column of A_{i+1} by $\mathbf{a}_{i+1}^{\perp}$, i.e.

$$B_{i+1} \quad := \quad \left[\begin{array}{cccc} \mathbf{a}_1 & \ldots & \mathbf{a}_i & \mathbf{a}_{i+1}^{\perp} \end{array} \right].$$

Then

$$B_{i+1}^T \quad = \quad \left[\begin{array}{c} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_i^T \\ (\mathbf{a}_{i+1}^{\perp})^T \end{array} \right] \quad = \quad \left[\begin{array}{c} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_i^T \\ \mathbf{a}_{i+1}^T - c_1\mathbf{a}_1^T - \cdots - c_i\mathbf{a}_i^T \end{array} \right].$$

So, B_{i+1}^T can be obtained from A_{i+1}^T via the following sequence of i elementary row operations:

- $R_{i+1} \to R_{i+1} - c_1R_1$;

  $\vdots$

- $R_{i+1} \to R_{i+1} - c_iR_i$.

Let $E_1, \ldots, E_i$ be the elementary matrices corresponding to these i elementary row operations, so that $B_{i+1}^T = E_i \ldots E_1 A_{i+1}^T$, and consequently, $B_{i+1} = A_{i+1}E_1^T \ldots E_i^T$. By Theorem 7.3.2(c), we see that $\det(E_1) = \cdots = \det(E_i) = 1$.[22] We now compute:

$$\begin{aligned} \det(B_{i+1}^T B_{i+1}) \quad &= \quad \det\Big((E_i \ldots E_1 A_{i+1}^T)(A_{i+1}E_1^T \ldots E_i^T)\Big) \\ &\overset{(*)}{=} \quad \det(E_i) \ldots \det(E_1)\det(A_{i+1}^T A_{i+1})\det(E_1^T) \ldots \det(E_i^T) \end{aligned}$$

[22]Indeed, for each $j \in \{1, \ldots, i\}$, the matrix E_j is obtained by performing the row operation $R_{i+1} \to R_{i+1} - c_jR_j$ on the identity matrix I_{i+1}, and so by Theorem 7.3.2(c), we have that $\det(E_j) = \det(I_{i+1}) = 1$.

$$\begin{aligned} &\overset{(**)}{=} \underbrace{\det(E_i)}_{=1} \dots \underbrace{\det(E_1)}_{=1} \det(A_{i+1}^T A_{i+1}) \underbrace{\det(E_1)}_{=1} \dots \underbrace{\det(E_i)}_{=1} \\ &= \det(A_{i+1}^T A_{i+1}), \end{aligned}$$

where (*) follows from Theorem 7.5.2, and (**) follows from Theorem 7.1.3. But note that $B_{i+1} = \left[\begin{array}{c:c} A_i & \mathbf{a}_{i+1}^\perp \end{array}\right]$, and so

$$\begin{aligned} B_{i+1}^T B_{i+1} &= \left[\begin{array}{c} A_i^T \\ \hdashline (\mathbf{a}_{i+1}^\perp)^T \end{array}\right] \left[\begin{array}{c:c} A_i & \mathbf{a}_{i+1}^\perp \end{array}\right] \\ &= \left[\begin{array}{c:c} A_i^T A_i & A_i^T \mathbf{a}_{i+1}^\perp \\ \hdashline (\mathbf{a}_{i+1}^\perp)^T A_i & (\mathbf{a}_{i+1}^\perp)^T \mathbf{a}_{i+1}^\perp \end{array}\right] \\ &\overset{(*)}{=} \left[\begin{array}{c:c} A_i^T A_i & \mathbf{0} \\ \hdashline \mathbf{0}^T & ||\mathbf{a}_{i+1}^\perp||^2 \end{array}\right], \end{aligned}$$

where in (*), we used the fact that $\mathbf{a}_{i+1}^\perp$ is orthogonal to the columns of A, and so $A^T \mathbf{a}_{i+1}^\perp = \mathbf{0}$,[23] and we also used the fact that $(\mathbf{a}_{i+1}^\perp)^T \mathbf{a}_{i+1}^\perp = \mathbf{a}_{i+1}^\perp \cdot \mathbf{a}_{i+1}^\perp = ||\mathbf{a}_{i+1}^\perp||$. We now compute:

$$\begin{aligned} \det(A_{i+1}^T A_{i+1}) &= \det(B_{i+1}^T B_{i+1}) \\ &= \left|\begin{array}{c:c} A_i^T A_i & \mathbf{0} \\ \hdashline \mathbf{0}^T & ||\mathbf{a}_{i+1}^\perp||^2 \end{array}\right| \\ &\overset{(*)}{=} (-1)^{(i+1)+(i+1)} \, ||\mathbf{a}_{i+1}^\perp||^2 \det(A_i^T A_i) \\ &= \det(A_i^T A_i) \, ||\mathbf{a}_{i+1}^\perp||^2 \\ &\overset{(**)}{=} V_i(\mathbf{a}_1, \dots, \mathbf{a}_i)^2 \, ||\mathbf{a}_{i+1}^\perp||^2 \\ &\overset{(***)}{=} V_{i+1}(\mathbf{a}_1, \dots, \mathbf{a}_i, \mathbf{a}_{i+1})^2, \end{aligned}$$

where (*) follows by Laplace expansion along the rightmost column, (**) follows from

[23] Indeed,

$$A^T \mathbf{a}_{i+1}^\perp = \left[\begin{array}{c} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_i^T \end{array}\right] \mathbf{a}_{i+1}^\perp = \left[\begin{array}{c} \mathbf{a}_1 \cdot \mathbf{a}_{i+1}^\perp \\ \vdots \\ \mathbf{a}_1 \cdot \mathbf{a}_{i+1}^\perp \end{array}\right].$$

Since $\mathbf{a}_{i+1}^\perp \in \mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_i)^\perp$, we see that $\mathbf{a}_1 \cdot \mathbf{a}_{i+1}^\perp = \dots = \mathbf{a}_i \cdot \mathbf{a}_{i+1}^\perp = 0$, and so $A^T \mathbf{a}_{i+1}^\perp = \mathbf{0}$.

the induction hypothesis, and (***) follows from the definition of $V_{i+1}(\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{a}_{i+1})$. Since $V_{i+1}(\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{a}_{i+1}) \geq 0$ (by Proposition 7.10.1), we may now take the square root of both sides to obtain

$$V_{i+1}(\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{a}_{i+1}) \quad = \quad \sqrt{\det(A_{i+1}^T A_{i+1})}.$$

This completes the induction. □

The following corollary of Theorem 7.10.2 gives a geometric interpretation of the determinant.

Corollary 7.10.3. *Let* $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$. *Then*

$$V_n(\mathbf{a}_1, \ldots, \mathbf{a}_n) \quad = \quad |det(\begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix})|.$$

Proof. First of all, we note that $A := \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$ is an $n \times n$ matrix (with entries in $\mathbb{R}$), and so it has a determinant. We now compute:

$$\begin{aligned} V_n(\mathbf{a}_1, \ldots, \mathbf{a}_n) &= \sqrt{\det(A^T A)} && \text{by Theorem 7.10.2} \\ &= \sqrt{\det(A^T)\det(A)} && \text{by Theorem 7.5.2} \\ &= \sqrt{\det(A)^2} && \text{by Theorem 7.1.3} \\ &= |\det(A)|. \end{aligned}$$

This completes the argument. □

Our next corollary states that the m-volume of an m-parallelepiped remains unchanged if we merely change the order of the vectors that determine our m-parallelepiped.

Corollary 7.10.4. *Let* $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ *and* $\sigma \in S_m$. *Then* $V_m(\mathbf{a}_1, \ldots, \mathbf{a}_m) = V_m(\mathbf{a}_{\sigma(1)}, \ldots, \mathbf{a}_{\sigma(m)})$.

Proof. Set $A := \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_m \end{bmatrix}$ and $A_\sigma := \begin{bmatrix} \mathbf{a}_{\sigma(1)} & \ldots & \mathbf{a}_{\sigma(m)} \end{bmatrix}$, and consider P_σ, the matrix of the permutation σ.[24] By Theorem 2.3.15(c), we have that $A_\sigma = AP_\sigma^T$, and by Proposition 7.1.1, we have that $\det(P_\sigma) = \text{sgn}(\sigma)$. But now

$$\begin{aligned} V_m(\mathbf{a}_{\sigma(1)}, \ldots, \mathbf{a}_{\sigma(m)}) &\overset{(*)}{=} \sqrt{\det(A_\sigma^T A_\sigma)} \\ &= \sqrt{\det\big((AP_\sigma^T)^T (AP_\sigma^T)\big)} \end{aligned}$$

[24] We discussed permutation matrices in subsection 2.3.7.

$$\begin{aligned} &= \sqrt{\det(P_\sigma A^T A P_\sigma^T)} \\ &\overset{(**)}{=} \sqrt{\det(P_\sigma)\det(A^T A)\det(P_\sigma^T)} \\ &\overset{(***)}{=} \sqrt{\det(P_\sigma)\det(A^T A)\det(P_\sigma)} \\ &= \sqrt{\mathrm{sgn}(\sigma)^2 \det(A^T A)} \\ &= \sqrt{\det(A^T A)} \\ &\overset{(*)}{=} V_m(\mathbf{a}_1, \ldots, \mathbf{a}_m), \end{aligned}$$

where both instances of (*) follow from Theorem 7.10.2, (**) follows from Theorem 7.5.2, and (***) follows from Theorem 7.1.3. □

Corollary 7.10.5. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$, and let $A \in \mathbb{R}^{n \times n}$. Then*

$$V_n(A\mathbf{v}_1, \ldots, A\mathbf{v}_n) = |det(A)|\, V_n(\mathbf{v}_1, \ldots, \mathbf{v}_n).$$

Proof. Set $B := \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{bmatrix}$ and $C := \begin{bmatrix} A\mathbf{v}_1 & \ldots & A\mathbf{v}_n \end{bmatrix} = AB$. Note that A, B, and $C = AB$ all belong to $\mathbb{R}^{n \times n}$, and so all three matrices have determinants. We now compute:

$$\begin{aligned} V_n(A\mathbf{v}_1, \ldots, A\mathbf{v}_n) &\overset{(*)}{=} \sqrt{\det(C^T C)} \\ &= \sqrt{\det\Big((AB)^T (AB)\Big)} \\ &= \sqrt{\det(B^T A^T A B)} \\ &\overset{(**)}{=} \sqrt{\det(B^T)\det(A^T)\det(A)\det(B)} \\ &\overset{(***)}{=} \sqrt{\det(A)^2 \det(B^T)\det(B)} \\ &\overset{(**)}{=} \sqrt{\det(A)^2\, \det(B^T B)} \\ &= |\det(A)| \sqrt{\det(B^T B)} \\ &\overset{(*)}{=} |\det(A)|\, V_n(\mathbf{v}_1, \ldots, \mathbf{v}_n), \end{aligned}$$

where both instances of (*) follow from Theorem 7.10.2, both instances of (**) follow from Theorem 7.5.2, and (***) follows from Theorem 7.1.3. □

Remark: For $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ $(m \neq n)$ and $A \in \mathbb{R}^{n \times n}$, the formula from Corollary 7.10.5 fails, i.e.

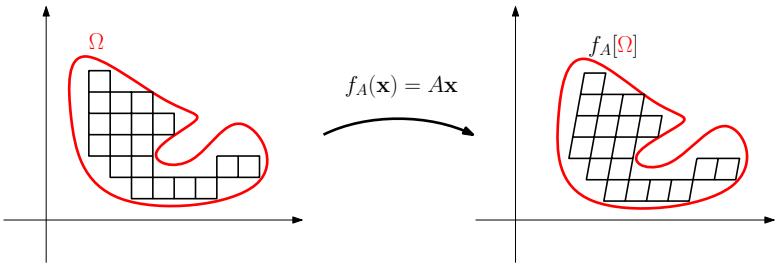
$$V_m(A\mathbf{v}_1, \ldots, A\mathbf{v}_m) \quad \not= \quad |\det(A)|\, V_m(\mathbf{v}_1, \ldots, \mathbf{v}_m).$$

For instance, for $m = 1$ and $n = 2$, we can take $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, so that $A\mathbf{v}_1 = \mathbf{v}_1$. Then $V_1(A\mathbf{v}_1) = V_1(\mathbf{v}_1) = ||\mathbf{v}_1|| = 1$, but $\det(A) = 0$, and so $V_1(A\mathbf{v}_1) \neq |\det(A)|\, V_1(\mathbf{v}_1)$.

Suppose that Ω is any object in $\mathbb{R}^n$ for which n-volume $V_n(\Omega)$ can be defined. We will not go into the technical details of how this can be done, but the idea is that we approximate Ω with ever smaller n-dimensional hypercubes; the sum of n-volumes of those n-hypercubes (which are simply n-parallelepipeds, and so we know how to compute their n-volume) will give us an ever better approximation of the n-volume of Ω that we wish to define. To obtain the actual n-volume of Ω, we take the limit of these ever-finer approximations. If the limit exists, then Ω will have an n-volume (defined to be this limit). If the limit does not exist, then n-volume is undefined for Ω. (It is actually pretty difficult to construct Ω for which volume is undefined! Any reasonably pretty object Ω will have a volume, although that volume may possibly be zero.) Now, suppose we are given a matrix $A \in \mathbb{R}^{n \times n}$. We consider the linear function $f_A : \mathbb{R}^n \to \mathbb{R}^n$ whose standard matrix is A (i.e. for all $\mathbf{x} \in \mathbb{R}^n$, we have $f_A(\mathbf{x}) = A\mathbf{x}$). Then each of the small n-hypercubes gets mapped onto a small n-parallelepiped; if the small n-hypercubes each had volume V, then by Corollary 7.10.5, the small n-parallelepipeds that these n-hypercubes get mapped onto via f_A will have volume $|\det(A)|\, V$. So, we get the following formula for the n-volume of the image of Ω under f_A:

$$V_n(f_A[\Omega]) \quad = \quad |\det(A)|\, V_n(\Omega).$$

For the case $n = 2$, see the picture below.



Example 7.10.6. *Let a and b be positive real numbers. Compute the area (i.e. 2-volume) of the region bounded by the ellipse whose equation is*

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \quad = \quad 1.$$

Solution. We need compute the area of the region

$$E := \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R},\ \tfrac{x_1^2}{a^2} + \tfrac{x_2^2}{b^2} \leq 1 \right\}.$$

Consider the unit disk

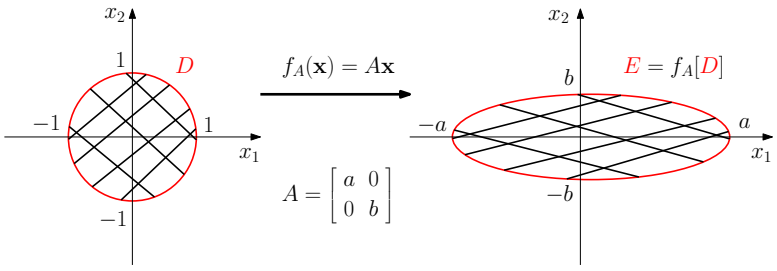
$$D := \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R},\ x_1^2 + x_2^2 \leq 1 \right\}$$

and the matrix

$$A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}.$$

Let $f_A : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear function whose standard matrix is A, so that for all $\begin{bmatrix} x_1 & x_2 \end{bmatrix}^T \in \mathbb{R}^2$, we have

$$f_A\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 \\ bx_2 \end{bmatrix}.$$



We now see that

$$\begin{aligned} f_A[D] &= \left\{ f_A\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) \mid x_1, x_2 \in \mathbb{R},\ x_1^2 + x_2^2 \leq 1 \right\} \\ &= \left\{ \begin{bmatrix} ax_1 \\ bx_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R},\ x_1^2 + x_2^2 \leq 1 \right\} \\ &= \left\{ \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \mid y_1, y_2 \in \mathbb{R},\ (\tfrac{y_1}{a})^2 + (\tfrac{y_2}{b})^2 \leq 1 \right\} \\ &= \left\{ \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \mid y_1, y_2 \in \mathbb{R},\ \tfrac{y_1^2}{a^2} + \tfrac{y_2^2}{b^2} \leq 1 \right\} \\ &= E. \end{aligned}$$

Therefore, the area of E is

$$\text{area}(E) = \underbrace{|\det(A)|}_{=ab} \underbrace{\text{area}(D)}_{=1^2\pi} = ab\pi.$$

□

7.11 Common roots of polynomials via determinants

In this section, we show how determinants can help us determine whether two polynomials have a common root, as long as the field that we are working over is **algebraically closed**. Algebraically closed fields were discussed in subsection 2.4.5. As we discussed in that subsection, of all the fields that we have seen in these lecture notes, only the field $\mathbb{C}$ is algebraically closed (although other algebraically closed fields do exist). The fact that $\mathbb{C}$ is algebraically closed follows immediately from the Fundamental Theorem of Algebra (see subsection 0.3.2).

As we saw in subsection 2.4.5, any non-constant polynomial with coefficients in an algebraically closed field has a root in that field.[25] However, there is no general formula for computing roots of such polynomials, even when the field that we are working over is the familiar field $\mathbb{C}$ of complex number. So, it may be surprising that, given arbitrary polynomials $p(x)$ and $q(x)$ with coefficients in an algebraically closed field $\mathbb{F}$, we can use determinants to determine whether $p(x)$ and $q(x)$ have a common root, i.e. whether there exists a number $x_0 \in \mathbb{F}$ for which we have $p(x_0) = 0$ and $q(x_0) = 0$ (see Theorem 7.11.1 below). However, the determinant in question will only tell us whether such a common root exists; it provides no information on how one might actually compute such a root. The theorem below is proven for arbitrary algebraically closed fields (and it becomes false if the field in question is not algebraically closed). However, if this level of abstraction bothers you, feel free to assume that the field in question is $\mathbb{C}$ (which is the only algebraically closed field that we will ever see in these lecture notes).

Theorem 7.11.1. *Let $\mathbb{F}$ be an **algebraically closed field**. Let m and n be positive integers, and let $p(x) = \sum_{i=0}^{m} a_i x^i$ ($a_m \neq 0$) and $q(x) = \sum_{i=0}^{n} b_i x^i$ ($b_n \neq 0$) be polynomials with coefficients in $\mathbb{F}$. Let P be the $n \times (n+m)$ matrix whose j-th row (for $j \in \{1, \ldots, n\}$) is*

$$\left[\underbrace{0 \ \ldots \ 0}_{j-1} \ a_m \ a_{m-1} \ \ldots \ a_0 \ \underbrace{0 \ \ldots \ 0}_{n-j} \right],$$

and let Q be the $m \times (n+m)$ matrix whose j-th row (for $j \in \{1, \ldots, m\}$) is

$$\left[\underbrace{0 \ \ldots \ 0}_{j-1} \ b_n \ b_{n-1} \ \ldots \ b_0 \ \underbrace{0 \ \ldots \ 0}_{m-j} \right].$$

[25] This does **not** hold in general if the field in question is not algebraically closed!

Then $p(x)$ and $q(x)$ have a common root in $\mathbb{F}$ if and only if

$$det\Big(\left[\begin{array}{c} P \\ \hdashline Q \end{array} \right] \Big) \quad = \quad 0.$$

Remark: For example, if $m = 3$ and $n = 5$, so that

- $p(x) = a_3x^3 + a_2x^2 + a_1x + a_0$,
- $q(x) = b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0$,

then we have

$$\left[\begin{array}{c} P \\ \hdashline Q \end{array} \right] \quad = \quad \left[\begin{array}{cccccccc} a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 & 0 \\ 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 \\ 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 \\ 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 \\ 0 & 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 \\ \hdashline b_5 & b_4 & b_3 & b_2 & b_1 & b_0 & 0 & 0 \\ 0 & b_5 & b_4 & b_3 & b_2 & b_1 & b_0 & 0 \\ 0 & 0 & b_5 & b_4 & b_3 & b_2 & b_1 & b_0 \end{array} \right]_{8\times 8}.$$

Proof of Theorem 7.11.1. We begin by proving a technical claim.

> **Claim.** Polynomials $p(x)$ and $q(x)$ have a common root in $\mathbb{F}$ if and only if there exist non-zero polynomials $r(x)$ and $s(x)$ with coefficients in $\mathbb{F}$ that satisfy the following:
>
> - $\deg\big(r(x)\big) \leq n - 1$;
> - $\deg\big(s(x)\big) \leq m - 1$;
> - $r(x)p(x) + s(x)q(x) = 0$.

Proof of the Claim. Suppose first that $p(x)$ and $q(x)$ have a common root in $\mathbb{F}$, say α. Then we set $r(x) := \frac{q(x)}{x-\alpha}$ and $s(x) := -\frac{p(x)}{x-\alpha}$, and we observe that $\deg\big(r(x)\big) = \deg\big(q(x)\big) - 1 = n - 1$, $\deg\big(s(x)\big) = \deg\big(p(x)\big) - 1 = m - 1$, and

$$r(x)p(x) + s(x)q(x) \quad = \quad \tfrac{q(x)p(x)}{x-\alpha} - \tfrac{p(x)q(x)}{x-\alpha} \quad = \quad 0.$$

Suppose conversely there exist non-zero polynomials $r(x)$ and $s(x)$ with coefficients in $\mathbb{F}$ such that $\deg\big(r(x)\big) \leq n-1$, $\deg\big(s(x)\big) \leq m-1$, and $r(x)p(x)+s(x)q(x) = 0$. Then $r(x)p(x)$ and $s(x)q(x)$ are non-constant polynomials with coefficients in $\mathbb{F}$, and they have exactly the same roots with the same corresponding multiplicities. Since $\deg\big(p(x)\big) = m$, we know that $p(x)$ has exactly m roots in $\mathbb{F}$ (when multiplicities are taken into account).[26] But $\deg\big(s(x)\big) \leq m - 1$, and so at least one of the roots of

[26] Here, we are using the fact that $\mathbb{F}$ is algebraically closed.

$p(x)$ either fails to be a root of $s(x)$, or is a root of $s(x)$ but has smaller multiplicity in $s(x)$ than in $p(x)$. This root of $p(x)$ must therefore be a root of $q(x)$.[27] ♦

In view of the Claim, it now suffices to determine if there exist non-zero polynomials $r(x) = \sum_{i=0}^{n-1} c_i x^i$ and $s(x) = \sum_{i=0}^{m-1} d_i x^i$ such that $r(x)p(x) + s(x)q(x) = 0$. So, we need to determine if there exist $c_0, \ldots, c_{n-1}, d_0, \ldots, d_{m-1} \in \mathbb{F}$ such that at least one of $c_0, \ldots, c_{n-1}$ is non-zero and at least one of $d_0, \ldots, d_{m-1}$ is non-zero, and such that

$$\underbrace{\Big(\sum_{i=0}^{n-1} c_i x^i\Big)}_{=r(x)}\underbrace{\Big(\sum_{i=0}^{m} a_i x^i\Big)}_{=p(x)} + \underbrace{\Big(\sum_{i=0}^{m-1} d_i x^i\Big)}_{=s(x)}\underbrace{\Big(\sum_{i=0}^{n} b_i x^i\Big)}_{=q(x)} \quad = \quad 0.$$

But obviously, if $c_0, \ldots, c_{n-1}$ are all zero, then $d_0, \ldots, d_{m-1}$ are all zero, and vice versa. So, we in fact need to determine if the above equality holds for some numbers $c_0, \ldots, c_{n-1}, d_0, \ldots, d_{m-1} \in \mathbb{F}$, at least one of which is non-zero. We now write the polynomial on the left-hand-side in the standard form, and we set all the coefficients that we obtain equal to zero.[28] This yields a system of $n + m$ linear equations in the variables $c_{n-1}, \ldots, c_0, d_{m-1}, \ldots, d_0$ (we treat $a_m, \ldots, a_0, b_n, \ldots, b_0$ as constants). In each equation, we arrange the variables $c_{n-1}, \ldots, c_0, d_{m-1}, \ldots, d_0$ in this order from left to right. We arrange the equations for the coefficients in front of $x^{n+m-1}, \ldots, x^1, x^0$ from top to bottom. We then rewrite this linear system as a matrix-vector equation

$$A \left[\begin{array}{ccc:ccc} c_{n-1} & \ldots & c_0 & d_{m-1} & \ldots & d_0 \end{array} \right]^T \quad = \quad \mathbf{0},$$

and we observe that the coefficient matrix A satisfies $A^T = \left[\begin{array}{c:c} P & \\ \hdashline & Q \end{array}\right]$.[29]

[27] We are using the fact that $r(x)p(x)$ and $s(x)q(x)$ have the same roots with the same corresponding multiplicities.

[28] We can do this since our polynomial is identically zero, i.e. it is zero as a polynomial. This means precisely that all its coefficients are zero.

[29] For example, if $m = 3$ and $n = 5$, so that

- $p(x) = a_3x^3 + a_2x^2 + a_1x + a_0$,
- $q(x) = b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0$,
- $r(x) = c_4x^4 + c_3x^3 + c_2x^2 + c_1x + c_0$,
- $s(t) = d_2x^2 + d_1x + d_0$,

then our equation becomes

$$\underbrace{\Big(\sum_{i=0}^{4} c_i x^i\Big)}_{=r(x)}\underbrace{\Big(\sum_{i=0}^{3} a_i x^i\Big)}_{=p(x)} + \underbrace{\Big(\sum_{i=0}^{2} d_i x^i\Big)}_{=s(x)}\underbrace{\Big(\sum_{i=0}^{5} b_i x^i\Big)}_{=q(x)} \quad = \quad 0,$$

which yields the system of linear equations below (we consider the coefficients in front of

We now have the following sequence of equivalent statements:

$$\begin{array}{lll} p(x) \text{ and } q(x) \text{ have} \\ \text{a common root in } \mathbb{F} & \iff & A\left[\begin{array}{cccc:cccc} c_{n-1} & \dots & c_0 & & d_{m-1} & \dots & d_0 \end{array}\right]^T = \mathbf{0} \\ & & \text{has a non-zero solution} \\ & \overset{(*)}{\iff} & A \text{ is non-invertible} \\ & \overset{(*)}{\iff} & A^T = \left[\begin{array}{c} P \\ \hdashline Q \end{array}\right] \text{ is non-invertible} \\ & \overset{(*)}{\iff} & \det\left(\left[\begin{array}{c} P \\ \hdashline Q \end{array}\right]\right) = 0, \end{array}$$

where all three instances of (*) follow from the Invertible Matrix Theorem (version 3;

$x^7, x^6, x^5, x^4, x^3, x^2, x^1, x^0$ from top to bottom, and we arrange the variables $c_4, c_3, c_2, c_1, c_0, d_2, d_1, d_0$ from left to right).

$$\begin{array}{c:ccccccccccccccccc} & c_4 & & c_3 & & c_2 & & c_1 & & c_0 & & d_2 & & d_1 & & d_0 & & \\ \hline x^7 & a_3c_4 & & & & & & & & & + & b_5d_2 & & & & & = & 0 \\ x^6 & a_2c_4 & + & a_3c_3 & & & & & & & + & b_4d_2 & + & b_5d_1 & & & = & 0 \\ x^5 & a_1c_4 & + & a_2c_3 & + & a_3c_2 & & & & & + & b_3d_2 & + & b_4d_1 & + & b_5d_0 & = & 0 \\ x^4 & a_0c_4 & + & a_1c_3 & + & a_2c_2 & + & a_3c_1 & & & + & b_2d_2 & + & b_3d_1 & + & b_4d_0 & = & 0 \\ x^3 & & & a_0c_3 & + & a_1c_2 & + & a_2c_1 & + & a_3c_0 & + & b_1d_2 & + & b_2d_1 & + & b_3d_0 & = & 0 \\ x^2 & & & & & a_0c_2 & + & a_1c_1 & + & a_2c_0 & + & b_0d_2 & + & b_1d_1 & + & b_2d_0 & = & 0 \\ x^1 & & & & & & & a_0c_1 & + & a_1c_0 & & & + & b_0d_1 & + & b_1d_0 & = & 0 \\ x^0 & & & & & & & & & a_0c_0 & & & & & + & b_0d_0 & = & 0 \end{array}$$

This linear system, in turn, translates into the following matrix-vector equation:

$$\left[\begin{array}{ccccc:ccc} a_3 & 0 & 0 & 0 & 0 & b_5 & 0 & 0 \\ a_2 & a_3 & 0 & 0 & 0 & b_4 & b_5 & 0 \\ a_1 & a_2 & a_3 & 0 & 0 & b_3 & b_4 & b_5 \\ a_0 & a_1 & a_2 & a_3 & 0 & b_2 & b_3 & b_4 \\ 0 & a_0 & a_1 & a_2 & a_3 & b_1 & b_2 & b_3 \\ 0 & 0 & a_0 & a_1 & a_2 & b_0 & b_1 & b_2 \\ 0 & 0 & 0 & a_0 & a_1 & 0 & b_0 & b_1 \\ 0 & 0 & 0 & 0 & a_0 & 0 & 0 & b_0 \end{array}\right] \left[\begin{array}{c} c_4 \\ c_3 \\ c_2 \\ c_1 \\ c_0 \\ \hdashline d_2 \\ d_1 \\ d_0 \end{array}\right] = \mathbf{0}.$$

Note that the transpose of the coefficient matrix that we obtained is precisely the matrix

$$\left[\begin{array}{c} P \\ \hdashline Q \end{array}\right] = \left[\begin{array}{cccccccc} a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 & 0 \\ 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 \\ 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 \\ 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 \\ 0 & 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 \\ \hdashline b_5 & b_4 & b_3 & b_2 & b_1 & b_0 & 0 & 0 \\ 0 & b_5 & b_4 & b_3 & b_2 & b_1 & b_0 & 0 \\ 0 & 0 & b_5 & b_4 & b_3 & b_2 & b_1 & b_0 \end{array}\right]_{8\times 8}$$

from the Remark following the statement of the theorem.

see subsection 7.4.1). This completes the argument. □

Example 7.11.2. *Determine whether the polynomials* $p(x) = 5x^3 - 2x^2 + x - 4$ *and* $q(x) = 7x^2 - 6x - 1$ *have a common complex root.*

Solution. In this case, it is easy to see that $p(1) = 0$ and $q(1) = 0$, and so 1 is a common root of $p(x)$ and $q(x)$. However, let us use Theorem 7.11.1, in order to illustrate how this theorem can be applied.

Using the notation of Theorem 7.11.1, we have that $m = 3$, $n = 2$, and the matrices P and Q are given by

- $P = \begin{bmatrix} 5 & -2 & 1 & -4 & 0 \\ 0 & 5 & -2 & 1 & -4 \end{bmatrix}$;
- $Q = \begin{bmatrix} 7 & -6 & -1 & 0 & 0 \\ 0 & 7 & -6 & -1 & 0 \\ 0 & 0 & 7 & -6 & -1 \end{bmatrix}$.

We now have that

$$\det\Big(\begin{bmatrix} P \\ \hdashline Q \end{bmatrix}\Big) = \left|\begin{array}{ccccc} 5 & -2 & 1 & -4 & 0 \\ 0 & 5 & -2 & 1 & -4 \\ \hdashline 7 & -6 & -1 & 0 & 0 \\ 0 & 7 & -6 & -1 & 0 \\ 0 & 0 & 7 & -6 & -1 \end{array}\right| = 0.$$

Theorem 7.11.1 now guarantees that $p(x)$ and $q(x)$ have a common complex root. □

Chapter 8

Eigenvalues and eigenvectors

Remark: In our study of eigenvalues and eigenvectors, we will often make reference to "algebraically closed fields." Algebraically closed fields were covered in subsection 2.4.5, which the reader may wish to review before reading the present chapter. In a nutshell, a field $\mathbb{F}$ is *algebraically closed* if every non-constant polynomial with coefficients in $\mathbb{F}$ has a root in $\mathbb{F}$. It can be shown that if $\mathbb{F}$ is an algebraically closed field, then any non-constant polynomial with coefficients in $\mathbb{F}$ can be factored into linear terms. Of all the fields that we have seen in these lecture notes, only $\mathbb{C}$ is algebraically closed. (Other algebraically closed fields exist, but they are not discussed in these lecture notes.) Fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{Z}_p$ (where p is a prime number) are not algebraically closed.

8.1 Eigenvectors, eigenvalues, and eigenspaces

8.1.1 Eigenvectors and eigenvalues of linear functions

Suppose that V is a vector spaces over a field $\mathbb{F}$, and that $f : V \to V$ is a linear function. An *eigenvector* of f is a vector $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ for which there exists a scalar $\lambda \in \mathbb{F}$, called the *eigenvalue* of f associated with the eigenvector $\mathbf{v}$, such that

$$f(\mathbf{v}) \quad = \quad \lambda\mathbf{v}.$$

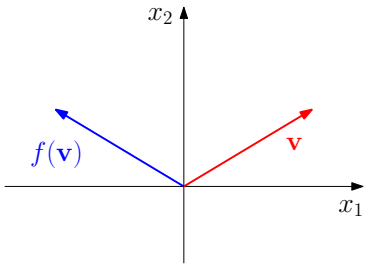
Under these circumstances, we also say that $\mathbf{v}$ is an eigenvector of f associated with the eigenvalue λ. So, the eigenvectors of f are those **non-zero** vectors in V that simply get scaled by f, and the eigenvalues are the scalars that the eigenvectors get scaled by. By definition, an eigenvector cannot be $\mathbf{0}$, but an eigenvalue may possibly be 0.

Remark: Note that eigenvectors and eigenvalues are only defined for those linear functions whose domain is the same as the codomain.

Example 8.1.1. *Consider the linear function* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *given by*

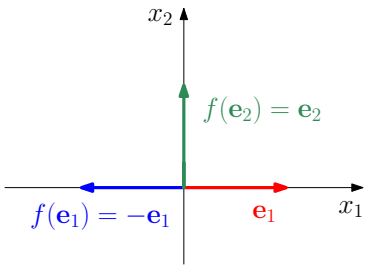
$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ x_2 \end{bmatrix}$$

for all $x_1, x_2 \in \mathbb{R}$. *So,* f *is the reflection about the* x_2*-axis (see the picture below), and its standard matrix is* $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.



As usual, $\mathbf{e}_1$ *and* $\mathbf{e}_2$ *are the standard basis vectors of* $\mathbb{R}^2$. *Then*

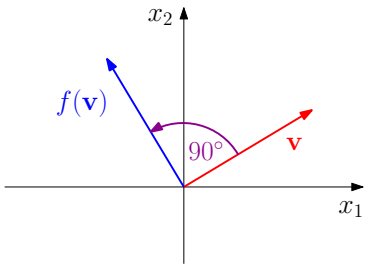
- $\mathbf{e}_1$ *is an eigenvector of* f *associated with the eigenvalue* $\lambda_1 := -1$, *since* $f(\mathbf{e}_1) = -\mathbf{e}_1 = \lambda_1 \mathbf{e}_1$;
- $\mathbf{e}_2$ *is an eigenvector of* f *associated with the eigenvalue* $\lambda_2 := 1$, *since* $f(\mathbf{e}_2) = \mathbf{e}_2 = \lambda_2 \mathbf{e}_2$.



Example 8.1.2. *Consider the linear function* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *given by*

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

for all $x_1, x_2 \in \mathbb{R}$. *So,* f *is the counterclockwise rotation by* $90°$ *about the origin (see the picture below), and its standard matrix is* $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. *This function has no eigenvectors (and consequently, it has no eigenvalues), since it does not simply scale any non-zero vector in* $\mathbb{R}^2$.



Example 8.1.3. *Consider the linear function $f : \mathbb{C}^2 \to \mathbb{C}^2$ given by*

$$f\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) \quad = \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad = \quad \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

for all $x_1, x_2 \in \mathbb{C}$. (This is the same formula as the one from Example 8.1.2, except that we are now working over $\mathbb{C}$, rather than over $\mathbb{R}$.) Then

- $\mathbf{v}_1 := \begin{bmatrix} i \\ 1 \end{bmatrix}$ *is an eigenvalue of f associated with the eigenvalue $\lambda_1 := i$, since*

$$f(\mathbf{v}_1) \quad = \quad \begin{bmatrix} -1 \\ i \end{bmatrix} \quad = \quad i \begin{bmatrix} i \\ 1 \end{bmatrix} \quad = \quad \lambda_1 \mathbf{v}_1;$$

- $\mathbf{v}_2 := \begin{bmatrix} -i \\ 1 \end{bmatrix}$ *is an eigenvector of f associated with the eigenvalue $\lambda_2 := -i$, since*

$$f(\mathbf{v}_2) \quad = \quad \begin{bmatrix} -1 \\ -i \end{bmatrix} \quad = \quad (-i) \begin{bmatrix} -i \\ 1 \end{bmatrix} \quad = \quad \lambda_2 \mathbf{v}_2.$$

Remark: It may be somewhat surprising that the linear function f from Example 8.1.2 has no eigenvectors and no eigenvalues, whereas the one from Example 8.1.3 has them. As we shall see once we learn how to actually compute eigenvalues and eigenvectors (this will involve finding roots of polynomials), the essential difference is that $\mathbb{C}$ is an algebraically closed field, whereas $\mathbb{R}$ is not.

Eigenspaces. For a linear function $f : V \to V$, where V is a vector space over a field $\mathbb{F}$, and for a scalar $\lambda \in \mathbb{F}$, we define

$$E_\lambda(f) \quad := \quad \{\mathbf{v} \in V \mid f(\mathbf{v}) = \lambda \mathbf{v}\}.$$

Note that $\mathbf{0} \in E_\lambda(f)$, since $f(\mathbf{0}) \overset{(*)}{=} \mathbf{0} = \lambda \mathbf{0}$, where (*) follows from Proposition 4.1.6 (since f is linear). The set $E_\lambda(f)$ can be defined for any scalar λ, but it is only interesting in the case when λ is an eigenvalue of V, in which case $E_\lambda(f)$ is called the *eigenspace* of f associated with the eigenvalue λ. Note that, for an eigenvalue λ of f, the elements of the eigenspace $E_\lambda(f)$ are precisely the zero vector and the

eigenvectors of f associated with λ.[1] On the other hand, if λ is not an eigenvalue of f, then we simply have that $E_\lambda(f) = \{\mathbf{0}\}$, and we do not refer to $E_\lambda(f)$ as an eigenspace.

Proposition 8.1.4. *Let V be a vector space over a field $\mathbb{F}$, and let $f : V \to V$ be a linear function. Then both the following hold:*

(a) for all scalars $\lambda \in \mathbb{F}$, $E_\lambda(f)$ is a subspace of V, and this subspace is non-trivial (i.e. contains at least one non-zero vector) if and only if λ is an eigenvalue of f;

(b) for all distinct scalars $\lambda_1, \lambda_2 \in \mathbb{F}$, we have that $E_{\lambda_1}(f) \cap E_{\lambda_2}(f) = \{\mathbf{0}\}$.

Proof. (a) Fix a scalar $\lambda \in \mathbb{F}$. We first show that $E_\lambda(f)$ is a subsapce of V. In view of Theorem 3.1.7, it suffices to show that $E_\lambda(f)$ contains $\mathbf{0}$ and is closed under vector addition and scalar multiplication.

We already saw in the discussion above that $\mathbf{0} \in E_\lambda(f)$.[2] Next, for vectors $\mathbf{v}_1, \mathbf{v}_2 \in E_\lambda(f)$, we have that

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2) &= f(\mathbf{v}_1) + f(\mathbf{v}_2) && \text{because } f \text{ is linear} \\ &= \lambda\mathbf{v}_1 + \lambda\mathbf{v}_2 && \text{because } \mathbf{v}_1, \mathbf{v}_2 \in E_\lambda(f) \\ &= \lambda(\mathbf{v}_1 + \mathbf{v}_2), \end{aligned}$$

and consequently, $\mathbf{v}_1 + \mathbf{v}_2 \in E_\lambda(f)$. Finally, for a vector $\mathbf{v} \in E_\lambda(f)$ and a scalar $\alpha \in \mathbb{F}$, we have that

$$\begin{aligned} f(\alpha\mathbf{v}) &= \alpha f(\mathbf{v}) && \text{because } f \text{ is linear} \\ &= \alpha(\lambda\mathbf{v}) && \text{because } \mathbf{v} \in E_\lambda(f) \\ &= \lambda(\alpha\mathbf{v}), \end{aligned}$$

and it follows that $\alpha\mathbf{v} \in E_\lambda(f)$. This proves that $E_\lambda(f)$ is indeed a subspace of V.

The fact that the subspace $E_\lambda(f)$ is non-trivial if and only if λ is an eigenvalue of f follows immediately from the appropriate definitions.[3]

[1] By definition, $\mathbf{0}$ cannot be an eigenvector.

[2] Here is the proof once again. We have that $f(\mathbf{0}) \overset{(*)}{=} \mathbf{0} = \lambda\mathbf{0}$, where (*) follows from Proposition 4.1.6 (since f is linear). So, by the definition of $E_\lambda(f)$, we have that $\mathbf{0} \in E_\lambda(f)$.

[3] Here is a detailed proof. Suppose first that the subspace $E_\lambda(f)$ is non-trivial, and fix some $\mathbf{v} \in E_\lambda(f) \setminus \{\mathbf{0}\}$. But then $\mathbf{v}$ is a non-zero vector in V that satisfies $f(\mathbf{v}) = \lambda\mathbf{v}$, which by definition means that $\mathbf{v}$ is an eigenvector of V, and that λ is the associated eigenvalue. On the other hand, if λ is an eigenvalue of V, then by definition, there exists a non-zero vector $\mathbf{v} \in V$ such that $f(\mathbf{v}) = \lambda\mathbf{v}$, which means that $\mathbf{v} \in E_\lambda(f)$, and so the subspace $E_\lambda(f)$ contains a non-zero vector and is therefore non-trivial.

(b) Fix distinct scalars $\lambda_1, \lambda_2 \in \mathbb{F}$. By (a), $E_{\lambda_1}(f)$ and $E_{\lambda_2}(f)$ are both subspaces of V, and consequently, $\mathbf{0} \in E_{\lambda_1}(f) \cap E_{\lambda_2}(f)$. Now, fix any $\mathbf{v} \in E_{\lambda_1}(f) \cap E_{\lambda_2}(f)$. Since $\mathbf{v} \in E_{\lambda_1}(f)$, we have that $f(\mathbf{v}) = \lambda_1\mathbf{v}$, and since $\mathbf{v} \in E_{\lambda_2}(f)$, we have that $f(\mathbf{v}) = \lambda_2\mathbf{v}$. So, $\lambda_1\mathbf{v} = \lambda_2\mathbf{v}$, and consequently, $(\lambda_1 - \lambda_2)\mathbf{v} = \mathbf{0}$. Since $\lambda_1 - \lambda_2 \neq 0$ (because $\lambda_1 \neq \lambda_2$), Proposition 3.1.3(c) guarantees that $\mathbf{v} = \mathbf{0}$. This proves that $E_{\lambda_1}(f) \cap E_{\lambda_2}(f) = \{\mathbf{0}\}$. □

Terminology: Suppose that V is a vector space over a field $\mathbb{F}$, and that λ is an eigenvalue of a linear function $f : V \to V$. The *geometric multiplicity* of the eigenvalue λ is defined to be $\dim\big(E_\lambda(f)\big)$. So, the geometric multiplicity of an eigenvalue is the dimension of the associated eigenspace.

Remark: Suppose that V is a vector space over a field $\mathbb{F}$, and that λ is an eigenvalue of a linear function $f : V \to V$. Then the elements of the eigenspace $E_\lambda(f)$ are precisely the vector $\mathbf{0}$ and the eigenvectors of f associated with the eigenvalue λ. Since no linearly independent set of vectors contains $\mathbf{0}$, it follows that all vectors in any linearly independent set of vectors in $E_\lambda(f)$ are eigenvectors of f associated with λ. Consequently, all vectors in any basis of $E_\lambda(f)$ are eigenvectors of f associated with the eigenvalue λ.

8.1.2 Eigenvectors and eigenvalues of square matrices

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. An *eigenvector* of A is a vector $\mathbf{v} \in \mathbb{F}^n \setminus \{\mathbf{0}\}$ for which there exists a scalar $\lambda \in \mathbb{F}$, called the *eigenvalue* of A associated with the eigenvector $\mathbf{v}$, such that

$$A\mathbf{v} \;=\; \lambda\mathbf{v}.$$

Under these circumstances, we also say that $\mathbf{v}$ is an eigenvector of A associated with the eigenvalue λ.

Remark: Note that eigenvectors and eigenvalues are only defined for **square** matrices. Eigenvectors are, by definition, non-zero, whereas eigenvalues may possibly be zero.

Eigenspaces. For a square matrix $A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is some field), and for a scalar $\lambda \in \mathbb{F}$, we define

$$E_\lambda(A) \;:=\; \{\mathbf{v} \in \mathbb{F}^n \mid A\mathbf{v} = \lambda\mathbf{v}\}.$$

If λ is an eigenvalue of A, then $E_\lambda(A)$ is called the *eigenspace* of A associated with the eigenvalue λ. Note that, for an eigenvalue λ of A, the elements of the eigenspace $E_\lambda(A)$ are precisely the zero vector and the eigenvectors of A associated with λ. On the other hand, if λ is not an eigenvalue of A, then we simply have that $E_\lambda(A) = \{\mathbf{0}\}$, and we do not refer to $E_\lambda(A)$ as an eigenspace.

Proposition 8.1.5. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^n \to \mathbb{F}^n$ be a linear function, and let A be the standard matrix of f. Then f and A have exactly the same eigenvalues and the associated eigenectors. Moreover, for all eigenvalues λ of f and A, we have that $E_\lambda(f) = E_\lambda(A)$.*

Proof. This follows immediately from the appropriate definitions. □

Propositions 8.1.4 and 8.1.5 immediately imply the following proposition.

Proposition 8.1.6. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then all the following hold:*

(a) for all scalars $\lambda \in \mathbb{F}$, $E_\lambda(A)$ is a subspace of $\mathbb{F}^n$, and this subspace is non-trivial (i.e. contains at least one non-zero vector) if and only if λ is an eigenvalue of A;

(b) for all distinct scalars $\lambda_1, \lambda_2 \in \mathbb{F}$, we have that $E_{\lambda_1}(A) \cap E_{\lambda_2}(A) = \{\mathbf{0}\}$.

Proof. Consider the function $f_A : \mathbb{F}^n \to \mathbb{F}^n$, given by $f_A(\mathbf{v}) = A\mathbf{v}$ for all vectors $\mathbf{v} \in \mathbb{F}^n$. Then f_A is linear (by Proposition 1.10.4), and moreover, A is the standard matrix of f_A. So, by Proposition 8.1.5, we have that for all $\lambda \in \mathbb{F}$, $E_\lambda(A) = E_\lambda(f_A)$. The result now follows immediately from Proposition 8.1.4. □

Terminology: Suppose that $\mathbb{F}$ is a field, and that λ is an eigenvalue of a square matrix $A \in \mathbb{F}^{n\times n}$. The *geometric multiplicity* of the eigenvalue λ is defined to be $\dim\big(E_\lambda(A)\big)$. So, the geometric multiplicity of an eigenvalue is the dimension of the associated eigenspace.

Proposition 8.1.7. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis of V, and let $f : V \to V$ be a linear function. Then for all $\lambda \in \mathbb{F}$, we have that*

$$E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}\Big) \;=\; \Big\{ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} \,\big|\, \mathbf{v} \in E_\lambda(f)\Big\}.$$

Consequently, the linear function f and the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ have exactly the same eigenvalues, with exactly the same corresponding geometric multiplicities.

Proof. Fix a scalar $\lambda \in \mathbb{F}$. We first note that, by Proposition 8.1.4, $E_\lambda(f)$ is a subspace of V, and by Proposition 8.1.6, $E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}\Big)$ is a subspace of $\mathbb{F}^n$. Now, we claim that

$$E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}\Big) \;=\; \Big\{ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} \,\big|\, \mathbf{v} \in E_\lambda(f)\Big\}.$$

For this, we must prove the following two inclusions:

(1) $E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}\Big) \subseteq \Big\{ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} \,\big|\, \mathbf{v} \in E_\lambda(f)\Big\}$;

(2) $\Big\{ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} \mid \mathbf{v} \in E_\lambda(f) \Big\} \subseteq E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big)$.

We first prove (1). Fix any vecor $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big)$. Set $\mathbf{v} := x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n$, so that $\big[\ \mathbf{v}\ \big]_{\mathcal{B}} = \mathbf{x}$. We must show that $\mathbf{v} \in E_\lambda(f)$. We compute:

$$\big[\ f(\mathbf{v})\ \big]_{\mathcal{B}} \;=\; {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \underbrace{\big[\ \mathbf{v}\ \big]_{\mathcal{B}}}_{=\mathbf{x}} \overset{(*)}{=} \lambda \underbrace{\big[\ \mathbf{v}\ \big]_{\mathcal{B}}}_{=\mathbf{x}} \overset{(**)}{=} \big[\ \lambda\mathbf{v}\ \big]_{\mathcal{B}},$$

where (*) follows from the fact that $\mathbf{x} \in E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big)$, and (**) follows from the linearity of $\big[\ \cdot\ \big]_{\mathcal{B}}$. Since $\big[\ \cdot\ \big]_{\mathcal{B}}$ is an isomorphism (and in particular, one-to-one), we see that $f(\mathbf{v}) = \lambda\mathbf{v}$. By definition, this means that $\mathbf{v} \in E_\lambda(f)$. This proves (1).

Let us now prove (2). Fix any $\mathbf{v} \in E_\lambda(f)$. Then

$$\begin{aligned} {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}\ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} &= \big[\ f(\mathbf{v})\ \big]_{\mathcal{B}} \\ &= \big[\ \lambda\mathbf{v}\ \big]_{\mathcal{B}} \qquad \text{because } \mathbf{v} \in E_\lambda(f) \\ &= \lambda\big[\ \mathbf{v}\ \big]_{\mathcal{B}} \qquad \text{because } \big[\ \cdot\ \big]_{\mathcal{B}} \text{ is linear,} \end{aligned}$$

and it follows that $\big[\ \mathbf{v}\ \big]_{\mathcal{B}} \in E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big)$. This proves (2).

We have now proven both (1) and (2), and it follows that

$$E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big) \;=\; \Big\{ \big[\ \mathbf{v}\ \big]_{\mathcal{B}} \mid \mathbf{v} \in E_\lambda(f) \Big\}.$$

Thus, $E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big)$ is the image of $E_\lambda(f)$ under the isomorphism $\big[\ \cdot\ \big]_{\mathcal{B}} : V \to \mathbb{F}^n$. So, by Proposition 4.4.7, we have that

$$\dim\Big(E_\lambda(f)\Big) \;=\; \dim\Big(E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big)\Big).$$

In particular, the subspace $E_\lambda(f)$ of V is non-trivial if and only if the subspace $E_\lambda\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big)$ of $\mathbb{F}^n$ is non-trivial, and so by Propositions 8.1.4 and 8.1.6, λ is an eigenvalue of f if and only if it is an eigenvalue of ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. Moreover, by what we just proved, if λ is an eigenvalue of f and ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$, then the corresponding eigenspaces have the same dimension, that is, the eigenvalue λ has the same geometric multiplicity with respect to the linear function f and with respect to the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. □

Remark: In view of Propositions 8.1.5 and 8.1.7, we see that the study of eigenvalues and eigenvectors of linear functions from a non-trivial, finite-dimensional vector space to itself is essentially equivalent to the study of eigenvalues and eigenvectors of square matrices. The computational tools that we develop for finding eigenvectors and eigenvalues will primarily be for square matrices. On the other hand, some of the theoretical results that we prove will be for linear functions instead, and we will obtain corresponding results for matrices as more or less immediate corollaries.

8.2 The characteristic polynomial

In this section, we introduce the "characteristic polynomial" of a square matrix (see subsection 8.2.1), and we show that its roots are precisely the eigenvalues of the matrix in question. We define the "algebraic multiplicity" of an eigenvalue to be its multiplicity as a root of the characteristic polynomial. As we shall see, the geometric multiplicity of an eigenvalue is no greater than its algebraic multiplicity (see Theorem 8.2.3). We also show that a square matrix is invertible if and only if 0 is **not** its eigenvalue (see Proposition 8.2.11), and we add this to our previous version of the Invertible Matrix Theorem (see subsection 7.4.1) to produce the fourth and final version of the Invertible Matrix Theorem (see subsection 8.2.6). Finally, in subsection 8.2.7, we introduce characteristic polynomial of a linear function (having the same non-trivial, finite-dimensional vector space both for its domain and codomain).

8.2.1 The characteristic polynomial of a square matrix

Given a field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n\times n}$, the *characteristic polynomial* of A is defined to be

$$p_A(\lambda) \;:=\; \det\big(\lambda I_n - A\big).$$

The *characteristic equation* of A is the equation

$$\det\big(\lambda I_n - A\big) \;=\; 0.$$

So, the roots of the characteristic polynomial of A are precisely the solutions of the characteristic equation of A.

Example 8.2.1. *Compute the characteristic polynomial of the following matrix in* $\mathbb{C}^{3\times 3}$*:*

$$A \;=\; \begin{bmatrix} 1 & -2 & 3 \\ -1 & 0 & 2 \\ 2 & -1 & -3 \end{bmatrix}.$$

Solution. The characteristic polynomial of A is:

$$p_A(\lambda) \;=\; \det(\lambda I_3 - A) \;=\; \begin{vmatrix} \lambda-1 & 2 & -3 \\ 1 & \lambda & -2 \\ -2 & 1 & \lambda+3 \end{vmatrix} \;=\; \lambda^3 + 2\lambda^2 - 9\lambda - 3.$$

□

Remark: For a field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n\times n}$, the characteristic polynomial $p_A(\lambda) = \det(\lambda I_n - A)$ is a polynomial of degree n, with leading coefficient 1, i.e. the coefficient in front of λ^n in $p_A(\lambda)$ is 1. In some texts, the characteristic polynomial

is defined to be $\det(A - \lambda I_n)$. By Proposition 7.2.3, we have that $\det(A - \lambda I_n) = (-1)^n \det(\lambda I_n - A)$, and so the polynomials $\det(\lambda I_n - A)$ and $\det(A - \lambda I_n)$ have exactly the same roots, with the same corresponding multiplicities, which is what we will actually care about when it comes to the characteristic polynomial. The main advantage of using $\det(\lambda I_n - A)$ rather than $\det(A - \lambda I_n)$ is that the former polynomial has leading coefficient 1, whereas the latter has leading coefficient $(-1)^n$, which is -1 if n is odd.

Theorem 8.2.2. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n \times n}$, and let $\lambda_0 \in \mathbb{F}$. Then*

$$E_{\lambda_0}(A) \;=\; Nul\big(\lambda_0 I_n - A\big) \;=\; Nul\big(A - \lambda_0 I_n\big).$$

Moreover, the following are equivalent:

(1) λ_0 is an eigenvalue of A;

(2) λ_0 is a root of the characteristic polynomial of A, i.e. $p_A(\lambda_0) = 0$;

(3) λ_0 is a solution of the characteristic equation of A, i.e. $det\big(\lambda_0 I_n - A\big) = 0$.

Proof. Obviously, for all $\mathbf{v} \in \mathbb{F}^n$, we have that $\big(\lambda_0 I_n - A\big)\mathbf{v} = \mathbf{0}$ if and only if $\big(A - \lambda_0 I_n\big)\mathbf{v} = \mathbf{0}$. So, $\mathrm{Nul}\big(\lambda_0 I_n - A\big) = \mathrm{Nul}\big(A - \lambda_0 I_n\big)$. Further, we compute:

$$\begin{aligned} E_{\lambda_0}(A) &= \big\{\mathbf{v} \in \mathbb{F}^n \mid A\mathbf{v} = \lambda_0 \mathbf{v}\big\} \\ &= \big\{\mathbf{v} \in \mathbb{F}^n \mid A\mathbf{v} = \lambda_0 I_n \mathbf{v}\big\} \\ &= \big\{\mathbf{v} \in \mathbb{F}^n \mid (\lambda_0 I_n - A)\mathbf{v} = \mathbf{0}\big\} \\ &= \mathrm{Nul}\big(\lambda_0 I_n - A\big). \end{aligned}$$

It remains to show that (1), (2), and (3) are equivalent. The fact that (2) and (3) are equivalent follows immediately from the appropriate definitions. It remains to prove that (1) and (3) are equivalent. For this, we have the following sequence of equivalent statements:

$$\begin{aligned} &\underbrace{\lambda_0 \text{ is an eigenvalue of } A}_{(1)} && \\ &\overset{(*)}{\iff} && E_{\lambda_0}(A) \neq \{\mathbf{0}\} \\ &\overset{(**)}{\iff} && \mathrm{Nul}\big(\lambda_0 I_n - A\big) \neq \{\mathbf{0}\} \\ &\overset{(***)}{\iff} && \text{the matrix } \lambda_0 I_n - A \\ &&& \text{is not invertible} \\ &\overset{(***)}{\iff} && \underbrace{\det\big(\lambda_0 I_n - A\big) = 0}_{(3)} \end{aligned}$$

where (*) follows from Proposition 8.1.6, (**) follows from the fact that $E_{\lambda_0}(A) = \mathrm{Nul}\big(\lambda_0 I_n - A\big)$ (which we proved above), and both instances of (***) follow from the Invertible Matrix Theorem (version 3; see subsection 7.4.1). This completes the argument. □

Algebraic multiplicities of eigenvalues. By Theorem 8.2.2, the eigenvalues of a square matrix are precisely the roots of its characteristic polynomial. With this in mind, we define the "algebraic multiplicity" of an eigenvalue as follows. For a field $\mathbb{F}$, a matrix $A \in \mathbb{F}^{n\times n}$, and an eigenvalue λ_0 of A, the *algebraic multiplicity* of the eigenvalue λ_0 is its multiplicity as a root of the characteristic polynomial of A, or in other words, it is the largest integer k such that $(\lambda - \lambda_0)^k \mid p_A(\lambda)$, i.e. such that $(\lambda - \lambda_0)^k$ divides the polynomial $p_A(\lambda)$.[4] Since $\deg\big(p_A(\lambda)\big) = n$, the sum of algebraic multiplicities of the eigenvalues of the matrix $A \in \mathbb{F}^{n\times n}$ is at most n; if the field $\mathbb{F}$ is algebraically closed, then the sum of algebraic multiplicities of the eigenvalues of A is exactly n.[5] In view of Theorem 8.2.2, this implies that if the field $\mathbb{F}$ is algebraically closed, then A has exactly n eigenvalues when algebraic multiplicities are taken into account. However, this is not necessarily true if $\mathbb{F}$ is not algebraically closed (in that case, it is even possible that A has no eigenvalues at all).

Theorem 8.2.3. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$. Then the geometric multiplicity of any eigenvalue of A is no greater than the algebraic multiplicity of that eigenvalue.*

Remark: Schematically, Theorem 8.2.3 states that for an eigenvalue λ of a matrix $A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is an arbitrary field), we have that:

$$\text{geometric multiplicity of } \lambda \quad \leq \quad \text{algebraic multiplicity of } \lambda.$$

We postpone the proof of Theorem 8.2.3 to the very end of this section (see subsection 8.2.8); no result of the present section relies on this theorem.

The spectrum of a square matrix. The *spectrum* of a square matrix $A \in \mathbb{F}^{n\times n}$ is the multiset of all eigenvalues of A, with algebraic multiplicities taken into account.[6] For example, if a matrix $A \in \mathbb{C}^{5\times 5}$ has eigenvalues 1 (with algebraic multiplicity 1), $1+i$ (with algebraic multiplicity 2), and $1-i$ (with algebraic multiplicity 2), then the spectrum of A is $\{1, 1+i, 1+i, 1-i, 1-i\}$. In general, the spectrum of a matrix

[4] In other words, k is the largest integer such that there exists some polynomial $q(\lambda)$ with coefficients in $\mathbb{F}$ such that $p_A(\lambda) = (\lambda - \lambda_0)^k q(\lambda)$.

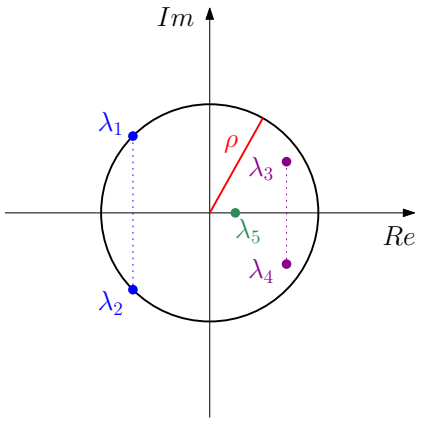
[5] Indeed, if $\mathbb{F}$ is algebraically closed, then the characteristic polynomial $p_A(\lambda)$ can be written as a product of linear factors, and there are n of those factors. If $\mathbb{F}$ is not algebraically closed, we might or might not be able to factor $p_A(\lambda)$ in this way, which is why the sum of algebraic multiplicities of the eigenvalues of A is at most n (possibly strictly smaller than n).

[6] This means that the number of times that an eigenvalue appears in the spectrum is equal to the algebraic multiplicity of that eigenvalue. The order in which we list the eigenvalues in the spectrum does not matter, but repetitions do matter.

$A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) has at most n elements; if the field $\mathbb{F}$ is algebraically closed, then the spectrum of A has exactly n elements.

The spectral radius. For a matrix $A \in \mathbb{C}^{n\times n}$, the *spectral radius* of A, denoted by $\rho(A)$, is the maximum absolute value of any eigenvalue of A. For example, if the spectrum of a matrix $A \in \mathbb{C}^{5\times 5}$ is $\{1, 1+i, 1+i, 1-i, 1-i\}$, then the spectral radius of A is $\rho(A) = \max\{|1|, |1+i|, |1+i|, |1-i|, |1-i|\} = \sqrt{2}$.[7]

In view of Theorems 0.3.6 and 8.2.2, we can visualize the complex eigenvalues of an $n \times n$ matrix A with **real** entries (however, we consider A to be a matrix in the vector space $\mathbb{C}^{n\times n}$, so that it can have complex eigenvalues). Its characteristic polynomial $p_A(\lambda)$ is of degree n and has real coefficients. By Theorem 0.3.6, the roots of this polynomial come in conjugate pairs,[8] and moreover, by Theorem 8.2.2, those roots are precisely the eigenvalues of A. The eigenvalues all lie in the complex plane, in the disk centered at the origin and with radius $\rho(A)$, and they are symmetric about the real axis. Visually, the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ of a matrix $A \in \mathbb{C}^{5\times 5}$ with real entries might appear as in the picture below (the conjugate pairs are color coded for emphasis).



8.2.2 Some numerical examples

Example 8.2.4. *Consider the following matrix in* $\mathbb{C}^{3\times 3}$*:*

$$A = \begin{bmatrix} 4 & 0 & -2 \\ 2 & 5 & 4 \\ 0 & 0 & 5 \end{bmatrix}.$$

(a) Compute the characteristic polynomial $p_A(\lambda)$ *of the matrix* A*.*

(b) Compute all the eigenvalues of A *and their algebraic multiplicities, and compute the spectrum of* A*.*

[7] Indeed, $|1| = 1$, $|1+i| = \sqrt{2}$, and $|1-i| = \sqrt{2}$. So, $\rho(A) = \sqrt{2}$.
[8] Each real root is its own conjugate pair.

(c) For each eigenvalue λ of A, compute a basis of the eigenspace $E_\lambda(A)$ and specify the geometric multiplicity of the eigenvalue λ.

Solution. (a) The characteristic polynomial of A is:

$$\begin{aligned} p_A(\lambda) &= \det(\lambda I_3 - A) \\ &= \begin{vmatrix} \lambda-4 & 0 & 2 \\ -2 & \lambda-5 & -4 \\ 0 & 0 & \lambda-5 \end{vmatrix} \\ &\overset{(*)}{=} (\lambda-4)(\lambda-5)^2 \\ &= \lambda^3 - 14\lambda^2 + 65\lambda - 100, \end{aligned}$$

where the easiest way to obtain (*) is via Laplace expansion along the second column.

Remark: We did not really need to expand in the last line. We only really care about the roots of the characteristic polynomial, and it is more convenient to have a form that is already factored. So, $p_A(\lambda) = (\lambda-4)(\lambda-5)^2$ is a "better" answer than $p_A(\lambda) = \lambda^3 - 14\lambda^2 + 65\lambda - 100$, although they are both correct.

(b) From part (a), we see that A has two eigenvalues, namely, the eigenvalue $\lambda_1 = 4$ (with algebraic multiplicity 1), and the eigenvalue $\lambda_2 = 5$ (with algebraic multiplicity 2). So, the spectrum of A is $\{4, 5, 5\}$.

(c) For each $i \in \{1, 2\}$, we have that

$$E_{\lambda_i}(A) = \mathrm{Nul}\big(\lambda_i I_3 - A\big),$$

which is precisely the set of all solutions of the characteristic equation

$$(\lambda_i I_3 - A)\mathbf{x} = \mathbf{0}.$$

Let us now compute a basis of each of the two eigenspaces.

For $\lambda_1 = 4$, we have that

$$\lambda_1 I_3 - A = \begin{bmatrix} \lambda_1-4 & 0 & 2 \\ -2 & \lambda_1-5 & -4 \\ 0 & 0 & \lambda_1-5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 2 \\ -2 & -1 & -4 \\ 0 & 0 & -1 \end{bmatrix},$$

and that

$$\mathrm{RREF}(\lambda_1 I_3 - A) = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Consequently, the general solution of the equation $(\lambda_1 I_3 - A)\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = \begin{bmatrix} -t/2 \\ t \\ 0 \end{bmatrix} = t \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} = \tfrac{t}{2} \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \qquad \text{with } t \in \mathbb{C}.$$

So, $\left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\}$ is a basis of the eigespace $E_{\lambda_1}(A) = \mathrm{Nul}\big(A - \lambda_1 I_n\big)$,[9] and we see that the eigenvalue $\lambda_1 = 4$ has geometric multiplicity 1.

For $\lambda_2 = 5$, we have that

$$\lambda_2 I_3 - A = \begin{bmatrix} \lambda_2 - 4 & 0 & 2 \\ -2 & \lambda_2 - 5 & -4 \\ 0 & 0 & \lambda_2 - 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ -2 & 0 & -4 \\ 0 & 0 & 0 \end{bmatrix},$$

and that

$$\mathrm{RREF}(\lambda_2 I_3 - A) = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Consequently, the general solution of the equation $(\lambda_2 I_3 - A)\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = \begin{bmatrix} -2t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \qquad \text{with } s, t \in \mathbb{C}.$$

So, $\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis of the eigenspace $E_{\lambda_2}(A) = \mathrm{Nul}\big(A - \lambda_2 I_n\big)$, and we see that the eigenvalue $\lambda_2 = 5$ has geometric multiplicity 2. □

Example 8.2.5. *Consider the following matrix in $\mathbb{R}^{2\times 2}$:*

$$B = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

(Note that B is the standard matrix of counterclockwise rotation by $45°$ about the origin in $\mathbb{R}^2$.)

(a) Compute the characteristic polynomial $p_B(\lambda)$ of the matrix B.

[9] It is also true that $\left\{ \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis of E_{λ_1}. However, it is nicer to get integers (when possible).

(b) Compute all the (real) eigenvalues of B and their algebraic multiplicities.

> ***Remark:*** *Since we consider B to be a matrix in $\mathbb{R}^{2\times 2}$, we need to look for* ***real*** *eigenvalues only.*

(c) For each eigenvalue λ of B, compute a basis of the eigenspace $E_\lambda(B)$ and specify the geometric multiplicity of the eigenvalue λ.

Solution. (a) The characteristic polynomial of B is:

$$\begin{aligned} p_B(\lambda) &= \det(\lambda I_2 - B) \\ &= \begin{vmatrix} \lambda - \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \lambda - \frac{1}{\sqrt{2}} \end{vmatrix} \\ &= (\lambda - \tfrac{1}{\sqrt{2}})(\lambda - \tfrac{1}{\sqrt{2}}) - \tfrac{1}{\sqrt{2}}(-\tfrac{1}{\sqrt{2}}) \\ &= \lambda^2 - \sqrt{2}\lambda + 1. \end{aligned}$$

(b,c) We need to find any real roots that the polynomial $p_B(\lambda)$ may have, i.e. any real solutions that the quadratic equation

$$\lambda^2 - \sqrt{2}\lambda + 1 = 0$$

may have. The discriminant of this quadratic equation is $(-\sqrt{2})^2 - 4 \cdot 1 \cdot 1 = -2 < 0$, and it follows that the equation has no real solutions. Therefore, B has no real eigenvalues, and it follows that the spectrum of B is empty.

Remark: The fact that B has no eigenvectors (and consequently, no eigenvalues) also follows from geometric considerations. Indeed, B is the standard matrix of counterclockwise rotation by 45° about the origin in $\mathbb{R}^2$. So, no non-zero vector in $\mathbb{R}^2$ simply gets scaled when we multiply it on the left by B, which by definition means that B has no eigenvectors. □

Example 8.2.6. *Consider the following matrix in $\mathbb{C}^{2\times 2}$:*

$$C = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

> ***Remark:*** *This is the same as the matrix B from Example 8.2.5, but this time, we consider the matrix to be in $\mathbb{C}^{2\times 2}$.*

(a) Compute the characteristic polynomial $p_C(\lambda)$ of the matrix C.

(b) Compute all the eigenvalues of C and their algebraic multiplicities.

Remark: *Since we consider C to be a matrix in $\mathbb{C}^{2\times 2}$, we need to look for* ***complex*** *eigenvalues. (Note that all real numbers are complex! So, if our eigenvalues ended up being real, they would still count as complex eigenvalues.)*

(c) For each eigenvalue λ of C, compute a basis of the eigenspace $E_\lambda(C)$ and specify the geometric multiplicity of the eigenvalue λ.

Solution. (a) The characteristic polynomial of C is the same as the characteristic polynomial of the matrix B from Example 8.2.5, since the two characteristic polynomials are computed in exactly the same way. Indeed,

$$\begin{aligned} p_C(\lambda) &= \det(\lambda I_2 - C) \\ &= \begin{vmatrix} \lambda - \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \lambda - \frac{1}{\sqrt{2}} \end{vmatrix} \\ &= (\lambda - \tfrac{1}{\sqrt{2}})(\lambda - \tfrac{1}{\sqrt{2}}) - \tfrac{1}{\sqrt{2}}(-\tfrac{1}{\sqrt{2}}) \\ &= \lambda^2 - \sqrt{2}\lambda + 1. \end{aligned}$$

(b) We need to find the (complex) roots of the characteristic polynomial $p_C(\lambda)$, together with their algebraic multiplicities. The quadratic equation

$$\underbrace{\lambda^2 - \sqrt{2}\lambda + 1}_{=p_C(\lambda)} = 0$$

has solutions

$$\lambda_{1,2} = \tfrac{-(-\sqrt{2}) \pm \sqrt{(-\sqrt{2})^2 - 4\cdot 1\cdot 1}}{2\cdot 1} = \tfrac{\sqrt{2} \pm \sqrt{-2}}{2} = \tfrac{1 \pm i}{\sqrt{2}},$$

that is,

$$\lambda_1 = \tfrac{1+i}{\sqrt{2}} \quad \text{and} \quad \lambda_2 = \tfrac{1-i}{\sqrt{2}}.$$

Complex numbers λ_1 and λ_2 are the eigenvalues of the matrix C, and they each have algebraic multiplicity 1, since $p_C(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)$. The spectrum of C is $\{\lambda_1, \lambda_2\} = \left\{\frac{1+i}{\sqrt{2}}, \frac{1-i}{\sqrt{2}}\right\}$.

(c) For each $i \in \{1, 2\}$, the eigenspace $E_{\lambda_i}(C)$ is precisely the set of solutions of the characteristic equation

$$(\lambda_i I_2 - C)\mathbf{x} = \mathbf{0}.$$

For $\lambda_1 = \frac{1+i}{\sqrt{2}}$, we have that

$$\lambda_1 I_2 - C = \begin{bmatrix} \lambda_1 - \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \lambda_1 - \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \end{bmatrix},$$

and that

$$\mathrm{RREF}(\lambda_1 I_2 - C) \;=\; \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}.$$

Consequently, the general solution of the equation $(\lambda_1 I_2 - C)\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} \;=\; \begin{bmatrix} it \\ t \end{bmatrix} \;=\; t \begin{bmatrix} i \\ 1 \end{bmatrix}, \qquad \text{with } t \in \mathbb{C}.$$

So, $\left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\}$ is a basis of the eigenspace $E_{\lambda_1}(C) = \mathrm{Nul}\big(\lambda_1 I_2 - C\big)$, and we see that the eigenvalue $\lambda_1 = \frac{1+i}{\sqrt{2}}$ has geometric multiplicity 1.

For $\lambda_2 = \frac{1-i}{\sqrt{2}}$, we have that

$$\lambda_2 I_2 - C \;=\; \begin{bmatrix} \lambda_2 - \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \lambda_2 - \frac{1}{\sqrt{2}} \end{bmatrix} \;=\; \begin{bmatrix} -\frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \end{bmatrix},$$

and that

$$\mathrm{RREF}(\lambda_2 I_2 - C) \;=\; \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}.$$

Consequently, the general solution of the equation $(\lambda_2 I_2 - C)\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} \;=\; \begin{bmatrix} -it \\ t \end{bmatrix} \;=\; t \begin{bmatrix} -i \\ 1 \end{bmatrix}, \qquad \text{with } t \in \mathbb{C}.$$

So, $\left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix} \right\}$ is a basis of the eigenspace $E_{\lambda_2}(C) = \mathrm{Nul}\big(\lambda_2 I_2 - C\big)$, and we see that the eigenvalue $\lambda_2 = \frac{1-i}{\sqrt{2}}$ has geometric multiplicity 1. □

Eigenvectors and eigenvalues of triangular matrices. By Proposition 7.3.1, the determinant of a triangular matrix is equal to the product of its entries on the main diagonal. This immediately yields the following proposition.

Proposition 8.2.7. *Let $\mathbb{F}$ be a field, and let $A = \big[\ a_{i,j}\ \big]_{n\times n}$ be a triangular matrix in $\mathbb{F}^{n\times n}$. Then the characteristic polynomial of A is*

$$p_A(\lambda) \;=\; \prod_{i=1}^{n} (\lambda - a_{i,i}) \;=\; (\lambda - a_{1,1})(\lambda - a_{2,2}) \dots (\lambda - a_{n,n}),$$

the eigenvalues of A are precisely the entries of A on its main diagonal, and moreover, the algebraic multiplicity of each eigenvalue is precisely the number of times that it appears on the main diagonal of A.[10] *Consequently, the spectrum of A is*

[10] However, the geometric multiplicity may possibly be smaller, as Example 8.2.8 shows. We note, however, that the geometric multiplicity will never be larger than this, as per Theorem 8.2.3 (which we have not proven yet).

$\{a_{1,1}, a_{2,2}, \ldots, a_{n,n}\}$, *i.e. the multiset formed precisely by the main diagonal entries of A, with each number appearing in the spectrum of A the same number of times as on the main diagonal of A.*

Proof. Since A is triangular, so is the matrix $\lambda I_n - A$; so, the determinant of $\lambda I_n - A$ can be computed simply by multiplying its entries on the main diagonal. It follows that the characteristic polynomial of A is

$$p_A(\lambda) \;=\; \det(\lambda I_n - A) \;=\; (\lambda - a_{1,1})(\lambda - a_{2,2}) \ldots (\lambda - a_{n,n}),$$

and the result follows. □

Example 8.2.8. *Consider the following matrix in $\mathbb{C}^{5\times 5}$:*

$$A \;=\; \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

(a) Compute the characteristic polynomial $p_A(\lambda)$ of the matrix A.

(b) Compute all the eigenvalues of A and their algebraic multiplicities, and compute the spectrum of A.

(c) For each eigenvalue λ of A, compute a basis of the eigenspace $E_\lambda(A)$ and specify the geometric multiplicity of the eigenvalue λ.

Solution. (a) The matrix A is upper triangular, and so its characteristic polynomial is

$$p_A(\lambda) \;=\; \det\big(\lambda I_5 - A\big) \;=\; \begin{vmatrix} \lambda - 1 & -2 & 0 & 0 & 0 \\ 0 & \lambda - 2 & 0 & 0 & 0 \\ 0 & 0 & \lambda - 1 & -1 & -3 \\ 0 & 0 & 0 & \lambda - 3 & -3 \\ 0 & 0 & 0 & 0 & \lambda - 3 \end{vmatrix}$$

$$= \; (\lambda - 1)^2(\lambda - 2)(\lambda - 3)^2.$$

(b) We see from part (a) that A has three eigenvalues, namely, $\lambda_1 = 1$ (with algebraic multiplicity 2), $\lambda_2 = 2$ (with algebraic multiplicity 1), and $\lambda = 3$ (with algebraic multiplicity 2).[11] So, the spectrum of A is $\{1, 1, 2, 3, 3\}$.

[11] We could also have obtained the same answer by noticing that A is triangular, and that 1 appears twice on the main diagonal of A, 2 appears once, and 3 appears twice.

(c) For each $i \in \{1, 2, 3\}$, the eigenspace $E_{\lambda_i}(A)$ is precisely the set of solutions of the characteristic equation

$$(\lambda_i I_5 - A)\mathbf{x} \;=\; \mathbf{0}.$$

For $\lambda_1 = 1$, we have that

$$\lambda_1 I_5 - A \;=\; \begin{bmatrix} 0 & -2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -3 \\ 0 & 0 & 0 & -2 & -3 \\ 0 & 0 & 0 & 0 & -2 \end{bmatrix},$$

and that

$$\mathrm{RREF}(\lambda_1 I_5 - A) \;=\; \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Consequently, the general solution of the equation $(\lambda_1 I_5 - A)\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} \;=\; \begin{bmatrix} s \\ 0 \\ t \\ 0 \\ 0 \end{bmatrix} \;=\; s \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \text{with } s, t \in \mathbb{C}.$$

So,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

is a basis of the eigenspace $E_{\lambda_1}(A) = \mathrm{Nul}\big(\lambda_1 I_5 - A\big)$, and we see that the eigenvalue $\lambda_1 = 1$ has geometric multiplicity 2.

For $\lambda_2 = 2$, we have that

$$\lambda_2 I_5 - A \;=\; \begin{bmatrix} 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & -1 & -3 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix},$$

and that

$$\mathrm{RREF}(\lambda_2 I_5 - A) = \begin{bmatrix} 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Consequently, the general solution of the equation $(\lambda_2 I_5 - A)\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = \begin{bmatrix} 2t \\ t \\ 0 \\ 0 \\ 0 \end{bmatrix} = t \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \text{with } t \in \mathbb{C}.$$

So,

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

is a basis of the eigenspace $E_{\lambda_2}(A) = \mathrm{Nul}\big(\lambda_2 I_5 - A\big)$, and we see that the eigenvalue $\lambda_2 = 2$ has geometric multiplicity 1.

For $\lambda_3 = 3$, we have that

$$\lambda_3 I_5 - A = \begin{bmatrix} 2 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & -3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and that

$$\mathrm{RREF}(\lambda_3 I_5 - A) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Consequently, the general solution of the equation $(\lambda_3 I_5 - A)\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ \frac{t}{2} \\ t \\ 0 \end{bmatrix} = t \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix} = \frac{t}{2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \qquad \text{with } t \in \mathbb{C}.$$

So,

$$\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 0 \end{bmatrix} \right\}$$

is a basis of the eigenspace $E_{\lambda_3}(A)$, and we see that the eigenvalue $\lambda_3 = 3$ has geometric multiplicity 1. □

8.2.3 Eigenvectors and eigenvalues of similar matrices

Recall from subsection 4.5.2 that matrices $A, B \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) are said to be *similar* if there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$. By Proposition 4.5.13, matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$. Moreover, by Theorem 4.5.16, two matrices in $\mathbb{F}^{n\times n}$ are similar if and only if they are matrices of the same linear function from an n-dimensional vector space (over $\mathbb{F}$) to itself, but possibly with respect to different bases. When it comes to eigenvalues and eigenvectors, we have the following result for similar matrices.

Theorem 8.2.9. *Let $\mathbb{F}$ be a field, and let $A, B \in \mathbb{F}^{n\times n}$ be similar matrices. Then A and B have the same characteristic polynomial, as well as the same eigenvalues, with the same corresponding algebraic multiplicities, and the same corresponding geometric multiplicities. Moreover, A and B have the same spectrum.*

Warning: Similar matrices A and B need not have the same eigenspaces, that is, for an eigenvalue λ of A and B:

$$E_\lambda(A) \;\not\asymp\; E_\lambda(B)$$

Proof. Let us first show that A and B have the same eigenvalues with the same corresponding geometric multiplicities. Since A and B are similar, Theorem 4.5.16 guarantees that there exists a linear function $f : \mathbb{F}^n \to \mathbb{F}^n$ and bases $\mathcal{A}$ and $\mathcal{B}$ of $\mathbb{F}^n$ such that $A = {}_{\mathcal{A}}\big[\ f\ \big]_{\mathcal{A}}$ and $B = {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. But then by Proposition 8.1.7, the linear function f and the matrix $A = {}_{\mathcal{A}}\big[\ f\ \big]_{\mathcal{A}}$ have exactly the same eigenvalues, with exactly the same corresponding geometric multiplicities, and the same holds for f and the matrix $B = {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. So, A and B have exactly the same eigenvalues with exactly the same corresponding geometric multiplicities.

It now remains to show that A and B have the same characteristic polynomial, since this will (by definition) imply that A and B have the same spectrum, and in particular, that the eigenvalues of A and B have the same corresponding algebraic multiplicities. Since A and B are similar, we know that there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$. We now compute:

$$\begin{aligned}
p_B(\lambda) &= \det\big(\lambda I_n - B\big) \\
&= \det\big(\lambda I_n - P^{-1}AP\big) \\
&= \det\big(P^{-1}(\lambda I_n - A)P\big) \\
&= \det(P^{-1})\det\big(\lambda I_n - A\big)\det(P) && \text{by Theorem 7.5.2} \\
&= \tfrac{1}{\det(P)}\det\big(\lambda I_n - A\big)\det(P) && \text{by Corollary 7.5.3} \\
&= \det\big(\lambda I_n - A\big) \\
&= p_A(\lambda).
\end{aligned}$$

This completes the argument. □

Remark: The converse of Theorem 8.2.9 is false: two matrices in $\mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) that have the same characteristic polynomial, as well as the same eigenvalues, with the same corresponding algebraic multiplicities, and the same corresponding geometric multiplicities, need not be similar. We will see examples of this when we study the "Jordan normal form" (see section 8.6).

8.2.4 A relationship between the spectrum, trace, and determinant of a matrix

The *trace* of a square matrix $A = \big[\ a_{i,j}\ \big]_{n\times n}$ with entries in some field $\mathbb{F}$ is defined to be $\text{trace}(A) := \sum_{i=1}^{n} a_{i,i}$, i.e. the trace of A is the sum of entries on the main diagonal of A. For example, for the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

in $\mathbb{C}^{3\times 3}$, we have that $\text{trace}(A) = 1 + 5 + 9 = 15$.

Theorem 8.2.10. *Let $\mathbb{F}$ be a field, let $A = \big[\ a_{i,j}\ \big]_{n\times n}$ be a matrix in $\mathbb{F}^{n\times n}$, and assume that $\{\lambda_1, \ldots, \lambda_n\}$ is the spectrum of A. Then*

(a) $det(A) = \lambda_1 \ldots \lambda_n$;

(b) $trace(A) = \lambda_1 + \cdots + \lambda_n$.

Warning: Theorem 8.2.10 only applies if the spectrum of the matrix $A \in \mathbb{F}^{n\times n}$ contains n eigenvalues (counting algebraic multiplicities)! This will always be the case if the field $\mathbb{F}$ is **algebraically closed** (for example, if $\mathbb{F} = \mathbb{C}$), but need not be the case otherwise.

Proof. By definition, we have that $p_A(\lambda) = \det(\lambda I_n - A)$. On the other hand, since $\{\lambda_1, \ldots, \lambda_n\}$ is the spectrum of A (and A is an $n \times n$ matrix), we see that $p_A(\lambda) = (\lambda - \lambda_1) \ldots (\lambda - \lambda_n)$.

(a) By setting $\lambda = 0$, we obtain

$$p_A(0) \quad = \quad (0 - \lambda_1) \ldots (0 - \lambda_n) \quad = \quad (-1)^n \lambda_1 \ldots \lambda_n.$$

On the other hand, we have that

$$p_A(0) \quad = \quad \det(0I_n - A) \quad = \det(-A) \quad \overset{(*)}{=} \quad (-1)^n \det(A),$$

where (*) follows from Proposition 7.2.3. It now follows that $(-1)^n \lambda_1 \ldots \lambda_n = (-1)^n \det(A)$, and consequently, $\det(A) = \lambda_1 \ldots \lambda_n$.

(b) We will compute the coefficient in front of λ^{n-1} in the characteristic polynomial $p_A(\lambda)$ in two ways.

First, since $p_A(\lambda) = (\lambda - \lambda_1) \ldots (\lambda - \lambda_n)$, it is clear that the coefficient in front of λ^{n-1} is $-\lambda_1 - \cdots - \lambda_n$.

On the other hand, we have that

$$p_A(\lambda) \quad = \quad \det(\lambda I_n - A) \quad = \quad \begin{vmatrix} \lambda - a_{1,1} & -a_{1,2} & \ldots & -a_{1,n} \\ -a_{2,1} & \lambda - a_{2,2} & \ldots & -a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n,1} & -a_{n,2} & \ldots & \lambda - a_{n,n} \end{vmatrix}.$$

We now use the definition of the determinant: the only permutation $\sigma \in S_n$ that produces a polynomial with λ^{n-1} appearing with it (with a possibly non-zero coefficient) is the identity permutation $\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ 1 & 2 & \ldots & n \end{pmatrix}$,[12] and clearly, this permutation is even, i.e. has sign 1. So, the coefficient in front of λ^{n-1} in $p_A(\lambda)$ is equal to the coefficient of λ^{n-1} in the product $(\lambda - a_{1,1})(\lambda - a_{2,2}) \ldots (\lambda - a_{n,n})$, and this coefficient is precisely $-a_{1,1} - a_{2,2} - \cdots - a_{n,n} = -\text{trace}(A)$.

We have now computed the coefficient in front of λ^{n-1} in the polynomial $p_A(\lambda)$ in two ways: we got $-\lambda_1 - \cdots - \lambda_n$ the first time, and we got $-\text{trace}(A)$ the second time. So, $-\lambda_1 - \cdots - \lambda_n = -\text{trace}(A)$, and it follows that $\text{trace}(A) = \lambda_1 + \cdots + \lambda_n$. □

8.2.5 Eigenvalues and invertibility

Proposition 8.2.11. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$. Then A is invertible if and only if 0 is **not** an eigenvalue of A.*

Proof. It suffices to show that 0 is an eigenvalue of A if and only if A is not invertible. We have the following sequence of equivalent statements:

[12] Note that the identity permutation encodes the selection of the entire main diagonal.

$$\begin{array}{lll} 0 \text{ is eigenvalue of } A & \iff & \det(0I_n - A) = 0 \qquad \text{by Theorem 8.2.2} \\ & \iff & \det(-A) = 0 \\ & \iff & (-1)^n \det(A) = 0 \qquad \text{by Proposition 7.2.3} \\ & \iff & \det(A) = 0 \\ & \iff & A \text{ is not invertible} \qquad \text{by Theorem 7.4.1.} \end{array}$$

This completes the argument. □

8.2.6 The Invertible Matrix Theorem (version 4)

In this subsection, we state and prove the fourth and final version of the Invertible Matrix Theorem. This version of the theorem is obtained from the previous one (see subsection 7.4.1) by adding the eigenvalue condition from Proposition 8.2.11 as the last item. Note that our final version of the Invertible Matrix Theorem uses up all 26 letters of the English alphabet!

The Invertible Matrix Theorem (version 4). *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$ be a **square** matrix. Further, let $f : \mathbb{F}^n \to \mathbb{F}^n$ be given by $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$.*[13] *Then the following are equivalent:*

(a) A is invertible (i.e. A has an inverse);

(b) A^T is invertible;

(c) $RREF(A) = I_n$;

(d) $RREF\left(\left[\begin{array}{c|c} A & I_n \end{array}\right]\right) = \left[\begin{array}{c|c} I_n & B \end{array}\right]$ for some matrix $B \in \mathbb{F}^{n \times n}$;

(e) $rank(A) = n$;

(f) $rank(A^T) = n$;

(g) A is a product of elementary matrices;

(h) the homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$);

(i) there exists some vector $\mathbf{b} \in \mathbb{F}^n$ such that the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;

(j) for all vectors $\mathbf{b} \in \mathbb{F}^n$, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution;

[13] Since f is a matrix transformation, Proposition 1.10.4 guarantees that f is linear. Moreover, A is the standard matrix of f.

(k) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *has at most one solution;*

(l) for all vectors $\mathbf{b} \in \mathbb{F}^n$*, the matrix-vector equation* $A\mathbf{x} = \mathbf{b}$ *is consistent;*

(m) f *is one-to-one;*

(n) f *is onto;*

(o) f *is an isomorphism;*

(p) there exists a matrix $B \in \mathbb{F}^{n\times n}$ *such that* $BA = I_n$ *(i.e.* A *has a left inverse);*

(q) there exists a matrix $C \in \mathbb{F}^{n\times n}$ *such that* $AC = I_n$ *(i.e.* A *has a right inverse);*

(r) the columns of A *are linearly independent;*

(s) the columns of A *span* $\mathbb{F}^n$ *(i.e.* $Col(A) = \mathbb{F}^n$*);*

(t) the columns of A *form a basis of* $\mathbb{F}^n$*;*

(u) the rows of A *are linearly independent;*

(v) the rows of A *span* $\mathbb{F}^{1\times n}$ *(i.e.* $Row(A) = \mathbb{F}^{1\times n}$*);*

(w) the rows of A *form a basis of* $\mathbb{F}^{1\times n}$*;*

(x) $Nul(A) = \{\mathbf{0}\}$ *(i.e.* $dim\big(Nul(A)\big) = 0$*);*

(y) $det(A) \neq 0$*;*

(z) 0 *is not an eigenvalue of* A*.*

Proof. Items (a)-(y) are the same as those from the Invertible Matrix Theorem (version 3) from subsection 7.4.1. The equivalence of (a) and (z) follows from Proposition 8.2.11. □

8.2.7 The characteristic polynomial and spectrum of a linear function

In section 7.5, we defined the determinant of a linear function whose domain and codomain and one and the same non-trivial, finite-dimensional vector space. Before proceeding, let us recall this definition. Suppose that V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and that $f : V \to V$ is a linear function. Then we define the *determinant* of f to be

$$\det(f) \;:=\; \det\Big({}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}\Big),$$

where $\mathcal{B}$ is any basis of V. As we explained in section 7.5, the reason that $\det(f)$ is well defined is because, by Theorem 4.5.16, all matrices of the form ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ are similar, and therefore (by Corollary 7.5.4) have the same determinant.

We can similarly define the characteristic polynomial of linear functions, as long as their domain and codomain are one and the same non-trivial, finite-dimensional vector space. So, let us once again suppose that V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$. As usual, Id_V is the identity function on V, i.e. it is the function $\mathrm{Id}_V : V \to V$ given by $\mathrm{Id}_V(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$. The *characteristic polynomial* of a linear function $f : V \to V$ is defined to be the polynomial

$$p_f(\lambda) \ := \ \det\big(\lambda \mathrm{Id}_V - f\big) \ = \ \det\Big({}_{\mathcal{B}}[\ \lambda \mathrm{Id}_V - f\ ]_{\mathcal{B}}\Big),$$

where $\mathcal{B}$ is **any** basis of V. As per our discussion above, the polynomial $p_f(\lambda)$ depends only on f, and not on the particular choice of the basis $\mathcal{B}$. The *characteristic equation* of f is the equation

$$\det\big(\lambda \mathrm{Id}_V - f\big) \ = \ 0.$$

So, the roots of the characteristic polynomial of f are precisely the solutions of the characteristic equation of f.

By Theorem 8.2.9, similar matrices have the same characteristic polynomial. In view of the characterization of similar matrices given by Theorem 4.5.16, it should not be surprising that the characteristic polynomial of a linear function is exactly the same as the characteristic polynomial of its matrices (as long as we use the same basis for the domain and codomain). More precisely, we have the following proposition.

Proposition 8.2.12. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, let $\mathcal{B}$ be any basis of V, let $f : V \to V$ be a linear function, and set $B := {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$. Then $p_f(\lambda) = p_B(\lambda)$.*

Proof. Set $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, and as usual, let $\mathcal{E}_n = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{F}^n$. By Theorem 4.5.1, we have that

$$\begin{aligned} {}_{\mathcal{B}}[\ \mathrm{Id}_V\ ]_{\mathcal{B}} \ &= \ [\ [\ \mathrm{Id}_V(\mathbf{b}_1)\ ]_{\mathcal{B}} \ \ \ldots \ \ [\ \mathrm{Id}_V(\mathbf{b}_n)\ ]_{\mathcal{B}}\ ] \\ &= \ [\ [\ \mathbf{b}_1\ ]_{\mathcal{B}} \ \ \ldots \ \ [\ \mathbf{b}_n\ ]_{\mathcal{B}}\ ] \\ &= \ [\ \mathbf{e}_1 \ \ \ldots \ \ \mathbf{e}_n\ ] \ = \ I_n. \end{aligned}$$

We now compute:

$$\begin{aligned} p_f(\lambda) \ &= \ \det\big(\lambda \mathrm{Id}_V - f\big) && \text{by definition} \\ &= \ \det\Big({}_{\mathcal{B}}[\ \lambda \mathrm{Id}_V - f\ ]_{\mathcal{B}}\Big) && \text{by definition} \end{aligned}$$

$$\begin{aligned} &= \det\Big(\lambda\, {}_{\mathcal{B}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}} - {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}\Big) && \text{by Theorem 4.5.3} \\ &= \det\big(\lambda I_n - B\big) \\ &= p_B(\lambda) && \text{by definition.} \end{aligned}$$

This completes the argument. □

We also have the following analog of Theorem 8.2.2.

Theorem 8.2.13. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, let $f : V \to V$ be a linear function, and let $\lambda_0 \in \mathbb{F}$. Then*

$$E_{\lambda_0}(f) \;=\; Ker\big(\lambda_0 Id_V - f\big) \;=\; Ker\big(f - \lambda_0 Id_V\big).$$

Moreover, the following are equivalent:

(1) λ_0 is an eigenvalue of f;

(2) λ_0 is a root of the characteristic polynomial of f, i.e. $p_f(\lambda_0) = 0$;

(3) λ_0 is a solution of the characteristic equation of f, i.e. $det\big(\lambda_0 Id_V - f\big) = 0$.

Proof. We proceed similarly as in the proof of Theorem 8.2.2. Obviously, for all $\mathbf{v} \in V$, we have that $\big(\lambda_0 \mathrm{Id}_V - f\big)(\mathbf{v}) = \mathbf{0}$ if and only if $\big(f - \lambda_0 \mathrm{Id}_V\big)(\mathbf{v}) = \mathbf{0}$. So, $\mathrm{Ker}\big(\lambda_0 \mathrm{Id}_V - f\big) = \mathrm{Ker}\big(f - \lambda_0 \mathrm{Id}_V\big)$. Further, we compute:

$$\begin{aligned} E_{\lambda_0}(A) &= \big\{\mathbf{v} \in V \mid f(\mathbf{v}) = \lambda_0 \mathbf{v}\big\} \\ &= \big\{\mathbf{v} \in V \mid f(\mathbf{v}) = (\lambda_0 \mathrm{Id}_V)(\mathbf{v})\big\} \\ &= \big\{\mathbf{v} \in \mathbb{F}^n \mid (\lambda_0 \mathrm{Id}_V - f)(\mathbf{v}) = \mathbf{0}\big\} \\ &= \mathrm{Ker}\big(\lambda_0 \mathrm{Id}_V - f\big). \end{aligned}$$

It remains to show that (1), (2), and (3) are equivalent. The fact that (2) and (3) are equivalent follows immediately from the definition of the characteristic polynomial and the characteristic equation of f. It remains to show that (1) and (2) are equivalent. For this, we fix any basis $\mathcal{B}$ of V, and set $B := {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. We then have the following sequence of equivalent statements:

$$\underbrace{\lambda_0 \text{ is an eigenvalue of } f}_{(1)} \qquad \overset{(*)}{\Longleftrightarrow} \qquad \lambda_0 \text{ is an eigenvalue of } B$$

$$\overset{(**)}{\iff} \quad \lambda_0 \text{ is a root of } p_B(\lambda)$$

$$\overset{(***)}{\iff} \quad \underbrace{\lambda_0 \text{ is a root of } p_f(\lambda)}_{(2)},$$

where (*) follows from Proposition 8.1.7, (**) follows from Theorem 8.2.2, and (***) follows from the fact that $p_f(\lambda) = p_B(\lambda)$ (by Proposition 8.2.12). □

Suppose that $f : V \to V$ is a linear function, where V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$. In view of Theorem 8.2.13, we may define the *algebraic multiplicity* of an eigenvalue λ_0 of f to be the largest positive integer k such that $(\lambda - \lambda_0)^k$ divides the polynomial $p_f(\lambda)$. The *spectrum* of f is the multiset of all the eigenvalues of f, with algebraic multiplicities taken into account.

Proposition 8.2.14. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, let $f : V \to V$ be a linear function, and let $\mathcal{B}$ be any basis of V. Then f and ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ have the same characteristic polynomial, and the same spectrum. Moreover, f and ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ have exactly the same eigenvalues, with exactly the same corresponding geometric multiplicities, and exactly the same corresponding algebraic multiplicities.*

Proof. The fact that f and ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ have the same eigenvalues, with the same geometric multiplicities, follows immediately from Proposition 8.1.7. The fact that they have the same characteristic polynomial (and consequently the same spectrum) follows immediately from Proposition 8.2.12. Since f and ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ have the same spectrum, their eigenvalues have the same algebraic multiplicities. □

As a special case for linear functions of the form $f : \mathbb{F}^n \to \mathbb{F}^n$ (where $\mathbb{F}$ is a field) and their standard matrices, we have the following proposition.

Proposition 8.2.15. *Let $\mathbb{F}$ be a field, let $f : \mathbb{F}^n \to \mathbb{F}^n$ be a linear function, and let A be the standard matrix of f. Then f and A have the same characteristic polynomial and the same spectrum. Moreover, for each eigenvalue λ of f and A, all the following hold:*

- *the algebraic multiplicity of λ as an eigenvalue of f is the same as the algebraic multiplicity of λ as an eigenvalue of A;*
- *the geometric multiplicity of λ as an eigenvalue of f is the same as the geometric multiplicity of λ as an eigenvalue of A;*
- $E_\lambda(f) = E_\lambda(A)$.

Proof. Since A is the standard matrix of f, we have that $A = {}_{\mathcal{E}_n}[\ f\ ]_{\mathcal{E}_n}$, where $\mathcal{E}_n$ is the standard basis of $\mathbb{F}^n$. The result now follows immediately from Propositions 8.1.5 and 8.2.14. □

Example 8.2.16. *Consider the function $f : \mathbb{P}^2_{\mathbb{R}} \to \mathbb{P}^2_{\mathbb{R}}$ given by*

$$f\big(a_2x^2 + a_1x + a_0\big) \;=\; (a_2 + a_0)x^2 - a_1x + (a_2 + a_0)$$

for all $a_0, a_1, a_2 \in \mathbb{R}$. Prove that f is linear, and compute its characteristic polynomial and spectrum. Identify the eigenvalues of f, and for each eigenvalue λ of f, determine its geometric and algebraic multiplicity, and compute a basis of the eigenspace $E_\lambda(f)$.

Solution. We first show that f is linear.

1. Fix polynomials $p(x), q(x) \in \mathbb{P}^2_{\mathbb{R}}$. Then there exist some $a_0, a_1, a_2, b_0, b_1, b_2 \in \mathbb{R}$ such that $p(x) = a_2x^2 + a_1x + a_0$ and $q(x) = b_2x^2 + b_1x + b_0$. We now compute:

$$\begin{aligned}
f\big(p(x) + q(x)\big) &= f\Big((a_2x^2 + a_1x + a_0) + (b_2x^2 + b_1x + b_0)\Big) \\
&= f\Big((a_2 + b_2)x^2 + (a_1 + b_1)x + (a_0 + b_0)\Big) \\
&= \big((a_2 + b_2) + (a_0 + b_0)\big)x^2 - (a_1 + b_1)x + \big((a_2 + b_2) + (a_0 + b_0)\big) \\
&= \big((a_2 + a_0)x^2 - a_1x + (a_2 + a_0)\big) + \big((b_2 + b_0)x^2 - b_1x + (b_2 + b_0)\big) \\
&= f\big(p(x)\big) + f\big(q(x)\big).
\end{aligned}$$

2. Fix a polynomial $p(x) \in \mathbb{P}^2_{\mathbb{R}}$ and a scalar $\alpha \in \mathbb{R}$. Since $p(x) \in \mathbb{P}^2_{\mathbb{R}}$, there exist some $a_0, a_1, a_2 \in \mathbb{R}$ such that $p(x) = a_2x^2 + a_1x + a_0$. We now compute:

$$\begin{aligned}
f\big(\alpha p(x)\big) &= f\Big(\alpha(a_2x^2 + a_1x + a_0)\Big) \\
&= f\Big((\alpha a_2)x^2 + (\alpha a_1)x + (\alpha a_0)\Big) \\
&= \big(\alpha a_2 + \alpha a_0\big)x^2 - (\alpha a_1)x + \big(\alpha a_2 + \alpha a_0\big) \\
&= \alpha\big((a_2 + a_0)x^2 - a_1x + (a_2 + a_0)\big) \\
&= \alpha f\big(p(x)\big).
\end{aligned}$$

From 1 and 2, we see that f is linear.

Now, consider the basis $\mathcal{A} = \{1, x, x^2\}$ of f. We compute:

$$\begin{aligned}
A &:= {}_{\mathcal{A}}\big[\ f\ \big]_{\mathcal{A}} \\
&\overset{(*)}{=} \Big[\ \big[\ f(1)\ \big]_{\mathcal{A}}\ \ \big[\ f(x)\ \big]_{\mathcal{A}}\ \ \big[\ f(x^2)\ \big]_{\mathcal{A}}\ \Big]
\end{aligned}$$

$$= \quad \big[\ \big[\ x^2+1\ \big]_{\mathcal{A}} \quad \big[\ -x\ \big]_{\mathcal{A}} \quad \big[\ x^2+1\ \big]_{\mathcal{A}}\ \big]$$

$$= \quad \begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

where (*) follows from Theorem 4.5.1. We can now compute the characteristic polynomial of f, as follows:

$$\begin{aligned} p_f(\lambda) \overset{(*)}{=} f_A(\lambda) &= \det\big(\lambda I_3 - A\big) \\ &= \begin{vmatrix} \lambda-1 & 0 & -1 \\ 0 & \lambda+1 & 0 \\ -1 & 0 & \lambda-1 \end{vmatrix} \\ &\overset{(**)}{=} (\lambda+1)\begin{vmatrix} \lambda-1 & -1 \\ -1 & \lambda-1 \end{vmatrix} \\ &= (\lambda+1)\big((\lambda-1)^2-1\big) \\ &= \lambda(\lambda-2)(\lambda+1), \end{aligned}$$

where (*) follows from Proposition 8.2.12, and (**) is obtained via Laplace expansion along the second column. Optionally, we can compute the product above to obtain $p_f(\lambda) = \lambda^3 - \lambda^2 - 2\lambda$, although the characteristic polynomial is more useful in factored form. The roots of the characteristic polynomial $p_f(\lambda)$ are $\lambda_1 = 0$, $\lambda_2 = 2$, and $\lambda_3 = -1$, each with multiplicity 1. So, the eigenvalues of f are $\lambda_1 = 0$, $\lambda_2 = 2$, and $\lambda_3 = -1$, each with algebraic multiplicity 1. The spectrum of f is $\{0, 2, -1\}$.

It remains to compute a basis of each of the three eigenspaces, and to determine the geometric multiplicity of each eigenvalue. For each index $i \in \{1, 2, 3\}$, we will first compute a basis of the eigenspace $E_{\lambda_i}(A)$, and then we will use Proposition 8.1.7, plus Theorem 4.4.4(c), to "translate" this bases into a basis of $E_{\lambda_i}(f)$.

We first deal with the eigenvalue $\lambda_1 = 0$. We compute:

$$\mathrm{RREF}\big(\lambda_1 I_3 - A\big) \quad = \quad \left(\begin{bmatrix} -1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & -1 \end{bmatrix}\right) \quad = \quad \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore, $\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis of $E_{\lambda_1}(A) = \mathrm{Nul}\big(\lambda_1 I_3 - A\big)$. But note that

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad = \quad \big[\ x^2 - 1\ \big]_{\mathcal{A}}.$$

By Proposition 8.1.7, $E_{\lambda_1}(A)$ is simply the image of $E_{\lambda_1}(f)$ under the isomorphism $\big[\ \cdot\ \big]_{\mathcal{A}}$. So, by Theorem 4.4.4(c), $\{x^2-1\}$ is a basis of $E_{\lambda_1}(f)$. In particular, the geometric multiplicity of $\lambda_1 = 0$ as an eigenvalue of f is 1.

We next deal with the eigenvalue $\lambda_2 = 2$. We compute:

$$\mathrm{RREF}\big(\lambda_2 I_3 - A\big) \;=\; \mathrm{RREF}\Big(\begin{bmatrix} 1 & 0 & -1 \\ 0 & 3 & 0 \\ -1 & 0 & 1 \end{bmatrix}\Big) \;=\; \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis of $E_{\lambda_2}(A) = \mathrm{Nul}\big(\lambda_2 I_3 - A\big)$. Note that

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \;=\; \big[\ x^2+1\ \big]_{\mathcal{A}}.$$

So, $\{x^2+1\}$ is basis of $E_{\lambda_2}(f)$. (The argument is analogous to the one we gave for λ_1.) In particular, the geometric multiplicity of $\lambda_2 = 2$ as an eigenvalue of f is 1.

Finally, we deal with the eigenvalue $\lambda_3 = -1$. We compute:

$$\mathrm{RREF}\big(\lambda_3 I_3 - A\big) \;=\; \mathrm{RREF}\Big(\begin{bmatrix} -2 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & -2 \end{bmatrix}\Big) \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore, $\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis of $E_{\lambda_3}(A) = \mathrm{Nul}\big(\lambda_3 I_3 - A\big)$. Note that

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \;=\; \big[\ x\ \big]_{\mathcal{A}}.$$

So, $\{x\}$ is basis of $E_{\lambda_3}(f)$. (The argument is analogous to the one we gave for λ_1.) In particular, the geometric multiplicity of $\lambda_3 = -1$ as an eigenvalue of f is 1.

We now summarize our results as follows.

- The characteristic polynomial of f is $p_f(\lambda) = \lambda(\lambda-2)(\lambda+1)$.
- The spectrum of f is $\{0, 2, -1\}$.
- The linear function f has three eigenvalues, namely $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = -1$, and each of these three eigenvalues has algebraic multiplicity 1 and geometric multiplicity 1.

- We have the following bases of the three eigenspaces of f:
  - $\{x^2-1\}$ is a basis of $E_{\lambda_1}(f) = E_0(f)$;
  - $\{x^2+1\}$ is a basis of $E_{\lambda_2}(f) = E_2(f)$;
  - $\{x\}$ is a basis of $E_{\lambda_3}(f) = E_{-1}(f)$.

Optional: Since it is easy to miscompute, it is not a bad idea to check that each vector in a basis of an eigenspace of f associated with λ_i (for $i \in \{1,2,3\}$) really is an eigenvector of f associated with λ_i. For this, we compute:

- $f(x^2-1) = (1-1)x^2 - 0x + (1-1) = 0 = 0(x^2-1) = \lambda_1(x^2-1)$;
- $f(x^2+1) = (1+1)x^2 + 0x + (1+1) = 2x^2+2 = 2(x^2+1) = \lambda_2(x^2+1)$;
- $f(x) = (0+0)x^2 - 1x + (0+0) = (-1)x = \lambda_3 x$.

□

8.2.8 Proof of Theorem 8.2.3

In this subsection, we prove Theorem 8.2.3, which states that the geometric multiplicity of an eigenvalue of a square matrix is no greater than the algebraic multiplicity of that eigenvalue. In fact, it will be a bit more convenient to first prove an analog of this theorem for linear functions (see Theorem 8.2.17 below), and to then derive Theorem 8.2.3 as in immediate corollary.

Theorem 8.2.17. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $f : V \to V$ be a linear function. Then the geometric multiplicity of any eigenvalue of f is no greater than the algebraic multiplicity of that eigenvalue.*

Proof. Suppose that λ_0 is an eigenvalue of f of geometric multiplicity k. We must show that the eigenvalue λ_0 has algebraic multiplicity at least k, that is, that $(\lambda-\lambda_0)^k \mid p_f(\lambda)$. The goal is to find a basis $\mathcal{B}$ of V for which it can easily be shown that $(\lambda-\lambda_0)^k$ divides the polynomial $p_B(\lambda)$, where $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$; this is enough because, by Proposition 8.2.12, $p_f(\lambda) = p_B(\lambda)$.

Since the geometric multiplicity of the eigenvalue λ_0 of f is k, we see that the eigenspace $E_{\lambda_0}(f)$ has a k-element basis, say $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$.[14] In particular, $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ is a linearly independent set of vectors in V, and so by Theorem 3.2.19, it can be extended to a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{b}_{k+1}, \ldots, \mathbf{b}_n\}$ of V. As usual $\mathcal{E}_n = \{\mathbf{e}_1^n, \ldots, \mathbf{e}_n^n\}$ is the standard basis of $\mathbb{F}^n$. We now compute:

[14] Obviously, $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are all eigenvectors of f associated with the eigenvalue λ_0, and they form a linearly independent set of vectors in $\mathbb{F}^n$.

$$
\begin{aligned}
B &:= {}_{\mathcal{B}}[\, f \,]_{\mathcal{B}} \\
&\overset{(*)}{=} \begin{bmatrix} [\, f(\mathbf{b}_1) \,]_{\mathcal{B}} & \dots & [\, f(\mathbf{b}_k) \,]_{\mathcal{B}} & [\, f(\mathbf{b}_{k+1}) \,]_{\mathcal{B}} & \dots & [\, f(\mathbf{b}_n) \,]_{\mathcal{B}} \end{bmatrix} \\
&\overset{(**)}{=} \begin{bmatrix} [\, \lambda_0 \mathbf{b}_1 \,]_{\mathcal{B}} & \dots & [\, \lambda_0 \mathbf{b}_k \,]_{\mathcal{B}} & [\, f(\mathbf{b}_{k+1}) \,]_{\mathcal{B}} & \dots & [\, f(\mathbf{b}_n) \,]_{\mathcal{B}} \end{bmatrix} \\
&= \begin{bmatrix} \lambda_0 \mathbf{e}_1^n & \dots & \lambda_0 \mathbf{e}_k^n & [\, f(\mathbf{b}_{k+1}) \,]_{\mathcal{B}} & \dots & [\, f(\mathbf{b}_n) \,]_{\mathcal{B}} \end{bmatrix} \\
&= \left[\begin{array}{c:ccc} \lambda_0 I_k & & & \\ \hline O_{(n-k)\times k} & [\, f(\mathbf{b}_{k+1}) \,]_{\mathcal{B}} & \dots & [\, f(\mathbf{b}_n) \,]_{\mathcal{B}} \end{array}\right],
\end{aligned}
$$

where (*) follows from Theorem 4.5.1, and (**) follows from the fact that $\mathbf{b}_1, \dots, \mathbf{b}_k \in E_{\lambda_0}(f)$.

If $k = n$, then we have that $B = \lambda_0 I_n$, and so by Proposition 8.2.7, $p_B(\lambda) = (\lambda - \lambda_0)^n$, and in particular, $(\lambda - \lambda_0)^k \mid p_B(\lambda)$.

From now on, we may assume that $k < n$. We then have that

$$
p_B(\lambda) = \det\big(\lambda I_n - B\big) = \left|\begin{array}{c:c} (\lambda - \lambda_0) I_k & \\ \hline O_{(n-k)\times k} & C \end{array}\right|,
$$

where

$$
C = \begin{bmatrix} \lambda \mathbf{e}_{k+1} - [\, f(\mathbf{b}_{k+1}) \,]_{\mathcal{B}} & \dots & \lambda \mathbf{e}_n - [\, f(\mathbf{b}_n) \,]_{\mathcal{B}} \end{bmatrix}_{(n-k)\times n}.
$$

Thus, $p_B(\lambda)$ is of the form

$$
p_B(\lambda) = \left|\begin{array}{cccc:cccc} \lambda - \lambda_0 & 0 & \dots & 0 & * & * & \dots & * \\ 0 & \lambda - \lambda_0 & \dots & 0 & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda - \lambda_0 & * & * & \dots & * \\ \hline 0 & 0 & \dots & 0 & * & * & \dots & * \\ 0 & 0 & \dots & 0 & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & * & * & \dots & * \end{array}\right|,
$$

where the red submatrix in the upper-left corner (to the left of the vertical dotted line, and above the horizontal dotted line) is of size $k \times k$. By iteratively performing Laplace expansion along the first column, we see that $p_B(\lambda)$ has a factor $(\lambda - \lambda_0)^k$. This completes the argument. □

We are now ready to prove Theorem 8.2.3, restated below for the reader's convenience.

Theorem 8.2.3. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$. Then the geometric multiplicity of any eigenvalue of A is no greater than the algebraic multiplicity of that eigenvalue.*

Proof. Let $f_A : \mathbb{F}^n \to \mathbb{F}^n$ be given by $f_A(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$. Then f_A is linear (by Proposition 1.10.4), and its standard matrix is A. By Proposition 8.2.15, A and f_A have exactly the same eigenvalues, with the same corresponding geometric multiplicities, and the same corresponding algebraic multiplicities. The result now follows from Theorem 8.2.17 applied to the linear function f_A. □

8.3 The Cayley-Hamilton theorem

The famous Cayley-Hamilton theorem essentially states that every square matrix is a root of its own characteristic polynomial. (Here, we need to treat the free coefficient of the characteristic polynomial as that coefficient times the identity matrix of the appropriate size.) For example, for the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$ or $\mathbb{C}$, we have that

$$p_A(\lambda) = \det(\lambda I_2 - A) = \begin{vmatrix} \lambda - 1 & -2 \\ -3 & \lambda - 4 \end{vmatrix} = \lambda^2 - 5\lambda - 2,$$

and we see that

$$\begin{aligned} A^2 - 5A - 2I_2 &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^2 - 5\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix} - \begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

We note that the proof of the Cayley-Hamilton theorem relies on the adjugate matrix (see section 7.8) and on Theorem 7.8.2. Let us now state and prove the Cayley-Hamilton theorem.

The Cayley-Hamilton theorem. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ be a square matrix, and let $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$ be the characteristic polynomial of A. Then*

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1A + a_0I_n = O_{n\times n}.$$

Proof. If $n = 1$, then the result is immediate.[15] So, assume that $n \geq 2$. By Theorem 7.8.2 applied to the matrix $\lambda I_n - A$ (where λ is a variable), we get that

$$(\lambda I_n - A)\operatorname{adj}(\lambda I_n - A) = \det(\lambda I_n - A) I_n.$$

Now, note that each cofactor of the matrix $\lambda I_n - A$ is a polynomial (in variable λ) of degree at most λ^{n-1}. Since the entries of $\operatorname{adj}(\lambda I_n - A)$ are precisely the cofactors of $\lambda I_n - A$, it follows that each entry of $\operatorname{adj}(\lambda I_n - A)$ is a polynomial (in the variable λ) of degree at most $n-1$. So, the matrix $\operatorname{adj}(\lambda I_n - A)$ can be expressed in the form

$$\operatorname{adj}(\lambda I_n - A) = \lambda^{n-1} B_{n-1} + \lambda^{n-2} B_{n-2} + \cdots + \lambda B_1 + B_0,$$

for some matrices $B_0, B_1, \ldots, B_{n-1} \in \mathbb{F}^{n \times n}$. Consequently,

$$\underbrace{(\lambda I_n - A)(\underbrace{\lambda^{n-1} B_{n-1} + \lambda^{n-2} B_{n-2} + \cdots + \lambda B_1 + B_0}_{=\operatorname{adj}(\lambda I_n - A)})}_{:=\text{LHS}} = \underbrace{\det(\lambda I_n - A) I_n}_{:=\text{RHS}}.$$

For the left-hand-side, we have

$$\begin{aligned} \text{LHS} &= (\lambda I_n - A)(\lambda^{n-1} B_{n-1} + \cdots + \lambda B_1 + B_0) \\ &= \lambda^n B_{n-1} + \lambda^{n-1}(B_{n-2} - AB_{n-1}) + \lambda^{n-2}(B_{n-3} - AB_{n-2}) + \\ &\quad + \cdots + \lambda(B_0 - AB_1) - AB_0. \end{aligned}$$

For the right-hand-side, we have

$$\begin{aligned} \text{RHS} &= \det(\lambda I_n - A) I_n = p_A(\lambda) I_n \\ &= (\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-1} + \cdots + a_1\lambda + a_0) I_n \\ &= \lambda^n I_n + \lambda^{n-1} a_{n-1} I_n + \lambda^{n-2} a_{n-2} I_n + \cdots + \lambda a_1 I_n + a_0 I_n. \end{aligned}$$

The corresponding coefficients in front of λ^i (for $i \in \{0, 1, \ldots, n\}$) must be equal on the left-hand-side (LHS) and the right-hand-side (RHS). This yields the following $n+1$ equations.

$$\begin{aligned} B_{n-1} &= I_n \\ B_{n-2} - AB_{n-1} &= a_{n-1} I_n \\ B_{n-3} - AB_{n-2} &= a_{n-2} I_n \\ &\vdots \\ B_0 - AB_1 &= a_1 I_n \\ -AB_0 &= a_0 I_n \end{aligned}$$

[15] Indeed, suppose that $n = 1$, and consider any matrix $A = \begin{bmatrix} a_{1,1} \end{bmatrix}$ in $\mathbb{F}^{1\times 1}$. Then $p_A(\lambda) = \det(\lambda I_1 - A) = \det(\begin{bmatrix} \lambda - a_{1,1} \end{bmatrix}) = \lambda - a_{1,1}$, and we see that $A - a_{1,1} I_1 = O_{1\times 1}$.

We now multiply the first (top) equation by A^n on the left, the second equation by A^{n-1} on the left, the third equation by A^{n-2} on the left, and so on. (The $(n+1)$-th equation, i.e. the bottom one, gets multiplied by $A^0 = I_n$ on the left, i.e. it remains unchanged). This yields the following.

$$\begin{aligned} A^n B_{n-1} &= A^n \\ A^{n-1}B_{n-2} - A^n B_{n-1} &= a_{n-1}A^{n-1} \\ A^{n-2}B_{n-3} - A^{n-1}B_{n-2} &= a_{n-2}A^{n-2} \\ &\vdots \\ AB_0 - A^2 B_1 &= a_1 A \\ -AB_0 &= a_0 I_n \end{aligned}$$

We now add up the equations that we obtained. On the left-hand-side, the sum is obviously $O_{n\times n}$. So, the right-hand-side must also sum up to $O_{n\times n}$, i.e.

$$A^n + a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \cdots + a_1 A + a_0 I_n = O_{n\times n}.$$

But this is precisely what we needed to show. □

Corollary 8.3.1. *Let $\mathbb{F}$ be a field. For all matrices $A \in \mathbb{F}^{n\times n}$, both the following hold:*

(a) $A^n \in \mathrm{Span}(I_n, A, A^2, \ldots, A^{n-1})$, i.e. A^n is a linear combination of the matrices $I_n, A, A^2, \ldots, A^{n-1}$;

(b) if A is invertible, then $A^{-1} \in \mathrm{Span}(I_n, A, A^2, \ldots, A^{n-1})$, i.e. A^{-1} is a linear combination of the matrices $I_n, A, A^2, \ldots, A^{n-1}$.

Proof. Fix a matrix $A \in \mathbb{F}^{n\times n}$, and consider its characteristic polynomial $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

(a) By the Cayley-Hamilton theorem, we have that

$$A^n + a_{n-1}A^{n-1} + \cdots + a_a A^2 + a_1 A + a_0 I_n = O_{n\times n}.$$

Consequently,

$$A^n = -a_0 I_n - a_1 A - a_2 A^2 - \cdots - a_{n-1}A^{n-1}.$$

Thus, A^n is a linear combination of the matrices $I_n, A, A^2, \ldots, A^{n-1}$.

(b) Assume that A is invertible. Proposition 8.2.11 then guarantees that 0 is not an eigenvalue of A. Since the eigenvalues of A are precisely the roots of the characteristic polynomial of A, we have that $p_A(0) \neq 0$; since $p_A(0) = a_0$, it follows that $a_0 \neq 0$.[16]

[16]Indeed, 0 is not a root of $p_A(\lambda)$, and so $p_A(0) \neq 0$. But $p_A(0) = a_0$, and it follows that $a_0 \neq 0$.

Now, by the Cayley-Hamilton theorem, we have that

$$A^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I_n \quad = \quad O_{n\times n}.$$

We multiply both sides of the equation by A^{-1} on the right, and we obtain

$$A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1I_n + a_0A^{-1} \quad = \quad O_{n\times n},$$

and consequently,

$$a_0A^{-1} \quad = \quad -a_1I_n - a_2A - \cdots - a_{n-1}A^{n-2} - A^{n-1}.$$

Since $a_0 \neq 0$, this implies that

$$A^{-1} \quad = \quad -\tfrac{a_1}{a_0}I_n - \tfrac{a_2}{a_0}A - \cdots - \tfrac{a_{n-1}}{a_0}A^{n-2} - \tfrac{1}{a_0}A^{n-1}.$$

So, A^{-1} is a linear combination of the matrices $I_n, A, A^2, \ldots, A^{n-1}$. □

8.4 Eigenvectors and linear independence. Eigenbases

For a finite-dimensional vector space V over a field $\mathbb{F}$ and a linear function $f : V \to V$, an *eigenbasis* of V associated with f is a basis $\mathcal{B}$ of V such that all vectors in $\mathcal{B}$ are eigenvectors of f. Similarly, for an field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n\times n}$, an *eigenbasis* of $\mathbb{F}^n$ associated with A is a basis $\mathcal{B}$ of $\mathbb{F}^n$ such that all vectors in $\mathcal{B}$ are eigenvectors of A. Eigenbases do not always exist, and one of our goals in this section is to determine when they do and do not exist. As we shall see in section 8.5, eigenbases play a crucial role in matrix "diagonalization."

Our first proposition of the section (Proposition 8.4.1 below) states that, for a linear function $f : V \to V$, where V is a vector space over a field $\mathbb{F}$, any (finite) set of eigenvectors of V associated with pairwise distinct eigenvalues is linearly independent.

Proposition 8.4.1. *Let V be a vector space over a field $\mathbb{F}$, let $f : V \to V$ be a linear function, let $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ be pairwise distinct eigenvalues of f, associated with eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, respectively. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly independent set.*

Proof. We will prove inductively that for all $i \in \{0, \ldots, k\}$, the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is linearly independent. For $i = 0$, we have that $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\} = \emptyset$, which is obviously a linearly independent set. Now, fix an index $i \in \{0, \ldots, k-1\}$, and assume inductively that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is linearly independent. We must show that $\{\mathbf{v}_1, \ldots, \mathbf{v}_i, \mathbf{v}_{i+1}\}$ is linearly independent. Fix scalars $\alpha_1, \ldots, \alpha_i, \alpha_{i+1} \in \mathbb{F}$ such that

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_i\mathbf{v}_i + \alpha_{i+1}\mathbf{v}_{i+1} \quad = \quad \mathbf{0}.$$

If we multiply both sides of the equation above by λ_{i+1}, we obtain

(1) $\lambda_{i+1}\alpha_1\mathbf{v}_1 + \cdots + \lambda_{i+1}\alpha_i\mathbf{v}_i + \lambda_{i+1}\alpha_{i+1}\mathbf{v}_{i+1} = \mathbf{0}$.

If, on the other hand, we apply the function f to both sides and also use the fact that $f(\mathbf{0}) = \mathbf{0}$ (by Proposition 4.1.6), then we obtain

(2) $f\big(\alpha_1\mathbf{v}_1 + \cdots + \alpha_i\mathbf{v}_i + \alpha_{i+1}\mathbf{v}_{i+1}\big) = \mathbf{0}$.

We now compute:

$$\begin{aligned}\mathbf{0} &\overset{(2)}{=} f\big(\alpha_1\mathbf{v}_1 + \cdots + \alpha_i\mathbf{v}_i + \alpha_{i+1}\mathbf{v}_{i+1}\big)\\ &\overset{(*)}{=} \alpha_1 f(\mathbf{v}_1) + \cdots + \alpha_i f(\mathbf{v}_i) + \alpha_{i+1} f(\mathbf{v}_{i+1})\\ &\overset{(**)}{=} \alpha_1\lambda_1\mathbf{v}_1 + \cdots + \alpha_i\lambda_i\mathbf{v}_i + \alpha_{i+1}\lambda_{i+1}\mathbf{v}_{i+1},\end{aligned}$$

where (*) follows from the linearity of f (and more precisely, from Proposition 4.1.5), and (**) follows from the fact that $\mathbf{v}_1, \ldots, \mathbf{v}_i, \mathbf{v}_{i+1}$ are eigenvectors of f associated with eigenvalues $\lambda_1, \ldots, \lambda_i, \lambda_{i+1}$, respectively. Combining this with (1), we obtain

$$\begin{aligned}&\alpha_1\lambda_1\mathbf{v}_1 + \cdots + \alpha_i\lambda_i\mathbf{v}_i + \alpha_{i+1}\lambda_{i+1}\mathbf{v}_{i+1}\\ = \;&\lambda_{i+1}\alpha_1\mathbf{v}_1 + \cdots + \lambda_{i+1}\alpha_i\mathbf{v}_i + \lambda_{i+1}\alpha_{i+1}\mathbf{v}_{i+1}.\end{aligned}$$

By subtracting one side from the other and factoring, we get

$$\alpha_1(\lambda_1 - \lambda_{i+1})\mathbf{v}_1 + \cdots + \alpha_i(\lambda_i - \lambda_{i+1})\mathbf{v}_i = \mathbf{0}$$

By the induction hypothesis, vectors $\mathbf{v}_1, \ldots, \mathbf{v}_i$ are linearly independent, and it follows that $\alpha_1(\lambda_1 - \lambda_{i+1}) = \cdots = \alpha_1(\lambda_i - \lambda_{i+1}) = 0$. Since $\lambda_1 - \lambda_{i+1}, \ldots, \lambda_i - \lambda_{i+1}$ are all non-zero (because $\lambda_1, \ldots, \lambda_i, \lambda_{i+1}$ are pairwise distinct), we deduce that $\alpha_1 = \cdots = \alpha_i = 0$. Plugging this into our equation $\alpha_1\mathbf{v}_1 + \cdots + \alpha_i\mathbf{v}_i + \alpha_{i+1}\mathbf{v}_{i+1} = \mathbf{0}$, we get

$$\alpha_{i+1}\mathbf{v}_{i+1} = \mathbf{0}.$$

But $\mathbf{v}_{i+1}$ is an eigenvector of f, and so by definition, $\mathbf{v}_{i+1} \neq \mathbf{0}$. So, $\alpha_{i+1} = 0$. We have now shown that $\alpha_1 = \cdots = \alpha_i = \alpha_{i+1} = 0$, and we deduce that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_i, \mathbf{v}_{i+1}\}$ is linearly independent. This completes the induction. □

Proposition 8.4.2. *Let V be a vector space over a field $\mathbb{F}$, let $f : V \to V$ be a linear function, and let $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$ be pairwise distinct eigenvalues of f. For each $i \in \{1, \ldots, k\}$, let $\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,t_i}$ be linearly independent eigenvectors of f associated with the eigenvalue λ_i. Then the eigenvectors*

$$\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,t_1}, \mathbf{v}_{2,1}, \ldots, \mathbf{v}_{2,t_2}, \ldots, \mathbf{v}_{k,1}, \ldots, \mathbf{v}_{k,t_k}$$

are linearly independent.

Proof. Fix scalars $\alpha_{1,1},\dots,\alpha_{1,t_1},\alpha_{2,1},\dots,\alpha_{2,t_2},\dots,\alpha_{k,1},\dots,\alpha_{k,t_k}\in\mathbb{F}$ such that

$$\sum_{i=1}^{k}\left(\alpha_{i,1}\mathbf{v}_{i,1}+\cdots+\alpha_{i,t_i}\mathbf{v}_{i,t_i}\right) \quad = \quad \mathbf{0}.$$

Now, for each $i\in\{1,\dots,k\}$, set $\mathbf{v}_i:=\alpha_{i,1}\mathbf{v}_{i,1}+\cdots+\alpha_{i,t_i}\mathbf{v}_{i,t_i}$, that is

- $\mathbf{v}_1:=\alpha_{1,1}\mathbf{v}_{1,1}+\cdots+\alpha_{1,t_1}\mathbf{v}_{1,t_1}$;
- $\mathbf{v}_2:=\alpha_{2,1}\mathbf{v}_{2,1}+\cdots+\alpha_{2,t_2}\mathbf{v}_{2,t_2}$;

  $\vdots$
- $\mathbf{v}_k:=\alpha_{k,1}\mathbf{v}_{k,1}+\cdots+\alpha_{k,t_k}\mathbf{v}_{k,t_k}$.

So,

$$\mathbf{v}_1+\mathbf{v}_2+\cdots+\mathbf{v}_k \quad = \quad \mathbf{0}.$$

Now, note that for each $i\in\{1,\dots,k\}$, the vector $\mathbf{v}_i$ is a linear combination of vectors in $E_{\lambda_i}(f)$;[17] since (by Proposition 8.1.4) $E_{\lambda_i}(f)$ is a subsapce of V and is therefore closed under linear combinations, it follows that $\mathbf{v}_i\in E_{\lambda_i}(f)$. Consequently, for each $i\in\{1,\dots,k\}$, $\mathbf{v}_i$ is either $\mathbf{0}$ or an eigenvector of f associated with the eigenvalue λ_i. We claim that $\mathbf{v}_1=\mathbf{v}_2=\cdots=\mathbf{v}_k=\mathbf{0}$. Suppose otherwise. After possibly permuting the order of the λ_i's and the corresponding $\mathbf{v}_i$'s, we may assume that there exists some $\ell\in\{1,\dots,k\}$ such that $\mathbf{v}_1,\dots,\mathbf{v}_\ell$ are all non-zero (and are consequently eigenvectors of f associated with $\lambda_1,\dots,\lambda_\ell$), while $\mathbf{v}_{\ell+1},\dots,\mathbf{v}_k$ are all zero. So,

$$\mathbf{v}_1+\cdots+\mathbf{v}_\ell \quad = \quad \mathbf{0},$$

and it follows that $\{\mathbf{v}_1,\dots,\mathbf{v}_\ell\}$ is a linearly dependent set. But this contradicts Proposition 8.4.1. We have now shown that $\mathbf{v}_1=\cdots=\mathbf{v}_k=\mathbf{0}$. So, for all indices $i\in\{1,\dots,k\}$, we have that $\alpha_{i,1}\mathbf{v}_{i,1}+\cdots+\alpha_{i,t_i}\mathbf{v}_{i,t_i}=\mathbf{0}$; since vectors $\mathbf{v}_{i,1},\dots,\mathbf{v}_{i,t_i}$ are linearly independent, it follows that $\alpha_{i,1}=\cdots=\alpha_{i,t_i}=0$. Since this holds for all indices $i\in\{1,\dots,k\}$, we deduce that the eigenvectors

$$\mathbf{v}_{1,1},\dots,\mathbf{v}_{1,t_1},\mathbf{v}_{2,1},\dots,\mathbf{v}_{2,t_2},\dots,\mathbf{v}_{k,1},\dots,\mathbf{v}_{k,t_k}$$

are linearly independent, which is what we needed to show. □

Theorem 8.4.3. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and set $n:=dim(V)$. Let $f:V\to V$ be a linear function, let $\lambda_1,\dots,\lambda_k$ be all the (distinct) eigenvalues of f, and let $\mathcal{B}_1,\dots,\mathcal{B}_k$ be bases of the associated eigenspaces $E_{\lambda_1}(f),\dots,E_{\lambda_k}(f)$, respectively. Set $\mathcal{B}:=\mathcal{B}_1\cup\cdots\cup\mathcal{B}_k$. Then all the following hold:*

[17] Indeed, since $\mathbf{v}_{i,1},\dots,\mathbf{v}_{i,t_i}$ are eigenvectors of f associated with the eigenvalue λ_i, they all belong to $E_{\lambda_i}(f)$. By construction, $\mathbf{v}_i$ is a linear combination of these eigenvectors.

(a) $\mathcal{B}$ *is a linearly independent set of eigenvectors of* f*;*

(b) $dim\big(E_{\lambda_1}(f)\big) + \cdots + dim\big(E_{\lambda_k}(f)\big) \leq n$*, i.e. the sum of geometric multiplicities of the eigenvalues of* f *is at most* n*;*

(c) V *has an eigenbasis associated with* f *if and only if the sum of geometric multiplicities of the eigenvalues of* f *is* n*, and in this case,* $\mathcal{B}$ *is such an eigenbasis;*

(d) V *has an eigenbasis associated with* f *if and only if the sum of algebraic multiplicities of the eigenvalues of* f *is* n*, and the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity; in this case,* $\mathcal{B}$ *is an eigenbasis of* V *associated with the linear function* f*.*

Proof. Part (a) follows immediately from Proposition 8.4.2. Part (b) follows from (a) and from the fact that, by Theorem 3.2.17(a), any linearly independent set of vectors in an n-dimensional vector space contains at most n vectors.

Let us prove (c). Suppose first that the sum of geometric multiplicities of the eigenvalues of f is equal to n. Then $\mathcal{B}$ is a linearly independent set of size n in the n-dimensional vector space V. So, by Corollary 3.2.20(a), $\mathcal{B}$ is a basis of V. Since all vectors in $\mathcal{B}$ are eigenvectors of f, it follows that $\mathcal{B}$ is an eigenbasis of V associated with f.

Suppose, conversely, that V has an eigenbasis $\mathcal{C}$ associated with f; since $\dim(V) = n$, we see that $|\mathcal{C}| = n$. Since all vectors in $\mathcal{C}$ are eigenvecors of f, we see that they all belong to $E_{\lambda_1}(f) \cup \cdots \cup E_{\lambda_k}(f)$. But since the basis $\mathcal{C}$ of V is, in particular, linearly independent, we see that it cannot contain more than $\dim\Big(E_{\lambda_i}(f)\Big)$ many vectors from $E_{\lambda_i}(f)$ for any index $i \in \{1, \ldots, k\}$.[18] So, $|\mathcal{C}| \leq \dim\big(E_{\lambda_1}(f)\big) + \cdots + \dim\big(E_{\lambda_k}(f)\big)$. But now we have that

$$n \;=\; |\mathcal{C}| \;\leq\; \dim\big(E_{\lambda_1}(f)\big) + \cdots + \dim\big(E_{\lambda_k}(f)\big) \;\overset{\text{(b)}}{\leq}\; n,$$

and it follows that $\dim\big(E_{\lambda_1}(f)\big) + \cdots + \dim\big(E_{\lambda_k}(f)\big) = n$, i.e. the sum of geometric multiplicities of the eigenvalues of f is n. This proves (c).

It remains to prove (d). If the sum of algebraic multiplicities of the eigenvalues of f is equal to n, and the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity, then obviously, the sum of geometric multiplicities of f is equal to n, and so by (c), V has an eigenbasis associated with f, and $\mathcal{B}$ is one such eigenbasis. For the converse, assume that V has an eigenbasis $\mathcal{C}$ associated with f. Let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of f, with geometric multiplicities $g_1, \ldots, g_k$, respectively, and algebraic multiplicities $a_1, \ldots, a_k$, respectively. By (c), we have that $g_1 + \cdots + g_k = n$. On the other hand, the characteristic polynomial of f is of degree n, we see that the sum of algebraic multiplicitis of f is at most n, i.e.

[18] Once again, we are using Theorem 3.2.17(a).

$a_1 + \cdots + a_k \leq n$. But by Theorem 8.2.17, the geometric multiplicity of an eigenvalue of f is no greater than the algebraic multiplicity of that eigenvalue, that is, $g_i \leq a_i$ for all indices $i \in \{1, \ldots, n\}$. We now have that

$$n \;=\; g_1 + \cdots + g_k \;\leq\; a_1 + \cdots + a_k \;\leq\; n,$$

and we deduce that $a_1 + \cdots + a_k = n$ and that $g_i = a_i$ for all $i \in \{1, \ldots, k\}$. This proves (d). □

Corollary 8.4.4. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and set $n := \dim(V)$. If a linear function $f : V \to V$ has n distinct eigenvalues, then V has an eigenbasis associated with f.*

Proof. Let $f : V \to V$ be a linear function that has n distinct eigenvalues, say $\lambda_1, \ldots, \lambda_n$. By the definition of an eigenvalue,[19] we have that $\dim\big(E_{\lambda_i}(f)\big) \geq 1$ for all $i \in \{1, \ldots, n\}$. Consequently, $\dim\big(E_{\lambda_1}(f)\big) + \cdots + \dim\big(E_{\lambda_n}(f)\big) \geq n$. On the other hand, Theorem 8.4.3(b) guarantees that $\dim\big(E_{\lambda_1}(f)\big) + \cdots + \dim\big(E_{\lambda_n}(f)\big) \leq n$. Thus, $\dim\big(E_{\lambda_1}(f)\big) + \cdots + \dim\big(E_{\lambda_n}(f)\big) = n$, and so by Theorem 8.4.3(c), V has an eigenbasis associated with f. □

Propositions 8.4.1 and 8.4.2 can easily be "translated" into the language of matrices, as can Theorem 8.4.3. The case of Propositions 8.4.1 and 8.4.2 is left as an easy exercise for the reader. In the case of Theorem 8.4.3, we obtain the Theorem 8.4.5 (below).

Theorem 8.4.5. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$. Let $\lambda_1, \ldots, \lambda_k$ be all the (distinct) eigenvalues of A, and let $\mathcal{B}_1, \ldots, \mathcal{B}_k$ be bases of the associated eigenspaces $E_{\lambda_1}(A), \ldots, E_{\lambda_k}(A)$, respectively. Set $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$. Then all the following hold:*

(a) $\mathcal{B}$ is a linearly independent set of eigenvectors of A;

(b) $\dim\big(E_{\lambda_1}(A)\big) + \cdots + \dim\big(E_{\lambda_k}(A)\big) \leq n$, i.e. the sum of geometric multiplicities of the eigenvalues of A is at most n;

(c) $\mathbb{F}^n$ has an eigenbasis associated with A if and only if the sum of geometric multiplicities of the eigenvalues of A is n, and in this case, $\mathcal{B}$ is such an eigenbasis;

(d) $\mathbb{F}^n$ has an eigenbasis associated with A if and only if the sum of algebraic multiplicities of the eigenvalues of A is n, and the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity; in this case, $\mathcal{B}$ is an eigenbasis of $\mathbb{F}^n$ associated with the matrix A.

Proof. Define $f_A : \mathbb{F}^n \to \mathbb{F}^n$ by setting $f_A(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$. Then f_A is linear (by Proposition 1.10.4), and moreover, A is the standard matrix of f_A. The result now follows immediately from Proposition 8.2.15 (applied to f_A and A) and Theorem 8.4.3 (applied to f_A). □

[19] Or alternatively: by Proposition 8.1.4.

Corollary 8.4.6. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$. If A has n distinct eigenvalues, then $\mathbb{F}^n$ has an eigenbasis associated with A.*

Proof. Define $f_A : \mathbb{F}^n \to \mathbb{F}^n$ by setting $f_A(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$. Then f_A is linear (by Proposition 1.10.4), and A is the standard matrix of A. The result now follows immediately from Proposition 8.1.5 and Corollary 8.4.4. □

8.5 Diagonalization

8.5.1 Diagonal matrices and their powers

For a field $\mathbb{F}$, a square matrix $D \in \mathbb{F}^{n\times n}$ is *diagonal* if all its entries off the main diagonal are zero (the entries on the main diagonal may or may not be zero). For scalars $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{F}$, $D(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the $n \times n$ matrix with $\lambda_1, \lambda_2, \ldots, \lambda_n$ on the main diagonal (appearing in that order) and 0's everywhere else, i.e.

$$D(\lambda_1, \lambda_2, \ldots, \lambda_n) := \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1\mathbf{e}_1 & \ldots & \lambda_n\mathbf{e}_n \end{bmatrix},$$

where as usual, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are the standard basis vectors of $\mathbb{F}^n$.

Note that diagonal matrices are, in particular, triangular. So, Propositions 7.3.1 and 8.2.7 apply. More precisely, for scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ (where $\mathbb{F}$ is a field), and for the diagonal matrix $D := D(\lambda_1, \ldots, \lambda_n)$, we have the following:

- $\det(D) = \lambda_1 \ldots \lambda_n$;
- $p_D(\lambda) = (\lambda - \lambda_1) \ldots (\lambda - \lambda_n)$.

Proposition 8.5.1. *Let $\mathbb{F}$ be a field, let $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ ($n \geq 1$) be arbitrary scalars, and set $D := D(\lambda_1, \ldots, \lambda_n)$. Then both the following hold:*

(a) for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$ in $\mathbb{F}^n$, we have that

$$D\mathbf{x} = \begin{bmatrix} \lambda_1 x_1 \\ \vdots \\ \lambda_n x_n \end{bmatrix};$$

(b) for all matrices $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$ in $\mathbb{F}^{m\times n}$, we have that

$$AD = \begin{bmatrix} \lambda_1\mathbf{a}_1 & \ldots & \lambda_n\mathbf{a}_n \end{bmatrix}.$$

Proof. (a) Fix a vector $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$ in $\mathbb{F}^n$. We then compute:

$$\begin{aligned} D\mathbf{x} &= \begin{bmatrix} \lambda_1\mathbf{e}_1 & \dots & \lambda_n\mathbf{e}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &\overset{(*)}{=} x_1(\lambda_1\mathbf{e}_1) + \cdots + x_n(\lambda_n\mathbf{e}_n) \\ &= (\lambda_1 x_1)\mathbf{e}_1 + \cdots + (\lambda_n x_n)\mathbf{e}_n \\ &= \begin{bmatrix} \lambda_1 x_1 \\ \vdots \\ \lambda_n x_n \end{bmatrix}, \end{aligned}$$

where (*) follows from the definition of matrix-vector multiplication.

(b) Fix a matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ in $\mathbb{F}^{m\times n}$. We then compute:

$$\begin{aligned} AD &= A\begin{bmatrix} \lambda_1\mathbf{e}_1 & \dots & \lambda_n\mathbf{e}_n \end{bmatrix} \\ &\overset{(*)}{=} \begin{bmatrix} A(\lambda_1\mathbf{e}_1) & \dots & A(\lambda_n\mathbf{e}_n) \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1(A\mathbf{e}_1) & \dots & \lambda_n(A\mathbf{e}_n) \end{bmatrix} \\ &\overset{(**)}{=} \begin{bmatrix} \lambda_1\mathbf{a}_1 & \dots & \lambda_n\mathbf{a}_n \end{bmatrix}, \end{aligned}$$

where (*) follows from the definition of matrix multiplication, and (**) follows from Proposition 1.4.4. □

Proposition 8.5.2 (below) states that if the product of two diagonal matrices is another diagonal matrix.

Proposition 8.5.2. *Let $\mathbb{F}$ be a field, and let $\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_n \in \mathbb{F}$ ($n \geq 1$) be arbitrary scalars. Then*

$$D(\lambda_1, \dots, \lambda_n)\, D(\mu_1, \dots, \mu_n) = D(\lambda_1\mu_1, \dots, \lambda_n\mu_n).$$

Proof. We compute:

$$\begin{aligned} D(\lambda_1, \dots, \lambda_n)\, D(\mu_1, \dots, \mu_n) &= \begin{bmatrix} \lambda_1\mathbf{e}_1 & \dots & \lambda_n\mathbf{e}_n \end{bmatrix} D(\mu_1, \dots, \mu_n) \\ &\overset{(*)}{=} \begin{bmatrix} \mu_1(\lambda_1\mathbf{e}_1) & \dots & \mu_n(\lambda_n\mathbf{e}_n) \end{bmatrix} \\ &= \begin{bmatrix} (\lambda_1\mu_1)\mathbf{e}_1 & \dots & (\lambda_n\mu_n)\mathbf{e}_n \end{bmatrix} \\ &= D(\lambda_1\mu_1, \dots, \lambda_n\mu_n), \end{aligned}$$

where (*) follows from Proposition 8.5.1(b). □

Proposition 8.5.3. *Let $\mathbb{F}$ be a field, let $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ ($n \geq 1$), and set $D := D(\lambda_1, \ldots, \lambda_n)$. Then both the following hold:*

(a) for all non-negative integers m, we have that $D^m = D(\lambda_1^m, \ldots, \lambda_n^m)$;

(b) D is invertible if and only if $\lambda_1, \ldots, \lambda_n$ are all non-zero, and in this case, we have that $D^m = D(\lambda_1^m, \ldots, \lambda_n^m)$ for all integers m.

Proof. Part (a) follows from Proposition 8.5.2 via an easy induction on m (the details are left as an exercise).

We now prove (b). By Theorem 7.4.1, we know that D is invertible if and only if $\det(D) \neq 0$. Since $\det(D) = \lambda_1 \ldots \lambda_n$ (because D is diagonal), we deduce that D is invertible if and only if $\lambda_1, \ldots, \lambda_n$ are all non-zero.

Now assume that D is invertible, so that $\lambda_1, \ldots, \lambda_n$ are all non-zero and therefore have multiplicative inverses. Then

$$\underbrace{D(\lambda_1, \ldots, \lambda_n)}_{=D} \, D(\lambda_1^{-1}, \ldots, \lambda_n^{-1}) \quad \overset{(*)}{=} \quad D(\lambda_1\lambda_1^{-1}, \ldots, \lambda_n\lambda_n^{-1}) \quad = \quad I_n,$$

where (*) follows from Proposition 8.5.2. Corollary 3.3.18 now guarantees that $D^{-1} = D(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$. To obtain (b), we simply apply (a) twice: first to the diagonal matrix $D = D(\lambda_1, \ldots, \lambda_n)$, and then to the diagonal matrix $D^{-1} = D(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$.[20] □

8.5.2 Eigenbases and diagonal matrices of linear functions

Suppose that V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and suppose that $f : V \to V$ is a linear function. Theorem 8.4.3 gave us a criterion for determining whether V has an eigenbasis associated with the linear function f. Theorem 8.5.4 (below) states that if such an eigenbasis exists, then the matrix of f with respect to that basis is diagonal. As we saw in subsection 8.5.1, diagonal matrices have particularly nice computational properties.

Theorem 8.5.4. *Let V be a non-trivial, finite-dimensional vector space, let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V, and let $f : V \to V$ be a linear function. Then $\mathcal{B}$ is an eigenbasis of V associated with f if and only if the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is diagonal. Moreover, in this case, we have that*

$${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \quad = \quad D(\lambda_1, \ldots, \lambda_n),$$

[20] Indeed, by applying (a) to the diagonal matrix $D = D(\lambda_1, \ldots, \lambda_n)$, we get that $D^m = D(\lambda_1^m, \ldots, \lambda_n^m)$ for all non-negative integers m. On the other hand, by applying (a) to the diagonal matrix $D^{-1} = D(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$, we get that $(D^{-1})^m = D\big((\lambda_1^{-1})^m, \ldots, (\lambda_n^{-1})^m\big)$, that is, $D^{-m} = D(\lambda_1^{-m}, \ldots, \lambda_n^{-m})$ for all non-negative integers m. Combined, these two facts tell us that $D^m = D(\lambda_1^m, \ldots, \lambda_n^m)$ for all integers m.

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of f associated with the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, respectively.

Proof. Suppose first that $\mathcal{B}$ is an eigenbasis of V associated with f. Then, by definition, vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are eigenvectors of f, and we let $\lambda_1, \ldots, \lambda_n$, respectively, be the associated eigenvalues. Then $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ for all indices $i \in \{1, \ldots, n\}$, and we have the following:

$$\begin{aligned}
{}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} &= \big[\ \big[\ f(\mathbf{v}_1)\ \big]_{\mathcal{B}} \ \ldots\ \big[\ f(\mathbf{v}_1)\ \big]_{\mathcal{B}}\ \big] \qquad \text{by Theorem 4.5.1} \\
&= \big[\ \big[\ \lambda_1 \mathbf{v}_1\ \big]_{\mathcal{B}} \ \ldots\ \big[\ \lambda_n \mathbf{v}_n\ \big]_{\mathcal{B}}\ \big] \\
&= \big[\ \lambda_1 \mathbf{e}_1 \ \ldots\ \lambda_n \mathbf{e}_n\ \big] \\
&= D(\lambda_1, \ldots, \lambda_n).
\end{aligned}$$

Conversely, suppose that the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is diagonal, and let $\lambda_1, \ldots, \lambda_n$ be the entries of this matrix on the main diagonal, so that

$${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} = D(\lambda_1, \ldots, \lambda_n) = \big[\ \lambda_1 \mathbf{e}_1 \ \ldots\ \lambda_n \mathbf{e}_n\ \big].$$

We will show that the basis vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are eigenvectors of f with associated eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively. Fix any index $i \in \{1, \ldots, n\}$; we must show that $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$. Since $\mathbf{v}_i$ is the i-th basis vector of $\mathcal{B}$, we have that $\big[\ \mathbf{v}_i\ \big]_{\mathcal{B}} = \mathbf{e}_i$. We now compute:

$$\begin{aligned}
\big[\ f(\mathbf{v}_i)\ \big]_{\mathcal{B}} &= {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \big[\ \mathbf{v}_i\ \big]_{\mathcal{B}} \\
&= \big[\ \lambda_1 \mathbf{e}_1 \ \ldots\ \lambda_n \mathbf{e}_n\ \big]\, \mathbf{e}_i \\
&\overset{(*)}{=} \lambda_i \mathbf{e}_i \\
&= \lambda_i \big[\ \mathbf{v}_i\ \big]_{\mathcal{B}} \\
&\overset{(**)}{=} \big[\ \lambda_i \mathbf{v}_i\ \big]_{\mathcal{B}},
\end{aligned}$$

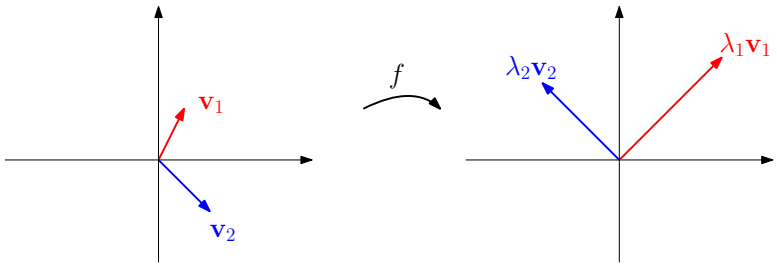
where (*) follows from Proposition 1.4.4, and (**) follows from the linearity of $\big[\ \cdot\ \big]_{\mathcal{B}}$. Since $\big[\ \cdot\ \big]_{\mathcal{B}}$ is an isomorphism (and in particular, one-to-one), it follows that $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$, which is what we needed to show. □

Remark: Suppose that V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$. By Theorems 4.3.2 and 8.5.4, linear functions from V to V that have a

diagonal matrix are precisely those that can be defined starting from some basis, and then scaling each of the basis elements. Indeed, suppose that $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is some basis of V, and that $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ are some scalars. By Theorem 4.3.2, there exists a unique linear function $f : V \to V$ such that $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ (for the special case of $\mathbb{R}^2$, see the picture below). But then by Theorem 8.5.4, we have that

$$ {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \quad = \quad D(\lambda_1, \ldots, \lambda_n). $$

By Theorem 8.5.4, the converse also holds.



Example 8.5.5. *Consider the function* $f : \mathbb{P}^2_{\mathbb{R}} \to \mathbb{P}^2_{\mathbb{R}}$ *given by*

$$ f\big(a_2x^2 + a_1x + a_0\big) \quad = \quad (a_2 + a_0)x^2 - a_1x + (a_2 + a_0) $$

for all $a_0, a_1, a_2 \in \mathbb{R}$. *We showed in Example 8.2.16 that* f *is linear. Moreover, in that example, we obtained the following:*

- *the characteristic polynomial of* f *is* $p_f(\lambda) = \lambda(\lambda - 2)(\lambda + 1)$.
- *the spectrum of* f *is* $\{0, 2, -1\}$.
- *the linear function* f *has three eigenvalues, namely* $\lambda_1 = 0$, $\lambda_2 = 1$, *and* $\lambda_3 = -1$, *and each of these three eigenvalues has algebraic multiplicity* 1 *and geometric multiplicity* 1.
- *we have the following bases of the three eigenspaces of* f*:*
  - $\{x^2 - 1\}$ *is a basis of* $E_{\lambda_1}(f) = E_0(f)$;
  - $\{x^2 + 1\}$ *is a basis of* $E_{\lambda_2}(f) = E_2(f)$;
  - $\{x\}$ *is a basis of* $E_{\lambda_3}(f) = E_{-1}(f)$.

Since $\dim(\mathbb{P}^2_{\mathbb{R}}) = 3$, *Theorem 8.4.3(c) guarantees that* $\mathcal{B} = \{x^2 - 1, x^2 + 1, x\}$ *is an eigenbasis of* $\mathbb{P}^2_{\mathbb{R}}$ *associated with the linear function* f. *Theorem 8.5.4 now guarantees that the matrix* ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ *is diagonal, and moreover, that*

$$ {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \quad = \quad D(\lambda_1, \lambda_2, \lambda_3) \quad = \quad D(0, 2, -1) \quad = \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}. $$

8.5.3 Matrix diagonalization

A matrix $A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) is *diagonalizable* if it is similar to a diagonal matrix. To *diagonalize* a diagonalizable matrix A means to compute a diagonal matrix D and an invertible matrix P such that $D = P^{-1}AP$ (equivalently: $A = PDP^{-1}$). Theorem 8.5.6 (below) gives a necessary and sufficient condition for a matrix to be diagonalizable, and it essentially gives us a recipe for actually diagonalizing such a matrix. We note that one reason why we care about diagonalizable matrices is that, using Propositions 4.5.15 and 8.5.3, we can easily compute a formula for an arbitrary power of a diagonalizable matrix A, at least provided we have actually diagonalized it first (see Examples 8.5.10 and 8.5.11).

Theorem 8.5.6. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a matrix. Then A is diagonalizable if and only if $\mathbb{F}^n$ has an eigenbasis associated with A. Moreover, if $\mathcal{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is any eigenbasis of $\mathbb{F}^n$ associated with A, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A associated with the eigenvectors $\mathbf{p}_1, \ldots, \mathbf{p}_n$, respectively, then*

$$D = P^{-1}AP \qquad \text{and} \qquad A = PDP^{-1},$$

where $D = D(\lambda_1, \ldots, \lambda_n)$ and $P = \begin{bmatrix} \mathbf{p}_1 & \cdots & \mathbf{p}_n \end{bmatrix}$.

Remark: We could, in principle, obtain Theorem 8.5.6 as a corollary of Theorem 8.5.4.[21] However, in this particular case, it is actually not much more difficult to prove Theorem 8.5.4 "from scratch," i.e. using matrices only.

Proof. Suppose first that $\mathbb{F}^n$ has an eigenbasis associated with A, and let $\mathcal{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ be such an eigenbasis. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A associated with the eigenvectors $\mathbf{p}_1, \ldots, \mathbf{p}_n$, so that $A\mathbf{p}_i = \lambda_i\mathbf{p}_i$ for all indices $i \in \{1, \ldots, n\}$. Further, set $D := D(\lambda_1, \ldots, \lambda_n)$ and $P := \begin{bmatrix} \mathbf{p}_1 & \cdots & \mathbf{p}_n \end{bmatrix}$, as in the statement of the theorem. Since the columns of P form a basis of $\mathbb{F}^n$, the Invertible Matrix Theorem (see subsection 8.2.6) guarantees that P is invertible. Now, it suffices to show that $PD = AP$, since this will imply that $D = P^{-1}AP$ and $A = PDP^{-1}$ (because P is invertible). We compute:

$$\begin{aligned}
PD &= \begin{bmatrix} \lambda_1\mathbf{p}_1 & \cdots & \lambda_n\mathbf{p}_n \end{bmatrix} && \text{by Proposition 8.5.1(b)} \\
&= \begin{bmatrix} A\mathbf{p}_1 & \cdots & A\mathbf{p}_n \end{bmatrix} && \text{because } A\mathbf{p}_i = \lambda_i\mathbf{p}_i \text{ for all } i \in \{1, \ldots, n\} \\
&= A\underbrace{\begin{bmatrix} \mathbf{p}_1 & \cdots & \mathbf{p}_n \end{bmatrix}}_{=P} && \text{by the definition of matrix multiplication} \\
&= AP.
\end{aligned}$$

[21] Try it!

Suppose, conversely, that A is diagonalizable, and fix matrices $D, P \in \mathbb{F}^{n\times n}$ such that D is diagonal, P is invertible, and $D = P^{-1}AP$. Set $D = D(\lambda_1, \ldots, \lambda_n)$ and $P = \begin{bmatrix} \mathbf{p}_1 & \ldots & \mathbf{p}_n \end{bmatrix}$. Since P is invertible, the Invertible Matrix Theorem (see subsection 8.2.6) guarantees that its columns form a basis of $\mathbb{F}^n$ (and in particular, $\mathbf{p}_1, \ldots, \mathbf{p}_n$ are all non-zero). Now, let us show that the columns of P are eigenvectors of A. Since $D = P^{-1}AP$, we have that $AP = PD$. But note that by the definition of matrix multiplication, we have that

$$AP = A\begin{bmatrix} \mathbf{p}_1 & \ldots & \mathbf{p}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{p}_1 & \ldots & A\mathbf{p}_n \end{bmatrix},$$

whereas by Proposition 8.5.1(b), we have that

$$PD = \begin{bmatrix} \lambda_1\mathbf{p}_1 & \ldots & \lambda_n\mathbf{p}_n \end{bmatrix}.$$

Since $AP = PD$, we deduce that

$$\begin{bmatrix} A\mathbf{p}_1 & \ldots & A\mathbf{p}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{p}_1 & \ldots & \lambda_n\mathbf{p}_n \end{bmatrix}.$$

But this implies that for all $i \in \{1, \ldots, n\}$, we have that $A\mathbf{p}_i = \lambda_i\mathbf{p}_i$, and so $\mathbf{p}_i$ is an eigenvector of A associated with the eigenvalue λ_i.[22] It now follows that $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is an eigenbasis of $\mathbb{F}^n$ associated with the matrix A. This completes the argument. □

Corollary 8.5.7. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$. If A has n distinct eigenvalues, then A is diagonalizable.*

Proof. Assume that A has n distinct eigenvalues. By Corollary 8.4.6, $\mathbb{F}^n$ has an eigenbasis associated with A. So, by Theorem 8.5.6, A is diagonalizable. □

Diagonalizing a matrix. Note that Theorems 8.4.5 and 8.5.6 together give us a recipe for determining whether a matrix $A \in \mathbb{F}^{n\times n}$ is diagonalizable, and if so, for diagonalizing it (i.e. for finding a diagonal matrix D and an invertible matrix P, both in $\mathbb{F}^{n\times n}$, such that $D = P^{-1}AP$). We proceed as follows.

1. We compute the characteristic polynomial $p_A(\lambda)$ and its roots. By Theorem 8.2.2, the roots of $p_A(\lambda)$ are the eigenvalues of A, and we can read off the algebraic multiplicities of those eigenvalues from the polynomial $p_A(\lambda)$.

    - Computing the roots of $p_A(\lambda)$ is the computationally tricky part, since there is no formula for computing the roots of a high-degree polynomial. If we cannot figure out how to compute the roots of $p_A(\lambda)$, then we are stuck: the matrix A may or may not be diagonalizable, but computationally, we cannot diagonalize it.

[22] Note that we are also using the fact that $\mathbf{p}_i \neq \mathbf{0}$.

2. If the sum of algebraic multiplicities of the eigenvalues of A is less than n, then by Theorem 8.4.5, $\mathbb{F}^n$ does not have an eigenbasis associated with A, and so by Theorem 8.5.6, A is not diagonalizable.

3. From now on, we assume that the sum of algebraic multiplicities of the eigenvalues of A, call them $\lambda_1, \ldots, \lambda_k$, is n. We then compute a basis $\mathcal{B}_i$ for each eigenspace $E_{\lambda_i}(A)$, which allows us to compute the geometric multiplicities of all the eigenvalues of A.

4. If the geometric multiplicity of some eigenvalue of A is smaller than its algebraic multiplicity,[23] then by Theorem 8.4.5, $\mathbb{F}^n$ does not have an eigenbasis associated with A, and so by Theorem 8.5.6, A is not diagonalizable.

5. From now on, we assume that the geometric multiplicity of each eigenvalue of A is equal to its algebraic multiplicity. Theorem 8.4.5 then guarantees that $\mathbb{F}^n$ has an eigenbasis associated with A, and moreover, that $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ is one such eigenbasis.

6. By Theorem 8.5.6, A is diagonalizable. We now follow the recipe from Theorem 8.5.6 to actually diagonalize A.

7. We form the matrix P whose columns are precisely the vectors in the eigenbasis $\mathcal{B}$. We form the diagonal matrix D, where on the main diagonal we place the eigenvalues of A, taking care that, for each $i \in \{1, \ldots, n\}$, the i-th entry on the main diagonal of D is the eigenvalue associated with the i-th column of P (which is, by construction, an eigenvector of A). Now $D = P^{-1}AP$.

Example 8.5.8. *Consider the following matrix in* $\mathbb{C}^{3\times 3}$*:*

$$A = \begin{bmatrix} 4 & 0 & -2 \\ 2 & 5 & 4 \\ 0 & 0 & 5 \end{bmatrix}.$$

Determine whether A is diagonalizable, and if so, diagonalize it.

Solution. The matrix A is precisely the matrix from Example 8.2.4. In that example, we determined that A has two eigenvalues, namely, $\lambda_1 = 4$ (with algebraic multiplicity 1 and geometric multiplicity 1) and $\lambda_2 = 5$ (with algebraic multiplicity 2 and geometric multiplicity 2). Since the sum of algebraic multiplicities of the eigenvalues of A is 3, and since the geometric multiplicity of each eigenvalue of A is equal to its algebraic multiplicity, we see that the 3×3 matrix A is indeed diagonalizable. In Example 8.2.4, we saw that $\left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\}$ is a basis of the eigespace $E_{\lambda_1}(A)$, and that

[23] By Theorem 8.2.3, the geometric multiplicity of an eigenvalue is either smaller than or equal to the algebraic multiplicity of that eigenvalue.

$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis of the eigenspace $E_{\lambda_2}(A)$. So, we set

$$D := \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad \text{and} \quad P := \begin{bmatrix} -1 & 0 & -2 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and we see that $D = P^{-1}AP$. □

Example 8.5.9. *Consider the following matrix in $\mathbb{C}^{5\times 5}$:*

$$A = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

Determine whether A is diagonalizable, and if so, diagonalize it.

Solution. The matrix A is precisely the matrix from Example 8.2.8. In that example, we determined that A has three eigenvalues, namely $\lambda_1 = 1$ (with algebraic multiplicity 2 and geometric multiplicity 2), $\lambda_2 = 2$ (with algebraic multiplicity 1 and geometric multiplicity 1), and $\lambda_3 = 3$ (with algebraic multiplicity 2 and geometric multiplicity 1). Since the geometric multiplicity of the eigenvalue $\lambda_3 = 3$ is strictly smaller than the algebraic multiplicity, we see that A is not diagonalizable. □

Matrix powers of diagonalizable matrices. If A is an arbitrary square matrix with entries in some field $\mathbb{F}$, then it is not easy to find a nice formula for arbitrary powers of A. On the other hand, if A happens to be diagonalizable, then this can easily be done using Propositions 4.5.15 and 8.5.3, at least provided we have actually diagonalized the matrix first.

Example 8.5.10. *Consider the following matrix in $\mathbb{C}^{3\times 3}$:*

$$A = \begin{bmatrix} 4 & 0 & -2 \\ 2 & 5 & 4 \\ 0 & 0 & 5 \end{bmatrix}.$$

Find a formula for A^m, where m is an arbitrary non-negative integer. Does the formula also work for negative integers m?

Solution. This is the matrix from Example 8.5.8. In that example, we computed matrices $D, P \in \mathbb{C}^{3\times 3}$ such that D is diagonal, P is invertible, and $D = P^{-1}AP$. The matrices in question were

$$D := \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad \text{and} \quad P := \begin{bmatrix} -1 & 0 & -2 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We then compute

$$P^{-1} = \begin{bmatrix} -1 & 0 & -2 \\ 2 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then for all non-negative integers m, we have the following:

$$\begin{aligned} A^m &\overset{(*)}{=} PD^mP^{-1} \\ &\overset{(**)}{=} \underbrace{\begin{bmatrix} -1 & 0 & -2 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{=P} \underbrace{\begin{bmatrix} 4^m & 0 & 0 \\ 0 & 5^m & 0 \\ 0 & 0 & 5^m \end{bmatrix}}_{=D^m} \underbrace{\begin{bmatrix} -1 & 0 & -2 \\ 2 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}}_{=P^{-1}} \\ &\overset{(***)}{=} \begin{bmatrix} 4^m & 0 & 2\cdot 4^m - 2\cdot 5^m \\ -2\cdot 4^m + 2\cdot 5^m & 5^m & -4^{m+1} + 4\cdot 5^m \\ 0 & 0 & 5^m \end{bmatrix} \\ &= \begin{bmatrix} 4^m & 0 & 2(4^m - 5^m) \\ 2(5^m - 4^m) & 5^m & 4(5^m - 4^m) \\ 0 & 0 & 5^m \end{bmatrix}, \end{aligned}$$

where (*) follows from Proposition 4.5.15, (**) follows from Proposition 8.5.3, and (***) follows by simple matrix multiplication.

It remains to check whether our formula for A^m also works for negative integers m. If A is not invertible, then A^m is not defined for negative integers m (and in particular, the formula does not work for negative m). On the other hand, if A is invertible, then Proposition 4.5.15 guarantees that our formula for A^m does in fact work for negative integers m. To see if A is invertible, we compute

$$\det(A) \overset{(*)}{=} \det(D) = 4\cdot 5\cdot 5 = 100 \neq 0$$

where (*) follows from Corollary 7.5.4, since matrices A and D are similar. Since $\det(A) \neq 0$, Theorem 7.4.1 guarantees that A is invertible. So, our formula for A^m does in fact work for negative numbers m as well.

To summarize, we have shown that

$$A^m = \begin{bmatrix} 4^m & 0 & 2(4^m - 5^m) \\ 2(5^m - 4^m) & 5^m & 4(5^m - 4^m) \\ 0 & 0 & 5^m \end{bmatrix}$$

for all integers m (positive, negative, and zero).

Optional: Since it is easy to miscompute, it is not a bad idea to check that the formula that we obtained is correct. We can do this by induction, as follows. For $m = 0$, we have

$$\begin{bmatrix} 4^0 & 0 & 2(4^0 - 5^0) \\ 2(5^0 - 4^0) & 5^0 & 4(5^0 - 4^0) \\ 0 & 0 & 5^0 \end{bmatrix} = I_3 = A^0.$$

Now, fix a non-negative integer m, and assume inductively that

$$A^m = \begin{bmatrix} 4^m & 0 & 2(4^m - 5^m) \\ 2(5^m - 4^m) & 5^m & 4(5^m - 4^m) \\ 0 & 0 & 5^m \end{bmatrix}.$$

We now compute:

$$\begin{aligned} A^{m+1} &= \underbrace{\begin{bmatrix} 4^m & 0 & 2(4^m - 5^m) \\ 2(5^m - 4^m) & 5^m & 4(5^m - 4^m) \\ 0 & 0 & 5^m \end{bmatrix}}_{\overset{(*)}{=} A^m} \underbrace{\begin{bmatrix} 4 & 0 & -2 \\ 2 & 5 & 4 \\ 0 & 0 & 5 \end{bmatrix}}_{=A} \\ &\overset{(**)}{=} \begin{bmatrix} 4^{m+1} & 0 & 8 \cdot 4^m - 10 \cdot 5^m \\ 10 \cdot 5^m - 8 \cdot 4^m & 5^{m+1} & 20 \cdot 5^m - 16 \cdot 4^m \\ 0 & 0 & 5^{m+1} \end{bmatrix} \\ &= \begin{bmatrix} 4^{m+1} & 0 & 2(4^{m+1} - 5^{m+1}) \\ 2(5^{m+1} - 4^{m+1}) & 5^{m+1} & 4(5^{m+1} - 4^{m+1}) \\ 0 & 0 & 5^{m+1} \end{bmatrix}, \end{aligned}$$

where (*) follows from the induction hypothesis, and (**) follows via simple matrix multiplication. This completes the induction and proves that our formula is correct. (Technically, we have only shown that our formula is correct for **non-negative** integers m. We could also prove inductively that the formula is true for negative integers m, but in practice, we need not bother. This is because this part is optional anyway, and it simply serves to increase our confidence that we didn't make any mistakes in our computation.) □

Example 8.5.11. *Consider the following matrix in $\mathbb{C}^{2\times 2}$:*

$$A = \begin{bmatrix} 6 & -2 \\ 6 & -1 \end{bmatrix}.$$

Find a formula for A^m, where m is an arbitrary non-negative integer. Does the formula also work for negative integers m?

Solution. We will first try to diagonalize A. We start by computing and factoring the characteristic polynomial of A:

$$\begin{aligned} p_A(\lambda) &= \det(\lambda I_2 - A) \\ &= \begin{vmatrix} \lambda - 6 & 2 \\ -6 & \lambda + 1 \end{vmatrix} \\ &= \lambda^2 - 5\lambda + 6 \\ &= (\lambda - 2)(\lambda - 3). \end{aligned}$$

So, A has two eigenvalues: $\lambda_1 = 2$ and $\lambda_2 = 3$, each with algebraic multiplicity 1. By Corollary 8.5.7, A is diagonalizable. So, let us diagonalize it.

First, we compute a basis of $E_{\lambda_1}(A)$. We have the following:

$$\mathrm{RREF}(\lambda_1 I_2 - A) = \mathrm{RREF}\Big(\begin{bmatrix} -4 & 2 \\ -6 & 3 \end{bmatrix}\Big) = \begin{bmatrix} 1 & -1/2 \\ 0 & 0 \end{bmatrix}.$$

So, $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ is a basis of $E_{\lambda_1}(A) = \mathrm{Nul}(\lambda_2 I_2 - A)$.

We now compute a basis of $E_{\lambda_2}(A)$. We have the following:

$$\mathrm{RREF}(\lambda_2 I_2 - A) = \mathrm{RREF}\Big(\begin{bmatrix} -3 & 2 \\ -6 & 4 \end{bmatrix}\Big) = \begin{bmatrix} 1 & -2/3 \\ 0 & 0 \end{bmatrix}.$$

So, $\left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$ is a basis of $E_{\lambda_2}(A) = \mathrm{Nul}(\lambda_2 I_2 - A)$.

Now, for

$$D := \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad P := \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix},$$

we have that $D = P^{-1}AP$. Moreover, we compute

$$\mathrm{RREF}\Big(\left[\begin{array}{c|c} P & I_2 \end{array}\right]\Big) = \left[\begin{array}{cc|cc} 1 & 0 & -3 & 2 \\ 0 & 1 & 2 & -1 \end{array}\right],$$

and we deduce that

$$P^{-1} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}.$$

We note that

$$\det(A) \overset{(*)}{=} \det(D) \overset{(**)}{=} 2 \cdot 3 = 6 \neq 0,$$

where (*) follows from Corollary 7.5.4 (because A and D are similar), and (**) follows from Proposition 7.3.1 (because D is diagonal, and in particular, triangular). By the

Invertible Matrix Theorem (see subsection 8.2.6), it follows that A is invertible. So, by Proposition 4.5.15, the following holds for all integers m:

$$\begin{aligned} A^m &= PD^mP^{-1} \\ &= \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 2^m & 0 \\ 0 & 3^m \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix} && \text{by Proposition 8.5.3, since } D \text{ is diagonal} \\ &= \begin{bmatrix} -3 \cdot 2^m + 4 \cdot 3^m & 2^{m+1} - 2 \cdot 3^m \\ -3 \cdot 2^{m+1} + 2 \cdot 3^{m+1} & 2^{m+2} - 3^{m+1} \end{bmatrix}. \end{aligned}$$

This is the formula that we need, and as we saw, it works for all integers m (positive, negative, and zero).

Optional: Let us check that our answer is correct, at least for non-negative integers m. For $m = 0$, we have the following:

$$\begin{bmatrix} -3 \cdot 2^0 + 4 \cdot 3^0 & 2^{0+1} - 2 \cdot 3^0 \\ -3 \cdot 2^{0+1} + 2 \cdot 3^{0+1} & 2^{0+2} - 3^{0+1} \end{bmatrix} = I_2 = A^0.$$

Now, fix a non-negative integer m, and assume inductively that

$$A^m = \begin{bmatrix} -3 \cdot 2^m + 4 \cdot 3^m & 2^{m+1} - 2 \cdot 3^m \\ -3 \cdot 2^{m+1} + 2 \cdot 3^{m+1} & 2^{m+2} - 3^{m+1} \end{bmatrix}.$$

We now compute:

$$\begin{aligned} A^{m+1} &= \underbrace{\begin{bmatrix} -3 \cdot 2^m + 4 \cdot 3^m & 2^{m+1} - 2 \cdot 3^m \\ -3 \cdot 2^{m+1} + 2 \cdot 3^{m+1} & 2^{m+2} - 3^{m+1} \end{bmatrix}}_{\overset{(*)}{=} A^m} \underbrace{\begin{bmatrix} 6 & -2 \\ 6 & -1 \end{bmatrix}}_{=A} \\ &\overset{(**)}{=} \begin{bmatrix} -6 \cdot 2^m + 12 \cdot 3^m & 4 \cdot 2^m - 6 \cdot 3^m \\ -6 \cdot 2^{m+1} + 6 \cdot 3^{m+1} & 4 \cdot 2^{m+1} - 3 \cdot 3^{m+1} \end{bmatrix} \\ &= \begin{bmatrix} -3 \cdot 2^{m+1} + 4 \cdot 3^{m+1} & 2^{m+2} - 2 \cdot 3^{m+1} \\ -3 \cdot 2^{m+2} + 2 \cdot 3^{m+2} & 2^{m+3} - 3^{m+2} \end{bmatrix}, \end{aligned}$$

where (*) follows from the induction hypothesis, and (**) follows via simple matrix multiplication. So, our formula is correct. □

Reading off the spectrum and bases of the eigenspaces of a square matrix from its diagonalization. Suppose that we have successfully diagonalized a square matrix $A \in \mathbb{F}^{n \times n}$ (where $\mathbb{F}$ is a field), that is, that we have computed a diagonal matrix

D and an invertible matrix P, both in $\mathbb{F}^{n\times n}$, such that $D = P^{-1}AP$. Then we can easily read off the spectrum and a basis of each eigenspace of A, as Proposition 8.5.12 (below) shows. We note that this proposition essentially summarizes various facts about diagonalizable matrices that we have proven already, but it is convenient to have them stated in one proposition.

Proposition 8.5.12. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$. Assume that $D = P^{-1}AP$, where $D = D(\lambda_1, \ldots, \lambda_n)$ is a diagonal and $P = \begin{bmatrix} \mathbf{p}_1 & \ldots & \mathbf{p}_n \end{bmatrix}$ an invertible matrix, both in $\mathbb{F}^{n\times n}$. Then the characteristic polynomial of A is*

$$p_A(\lambda) \quad = \quad \prod_{i=1}^{n}(\lambda - \lambda_i) \quad = \quad (\lambda - \lambda_1)\ldots(\lambda - \lambda_n),$$

and the spectrum of A is $\{\lambda_1, \ldots, \lambda_n\}$. Moreover, for each eigenvalue λ_0 of A,[24] *the algebraic and geometric multiplicity of λ_0 are both equal to the number of times that λ_0 appears on the main diagonal of D, and moreover, if λ_0 appears precisely in positions $i_1, \ldots, i_k$ of the main diagonal of D, then the corresponding columns of P (i.e. vectors $\mathbf{p}_{i_1}, \ldots, \mathbf{p}_{i_k}$) form a basis of the eigenspace $E_{\lambda_0}(A)$. Finally, $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is an eigenbasis of $\mathbb{F}^n$ associated with the matrix A.*

Proof. First, we we note that

$$p_A(\lambda) \quad \overset{(*)}{=} \quad p_D(\lambda) \quad \overset{(**)}{=} \quad (\lambda - \lambda_1)\ldots(\lambda - \lambda_n)$$

where (*) follows from Theorem 8.2.9 (because A and D are similar), and (**) follows from Proposition 8.2.7 (because D is diagonal, and in particular, triangular). It now immediately follows that the spectrum of A is $\{\lambda_1, \ldots, \lambda_n\}$.

> **Claim.** For all indices $i \in \{1, \ldots, n\}$, $\mathbf{p}_i$ is an eigenvector of A associated with the eigenvalue λ_i.

Proof of the Claim. See the proof of Theorem 8.5.6.[25] ♦

[24] So, $\lambda_0 \in \{\lambda_1, \ldots, \lambda_n\}$, since $\{\lambda_1, \ldots, \lambda_n\}$ is the spectrum of A.

[25] For the sake of completeness, here is the proof again, copy-pasted from the proof of Theorem 8.5.6. Since $D = P^{-1}AP$, we have that $AP = PD$. But note that by the definition of matrix multiplication, we have that

$$AP \quad = \quad A\begin{bmatrix} \mathbf{p}_1 & \ldots & \mathbf{p}_n \end{bmatrix} \quad = \quad \begin{bmatrix} A\mathbf{p}_1 & \ldots & A\mathbf{p}_n \end{bmatrix},$$

whereas by Proposition 8.5.1(b), we have that

$$PD \quad = \quad \begin{bmatrix} \lambda_1\mathbf{p}_1 & \ldots & \lambda_n\mathbf{p}_n \end{bmatrix}.$$

Since $AP = PD$, we deduce that

$$\begin{bmatrix} A\mathbf{p}_1 & \ldots & A\mathbf{p}_n \end{bmatrix} \quad = \quad \begin{bmatrix} \lambda_1\mathbf{p}_1 & \ldots & \lambda_n\mathbf{p}_n \end{bmatrix}.$$

But this implies that for all $i \in \{1, \ldots, n\}$, we have that $A\mathbf{p}_i = \lambda_i\mathbf{p}_i$, and so $\mathbf{p}_i$ is an eigenvector of A associated with the eigenvalue λ_i.

Using the Claim, we can easily show that $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is an eigenbasis of $\mathbb{F}^n$ associated with the matrix A. Indeed, since P is invertible, the Invertible Matrix Theorem (see subsection 8.2.6) implies that the columns of P form a basis of $\mathbb{F}^n$. On the other hand, by the Claim, each column of P is an eigenvector of A. So, $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is indeed an eigenbasis of $\mathbb{F}^n$ associated with the matrix A.

Now, suppose that λ_0 is any eigenvalue of A. Since $\mathbb{F}^n$ has an eigenbasis (namely, $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$) associated with A, Theorem 8.4.5 guarantees that the geometric multiplicity of each eigenvalue of A is equal to its algebraic multiplicity. In particular, the geometric multiplicity of λ_0 is equal to its algebraic multiplicity, which is precisely the number of times that it appears on the main diagonal of D.

Let us suppose that the eigenvalue λ_0 appears precisely k times on the main diagonal of D (so that both the algebraic and geometric multiplicity of the eigenvalue λ_0 is k), and that it appears precisely in positions $i_1, \ldots, i_k$, so that $\lambda_0 = \lambda_{i_1} = \cdots = \lambda_{i_k}$. We must show that $\mathcal{B}_0 := \{\mathbf{p}_{i_1}, \ldots, \mathbf{p}_{i_k}\}$ is a basis of the eigenspace $E_{\lambda_0}(A)$. First of all, by the Claim, all vectors in $\mathcal{B}_0$ are eigenvectors of A associated with the eigenvalue λ_0, and so they belong to $E_{\lambda_0}(A)$. Moreover, since $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is a basis of $\mathbb{F}^n$ (proven above), the set $\mathcal{B}_0$ is linearly independent. So, $\mathcal{B}_0$ is a set of k linearly independent vectors in the eigenspace $E_{\lambda_0}(A)$. Since $\dim\big(E_{\lambda_0}(A)\big) = k$ (because the geometric multiplicity of the eigenvalue λ_0 is k), Corollary 3.2.20 guarantees that $\mathcal{B}_0$ is in fact a basis of $E_{\lambda_0}(A)$. □

Example 8.5.13. *Consider the following matrices in $\mathbb{C}^{6\times 6}$ (color coded for emphasis):*

$$D = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & 3 & 8 & 8 & 3 & 4 \\ 2 & 8 & 0 & 0 & 0 & 2 \\ 5 & 4 & 6 & 4 & 5 & 0 \\ 0 & 5 & 8 & 5 & 4 & 3 \\ 1 & 0 & 8 & 0 & 3 & 0 \\ 0 & 2 & 0 & 3 & 0 & 2 \end{bmatrix}.$$

It can be checked that P is invertible (for example, we can compute that $det(P) = -1020 \neq 0$, and so by Theorem 7.4.1, P is invertible). We now set $A = PDP^{-1}$, so that $D = P^{-1}AP$. Then by Proposition 8.5.12, all the following hold:

- *the characteristic polynomial of A is*
$$p_A(\lambda) = (\lambda - 3)(\lambda - 4)^3(\lambda - 5)^2;$$
- *the spectrum of A is $\{5, 4, 5, 3, 4, 4\}$, which we can optionally reorder as $\{3, 4, 4, 4, 5, 5\}$;*
- *the eigenvalues of A are 3 (with algebraic and geometric multiplicity 1), 4 (with algebraic and geometric multiplicity 3), and 5 (with algebraic and geometric multiplicity 2);*

- *we can read off bases of the eigenspaces $E_3(A)$, $E_4(A)$, and $E_5(A)$, as follows:*
  - *a basis of $E_3(A)$ is*
$$\left\{ \begin{bmatrix} 8 \\ 0 \\ 4 \\ 5 \\ 0 \\ 3 \end{bmatrix} \right\},$$
  - *a basis of $E_4(A)$ is*
$$\left\{ \begin{bmatrix} 3 \\ 8 \\ 4 \\ 5 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 5 \\ 4 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 0 \\ 3 \\ 0 \\ 2 \end{bmatrix} \right\},$$
  - *a basis of $E_5(A)$ is*
$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 8 \\ 0 \\ 6 \\ 8 \\ 8 \\ 0 \end{bmatrix} \right\};$$
- *the columns of P form an eigenbasis of $\mathbb{C}^n$ associated with the matrix A.*

8.6 The Jordan normal form

This section is by far the most technical one of these lecture notes. So, some reading instructions are in order.

In subsection 8.6.1, we introduce matrices in "Jordan normal form" (also known as "Jordan matrices"), which are upper triangular matrices of a particular kind. In subsection 8.6.1, we also state our main results concerning such matrices. In particular, we will see that if $\mathbb{F}$ is an **algebraically closed field**,[26] then every matrix $A \in \mathbb{F}^{n \times n}$ is similar to some Jordan matrix $J \in \mathbb{F}^{n \times n}$ (see Theorem 8.6.2); the matrix J is in some sense unique, and it is referred to it as the "Jordan normal form of A." Moreover, Theorem 8.6.6 gives a recipe for computing the Jordan normal form of a given square matrix (with entries in an algebraically closed field $\mathbb{F}$), and subsection 8.6.2 contains a couple of worked out examples applying this theorem.

[26] Algebraically closed fields were discussed in subsection 2.4.5.

In subsection 8.6.3, we give an outline of the proof of our main theorems. For most readers, it is enough to read subsections 8.6.1, 8.6.2, and 8.6.3.

The remaining subsections (namely, subsections 8.6.4, 8.6.5, 8.6.6, and 8.6.7) are highly technical and can be considered "optional reading" for particularly ambitious students. Subsections 8.6.4, 8.6.5, 8.6.6 together give a formal proof of our main theorems (the theorems stated in subsection 8.6.1). Subsection 8.6.7 is computational, and it is essentially a more ambitious version of subsection 8.6.2. Suppose we are given a square matrix $A \in \mathbb{F}^{n\times n}$, where $\mathbb{F}$ is an algebraically closed field. By Theorem 8.6.2, the matrix A is similar to some matrix in Jordan normal form. In other words, there exists a Jordan matrix J and an invertible matrix P, both in $\mathbb{F}^{n\times n}$, such that $J = P^{-1}AP$. In subsection 8.6.2, we give a recipe for computing the Jordan matrix J. In subsection 8.6.7, we give a recipe for computing both J and P; the correctness of this recipe essentially follows from the proofs given in subsections 8.6.4, 8.6.5, and 8.6.6. We note, however, that readers who just wish to learn how to compute J and P mechanically, without necessarily understanding why the procedure works, can simply follow the steps described in subsection 8.6.7 (without having read the proofs from subsections 8.6.4, 8.6.5, and 8.6.6 first).

8.6.1 The Jordan normal form: definitions and statements of main theorems

In what follows, it will be notationally useful to define the "direct sum" of matrices. So, suppose that $\mathbb{F}$ is a field and $A \in \mathbb{F}^{n_1\times n_1}$ and $B \in \mathbb{F}^{n_2\times n_2}$ are **square** matrices. Then the *direct sum* of A and B is the $(n_1+n_2)\times(n_1+n_2)$ matrix

$$A \oplus B := \left[\begin{array}{c|c} A & O_{n_1\times n_2} \\ \hline O_{n_2\times n_1} & B \end{array}\right].$$

More generally, for square matrices $A_1 \in \mathbb{F}^{n_1\times n_1}, A_2 \in \mathbb{F}^{n_2\times n_2}, \ldots, A_k \in \mathbb{F}^{n_k\times n_k}$, we define the *direct sum* of $A_1, A_2 \ldots, A_k$ to be the $(n_1+n_2+\cdots+n_k)\times(n_1+n_2+\cdots+n_k)$ matrix

$$A_1 \oplus A_2 \oplus \cdots \oplus A_k := \left[\begin{array}{c|c|c|c} A_1 & O_{n_1\times n_2} & \ldots & O_{n_1\times n_k} \\ \hline O_{n_2\times n_1} & A_2 & \ldots & O_{n_2\times n_k} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline O_{n_k\times n_1} & O_{n_k\times n_2} & \ldots & A_k \end{array}\right].$$

For example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \oplus \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 4 & 5 & 6 & 0 \\ 0 & 0 & 7 & 8 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For a field $\mathbb{F}$, a scalar $\lambda_0 \in \mathbb{F}$, and a positive integer t, the *Jordan block* $J_t(\lambda_0)$ is defined to be following $t \times t$ matrix (with entries understood to be in $\mathbb{F}$):

$$J_t(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_0 \end{bmatrix}_{t \times t}.$$

Thus, $J_t(\lambda_0)$ is a matrix in $\mathbb{F}^{t \times t}$, it has all λ_0's on the main diagonal, all 1's on the diagonal right above the main diagonal, and 0's everywhere else. For example:

- $J_1(\lambda_0) = \begin{bmatrix} \lambda_0 \end{bmatrix}$;
- $J_2(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{bmatrix}$;
- $J_3(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 \\ 0 & \lambda_0 & 1 \\ 0 & 0 & \lambda_0 \end{bmatrix}$;
- $J_4(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 & 0 \\ 0 & \lambda_0 & 1 & 0 \\ 0 & 0 & \lambda_0 & 1 \\ 0 & 0 & 0 & \lambda_0 \end{bmatrix}$;
- $J_5(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 & 0 & 0 \\ 0 & \lambda_0 & 1 & 0 & 0 \\ 0 & 0 & \lambda_0 & 1 & 0 \\ 0 & 0 & 0 & \lambda_0 & 1 \\ 0 & 0 & 0 & 0 & \lambda_0 \end{bmatrix}$.

A *Jordan matrix* (also called a matrix in *Jordan normal form*) is any matrix that is a direct sum of one or more Jordan blocks. Thus, a Jordan matrix is a matrix of the form

$$J_{t_1}(\lambda_1) \oplus J_{t_2}(\lambda_2) \oplus \cdots \oplus J_{t_\ell}(\lambda_\ell) = \left[\begin{array}{c|c|c|c} J_{t_1}(\lambda_1) & O & \ldots & O \\ \hline O & J_{t_2}(\lambda_2) & \ldots & O \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline O & O & \ldots & J_{t_\ell}(\lambda_\ell) \end{array}\right],$$

where $\lambda_1, \ldots, \lambda_\ell$ are scalars in $\mathbb{F}$, $t_1, \ldots, t_\ell$ are positive integers, and the O's are zero matrices of appropriate sizes. For instance, the following is a Jordan matrix with four Jordan blocks, namely $J_3(5)$, $J_2(2)$, $J_1(2)$, and $J_3(5)$:

$$J_3(5) \oplus J_2(2) \oplus J_1(2) \oplus J_3(5) = \begin{bmatrix} 5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}.$$

Remark: Every diagonal matrix is a Jordan matrix. Moreover, note that a Jordan matrix is diagonal if and only if all its Jordan blocks are of size 1×1. On the other hand, if some Jordan block of a Jordan matrix J is of larger size (i.e. is of size $t \times t$ for some $t \geq 2$), then J will have at least one 1 on the diagonal right above the main diagonal.

Remark: Not all matrices that have an arbitrary main diagonal, all 0's and 1's on the diagonal right above the main one, and 0's everywhere else, are Jordan matrices. For example, the matrix

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

is **not** a Jordan matrix (because it is not a direct sum of Jordan blocks).

Theorem 8.6.1. *Let $\mathbb{F}$ be a field, and let $J_1, J_2 \in \mathbb{F}^{n \times n}$ be Jordan matrices. Then J_1 and J_2 are similar if and only if they have exactly the same Jordan blocks (counting repetitions, but not counting the order in which the blocks appear in the two matrices).*

The proof of Theorem 8.6.1 is postponed to subsection 8.6.4. For now, let us point out that the "if" ("$\Longleftarrow$") part is fairly easy, whereas the "only if" ("$\Longrightarrow$") part requires some work. The "if" part, i.e. the fact that two Jordan matrices that have the same Jordan blocks, counting repetitions, are indeed similar essentially follows from the fact that similar matrices represent the same linear function, only with respect to (possibly) different bases (see Theorem 4.5.16). A change in the order of Jordan blocks corresponds to a change in the order of basis vectors. For a formal proof, see Proposition 8.6.12. For now, let us take a look at a special case in order to gain some intuition. Suppose that V is a finite-dimensional vector space over a field $\mathbb{F}$, that $f : V \to V$ is a linear function, and that $\mathcal{B} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{t_1}, \mathbf{b}_1, \ldots, \mathbf{b}_{t_2}, \mathbf{c}_1, \ldots, \mathbf{c}_{t_3}, \mathbf{d}_1, \ldots, \mathbf{d}_{t_4}\}$ (with $t_1, t_2, t_3, t_4 \geq 1$) is a basis of V such that

$${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}} = \left[\begin{array}{c|c|c|c} J_{t_1}(\lambda_1) & O & O & O \\ \hline O & J_{t_2}(\lambda_2) & O & O \\ \hline O & O & J_{t_3}(\lambda_3) & O \\ \hline O & O & O & J_{t_4}(\lambda_4) \end{array}\right].$$

Then for the basis $\mathcal{C} = \{\mathbf{b}_1, \ldots, \mathbf{b}_{t_2}, \mathbf{d}_1, \ldots, \mathbf{d}_{t_4}, \mathbf{a}_1, \ldots, \mathbf{a}_{t_1}, \mathbf{c}_1, \ldots, \mathbf{c}_{t_3}\}$ of V, we have the following:

$$ {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}} \ = \ \left[\begin{array}{c|c|c|c} J_{t_2}(\lambda_2) & O & O & O \\ \hline O & J_{t_4}(\lambda_4) & O & O \\ \hline O & O & J_{t_1}(\lambda_1) & O \\ \hline O & O & O & J_{t_3}(\lambda_3) \end{array}\right]. $$

By Theorem 4.5.16, matrices ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ and ${}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$ are similar, and so the two Jordan matrices above are similar.

We now state two theorems involving Jordan matrices, namely, Theorems 8.6.2 and 8.6.4 below. As we shall see, the two theorems are equivalent (in the sense that either one easily implies the other). The proofs of the two theorems are long and technical; a proof outline is given in subsection 8.6.3, and a full proof is given in subsections 8.6.4, 8.6.5, 8.6.6.

Theorem 8.6.2. *Assume that $\mathbb{F}$ is an **algebraically closed field**, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then A is similar to a matrix J in Jordan normal form. Moreover, this matrix J is unique up to a reordering of the Jordan blocks.*

Terminology/Remark: Suppose that $A \in \mathbb{F}^{n\times n}$ is a matrix, where $\mathbb{F}$ is some algebraically closed field. Then any Jordan matrix that is similar to A is called a *Jordan normal form* of A. As we saw above, reordering the Jordan blocks of a Jordan matrix produces a Jordan matrix that is similar to the original one. So, if J is a Jordan normal form of A, then any Jordan matrix obtained from J by merely rearranging the order in which the Jordan blocks appear along the main diagonal is also a Jordan normal form of A. However, by the uniqueness part of Theorem 8.6.2, this exhausts the possibilities for different Jordan normal forms of A: any two Jordan normal forms of A have exactly the same Jordan blocks (with repetitions taken into account).

The following is an immediate corollary of Theorems 8.6.1 and 8.6.2 (plus Proposition 4.5.13).

Corollary 8.6.3. *Let $\mathbb{F}$ be an **algebraically closed field**, and let $A, B \in \mathbb{F}^{n\times n}$. Then A and B are similar if and only if they have the same Jordan normal form. More precisely, the following are equivalent:*

(a) A and B are similar;

(b) there exists a Jordan matrix $J \in \mathbb{F}^{n\times n}$ such that both A and B are similar to J;

(c) there exist Jordan matrices $J_A, J_B \in \mathbb{F}^{n\times n}$ such that A is similar to J_A, B is similar to J_B, and the Jordan matrices J_A and J_B can be obtained from each other by possibly rearranging the order of the Jordan blocks.

Proof (assuming Theorems 8.6.1 and 8.6.2). In what follows, we will use the fact that, by Proposition 4.5.13, matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$. We will prove "(a) $\implies$ (b) $\implies$ (c) $\implies$ (a)."

We first assume (a) and prove (b). By Theorem 8.6.2, A is similar to a Jordan matrix $J \in \mathbb{F}^{n\times n}$. Since matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$, it follows that B is similar to J.[27] This proves (b).

Next, we assume (b) and prove (c). By (b), A and B are both similar to the same Jordan matrix $J \in \mathbb{F}^{n\times n}$. But then we simply set $J_A := J$ and $J_B := J$, and (c) follows.

Finally, we assume (c) and prove (a). Let J_A and J_B be as in part (c). Then by Theorem 8.6.1, J_A and J_B are similar. Since matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$, we deduce that A is similar to B,[28] i.e. (a) holds. □

Remark: As we know, the field $\mathbb{C}$ is algebraically closed, and so Corollary 8.6.3 applies to matrices in $\mathbb{C}^{n\times n}$. On the other hand, $\mathbb{R}$ is **not** algebraically closed, and so we cannot apply Corollary 8.6.3 to matrices in $\mathbb{R}^{n\times n}$, or at least not directly. For a way around this, see Theorem 8.6.7.

Theorem 8.6.4. *Let V be a non-trivial, finite-dimensional vector space over an* ***algebraically closed field*** *$\mathbb{F}$, and let $f : V \to V$ be a linear function. Then there exists a basis $\mathcal{B}$ such that the matrix ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ is in Jordan normal form. Moreover, this matrix is unique in the following sense: if $\mathcal{B}_1$ and $\mathcal{B}_2$ are bases of V such that both ${}_{\mathcal{B}_1}[\ f\ ]_{\mathcal{B}_1}$ and ${}_{\mathcal{B}_2}[\ f\ ]_{\mathcal{B}_2}$ are in Jordan normal form, then these two matrices are the same up to a reordering of the Jordan blocks.*

Remarks:

1. Theorems 8.6.2 and 8.6.4 only hold for **algebraically closed fields**. The only algebraically closed field that we have seen is $\mathbb{C}$, but others do exist.

2. Theorem 4.5.16 essentially states that two $n \times n$ matrices are similar if and only if they represent the same linear function from an n-dimensional vector space to itself, only possibly with respect to different bases. It is then easy to show that Theorems 8.6.2 and Theorems 8.6.4 are equivalent in the sense that either one of them (combined with Theorem 4.5.16) readily implies the other. The details are left as an exercise.

3. As we saw in section 8.5, not all square matrices are diagonalizable, i.e. there are square matrices that are not similar to any diagonal matrix. However, as long as we are working over an **algebraically closed field**, Theorem 8.6.2 guarantees that any square matrix is similar to a matrix that is "almost

[27] Indeed, we have that A is similar to both B and and J. So, B is similar to J.

[28] Indeed, we have that A is similar to J_A, that J_A is similar to J_B, and that J_B is similar to B. So, A is similar to B.

diagonal," namely to its Jordan normal form. However, in the special case when a square matrix A is diagonalizable, the Jordan normal form of A is any diagonal matrix D that is similar to A.[29]

4. Since every Jordan matrix is upper triangular, its eigenvalues, together with their algebraic multiplicities, can easily be read off from the Jordan matrix itself (see Proposition 8.2.7): the eigenvalues are precisely the entries along the main diagonal of the Jordan matrix, and the algebraic multiplicity of each eigenvalue is the number of times that it appears on the main diagonal. For instance, the eigenvalues of the Jordan matrix

$$J_3(5) \oplus J_2(2) \oplus J_1(2) \oplus J_3(5) = \begin{bmatrix} 5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

are 5 (with algebraic multiplicity 6) and 2 (with algebraic multiplicity 3).

5. Perhaps more interestingly, the geometric multiplicity of each eigenvalue of a Jordan matrix J can also be read off quite easily: the geometric multiplicity of each eigenvalue λ is precisely the number of Jordan blocks of the form $J_t(\lambda)$ that appear along the main diagonal of J.[30] For instance, for the Jordan matrix above, the geometric multiplicity of the eigenvalue 5 is 2, and the geometric multiplicity of the eigenvalue 2 is also 2.

6. By Theorem 8.2.9, similar matrices have the same eigenvalues, with the same corresponding algebraic multiplicities, and the same corresponding geometric multiplicities. So, if we know the Jordan normal form of a matrix A, then we can easily read off the eigenvalues of A, together with their algebraic and geometric multiplicities. We note, however, that two square matrices of the same size, and with exactly the same eigenvalues, with the same corresponding algebraic and geometric multiplicities, need not be similar. Indeed, it is easy to construct two Jordan matrices that have different Jordan blocks, but have the same eigenvalues with the same corresponding algebraic and geometric multiplicities. By Theorem 8.6.1, such matrices are **not** similar. For a concrete example, consider the Jordan matrices $J_2(\lambda) \oplus J_2(\lambda)$ and $J_3(\lambda) \oplus J_1(\lambda)$, where

[29] All such diagonal matrices D have the spectrum of A on the main diagonal (in some order), and they are all similar to each other.

[30] Check this!

λ is an arbitrary scalar from the field in question; these two matrices have only one eigenvalue, namely λ, with algebraic multiplicity 4 and geometric multiplicity 2, but they have different Jordan blocks and are therefore not similar.

Example 8.6.5. *Let $A_1, A_2, A_3 \in \mathbb{C}^{7\times 7}$ be matrices whose Jordan normal forms are J_1, J_2, J_3, respectively, as follows:*

- $J_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$;
- $J_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$;
- $J_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

Determine which (if any) of A_1, A_2, A_3 are similar. Then, for each $i \in \{1, 2, 3\}$, compute its characteristic polynomial and spectrum, and find all the eigenvalues of A_i, along with their algebraic and geometric multiplicities.

Solutution. We first identify the Jordan blocks of the the three Jordan matrices. In each matrix, we use colors to indicate the Jordan blocks.

- $J_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = J_1(0) \oplus J_3(1) \oplus J_1(1) \oplus J_2(0)$;

- $J_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = J_1(1) \oplus J_2(0) \oplus J_1(0) \oplus J_3(1);$
- $J_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = J_1(0) \oplus J_2(1) \oplus J_2(1) \oplus J_2(0).$

We see that J_1 and J_2 have the same Jordan blocks (counting repetitions), and so A_1 and A_2 are similar. On the other hand, the Jordan blocks of the matrix J_3 are different from those of J_1 and J_2, and so A_3 is not similar to A_1 and A_2.

For each $i \in \{1, 2, 3\}$, we see that the characteristic polynomial of A_i is

$$p_{A_i}(\lambda) \overset{(*)}{=} p_{J_i}(\lambda) \overset{(**)}{=} \lambda^3(\lambda - 1)^4,$$

where (*) follows from the fact that A_i and J_i are similar (we are using Proposition 8.2.9), and (**) from the fact that the Jordan matrix J_i is upper triangular (we are using Proposition 8.2.7). Finally, we see from the matrices J_1, J_2, J_3, that A_1, A_2, A_3 all have spectrum $\{0, 0, 0, 1, 1, 1, 1\}$, and that they all have exactly two eigenvalues: the eigenvalue 0 with algebraic multiplicity 3 and geometric multiplicity 2, and the eigenvalue 1 with algebraic multiplicity 4 and geometric multiplicity 2. □

Computing the Jordan normal form. The following theorem allows us to actually compute the Jordan normal form of a square matrix (with entries in an algebraically closed field $\mathbb{F}$).

Theorem 8.6.6. *Let $\mathbb{F}$ be an **algebraically closed field**, let $A \in \mathbb{F}^{n \times n}$, and let*

$$(\underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{m_k})$$

be the spectrum of A, where $\lambda_1, \ldots, \lambda_k$ are pairwise distinct eigenvalues of f and $m_1, \ldots, m_k$ are positive integers.[31] *Then A is similar to a matrix $J \in \mathbb{F}^{n \times n}$ in Jordan normal form that has the following properties:*

[31] Since $\mathbb{F}$ is algebraically closed, we know that $m_1 + \cdots + m_k = n$.

(i) each Jordan block of the Jordan matrix J is of the form $J_t(\lambda_i)$ for some $i \in \{1, \ldots, k\}$ and $t \in \{1, \ldots, m_i\}$;

(ii) for each $i \in \{1, \ldots, k\}$ and each positive integer r, the Jordan matrix J has exactly

$$\mathit{rank}\big((A - \lambda_i I_n)^{r-1}\big) - \mathit{rank}\big((A - \lambda_i I_n)^{r}\big)$$

many Jordan blocks $J_t(\lambda_i)$ satisfying $t \geq r$.

Moreover, A is similar to any Jordan matrix in $\mathbb{F}^{n \times n}$ that satisfies conditions (i) and (ii) above.

The proof of Theorem 8.6.6 is given in subsection 8.6.6 (the proof relies on the rather technical results of subsections 8.6.4 and 8.6.5). For now, let us just note that Theorem 8.6.6 does indeed allow us to compute the Jordan normal form of a square matrix A with entries in an algebraically closed field $\mathbb{F}$, as long as we are able to factor its characteristic polynomial into linear terms.[32] Indeed, condition (i) of Theorem 8.6.6 tells us what sorts of Jordan blocks the Jordan normal form of A may possibly have. Condition (ii) gives us an easy way to compute the number of Jordan blocks of each type. Indeed, using the set-up and notation from Theorem 8.6.6, we consider an eigenvalue λ_i of A, and we fix a positive integer r. Then the number of Jordan blocks $J_r(\lambda_i)$ in the Jordan normal form of A is exactly

$$\underbrace{\Big(\operatorname{rank}\big((A - \lambda_i I_n)^{r-1}\big) - \operatorname{rank}\big((A - \lambda_i I_n)^{r}\big)\Big)}_{\substack{= \text{ number of Jordan blocks} \\ J_t(\lambda_i) \text{ satisfying } t \geq r}} - \underbrace{\Big(\operatorname{rank}\big((A - \lambda_i I_n)^{r}\big) - \operatorname{rank}\big((A - \lambda_i I_n)^{r+1}\big)\Big)}_{\substack{= \text{ number of Jordan blocks} \\ J_t(\lambda_i) \text{ satisfying } t \geq r+1}}.$$

So, we can compute both the possible types of Jordan blocks that the Jordan normal form of A may have, and the exact number of blocks of each possible type. The reader may have noticed that we in fact get an exact formula

$$\operatorname{rank}\big((A - \lambda_i I_n)^{r-1}\big) + \operatorname{rank}\big((A - \lambda_i I_n)^{r+1} - 2\operatorname{rank}\big((A - \lambda_i I_n)^{r}\big)$$

for the number of Jordan blocks $J_r(\lambda_i)$ in the Jordan normal form of A. However, it is arguably easier to memorize the formula for the number of Jordan blocks of the form $J_t(\lambda_i)$ satisfying $t \geq r$. For a couple of numerical examples, see subsection 8.6.2.

Matrix similarity over a field. A field $\mathbb{F}_1$ is a *subfield* of a field $\mathbb{F}_2$ if the following three conditions are satisfied:

[32] Any non-constant polynomial with coefficients in an algebraically closed field $\mathbb{F}$ can be factored into linear terms (with coefficients in $\mathbb{F}$). However, this is merely an existence result: actually computing the linear factors may be extremely difficult or even impossible. If we get stuck factoring the characteristic polynomial into linear terms, then Theorem 8.6.6 is of no use to us (computationally speaking).

- $\mathbb{F}_1 \subseteq \mathbb{F}_2$;
- for all $a, b \in \mathbb{F}_1$, the sum $a + b$ is the same in $\mathbb{F}_1$ and in $\mathbb{F}_2$;
- for all $a, b \in \mathbb{F}_1$, the product ab is the same in $\mathbb{F}_1$ and in $\mathbb{F}_2$.

For example, $\mathbb{Q}$ is a subfield of both $\mathbb{R}$ and $\mathbb{C}$, and $\mathbb{R}$ is a subfield of $\mathbb{C}$. On the other hand, for distinct prime numbers p and q, $\mathbb{Z}_p$ is **not** a subfield of $\mathbb{Z}_q$ (even if $p < q$). Moreover, for a prime number p, $\mathbb{Z}_p$ is **not** a subfield of any one of $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$. It can be shown that any field is a subfield of some algebraically closed field, but the proof of this fact is beyond the scope of these lecture notes, and we omit it. However, let us just point out that the field $\mathbb{R}$ is a subfield of the algebraically closed field $\mathbb{C}$.[33]

For a field $\mathbb{F}$, we say that $n \times n$ matrices A and B with entries in $\mathbb{F}$ are *similar over* $\mathbb{F}$ if and only if there exists an invertible matrix $P \in \mathbb{F}^{n \times n}$ such that $B = P^{-1}AP$. The reader will have noticed that this is simply our usual definition of matrix similarity in $\mathbb{F}^{n \times n}$. However, if $\mathbb{F}$ is a subfield of some larger field $\widetilde{\mathbb{F}}$, then it makes sense to speak of A and B being (or not being) similar over $\mathbb{F}$, or of them being (or not being) similar over $\widetilde{\mathbb{F}}$. In fact, it can be shown that the two notions are equivalent. More precisely, it can be shown that if $\mathbb{F}$ is a subfield of $\widetilde{\mathbb{F}}$, then $n \times n$ matrices A and B, with entries in $\mathbb{F}$, are similar over $\mathbb{F}$ if and only if they are similar over $\widetilde{\mathbb{F}}$, that is, the following are equivalent:

- there exists an invertible matrix $P \in \mathbb{F}^{n \times n}$ such that $B = P^{-1}AP$;
- there exists an invertible matrix $P \in \widetilde{\mathbb{F}}^{n \times n}$ such that $B = P^{-1}AP$.

We will not prove this in full generality, since it would involve theory that is beyond the scope of these lecture notes. However, for the special case of $\mathbb{R}$ and $\mathbb{C}$, we can give a proof that that is both fully formal and fairly simple (see Theorem 8.6.7 below). Before turning to the special case of $\mathbb{R}$ and $\mathbb{C}$, let us explain what similarity over different fields had to do with the Jordan normal form. Suppose that we need to check if two $n \times n$ matrices, call them A and B, with entries in some field $\mathbb{F}$, are similar (over $\mathbb{F}$). We first extend $\mathbb{F}$ to an algebraically closed field $\widetilde{\mathbb{F}}$. Then the following are equivalent:

- A and B are similar over $\mathbb{F}$;
- A and B are similar over $\widetilde{\mathbb{F}}$;
- A and B have the same Jordan normal form in $\widetilde{\mathbb{F}}^{n \times n}$ (up to a reordering of the Jordan blocks).

(The equivalence of the second and third item above follows from Corollary 8.6.3.) So, if we can compute the Jordan normal forms of A and B in $\widetilde{\mathbb{F}}^{n \times n}$, then we can immediately determine if A and B are similar over $\mathbb{F}$. Of course, actually computing

[33] Of course, $\mathbb{Q}$ is also a subfield of the algebraically closed field $\mathbb{C}$.

the Jordan normal forms of A and B (in $\widetilde{\mathbb{F}}^{n\times n}$) may be very difficult or even impossible, essentially because we might not succeed in factoring the characteristic polynomials $p_A(\lambda)$ and $p_B(\lambda)$.[34]

We now turn to the special case of $\mathbb{R}$ and $\mathbb{C}$. As we pointed out above, we have all the tools that we need to prove that two $n \times n$ matrices with real entries are similar over $\mathbb{R}$ if and only if they are similar over $\mathbb{C}$.

Theorem 8.6.7. *Two $n \times n$ matrices with real entries are similar over $\mathbb{R}$ if and only if they are similar over $\mathbb{C}$.*

Proof. Fix $n \times n$ matrices A and B with real entries. We must show that the following are equivalent:

- A and B are similar over $\mathbb{R}$, that is, there exists an invertible matrix $P \in \mathbb{R}^{n\times n}$ such that $B = P^{-1}AP$;
- A and B are similar over $\mathbb{C}$, that is, there exists an invertible matrix $P \in \mathbb{C}^{n\times n}$ such that $B = P^{-1}AP$.

If A and B are similar over $\mathbb{R}$, then they are obviously similar over $\mathbb{C}$. For the converse, we assume that A and B are similar over $\mathbb{C}$, and we prove that they are similar over $\mathbb{R}$. Fix an invertible matrix $P \in \mathbb{C}^{n\times n}$ such that $B = P^{-1}AP$, so that $PB = AP$. Obviously, there exist $n \times n$ matrices R and Q with real entries such that $P = R + iQ$.[35] So, $(R + iQ)B = A(R + iQ)$, and consequently, $RB + i(QB) = AR + i(AQ)$. By separating the real and imaginary parts (and relying on the fact A, B, R, Q all have only real entries), we get that $RB = AR$ and $QB = AQ$.[36] Therefore,

$$(R + cQ)B \quad = \quad A(R + cQ)$$

for all $c \in \mathbb{C}$. It now suffices to show that there exists a **real** number c such that the matrix $R + cQ$ is invertible; we will then have that $B = (R + cQ)^{-1}A(R + cQ)$, so that A and B are similar over $\mathbb{R}$ (because all entries of $R + cQ$ are real), which is what we need.

Consider the polynomial

$$q(x) \quad := \quad \det(R + xQ).$$

[34] Since $\widetilde{\mathbb{F}}$ is algebraically closed, the characteristic polynomials $p_A(\lambda)$ and $p_B(\lambda)$ can be factored into linear terms with coefficients in $\widetilde{\mathbb{F}}$. However, as we have pointed out a number of times already, this is only an existence statement: we have no general recipe for factoring.

[35] For example, if $P = \begin{bmatrix} 2+i & -3i \\ 1-2i & -7 \end{bmatrix}$, then we have that $P = R + iQ$ for $R := \begin{bmatrix} 2 & 0 \\ 1 & -7 \end{bmatrix}$ and $Q := \begin{bmatrix} 1 & -3 \\ -2 & 0 \end{bmatrix}$.

[36] If R is invertible, then we have that $B = R^{-1}AR$, and so A and B are similar over $\mathbb{R}$, and we are done. We are similarly done if Q is invertible. Unfortunately, it is possible that neither R nor Q is invertible, which is why we are not done yet.

Since R and Q are $n \times n$ matrices with real entries, we see that $q(x)$ is a polynomial with real coefficients and of degree at most n. Now, since P is invertible, the Invertible Matrix Theorem (see subsection 8.2.6) guarantees that $\det(P) \neq 0$. So, $q(i) = \det(R + iQ) = \det(P) \neq 0$. We have now shown that $q(x)$ is a non-zero polynomial with real coefficients and of degree at most n. Therefore, $q(x)$ has at most n complex roots, and in particular, it has at most n real roots. Thus, there exists a real number c such that $q(c) \neq 0$.[37] But then $\det(R+cQ) = q(c) \neq 0$, and so the Invertible Matrix Theorem (see subsection 8.2.6) guarantees $R + cQ$ is invertible. This completes the argument. □

Remark: In view of Theorem 8.6.7, we can rely on the Jordan normal form to check whether two $n \times n$ matrices with real entries are similar over $\mathbb{R}$, even though the field $\mathbb{R}$ is **not** algebraically closed. Indeed, for $n \times n$ matrices A and B with real entries, Corollary 8.6.3 and Theorem 8.6.7 together guarantee that the following are equivalent:

- A and B are similar over $\mathbb{R}$;
- A and B are similar over $\mathbb{C}$;
- A and B have the same Jordan normal form in $\mathbb{C}^{n\times n}$ (up to a reordering of the Jordan blocks).

If we manage to factor the characteristic polynomials of A and B, then we can compute the Jordan normal forms of A and B (seen as matrices in $\mathbb{C}^{n\times n}$) using Theorem 8.6.6. Importantly, the Jordan normal form of a square matrix with real entries may have entries that are not real.

8.6.2 Computing the Jordan normal form of a square matrix

In this subsection, we show how Theorem 8.6.6 can be used to compute the Jordan normal form of a matrix A in $\mathbb{C}^{n\times n}$. We give two fully worked out examples (Examples 8.6.8 and 8.6.9 below). In order to get a full picture of how Theorem 8.6.6 is used in the general case, we need our matrix A to be large enough. So, in our first example, we find the Jordan normal form of a 10×10 matrix, and in our second example, we do this for a 13×13 matrix. We note that the matrix powers and ranks in these examples are all computed with the help of a calculator (this type of computation would take a very long time if we were to do everything by hand).

[37] We can even choose c to be one of $0, 1, \ldots, n$. Indeed, since $q(x)$ has at most n complex roots, we know that at least one of $0, 1, \ldots, n$ is **not** a root of $q(x)$. We can choose c to be this non-root.

Example 8.6.8. *Consider the following matrix in* $\mathbb{C}^{10\times 10}$*:*

$$A := \begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -3 & 1 & 5 & 2 & -2 & -4 & -7 & 4 & -1 & 3 \\ 0 & 1 & 3 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -2 & -1 & 3 & 4 & -1 & -2 & -3 & 2 & -1 & 2 \\ -1 & 0 & 2 & 1 & 2 & -2 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 2 & -1 & 0 & 0 & 1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 7 & -2 & 1 & -2 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 3 & -1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & 0 \end{bmatrix}.$$

Using Theorem 8.6.6, compute the Jordan normal form of A.

Solution. First of all, we compute the characteristic polynomial of A, and we factor it into linear terms:

$$p_A(\lambda) = \det\big(\lambda I_{10} - A\big) = (\lambda-3)^8(\lambda-2)^2.$$

We see that the eigenvalues of A are $\lambda_1 = 3$ (with algebraic multiplicity 8) and $\lambda_2 = 2$ (with algebraic multiplicity 2). So, all of our Jordan blocks will be of the form $J_t(3)$ and $J_t(2)$ for various positive integers t. We now deal with the two eigenvalues separately, as follows.

We first deal with the eigenvalue $\lambda_1 = 3$. We compute the matrices $(A-\lambda_1 I_{10})^r$ for $r = 0, 1, 2, 3, \ldots$ along with their ranks. We keep computing until we get the same rank twice in a row. We obtain:

- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^0\Big) = 10$;[38]
- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^1\Big) = 7$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^2\Big) = 4$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^3\Big) = 2$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^4\Big) = 2$.

We have now obtained the same rank twice in a row, and so we can stop. We compute:

- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^0\Big) - \operatorname{rank}\Big((A-\lambda_1 I_{10})^1\Big) = 3$;

[38] By definition, we have that $(A-\lambda_1 I_{10})^0 = I_{10}$, and obviously, $\operatorname{rank}(I_{10}) = 10$.

- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^1\Big) - \operatorname{rank}\Big((A-\lambda_1 I_{10})^2\Big) = 3$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^2\Big) - \operatorname{rank}\Big((A-\lambda_1 I_{10})^3\Big) = 2$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{10})^3\Big) - \operatorname{rank}\Big((A-\lambda_1 I_{10})^4\Big) = 0$.

By Theorem 8.6.6, the Jordan normal form of A will contain:

- three Jordan blocks $J_t(\lambda_1) = J_t(3)$ with $t \geq 1$;
- three Jordan blocks $J_t(\lambda_1) = J_t(3)$ with $t \geq 2$;
- two Jordan blocks $J_t(\lambda_1) = J_t(3)$ with $t \geq 3$;
- zero Jordan blocks $J_t(\lambda_1) = J_t(3)$ with $t \geq 4$.

Keeping in mind that for any positive integer r, the number of Jordan blocks $J_r(\lambda_1) = J_r(3)$ in the Jordan normal form of A is equal to

$$\left(\begin{array}{l}\text{number of Jordan blocks}\\ J_t(\lambda_1)\text{ satisfying } t \geq r\end{array}\right) - \left(\begin{array}{l}\text{number of Jordan blocks}\\ J_t(\lambda_1)\text{ satisfying } t \geq r+1\end{array}\right),$$

we conclude that the Jordan normal form of A will contain exactly two Jordan blocks $J_3(\lambda_1) = J_3(3)$,[39] and exactly one Jordan block $J_2(\lambda_1) = J_2(3)$. The Jordan normal form of A contains no other Jordan blocks of the form $J_t(\lambda_1) = J_t(3)$.

It remains to deal with the eigenvalue $\lambda_2 = 2$. We compute the matrices $(A-\lambda_2 I_{10})^r$ for $r = 0, 1, 2, 3, \ldots$ along with their ranks. We keep computing until we get the same rank twice in a row. We obtain:

- $\operatorname{rank}\Big((A-\lambda_2 I_{10})^0\Big) = 10$;
- $\operatorname{rank}\Big((A-\lambda_2 I_{10})^1\Big) = 9$;
- $\operatorname{rank}\Big((A-\lambda_2 I_{10})^2\Big) = 8$;
- $\operatorname{rank}\Big((A-\lambda_2 I_{10})^3\Big) = 8$.

We have now obtained the same rank twice in a row, and so we can stop. We compute:

- $\operatorname{rank}\Big((A-\lambda_2 I_{10})^0\Big) - \operatorname{rank}\Big((A-\lambda_2 I_{10})^1\Big) = 1$;

[39] Indeed, it contains two Jordan blocks $J_t(3)$ with $t \geq 3$, but zero Jordan blocks $J_t(3)$ with $t \geq 4$. So, the number of Jordan blocks $J_t(3)$ with $t = 3$ is $2 - 0 = 2$.

- $\operatorname{rank}\Big((A-\lambda_2 I_{10})^1\Big) - \operatorname{rank}\Big((A-\lambda_2 I_{10})^2\Big) = 1;$
- $\operatorname{rank}\Big((A-\lambda_2 I_{10})^2\Big) - \operatorname{rank}\Big((A-\lambda_2 I_{10})^3\Big) = 0.$

By Theorem 8.6.6, the Jordan normal form of A will contain:

- one Jordan block $J_t(\lambda_2) = J_t(2)$ with $t \geq 1$;
- one Jordan block $J_t(\lambda_2) = J_t(2)$ with $t \geq 2$;
- zero Jordan blocks $J_t(\lambda_2) = J_t(2)$ with $t \geq 3$.

Consequently, the Jordan normal form of A will contain exactly one Jordan block $J_2(\lambda_2) = J_2(2)$, and it will contain no other Jordan blocks of the form $J_t(\lambda_2) = J_t(2)$.

Putting everything together, we get that the Jordan normal form of A is the following (color coded for ease of reading):

$$\begin{aligned} J &:= J_3(\lambda_1) \oplus J_3(\lambda_1) \oplus J_2(\lambda_1) \oplus J_2(\lambda_2) \\ &= J_3(3) \oplus J_3(3) \oplus J_2(3) \oplus J_2(2) \\ &= \begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}. \end{aligned}$$

We remark that we could have written our Jordan blocks in a different order, but in any case, the Jordan blocks would have to be the same as above (counting repetitions). For instance, $J_2(3) \oplus J_3(3) \oplus J_2(2) \oplus J_3(3)$ is also a Jordan normal form of A.

Remark: It is acceptable to leave $J_3(3) \oplus J_3(3) \oplus J_2(3) \oplus J_2(2)$ (color coded or not) as a final answer, without exhibiting the actual 10×10 matrix with its 100 entries. It is **not** acceptable to leave $J_3(\lambda_1) \oplus J_3(\lambda_1) \oplus J_2(\lambda_1) \oplus J_2(\lambda_2)$ as a final answer. □

Example 8.6.9. *Consider the following matrix in* $\mathbb{C}^{13\times 13}$*:*

$$A := \begin{bmatrix} 4 & 0 & 0 & 3 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 5 & 1 & -3 & 0 & 1 & -2 & 1 & 1 & 0 & 1 & -2 & 0 \\ -1 & -2 & 3 & 4 & 0 & -6 & 3 & 1 & -2 & 2 & -3 & 4 & -2 \\ 0 & -3 & 0 & 10 & -1 & -2 & 4 & 0 & -2 & -1 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 & 4 & 3 & -1 & 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & -12 & 1 & 2 & -4 & 0 & 2 & 1 & -2 & -6 & 0 \\ 1 & 1 & 1 & -1 & 0 & 6 & -1 & 3 & 1 & -2 & 3 & -2 & 2 \\ 0 & -1 & 0 & 3 & 0 & -1 & 2 & 0 & 3 & 0 & 0 & 2 & 0 \\ -1 & -1 & -1 & 4 & 0 & -4 & 3 & -1 & -1 & 4 & -2 & 3 & -2 \\ 0 & 1 & 0 & -3 & 0 & 1 & -2 & 0 & 1 & 0 & 4 & -2 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Using Theorem 8.6.6, compute the Jordan normal form of A.

Solution. We proceed similarly as in Example 8.6.8. First, we compute the characteristic polynomial of A, and we factor it into linear terms:

$$p_A(\lambda) \;=\; \det\big(\lambda I_{13} - A\big) \;=\; (\lambda-4)^{10}(\lambda-2)^3.$$

We see that the eigenvalues of A are $\lambda_1 = 4$ (with algebraic multiplicity 10) and $\lambda_2 = 2$ (with algebraic multiplicity 3). So, all of our Jordan blocks will be of the form $J_t(4)$ and $J_t(2)$ for various positive integers t. We now deal with the two eigenvalues separately, as follows.

We first deal with the eigenvalue $\lambda_1 = 4$. We compute the matrices $(A - \lambda_1 I_{13})^r$ for $r = 0, 1, 2, 3, \ldots$ along with their ranks. We keep computing until we get the same rank twice in a row. We obtain:

- $\operatorname{rank}\Big((A-\lambda_1 I_{13})^0\Big) = 13$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{13})^1\Big) = 9$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{13})^2\Big) = 6$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{13})^3\Big) = 5$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{13})^4\Big) = 4$;
- $\operatorname{rank}\Big((A-\lambda_1 I_{13})^5\Big) = 3$;

- $\text{rank}\Big((A-\lambda_1 I_{13})^6\Big) = 3.$

We have now obtained the same rank twice in a row, and so we can stop. We compute:

- $\text{rank}\Big((A-\lambda_1 I_{13})^0\Big) - \text{rank}\Big((A-\lambda_1 I_{13})^1\Big) = 4;$
- $\text{rank}\Big((A-\lambda_1 I_{13})^1\Big) - \text{rank}\Big((A-\lambda_1 I_{13})^2\Big) = 3;$
- $\text{rank}\Big((A-\lambda_1 I_{13})^2\Big) - \text{rank}\Big((A-\lambda_1 I_{13})^3\Big) = 1;$
- $\text{rank}\Big((A-\lambda_1 I_{13})^3\Big) - \text{rank}\Big((A-\lambda_1 I_{13})^4\Big) = 1;$
- $\text{rank}\Big((A-\lambda_1 I_{13})^4\Big) - \text{rank}\Big((A-\lambda_1 I_{13})^5\Big) = 1;$
- $\text{rank}\Big((A-\lambda_1 I_{13})^5\Big) - \text{rank}\Big((A-\lambda_1 I_{13})^6\Big) = 0.$

By Theorem 8.6.6, the Jordan normal form of A will contain:

- four Jordan blocks $J_t(\lambda_1) = J_t(4)$ with $t \geq 1$;
- three Jordan blocks $J_t(\lambda_1) = J_t(4)$ with $t \geq 2$;
- one Jordan block $J_t(\lambda_1) = J_t(4)$ with $t \geq 3$;
- one Jordan block $J_t(\lambda_1) = J_t(4)$ with $t \geq 4$;
- one Jordan block $J_t(\lambda_1) = J_t(4)$ with $t \geq 5$;
- zero Jordan blocks $J_t(\lambda_1) = J_t(4)$ with $t \geq 6$.

Keeping in mind that for any positive integer r, the number of Jordan blocks $J_r(\lambda_1) = J_r(4)$ in the Jordan normal form of A is equal to

$$\left(\begin{array}{l}\text{number of Jordan blocks}\\ J_t(\lambda_1)\text{ satisfying } t \geq r\end{array}\right) - \left(\begin{array}{l}\text{number of Jordan blocks}\\ J_t(\lambda_1)\text{ satisfying } t \geq r+1\end{array}\right),$$

we conclude that the Jordan normal form of A will contain exactly one Jordan block $J_5(\lambda_1) = J_5(4)$, two Jordan blocks $J_2(\lambda_1) = J_2(4)$, and one Jordan block $J_1(\lambda_1) = J_1(4)$. The Jordan normal form of A contains no other Jordan blocks of the form $J_t(\lambda_1) = J_t(4)$.

It remains to deal with the eigenvalue $\lambda_2 = 2$. We compute the matrices $(A-\lambda_2 I_{13})^r$ for $r = 0, 1, 2, 3, \ldots$ along with their ranks. We keep computing until we get the same rank twice in a row. We obtain:

- $\text{rank}\Big((A-\lambda_2 I_{13})^0\Big) = 13$;
- $\text{rank}\Big((A-\lambda_2 I_{13})^1\Big) = 11$;
- $\text{rank}\Big((A-\lambda_2 I_{13})^2\Big) = 10$;
- $\text{rank}\Big((A-\lambda_2 I_{13})^3\Big) = 10$.

We have now obtained the same rank twice in a row, and so we can stop. We compute:

- $\text{rank}\Big((A-\lambda_2 I_{13})^0\Big) - \text{rank}\Big((A-\lambda_2 I_{13})^1\Big) = 2$;
- $\text{rank}\Big((A-\lambda_2 I_{13})^1\Big) - \text{rank}\Big((A-\lambda_2 I_{13})^2\Big) = 1$;
- $\text{rank}\Big((A-\lambda_2 I_{13})^2\Big) - \text{rank}\Big((A-\lambda_2 I_{13})^3\Big) = 0$.

By Theorem 8.6.6, the Jordan normal form of A will contain:

- two Jordan blocks $J_t(\lambda_2) = J_t(2)$ with $t \geq 1$;
- one Jordan block $J_t(\lambda_2) = J_t(2)$ with $t \geq 2$;
- zero Jordan block $J_t(\lambda_2) = J_t(2)$ with $t \geq 3$.

Consequently, the Jordan normal form of A will contain exactly one Jordan block $J_2(\lambda_2) = J_2(2)$, one Jordan block $J_1(\lambda_2) = J_1(2)$, and no other Jordan blocks of the form $J_t(\lambda_2) = J_t(2)$.

Putting everything together, we get that the Jordan normal form of A is the following (color coded for ease of reading):

$$\begin{aligned} J &:= J_5(\lambda_1) \oplus J_2(\lambda_1) \oplus J_2(\lambda_1) \oplus J_1(\lambda_1) \oplus J_2(\lambda_2) \oplus J_1(\lambda_2) \\ &= J_5(4) \oplus J_2(4) \oplus J_2(4) \oplus J_1(4) \oplus J_2(2) \oplus J_1(2) \end{aligned}$$

$$= \begin{bmatrix} 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

As in Example 8.6.8, we could have written our Jordan blocks in a different order, as long as we preserved any repetitions. For example, $J_1(2) \oplus J_2(4) \oplus J_1(4) \oplus J_2(2) \oplus J_5(4) \oplus J_2(4)$ is also a Jordan normal form of A. □

Remark: Suppose we are given a matrix $A \in \mathbb{F}^{n\times n}$, where $\mathbb{F}$ is an algebraically closed field. We saw above how we can compute the Jordan normal form of A, that is, how we can find a Jordan matrix $J \in \mathbb{F}^{n\times n}$ that is similar to A. Could we also compute an invertible matrix $P \in \mathbb{F}^{n\times n}$ for which $J = P^{-1}AP$? This is indeed possible, but it is significantly more complicated than just computing a suitable Jordan matrix J. Unfortunately, any example that illustrates the procedure in full generality (more or less) requires a great deal of long and laborious computation.[40] For sufficiently brave readers, a recipe and a couple of examples are given in subsection 8.6.7.

8.6.3 A brief outline of the proof of Theorems 8.6.2 and 8.6.4

As we pointed out in subsection 8.6.1, Theorems 8.6.2 and 8.6.4 are equivalent, i.e. in view of Theorem 4.5.16, either one of these two theorems easily implies the other. More precisely, the existence parts of Theorems 8.6.2 and 8.6.4 are equivalent, as are the uniqueness parts of the two theorems. In what follows, we will outline the proof of the existence part of Theorem 8.6.4, and the proof of the uniqueness part of Theorem 8.6.2. (In passing, we will also say a few words about the proof of Theorem 8.6.6.) The uniqueness part is easier, and so we outline that first.

Uniqueness. Let us outline the proof of the uniqueness part of Theorem 8.6.2, restated below for the reader's convenience.

[40] It really is quite long and quite laborious, even if we use a calculator throughout.

Theorem 8.6.2. *Assume that $\mathbb{F}$ is an **algebraically closed field**, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then A is similar to a matrix J in Jordan normal form. Moreover, this matrix J is unique up to a reordering of the Jordan blocks.*

So, let us suppose that $\mathbb{F}$ is a field,[41] and assume that a matrix $A \in \mathbb{F}^{n\times n}$ is similar to a Jordan matrix $J \in \mathbb{F}^{n\times n}$. Clearly, it suffices to show that the types of Jordan blocks that appear in J are fully determined by A, as is the number of Jordan blocks of each type. (For a formal statement of what we are trying to prove, see the somewhat lengthy statement of Proposition 8.6.17.)

First of all, since A and J are similar, Theorem 8.2.9 guarantees that these two matrices have the same characteristic polynomial and the same spectrum. Since the Jordan matrix J is upper triangular, it has precisely its spectrum (equivalently: the spectrum of A) on the main diagonal. In particular, all Jordan blocks of J are of the form $J_t(\lambda)$, where λ is an eigenvalue of A, and t is a positive integer no greater than the algebraic multiplicity of λ as an eigenvalue of A. For each eigenvalue λ and positive integer r, we would like to compute the number of Jordan blocks $J_t(\lambda)$ that appear in J and satisfy $t \geq r$. If we can show that this number depends only on A (and not on the particular choice of J), then we are done, since for fixed λ and r, the number of Jordan blocks $J_r(\lambda)$ in J is equal to

$$\left(\begin{array}{l}\text{number of Jordan blocks}\\ J_t(\lambda) \text{ satisfying } t \geq r\end{array}\right) - \left(\begin{array}{l}\text{number of Jordan blocks}\\ J_t(\lambda) \text{ satisfying } t \geq r+1\end{array}\right).$$

We now fix an eigenvalue λ (say, of algebraic multiplicity m) of A, and we consider the matrix $J - \lambda I_n$; this matrix is a Jordan matrix (because J is), and moreover, each Jordan block $J_t(\lambda')$ of J corresponds to a Jordan block $J_t(\lambda' - \lambda)$ of $J - \lambda I_n$ in the obvious way. In particular, blocks $J_t(\lambda)$ of J correspond to the blocks $J_t(0)$ of $J - \lambda I_n$ in the natural way. So, we just need to count the number of blocks $J_t(0)$ in $J - \lambda I_n$ satisfying $t \geq r$, for all possible values of r. Here, it will be important to note that the sizes of the Jordan blocks $J_t(0)$ in $J - \lambda I_n$ sum up to m, whereas the sizes of the remaining Jordan blocks of $J - \lambda I_n$ sum up to $n - m$. Now, the idea is to compute $\operatorname{rank}\big((J - \lambda I_n)^r\big)$ for $r = 0, 1, 2, 3, \ldots$.

We first make a couple of key observations. The first observation is that if $A_1, \ldots, A_k$ are square matrices, then

- $\operatorname{rank}(A_1 \oplus \cdots \oplus A_k) = \operatorname{rank}(A_1) + \cdots + \operatorname{rank}(A_k)$, and
- $(A_1 \oplus \cdots \oplus A_k)^r = A_1^r \oplus \cdots \oplus A_k^r$ for all non-negative integers r.

So, the r-th power of a Jordan matrix is equal to the direct sum of the r-th powers of its Jordan blocks, and the rank of the r-th power of a Jordan matrix is equal to the sum of ranks of the r-th powers of its Jordan blocks.

[41] For the uniqueness part, we do not need $\mathbb{F}$ to be algebraically closed. Algebraic closure matters only for the existence part.

The second observation is that for any matrix A with t columns and entries in $\mathbb{F}$, the matrix $AJ_t(0)$ is obtained from A by first adding a zero column to the left, and then deleting the rightmost column of the resulting matrix; it then follows by an easy induction on r that

$$\big(J_t(0)\big)^r = \begin{cases} I_t & \text{if } r = 0 \\ \left[\begin{array}{c|c} O_{(t-r)\times r} & I_{t-r} \\ \hline O_{r\times r} & O_{r\times(t-r)} \end{array}\right] & \text{if } 1 \leq r \leq t-1 \\ O_{t\times t} & \text{if } r \geq t \end{cases}$$

for all positive integers t and r (see Proposition 8.6.13). From here, we can easily read off $\operatorname{rank}\Big(\big(J_t(0)\big)^r\Big)$.[42] Meanwhile, the remaining Jordan blocks of $J - \lambda I_n$ (i.e. those that have a number other than 0 on the main diagonal) are invertible matrices (because their determinant is non-zero);[43] therefore, all powers of these matrices are invertible and have full rank.[44] So, for any positive integer r, the r-th powers of the Jordan blocks of $J - \lambda I_n$ that have a non-zero on the main diagonal end up contributing $n - m$ to the rank of $(J - \lambda I_n)^r$. Meanwhile, the contribution that the r-th power of a Jordan block $J_t(0)$ of $J - \lambda I_n$ makes to the rank of $(J - \lambda I_n)^r$ can easily be read off from our formula for $\big(J_t(0)\big)^r$ above.

Thus, we can in fact obtain a formula for $\operatorname{rank}\big((J - \lambda I_n)^r\big)$ that depends on the number of Jordan blocks $J_t(0)$ for various values of t. With a little bit of computation, it can be shown that, for each positive integer r, the number of Jordan blocks $J_t(0)$ satisfying $t \geq r$ in the Jordan matrix $J - \lambda I_n$ is precisely

$$\operatorname{rank}\big((J - \lambda I_n)^{r-1}\big) - \operatorname{rank}\big((J - \lambda I_n)^r\big).$$

As discussed above, this is precisely the number of Jordan blocks $J_t(\lambda)$ satisfying $t \geq r$ in J. But since A and J are similar, so are $A - \lambda I_n$ and $A - \lambda I_n$, and by Proposition 4.5.15, so are all of their corresponding powers. By Corollary 4.5.17, similar matrices have the same rank. So, the number of Jordan blocks $J_t(\lambda)$ satisfying $t \geq r$ in J is in fact

$$\operatorname{rank}\big((A - \lambda I_n)^{r-1}\big) - \operatorname{rank}\big((A - \lambda I_n)^r\big).$$

This essentially completes the proof of the uniqueness part of Theorems 8.6.2. For the full details, see subsection 8.6.4.

[42] Indeed, we get that $\operatorname{rank}\Big(\big(J_t(0)\big)^r\Big) = t - r$ if $r \leq t - 1$, and that $\operatorname{rank}\Big(\big(J_t(0)\big)^r\Big) = 0$ if $r \geq t$.

[43] Any Jordan block is an upper triangular matrix, and so (by Proposition 7.3.1) its determinant can be computed by multiplying the entries on the main diagonal. If these entries are non-zero, then the determinant is non-zero. By the Invertible Matrix Theorem (see subsection 8.2.6), any square matrix whose determinant is non-zero is invertible.

[44] By Proposition 1.11.8(f), all powers of an invertible matrix are invertible, and by the Invertible Matrix Theorem (see subsection 8.2.6), invertible matrices have full rank.

Remark: The formula that we obtained above is precisely the same as the one from Theorem 8.6.6. However, the argument above (fortified with all the technical details) does not actually prove Theorem 8.6.6. It only proves that **if** a square matrix A is similar to a Jordan matrix J, **then** the Jordan blocks of J are as specified in Theorem 8.6.6. To prove that a matrix $A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is an algebraically closed field) is indeed similar to at least one Jordan matrix,[45] we need the existence part of Theorem 8.6.2, to which we now turn.

Existence. We now give a brief outline of the proof of the existence part of Theorem 8.6.4, restated below for the reader's convenience.

Theorem 8.6.4. *Let V be a non-trivial, finite-dimensional vector space over an* ***algebraically closed field*** *$\mathbb{F}$, and let $f : V \to V$ be a linear function. Then there exists a basis $\mathcal{B}$ such that the matrix ${}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ is in Jordan normal form. Moreover, this matrix is unique in the following sense: if $\mathcal{B}_1$ and $\mathcal{B}_2$ are bases of V such that both ${}_{\mathcal{B}_1}[\ f\ ]_{\mathcal{B}_1}$ and ${}_{\mathcal{B}_2}[\ f\ ]_{\mathcal{B}_2}$ are in Jordan normal form, then these two matrices are the same up to a reordering of the Jordan blocks.*

First, we need a couple of definitions. For a function $f : A \to A$ (where A is some set), we set $f^0 := \mathrm{Id}_A$ (the identity function on A), and for all non-negative integers k, we set $f^{k+1} := f^k \circ f$. So, for all positive integers k, we have that

$$f^k := \underbrace{f \circ \cdots \circ f}_{k}.$$

The function f^k is called the *k-th iterate* of f.

Next, for a linear function $f : V \to V$ (where V is a vector space over a field $\mathbb{F}$), a subspace U of V is said to be *f-invariant* if $f[U] \subseteq U$, i.e. if for all $\mathbf{u} \in U$, we have that $f(\mathbf{u}) \in U$. Under these circumstances, it is possible to restrict both the domain and the codomain of f to U. More precisely, we can define $f|_U : U \to U$ by setting $f|_U(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in U$; obviously, $f|_U$ is linear (since f is).

Here is an outline of the proof of the existence part of Theorem 8.6.4. Let $f : V \to V$ be a linear function, where V is a non-trivial, finite-dimensional vector space over an algebraically closed field $\mathbb{F}$, as in the statement of Theorem 8.6.4. Let

$$(\underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{m_k})$$

be the spectrum of f, where $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are pairwise distinct, and where $m_1, \ldots, m_k$ are positive integers. Since $\mathbb{F}$ is algebraically closed, we know that $m_1 + \cdots + m_k = \dim(V) =: n$.

[45] Here, it is important that $\mathbb{F}$ is algebraically closed: the statement becomes false otherwise.

By Theorem 8.2.13, for any eigenvalue λ of f, the eigenspace of f associated with λ is $E_\lambda(f) = \mathrm{Ker}\big(f - \lambda \mathrm{Id}_V\big)$. We can generalize this as follows. For each eigenvalue λ of f, we define the *generalized eigenspace* of f associated with λ to be the set

$$\begin{aligned} G_\lambda(f) &:= \big\{\mathbf{v} \in V \mid \exists r \in \mathbb{N}_0 \text{ s.t. } (f - \lambda \mathrm{Id}_V)^r(\mathbf{v}) = \mathbf{0}\big\} \\ &= \bigcup_{r=0}^{\infty} \mathrm{Ker}\Big(\big(f - \lambda \mathrm{Id}_V\big)^r\Big). \end{aligned}$$

It is easy to show that $G_\lambda(f)$ is a subspace of V (we simply check that it satisfies the three conditions from Theorem 3.1.7), and obviously, eigenspace $E_\lambda(f) = \mathrm{Ker}(f - \lambda I_n)$ is a subspace of the generalized eigenspace of $G_\lambda(f)$. Now, in the first part of our proof, we show that all generalized eigenspaces of f are f-invariant subspaces of V, and moreover, that V is the direct sum of the generalized eigenspaces of f, that is,

$$V = G_{\lambda_1}(f) \oplus \cdots \oplus G_{\lambda_k}(f).$$

We now consider the restrictions $f_1, \ldots, f_k$ of our linear function f to the generalized eigenspaces $G_{\lambda_1}(f), \ldots, G_{\lambda_k}(f)$, respectively.[46] It can be shown that for each index $i \in \{1, \ldots, k\}$, f_i has exactly one eigenvalue, namely λ_i, and the algebraic multiplicity of this eigenvalue is m_i. Further, it is easy to show that for any bases $\mathcal{B}_1, \ldots, \mathcal{B}_k$ of $G_{\lambda_1}(f), \ldots, G_{\lambda_k}(f)$, respectively, we have that $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ is a basis of V, and moreover, that

$${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} = {}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1} \oplus \cdots \oplus {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}.$$

If we can choose our bases $\mathcal{B}_1, \ldots, \mathcal{B}_k$ so that the matrices ${}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1}, \ldots, {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}$ are Jordan matrices, then the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ will also be a Jordan matrix. So, we may handle the restrictions $f_1, \ldots, f_k$ of f separately. Since each of these restrictions has only one eigenvalue, this effectively reduces the problem to the case when the linear function $f : V \to V$ has only one eigenvalue.

Now, note that if λ is the only eigenvalue of $f : V \to V$, then 0 is the only eigenvalue of $f - \lambda \mathrm{Id}_V$, and moreover, for any basis $\mathcal{B}$ of V, we have that

$${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} = {}_{\mathcal{B}}\big[\ f - \lambda \mathrm{Id}_V\ \big]_{\mathcal{B}} + \lambda I_n.$$

Thus, ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is a Jordan matrix if and only if ${}_{\mathcal{B}}\big[\ f - \lambda \mathrm{Id}_V\ \big]_{\mathcal{B}}$ is a Jordan matrix. So, we may consider $f - \lambda \mathrm{Id}_V$ instead of f. In other words, we have reduced our problem to the case when the only eigenvalue of f is 0.

From now on, we deal only with the case when f has exactly one eigenvalue, namely 0. Because $\mathbb{F}$ is algebraically closed, this eigenvalue has algebraic multiplicity $n = \dim(V)$, i.e. $p_f(\lambda) = \lambda^n$. It can then be shown that f is "nilpotent," that is,

[46] In other words, for each $i \in \{1, \ldots, k\}$, we set $f_i := f|_{G_{\lambda_i}(f)}$, which is well defined because $G_{\lambda_i}(f)$ is f-invariant.

that some iterate f^p of f is the zero function,[47] or equivalently, that $\operatorname{Ker}(f^p) = V$. Thus, we have reduced the problem of proving the existence part of Theorem 8.6.4 to the case when f is nilpotent.

The goal is now to construct a basis $\mathcal{B}$ of V such that the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is in Jordan normal form. Our basis will be the union of suitably chosen "Jordan chains" of the form $\{f^{r-1}(\mathbf{u}), f^{r-2}(\mathbf{u}), \ldots, f(\mathbf{u}), \mathbf{u}\}$, where $\mathbf{u} \in \operatorname{Ker}(f^r) \setminus \operatorname{Ker}(f^{r-1})$, so that $f^r(\mathbf{u}) = \mathbf{0}$, but $f^{r-1}(\mathbf{u}) \neq \mathbf{0}$. We will say that such a Jordan chain is *started by* $\mathbf{u}$.[48] So, our basis $\mathcal{B}$ will be of the form

$$\mathcal{B} \;=\; \Big\{ f^{a_1-1}(\mathbf{u}_1), \ldots, f(\mathbf{u}_1), \mathbf{u}_1, \ldots, f^{a_\ell-1}(\mathbf{u}_\ell), \ldots, f(\mathbf{u}_\ell), \mathbf{u}_\ell \Big\}$$

for some vectors $\mathbf{u}_1, \ldots, \mathbf{u}_\ell \in V$ and positive integers $a_1, \ldots, a_\ell$ such that for all indices $i \in \{1, \ldots, \ell\}$, we have that $f^{a_i}(\mathbf{u}_i) = \mathbf{0}$ and $f^{a_i-1}(\mathbf{u}_i) \neq \mathbf{0}$. These Jordan chains need to be chosen with care, so that they together produce a basis. We will say a few words about this below, but let us first explain why a basis of this type is actually useful. The point is that each Jordan chain $\{f^{a_i-1}(\mathbf{u}_i), f^{a_i-2}(\mathbf{u}), \ldots, f(\mathbf{u}_i), \mathbf{u}_i\}$ in our basis $\mathcal{B}$ will correspond to a Jordan block $J_{a_i}(0)$ in the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. To gain some intuition for how the Jordan chains behave, let us suppose that $\mathcal{B} = \{f^{r-1}(\mathbf{u}), f^{r-2}(\mathbf{u}), \ldots, f(\mathbf{u}), \mathbf{u}\}$ is our whole basis, i.e. our basis $\mathcal{B}$ consists of only one Jordan chain; in this particular case, we would in fact have that $r = \dim(V) = n$, and so $\mathcal{B} = \{f^{n-1}(\mathbf{u}), f^{n-2}(\mathbf{u}), \ldots, f(\mathbf{u}), \mathbf{u}\}$. Then, using the formula from Theorem 4.5.1, we would obtain

$$\begin{aligned}
{}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} &= \Big[\ \big[\ f\big(f^{n-1}(\mathbf{u})\big)\ \big]_{\mathcal{B}}\ \ \big[\ f\big(f^{n-2}(\mathbf{u})\big)\ \big]_{\mathcal{B}}\ \ \cdots\ \ \big[\ f\big(f(\mathbf{u})\big)\ \big]_{\mathcal{B}}\ \ \big[\ f(\mathbf{u})\ \big]_{\mathcal{B}}\ \Big] \\
&= \Big[\ \big[\ f^n(\mathbf{u})\ \big]_{\mathcal{B}}\ \ \big[\ f^{n-1}(\mathbf{u})\ \big]_{\mathcal{B}}\ \ \cdots\ \ \big[\ f^2(\mathbf{u})\ \big]_{\mathcal{B}}\ \ \big[\ f(\mathbf{u})\ \big]_{\mathcal{B}}\ \Big] \\
&= \Big[\ \big[\ \mathbf{0}\ \big]_{\mathcal{B}}\ \ \big[\ f^{n-1}(\mathbf{u})\ \big]_{\mathcal{B}}\ \ \cdots\ \ \big[\ f^2(\mathbf{u})\ \big]_{\mathcal{B}}\ \ \big[\ f(\mathbf{u})\ \big]_{\mathcal{B}}\ \Big] \\
&= \Big[\ \mathbf{0}\ \ \mathbf{e}_1^n\ \ \cdots\ \ \mathbf{e}_{n-2}^n\ \ \mathbf{e}_{n-1}^n\ \Big] \\
&= J_n(0).
\end{aligned}$$

The computation above works when $\mathcal{B}$ consists of only one Jordan chain. If $\mathcal{B}$ consists of multiple Jordan chains, then we simply get multiple Jordan blocks.

We now briefly outline the construction of the basis $\mathcal{B}$ discussed above. The argument proceeds by (strong) induction on $\dim(V)$. Let p be an integer such that f^p is a zero function, i.e. $\operatorname{Ker}(f^p) = V$. The goal is to show that for all vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$ that are "linearly independent over $\operatorname{Ker}(f^{p-1})$" (which means that they

[47] This essentially follows from the Cayley-Hamilton theorem (stated and proven in section 8.3). For the details, see Proposition 8.6.26.

[48] Granted, this terminology is slightly unfortunate, since the Jordan chain rather "ends" with $\mathbf{u}$. The point, however, is that once we have chosen $\mathbf{u}$, the Jordan chain is uniquely determined.

are linearly independent in the ordinary sense, and in addition, $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap \mathrm{Ker}(f^{p-1}) = \{\mathbf{0}\}$), there exists a basis $\mathcal{B}$ of V that is the union of some number of pairwise disjoint Jordan chains, and t of those chains are started by our preselected vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$. With a little bit of preprocessing, we may assume that f^{p-1} is not a zero function, i.e. $\mathrm{Ker}(f^{p-1}) \subsetneq \mathrm{Ker}(f^p) = V$. We may further assume that $\mathbf{v}_1, \ldots, \mathbf{v}_t$ is a maximal list of linearly independent vectors over $\mathrm{Ker}(f^{p-1})$, since otherwise, we simply extend this list to a maximal one (this is possible because V is finite-dimensional).

Now, if $p = 1$, then any basis of V that extends $\{\mathbf{v}_1, \ldots, \mathbf{v}_t\}$ will do.[49] Otherwise, it can be shown that $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_t)$ are linearly independent over $\mathrm{Ker}(f^{p-2})$. Then, we apply the induction hypothesis to the vector space $\mathrm{Ker}(f^{p-1})$, the linear function f (more precisely: its restriction to $\mathrm{Ker}(f^{p-1})$), and the vectors $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_t)$. We obtain a basis of $\mathrm{Ker}(f^{p-1})$ that is the union of pairwise disjoint Jordan chains, t of which are started by $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_t)$. This basis is of the form

$$\begin{array}{ll} & \{f^{p-1}(\mathbf{v}_1), f^{p-2}(\mathbf{v}_1), \ldots, f^2(\mathbf{v}_1), f(\mathbf{v}_1)\} \\ & \vdots \\ \bigcup & \{f^{p-1}(\mathbf{v}_t), f^{p-2}(\mathbf{v}_t), \ldots, f^2(\mathbf{v}_t), f(\mathbf{v}_t)\} \\ \bigcup & \{f^{a_{t+1}-1}(\mathbf{v}_{t+1}), f^{a_{t+1}-2}(\mathbf{v}_{t+1}), \ldots, f^2(\mathbf{v}_{t+1}), f(\mathbf{v}_{t+1}), \mathbf{v}_{t+1}\} \\ & \vdots \\ \bigcup & \{f^{a_{t+s}-1}(\mathbf{v}_{t+s}), f^{a_{t+s}-2}(\mathbf{v}_{t+s}), \ldots, f^2(\mathbf{v}_{t+s}), f(\mathbf{v}_{t+s}), \mathbf{v}_{t+s}\}, \end{array}$$

for some vectors $\mathbf{v}_{t+1}, \ldots, \mathbf{v}_{t+s} \in \mathrm{Ker}(f^{p-1})$ and positive integers $a_{t+1}, \ldots, a_{t+s}$ such that $f^{a_{t+i}}(\mathbf{v}_i) = \mathbf{0}$ and $f^{a_{t+i}-1}(\mathbf{v}_i) \neq \mathbf{0}$ for all $i \in \{1, \ldots, s\}$. We extend the chains started by $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_t)$ by adding to them the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$, respectively. These "extended" chains, plus the remaining ("non-extended") chains form the basis $\mathcal{B}$ of V that we need:

$$\mathcal{B} := \left(\begin{array}{ll} & \{f^{p-1}(\mathbf{v}_1), f^{p-2}(\mathbf{v}_1), \ldots, f^2(\mathbf{v}_1), f(\mathbf{v}_1), \mathbf{v}_1\} \\ & \vdots \\ \bigcup & \{f^{p-1}(\mathbf{v}_t), f^{p-2}(\mathbf{v}_t), \ldots, f^2(\mathbf{v}_t), f(\mathbf{v}_t), \mathbf{v}_t\} \\ \bigcup & \{f^{a_{t+1}-1}(\mathbf{v}_{t+1}), f^{a_{t+1}-2}(\mathbf{v}_{t+1}), \ldots, f^2(\mathbf{v}_{t+1}), f(\mathbf{v}_{t+1}), \mathbf{v}_{t+1}\} \\ & \vdots \\ \bigcup & \{f^{a_{t+s}-1}(\mathbf{v}_{t+s}), f^{a_{t+s}-2}(\mathbf{v}_{t+s}), \ldots, f^2(\mathbf{v}_{t+s}), f(\mathbf{v}_{t+s}), \mathbf{v}_{t+s}\} \end{array} \right).$$

Remark: This completes our outline of the proofs of Theorems 8.6.2 and 8.6.4. The full technical details are given in subsections 8.6.4, 8.6.5, and 8.6.6.

[49] This is because, if $p = 1$, then every non-zero vector $\mathbf{u}$ starts the one-element Jordan chain $\{\mathbf{u}\}$.

8.6.4 Invariant subspaces and the uniqueness of the Jordan normal form: the proof of Theorem 8.6.1

In this subsection, we prove Theorem 8.6.1. We note that this in fact proves the uniqueness (but not the existence) part of Theorems 8.6.2 and 8.6.4. We further note that no result of this subsection requires that the field that we are working over be algebraically closed. Algebraic closure will become relevant in subsection 8.6.5, which deals with the existence part of Theorems 8.6.2 and 8.6.4.

The proof of Theorem 8.6.1 has two parts. First, we prove that any if two matrices are the direct sums of the same matrices, but possibly arranged in a different order, then those two matrices are in fact similar (see Proposition 8.6.12); this immediately implies that two Jordan matrices that have exactly the same Jordan blocks (counting repetitions) are similar. Then, we show that if a square matrix A is similar to a Jordan matrix J, then the types of Jordan blocks appearing in J depend only on A, as does the number of Jordan blocks of each possible type (see Proposition 8.6.17).[50] Thus, any two Jordan matrices similar to a given square matrix A have exactly the same number of Jordan blocks (counting repetitions, but not counting the order in which they appear). Since matrix similarity is an equivalence relation, this, in particular, implies that any two similar Jordan matrices have exactly the same Jordan blocks (counting repetitions, but not counting the order in which the Jordan blocks appear in the two matrices).

Before turning to the proofs, we note that the formula for the number of Jordan blocks of each type from Proposition 8.6.17 is precisely the one given in Theorem 8.6.6 (stated in subsection 8.6.1). Thus, it may seem that Proposition 8.6.17 immediately implies Theorem 8.6.6. This, however, is not the case. Indeed, Proposition 8.6.17 does **not** state that any matrix $A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field) is similar to a Jordan matrix $J \in \mathbb{F}^{n\times n}$. For general fields $\mathbb{F}$, the statement is not even true. For algebraically closed fields $\mathbb{F}$, we do indeed get the existence of such a Jordan matrix J, but this requires a lot more work (see subsections 8.6.5 and 8.6.6).

Invariant subspaces. Suppose that V is a vector space over a field $\mathbb{F}$ and that $f : V \to V$ is a linear function. A subspace U of V is said to be *f-invariant* (or *invariant for f*) if $f[U] \subseteq U$, that is, if for all $\mathbf{u} \in U$, we have that $f(\mathbf{u}) \in U$. Under these circumstances (i.e. if U is an f-invariant subspace of V), we may define the function $f|_U : U \to U$ given by $f|_U(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in U$; thus, $f|_U$ is obtained by restricting both the domain and the codomain of f to U (which we can do because U is f-invariant), and obviously, $f|_U$ is linear (because f is linear).

Recall from subsection 3.2.6 that if V is a vector space over a field $\mathbb{F}$ and U_1 and U_2 are its subspaces such that $U_1 \cap U_2 = \{\mathbf{0}\}$ and $V = U_1 + U_2$, then we say that V is the *direct sum* of U_1 and U_2, and we write $V = U_1 \oplus U_2$. If

[50] We note that Proposition 8.6.17 immediately implies the uniqueness part of Theorem 8.6.2, as outlined in subsection 8.6.3.

$V = U_1 \oplus U_2$ is also finite-dimensional, then Theorem 3.2.23 immediately implies that $\dim(V) = \dim(U_1) + \dim(U_2)$.[51]

Obviously, the direct sum of subspaces can be generalized to more than two subspaces. Suppose that V is a vector space over a field $\mathbb{F}$, and suppose that $U_1, \ldots, U_k$ ($k \geq 1$) are its subspaces such that $U_i \cap U_j = \{\mathbf{0}\}$ for all distinct $i, j \in \{1, \ldots, k\}$ and such that $V = U_1 + \cdots + U_k$.[52] Under these circumstances, we say that V is the *direct sum* of $U_1, \ldots, U_k$, and we write $V = U_1 \oplus \cdots \oplus U_k$.

Proposition 8.6.10. *Let V be a finite-dimensional vector space over a field $\mathbb{F}$, let $\mathcal{B}$ be a basis of V, and let $(\mathcal{B}_1, \ldots, \mathcal{B}_k)$ be a partition of $\mathcal{B}$, i.e. assume that $\mathcal{B}_1, \ldots, \mathcal{B}_k$ are pairwise disjoint and that $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$.*[53] *For all $i \in \{1, \ldots, k\}$, set $U_i := \mathit{Span}(\mathcal{B}_i)$, so that $\mathcal{B}_i$ is a basis of U_i.*[54] *Then $V = U_1 \oplus \cdots \oplus U_k$.*

Proof. This essentially follows from the appropriate definitions. The details are left as a straightforward exercise. □

Proposition 8.6.11. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $f : V \to V$ be a linear function. Assume that $V = U_1 \oplus \cdots \oplus U_k$ ($k \geq 1$), where $U_1, \ldots, U_k$ are non-trivial, f-invariant subspaces of V. To simplify notation, we set $f_i := f|_{U_i}$ for all $i \in \{1, \ldots, k\}$. Then for any bases $\mathcal{B}_1, \ldots, \mathcal{B}_k$ of $U_1, \ldots, U_k$, respectively, we have that $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ is a basis of V, and moreover,*

$$ {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \quad = \quad {}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1} \oplus \cdots \oplus {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}. $$

Consequently,

$$ p_f(\lambda) \quad = \quad p_{f_1}(\lambda) \ldots p_{f_k}(\lambda), $$

i.e. the characteristic polynomial of f is equal to the product of the characteristic polynomials of $f_1, \ldots, f_k$.

Proof. For each index $i \in \{1, \ldots, k\}$, fix a basis $\mathcal{B}_i = \{\mathbf{b}_{i,1}, \ldots, \mathbf{b}_{i,m_i}\}$ of U_i. Set $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k = \{\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{1,m_1}, \ldots, \mathbf{b}_{k,1}, \ldots, \mathbf{b}_{k,m_k}\}$.

Claim 1. $\mathcal{B}$ is a basis of V.

Proof of Claim 1. Since V is finite-dimensional and satisfies $V = U_1 \oplus \cdots \oplus U_k$, we know that $\dim(V) = \dim(U_1) + \cdots + \dim(U_k) = m_1 + \cdots + m_k$.[55] Since $\mathcal{B}$ contains

[51] This is because $\dim(U_1 \cap U_2) = 0$.

[52] Here, $U_1 + \cdots + U_k$ is defined in the obvious way:

$$ U_1 + \cdots + U_k \quad := \quad \{\mathbf{u}_1 + \cdots + \mathbf{u}_k \mid \mathbf{u}_1 \in U_1, \ldots, \mathbf{u}_k \in U_k\}. $$

[53] Here, we allow $\mathcal{B}_1, \ldots, \mathcal{B}_k$ to possibly be empty, although in practice, we will apply the proposition only to the case when they are all non-empty.

[54] The fact that $\mathcal{B}_i$ is linearly independent follows from the fact that $\mathcal{B}$ is linearly independent (because it is a basis of V), and by construction, $\mathcal{B}_i$ is a spanning set of U_i.

[55] This readily follows from Theorem 3.2.23 via an easy induction on k.

precisely $m_1 + \cdots + m_k$ many vectors, it now suffices to show that $\mathcal{B}$ is a spanning set of V, for Theorem 3.2.20(b) will then immediately imply that $\mathcal{B}$ is a basis of V. Fix any $\mathbf{v} \in V$. Since $V = U_1 \oplus \cdots \oplus U_2$, there exist vectors $\mathbf{u}_1 \in U_1, \ldots, \mathbf{u}_k \in U_k$ such that $\mathbf{v} = \mathbf{u}_1 + \cdots + \mathbf{u}_k$. For each $i \in \{1, \ldots, k\}$, we fix scalars $\alpha_{i,1}, \ldots, \alpha_{i,m_i} \in \mathbb{F}$ such that

$$\mathbf{u}_i \;=\; \alpha_{i,1}\mathbf{b}_{i,1} + \cdots + \alpha_{i,m_i}\mathbf{b}_{i,m_i};$$

the existence of such scalars $\alpha_{i,1}, \ldots, \alpha_{i,m_i}$ follows from the fact that $\mathbf{u}_i \in U_i$ and the fact that $\mathcal{B}_i = \{\mathbf{b}_{i,1}, \ldots, \mathbf{b}_{i,m_i}\}$ is a basis of U_i. But now

$$\mathbf{v} \;=\; \sum_{i=1}^{k} \mathbf{u}_i \;=\; \sum_{i=1}^{k} (\alpha_{i,1}\mathbf{b}_{i,1} + \cdots + \alpha_{i,m_i}\mathbf{b}_{i,m_i}),$$

and we see that $\mathbf{v}$ is a linear combination of the vectors in $\mathcal{B}$. We now conclude that $\mathcal{B}$ is indeed a spanning set of V, and is therefore (as we discussed above) a basis of V. ♦

Claim 2. For all $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, m_i\}$, we have that

$$\left[\; f(\mathbf{b}_{i,j}) \;\right]_{\mathcal{B}} \;=\; \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hdashline \left[\; f_i(\mathbf{b}_{i,j}) \;\right]_{\mathcal{B}_i} \\ \hdashline 0 \\ \vdots \\ 0 \end{bmatrix},$$

with exactly $m_1 + \cdots + m_{i-1}$ many 0's in the top block,[56] and exactly $m_{i+1} + \cdots + m_k$ many 0's in the bottom block.[57]

Proof of Claim 2. Fix indices $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, m_i\}$. Since U_i is f-invariant, we have that $f(\mathbf{b}_{i,j}) \in U_i$. Therefore, there exist scalars $\alpha_{i,1}, \ldots, \alpha_{i,m_i}$ such that

$$f_i(\mathbf{b}_{i,j}) \;= f(\mathbf{b}_{i,j}) \;=\; \sum_{\ell=1}^{m_i} \alpha_{i,\ell}\mathbf{b}_{i,\ell}.$$

and consequently,

$$\left[\; f_i(\mathbf{b}_{i,j}) \;\right]_{\mathcal{B}_i} \;=\; \left[\; \alpha_{i,1} \;\; \ldots \;\; \alpha_{i,m_i} \;\right]^T$$

and

$$\left[\; f(\mathbf{b}_{i,j}) \;\right]_{\mathcal{B}} \;=\; \left[\; 0 \;\; \ldots \;\; 0 \;\vdots\; \alpha_{i,1} \;\; \ldots \;\; \alpha_{i,m_i} \;\vdots\; 0 \;\; \ldots \;\; 0 \;\right]^T,$$

[56] Meaning: above the block $\left[\; f(\mathbf{b}_{i,j}) \;\right]_{\mathcal{B}_i}$.

[57] Meaning: below the block $\left[\; f(\mathbf{b}_{i,j}) \;\right]_{\mathcal{B}_i}$.

with $m_1 + \cdots + m_{i-1}$ many 0's on the left (these 0's correspond to the coefficients 0 in front of the basis vectors $\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{1,m_1}, \ldots, \mathbf{b}_{i-1,1}, \ldots, \mathbf{b}_{i-1,m_{i-1}}$), and with $m_{i+1} + \cdots + m_k$ many 0's on the right (these 0's correspond to the coefficients 0 in front of the basis vectors $\mathbf{b}_{i+1,1}, \ldots, \mathbf{b}_{i+1,m_{i+1}}, \ldots, \mathbf{b}_{k,1}, \ldots, \mathbf{b}_{k,m_k}$). The result is now immediate. ♦

Claim 3. ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} = {}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1} \oplus \cdots \oplus {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}$.

Proof of Claim 3. This follows immediately from Theorem 4.5.1 and from Claim 2. ♦

Now, set $n := \dim(V)$. By Claim 1, we have that $n = m_1 + \cdots + m_k$, and consequently, $I_n = I_{m_1} \oplus \cdots \oplus I_{m_k}$. Further, set $B = {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. Then

$$\begin{aligned}
\lambda I_n - B &= \lambda I_n - {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \\
&\overset{(*)}{=} \Big((\lambda I_{m_1}) \oplus \cdots \oplus (\lambda I_{m_k})\Big) - \Big({}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1} \oplus \cdots \oplus {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}\Big) \\
&= \Big(\lambda I_{m_1} - {}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1}\Big) \oplus \cdots \oplus \Big(\lambda I_{m_k} - {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}\Big)
\end{aligned}$$

where (*) follows from Claim 3. Consequently,

$$\begin{aligned}
p_f(\lambda) \overset{(*)}{=} p_B(\lambda) &= \det\big(\lambda I_n - B\big) \\
&= \det\left(\Big(\lambda I_{m_1} - {}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1}\Big) \oplus \cdots \oplus \Big(\lambda I_{m_k} - {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}\Big)\right) \\
&\overset{(**)}{=} \prod_{i=1}^{k} \det\Big(\lambda I_{m_i} - {}_{\mathcal{B}_i}\big[\ f_i\ \big]_{\mathcal{B}_i}\Big) \overset{(*)}{=} \prod_{i=1}^{k} p_{f_i}(\lambda),
\end{aligned}$$

where both instances of (*) follow from Proposition 8.2.12, and (**) follows from Corollary 7.6.7. This completes the argument. □

Proposition 8.6.12. *Let $A_1, \ldots, A_k$ be square matrices with entries in some field $\mathbb{F}$, and let $\sigma \in S_k$. Then matrices $A := A_1 \oplus \cdots \oplus A_k$ and $A_\sigma := A_{\sigma(1)} \oplus \cdots \oplus A_{\sigma(k)}$ are similar.*

Remark: Proposition 8.6.12 immediately implies that any two Jordan matrices that have exactly the same Jordan blocks (counting repetitions) are similar.

Proof. For each index $i \in \{1, \ldots, k\}$, assume that the square matrix A_i is of size $m_i \times m_i$. Set $m := m_1 + \cdots + m_k$, so that A is of size $m \times m$, and let $\mathcal{E}_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ be the standard basis of $\mathbb{F}^m$. Let $f : \mathbb{F}^m \to \mathbb{F}^m$ be given by $f(\mathbf{x}) = A\mathbf{x}$, so that f linear and A is the standard matrix of f, i.e. $A = {}_{\mathcal{E}_m}\big[\ f\ \big]_{\mathcal{E}_m}$. For each index $i \in \{1, \ldots, k\}$,

we set $\mathcal{B}_i := \{\mathbf{e}_{m_1+\cdots+m_{i-1}+1}, \ldots, \mathbf{e}_{m_1+\cdots+m_{i-1}+m_i}\}$,[58] and we set $U_i := \mathrm{Span}(\mathcal{B}_i)$, so that $\mathcal{B}_i$ is a basis of U_i. It is then easy to see that for each $i \in \{1, \ldots, k\}$, the subspace U_i of $\mathbb{F}^m$ is f-invariant,[59] and moreover, that $A_i = {}_{\mathcal{B}_i}\big[\ f_i\ \big]_{\mathcal{B}_i}$, where $f_i := f|_{U_i}$.[60] Now, set $\mathcal{E}_m^\sigma := \mathcal{B}_{\sigma(1)} \cup \cdots \cup \mathcal{B}_{\sigma(k)}$; obviously, $\mathcal{E}_m^\sigma$ is a basis of $\mathbb{F}^m$, since it was obtained by simply permuting the vectors of the standard basis $\mathcal{E}_m$ of $\mathbb{F}^m$. It then follows from Proposition 8.6.11 that

$$ {}_{\mathcal{E}_m^\sigma}\big[\ f\ \big]_{\mathcal{E}_m^\sigma} \quad = \quad \underbrace{{}_{\mathcal{B}_{\sigma(1)}}\big[\ f_{\sigma(1)}\ \big]_{\mathcal{B}_{\sigma(1)}}}_{=A_{\sigma(1)}} \oplus \cdots \oplus \underbrace{{}_{\mathcal{B}_{\sigma(k)}}\big[\ f_{\sigma(k)}\ \big]_{\mathcal{B}_{\sigma(k)}}}_{=A_{\sigma(k)}} \quad = \quad A_\sigma. $$

By Theorem 4.5.16, matrices $A = {}_{\mathcal{E}_m}\big[\ f\ \big]_{\mathcal{E}_m}$ and $A_\sigma = {}_{\mathcal{E}_m^\sigma}\big[\ f\ \big]_{\mathcal{E}_m^\sigma}$ are similar. This completes the argument. □

Counting the number of Jordan blocks. Suppose we are are given a matrix $A \in \mathbb{F}^{n\times n}$ (where $\mathbb{F}$ is a field), and that we know that this matrix is similar to a Jordan matrix $J \in \mathbb{F}^{n\times n}$. In what follows, we would like to determine the types of Jordan blocks that the Jordan matrix J can have, and also to compute the number of blocks of each possible type (see Proposition 8.6.17 below); combined with Proposition 8.6.12, this will allow us to prove Theorem 8.6.1, which states that two Jordan matrices are similar if and only if they have exactly the same Jordan blocks (counting repetitions). This, in turn, will allow us to prove uniqueness part of Theorems 8.6.2 and 8.6.4. Proposition 8.6.17 will also be one of the ingredients of the proof of Theorem 8.6.6. We begin with four simple technical propositions (Propositions 8.6.13, 8.6.14, 8.6.15, and 8.6.16).

Proposition 8.6.13. *Let $\mathbb{F}$ be a field, and let t be a positive integer. Then both the following hold:*

(a) for all matrices $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_t \end{bmatrix}$ in $\mathbb{F}^{s\times t}$, we have that

$$ AJ_t(0) \quad = \quad \begin{bmatrix} \mathbf{0} & \mathbf{a}_1 & \ldots & \mathbf{a}_{t-1} \end{bmatrix}, $$

i.e. $AJ_t(0)$ is the $s \times t$ matrix obtained from A by first adding a zero column to the left, and then deleting the rightmost column of the resulting matrix;

[58] So, the vectors of $\mathcal{B}_1$ are the first m_1 vectors of $\mathcal{E}_m$, the vectors of $\mathcal{B}_2$ are the subsequent m_2 vectors of $\mathcal{E}_m$, and so on.

[59] Here are the details. Fix an index $i \in \{1, \ldots, k\}$. Then for all $\mathbf{e}_j \in \mathcal{B}_i$, $A\mathbf{e}_j$ is the j-th column of A, and we see from the construction of A that this column is a linear combination of the vectors in $\mathcal{B}_i$, i.e. $f(\mathbf{e}_j) = A\mathbf{e}_j \in \mathrm{Span}(\mathcal{B}_i) = U_i$. Since $U_i = \mathrm{Span}(\mathcal{B}_i)$ and since f is linear, it readily follows that $f[U_i] \subseteq U_i$.

[60] Details?

(b) we have that

$$\big(J_t(0)\big)^r \;=\; \begin{cases} I_t & \text{if } \; r = 0 \\ \left[\begin{array}{c|c} O_{(t-r)\times r} & I_{t-r} \\ \hline O_{r\times r} & O_{r\times(t-r)} \end{array}\right] & \text{if } \; 1 \leq r \leq t-1 \\ O_{t\times t} & \text{if } \; r \geq t \end{cases}$$

for all non-negative integers r.

Proof. Clearly, (b) follows from (a) via an easy induction on r. So, let us prove (a). Fix a positive integer t and a matrix $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_t \end{bmatrix}$ in $\mathbb{F}^{s\times t}$. By definition, we have that

$$J_t(0) \;=\; \begin{bmatrix} \mathbf{0} & \mathbf{e}_1 & \ldots & \mathbf{e}_{t-1} \end{bmatrix},$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_t$ are the standard basis vectors of $\mathbb{F}^t$. We now compute:

$$\begin{aligned} AJ_t(0) \;&=\; A\begin{bmatrix} \mathbf{0} & \mathbf{e}_1 & \ldots & \mathbf{e}_{t-1} \end{bmatrix} \\ &=\; \begin{bmatrix} A\mathbf{0} & A\mathbf{e}_1 & \ldots & A\mathbf{e}_{t-1} \end{bmatrix} && \text{by the definition of matrix multiplication} \\ &=\; \begin{bmatrix} \mathbf{0} & \mathbf{a}_1 & \ldots & \mathbf{a}_{t-1} \end{bmatrix} && \text{by Proposition 1.4.4.} \end{aligned}$$

This proves (a), and we are done. □

Proposition 8.6.14. *Let $\mathbb{F}$ be a field. Then*

$$\mathit{rank}\Big(\big(J_t(\lambda)\big)^r\Big) \;=\; \begin{cases} t & \text{if } \; \lambda \neq 0 \\ t-r & \text{if } \; \lambda = 0 \text{ and } r \leq t-1 \\ 0 & \text{if } \; \lambda = 0 \text{ and } r \geq t \end{cases}$$

for all $\lambda \in \mathbb{F}$ and all positive integers t and non-negative integers r.

Proof. Fix $\lambda \in \mathbb{F}$. If $\lambda = 0$, then the result follows immediately from Proposition 8.6.13. So, assume that $\lambda \neq 0$, and fix a positive integer t and a non-negative integer r. The Jordan block $J_t(\lambda)$ is an upper triangular $t \times t$ matrix with all λ's on the main diagonal, and so by Proposition 7.3.1, $\det\big(J_t(\lambda)\big) = \lambda^t \neq 0$. So, by the Invertible Matrix Theorem (see subsection 8.2.6), the matrix $J_t(\lambda)$ is invertible. Therefore, by Proposition 1.11.8(f), $\big(J_t(\lambda)\big)^r$ is also invertible, and consequently (by the Invertible Matrix Theorem) has rank t. □

Proposition 8.6.15. *Let $A_1, \ldots, A_k$ be square matrices with entries in some field $\mathbb{F}$. Then for all non-negative integers r, we have that*

$$(A_1 \oplus \cdots \oplus A_k)^r \;=\; A_1^r \oplus \cdots \oplus A_k^r.$$

Proof. This readily follows from the definition of matrix multiplication. □

Proposition 8.6.16. *Let $A_1, \ldots, A_k$ be square matrices with entries in some field $\mathbb{F}$. Then*

$$\operatorname{rank}(A_1 \oplus \cdots \oplus A_k) \quad = \quad \operatorname{rank}(A_1) + \cdots + \operatorname{rank}(A_k).$$

Proof. For each index $i \in \{1, \ldots, k\}$, assume that the square matrix A_i is of size $m_i \times m_i$. Set $A := A_1 \oplus \cdots \oplus A_k$ and $m := m_1 + \cdots + m_k$, so that A is of size $m \times m$. We first perform row reduction on the top m_1 many rows of A (while ignoring the remaining rows) in order to transform the matrix formed by the top m_1 many rows of A into one in row echelon form. Then, we row reduce the matrix formed by the subsequent m_2 many rows of A in order to turn that submatrix of A into one in row echelon form. We continue like this until we reach the bottom of our matrix A. This produces a matrix that contains exactly $\operatorname{rank}(A_1) + \cdots + \operatorname{rank}(A_k)$ many non-zero rows. Moreover, after "moving" any zero rows to the bottom of this matrix, we obtain a matrix in row echelon form that still has precisely $\operatorname{rank}(A_1) + \cdots + \operatorname{rank}(A_k)$ many non-zero rows. This matrix is row equivalent to A, and so (by Proposition 1.6.2), we have that $\operatorname{rank}(A) = \operatorname{rank}(A_1) + \cdots + \operatorname{rank}(A_k)$. □

Proposition 8.6.17. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n \times n}$, and let*

$$\big\{ \underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{m_k} \big\}$$

be the spectrum of A, where $\lambda_1, \ldots, \lambda_k$ are pairwise distinct eigenvalues of f and $m_1, \ldots, m_k$ are positive integers. Assume that A is similar to a Jordan matrix $J \in \mathbb{F}^{n \times n}$. Then

- *each Jordan block of the Jordan matrix J is of the form $J_t(\lambda_i)$ for some $i \in \{1, \ldots, k\}$ and $t \in \{1, \ldots, m_i\}$;*
- *for each $i \in \{1, \ldots, k\}$ and each positive integer r, the Jordan matrix J has exactly*

  $$\operatorname{rank}\big((A - \lambda_i I_n)^{r-1}\big) - \operatorname{rank}\big((A - \lambda_i I_n)^r\big)$$

  many Jordan blocks $J_t(\lambda_i)$ satisfying $t \geq r$.

Proof. Since A and J are similar, Theorem 8.2.9 guarantees that they have the same spectrum. Since the matrix J is upper triangular (because it is a Jordan matrix), Proposition 8.2.7 guarantees that J has the following entries on the main diagonal (counting repetitions, but possibly in a different order):

$$\underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{m_k}.$$

Note that this implies that $m_1 + \cdots + m_k = n$. It also proves that each Jordan block of J is of the form $J_t(\lambda_i)$ for some $i \in \{1, \ldots, k\}$ and $t \in \{1, \ldots, m_i\}$.

It remains to compute the number of Jordan blocks of each type in the matrix J. We begin with a simple technical claim.

> **Claim 1.** For all scalars $\lambda \in \mathbb{F}$ and positive integers r, matrices $(A-\lambda I_n)^r$ and $(J-\lambda I_n)^r$ are similar and therefore have the same rank.

Proof of Claim 1. Fix $\lambda \in \mathbb{F}$. First of all, since A and J are similar, we know that there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $J = P^{-1}AP$. Then

$$P^{-1}(A-\lambda I_n)P \;=\; P^{-1}AP - \lambda P^{-1}I_nP \;=\; J-\lambda_i I_n,$$

and so $A-\lambda I_n$ and $J-\lambda I_n$ are similar. Now, fix a positive integer r. By Proposition 4.5.15, matrices $(A-\lambda I_n)^r$ and $(J-\lambda I_n)^r$ are similar, and so by Corollary 4.5.17, these two matrices have the same rank. ♦

Now, fix an index $i \in \{1,\ldots,k\}$. We must show that for each positive integer r, the Jordan matrix J has exactly $\operatorname{rank}\big((A-\lambda_i I_n)^{r-1}\big) - \operatorname{rank}\big((A-\lambda_i I_n)^r\big)$ many Jordan blocks $J_t(\lambda_i)$ satisfying $t \geq r$. Let $a_1,\ldots,a_\ell$ be a non-decreasing sequence of positive integers such that the Jordan blocks of J of the form $J_t(\lambda_i)$ are precisely the blocks $J_{a_1}(\lambda_i),\ldots,J_{a_\ell}(\lambda_i)$, counting repetitions and appearing in some order along the main diagonal of J.[61] Clearly, $a_1+\cdots+a_\ell = m_i$.

> **Claim 2.** For all non-negative integers r and indices $q \in \{0,1,\ldots,k\}$, if $a_1 \geq \cdots \geq a_q \geq r \geq a_{q+1} \geq \cdots \geq a_k$,[62] then
>
> $$\operatorname{rank}\big((A-\lambda_i I_n)^r\big) \;=\; (n-m_i) + \sum_{j=1}^{q}(a_j - r).$$

Proof of Claim 2. First of all, since J is a Jordan matrix, so is the matrix $J-\lambda_i I_n$, and moreover, any Jordan block $J_t(\lambda_j)$ of J (with $j \in \{1,\ldots,k\}$) corresponds to a Jordan block $J_t(\lambda_j-\lambda_i)$ of $J-\lambda_i I_n$ in the obvious way. In particular, the Jordan blocks $J_t(\lambda_i)$ of J correspond to the Jordan blocks $J_t(0)$ of $J-\lambda_i I_n$ in the obvious way, and we see that the Jordan blocks of $J-\lambda_i I_n$ of the form $J_t(0)$ are precisely the blocks $J_{a_1}(0),\ldots,J_{a_\ell}(0)$ (counting repetitions, and appearing in some order in $J-\lambda_i I_n$). Let $J_1,\ldots,J_s$ be the Jordan blocks of the Jordan matrix $J-\lambda_i I_n$ other than $J_{a_1}(0),\ldots,J_{a_k}(0)$ (counting repetitions). So, $J-\lambda_i I_n$ is the direct sum of the Jordan blocks $J_1,\ldots,J_s,J_{a_1}(0),\ldots,J_{a_k}(0)$, arranged in some order.

[61] Of course, the Jordan matrix J may contain other Jordan blocks as well, but those other Jordan blocks do not have the eigenvalue λ_i on their main diagonal.

[62] If $q=0$, then this means that $r \geq a_1 \geq \cdots \geq a_k$. On the other hand, if $q=k$, then we have that $a_1 \geq \cdots \geq a_k \geq r$.

Now, fix a positive integer r and an index $q \in \{0, 1, \ldots, k\}$ such that $a_1 \geq \cdots \geq a_q \geq r \geq a_{q+1} \geq \cdots \geq a_k$. Then

$$\begin{aligned} \operatorname{rank}\big((A - \lambda_i I_n)^r\big) &\overset{(*)}{=} \operatorname{rank}\big((J - \lambda_i I_n)^r\big) \\ &\overset{(**)}{=} \Big(\sum_{j=1}^{s} \operatorname{rank}(J_j^r)\Big) + \Big(\sum_{j=1}^{k} \operatorname{rank}\Big(\big(J_{a_j}(0)\big)^r\Big)\Big), \end{aligned}$$

where (*) follows from Claim 1, and (**) follows from Propositions 8.6.15 and 8.6.16. By Proposition 8.6.14 we know that for all $\ell \in \{1, \ldots, s\}$, J_s^r is a matrix of full rank, and in particular, we have that

(1) $\operatorname{rank}(J_1^r) + \cdots + \operatorname{rank}(J_s^r) = n - m_i$.[63]

Proposition 8.6.14 further implies the following:

(2) for all $j \in \{1, \ldots, q\}$, we have that $\operatorname{rank}\Big(\big(J_{a_j}(0)\big)^r\Big) = a_j - r$ (because $r \leq a_j$);

(3) for all $j \in \{q+1, \ldots, k\}$, we have that $\operatorname{rank}\Big(\big(J_{a_j}(0)\big)^r\Big) = 0$ (because $r \geq a_j$).

It now follows that

$$\begin{aligned} \operatorname{rank}\big((A - \lambda_i I_n)^r\big) &= \Big(\sum_{j=1}^{s} \operatorname{rank}(J_j^r)\Big) + \Big(\sum_{j=1}^{k} \operatorname{rank}\Big(\big(J_{a_j}(0)\big)^r\Big)\Big) \\ &\overset{(*)}{=} (n - m_i) + \sum_{j=1}^{q} (a_j - r), \end{aligned}$$

where (*) follows from (1), (2), and (3). ♦

Now, fix a positive integer r, and let q be the number of Jordan blocks $J_t(\lambda_i)$ of J satisfying $t \geq r$. Then $a_1 \geq \cdots \geq a_q \geq r > a_{q+1} \geq \cdots \geq a_k$,[64] and consequently, $a_1 \geq \cdots \geq a_q \geq r - 1 \geq a_{q+1} \geq \cdots \geq a_k$. By applying Claim 2 first to r and then to $r - 1$, we obtain the following:

- $\operatorname{rank}\big((A - \lambda_i I_n)^r\big) = (n - m_i) + \sum_{j=1}^{q} (a_j - r)$;
- $\operatorname{rank}\big((A - \lambda_i I_n)^{r-1}\big) = (n - m_i) + \sum_{j=1}^{q} (a_j - (r-1))$.

Consequently, $\operatorname{rank}\big((A - \lambda_i I_n)^{r-1}\big) - \operatorname{rank}\big((A - \lambda_i I_n)^r\big) = q$, and we are done. □

[63] We are using the fact that $a_1 + \cdots + a_\ell = m_i$.

[64] If $q = 0$, then we simply have that $r > a_1 \geq \cdots \geq a_k$. On the other hand, if $q = k$, then $a_1 \geq \cdots \geq a_k \geq r$.

We are now ready to prove Theorem 8.6.1, restated below for the reader's convenience.

Theorem 8.6.1. *Let $\mathbb{F}$ be a field, and let $J_1, J_2 \in \mathbb{F}^{n\times n}$ be Jordan matrices. Then J_1 and J_2 are similar if and only if they have exactly the same Jordan blocks (counting repetitions, but not counting the order in which the blocks appear in the two matrices).*

Proof. If J_1 and J_2 have the same Jordan blocks (counting repetitions), then Proposition 8.6.12 guarantees that they are similar. On the other hand, if J_1 and J_2 are similar, then Proposition 8.6.17 guarantees that they have exactly the same Jordan blocks (counting repetitions).[65] □

8.6.5 Generalized eigenspaces, nilpotent linear functions, and the existence of the Jordan normal form of a square matrix

The main goal of this subsection is to prove Theorem 8.6.30, which is the existence part of Theorem 8.6.4. As an immediate consequence of Theorem 8.6.30, we obtain Corollary 8.6.31, which is the existence part of 8.6.2.

Iterated linear functions. Suppose that A is a set and $f : A \to A$ is a function.[66] We define $f^0 := \mathrm{Id}_A$,[67] and for all non-negative integers k, $f^{k+1} := f^k \circ f$. So, for all positive integers k, we have that

$$f^k = \underbrace{f \circ \cdots \circ f}_{k}.$$

The function f^k is called the *k-th iterate* of f.

Proposition 8.6.18. *Let V be a vector space over a field $\mathbb{F}$, and let $f : V \to V$ be a linear function. Then all the following hold:*

(a) functions $f^0, f^1, f^2, f^3, \ldots$ are all linear;[68]

(b) $\{\mathbf{0}\} = Ker(f^0) \subseteq Ker(f^1) \subseteq Ker(f^2) \subseteq Ker(f^3) \subseteq \ldots$;

(c) for all non-negative integers r, if $Ker(f^r) = Ker(f^{r+1})$, then $Ker(f^r) = Ker(f^{r+1}) = Ker(f^{r+2}) = Ker(f^{r+3}) = \ldots$;

(d) for all non-negative integers r and all scalars $\lambda \in \mathbb{F}$, both $Ker(f^r)$ and $Im(f^r)$ are $(f + \lambda Id_V)$-invariant subspaces of V;[69]

[65] Indeed, we set $A := J_1$. Then A is similar both to J_1 and to J_2, and we simply apply Proposition 8.6.17 twice: first to A and J_1, and then to A and J_2.

[66] Here, it is important that the domain and the codomain of f are the same.

[67] As usual, Id_A is the identity function on A, that is, $\mathrm{Id}_A : A \to A$ is given by $\mathrm{Id}_A(a) = a$ for all $a \in A$.

[68] Note that this implies that $\mathrm{Ker}(f^0), \mathrm{Ker}(f^1), \mathrm{Ker}(f^2), \mathrm{Ker}(f^3), \ldots$ are all defined.

[69] Since f and Id_V are linear, Proposition 4.1.7 guarantees that $f + \lambda \mathrm{Id}_V$ is linear for all $\lambda \in \mathbb{F}$.

(e) for all positive integers r, and all $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V$, if $\mathbf{v}_i \in Ker(f^i) \setminus Ker(f^{i-1})$ for all $i \in \{1, \ldots, r\}$, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent;

(f) for all non-negative integers r, all $i \in \{0, \ldots, r\}$, and all $\mathbf{v} \in V$, we have that $\mathbf{v} \in Ker(f^r)$ if and only if $f^{r-i}(\mathbf{v}) \in Ker(f^i)$;

(g) for all positive integers r and vectors $\mathbf{v} \in Ker(f^r) \setminus Ker(f^{r-1})$, both the following hold:

- *for all $i \in \{1, \ldots, r\}$, $f^{r-i}(\mathbf{v}) \in Ker(f^i) \setminus Ker(f^{i-1})$;*
- *the set $\left\{f^{r-1}(\mathbf{v}), f^{r-2}(\mathbf{v}), \ldots, f^2(\mathbf{v}), f(\mathbf{v}), \mathbf{v}\right\}$ is linearly independent.*

Proof. We prove (a)-(g) in order, with one exception: we prove (c) last, because our proof of (c) relies on (f).

(a) The function $f^0 = \mathrm{Id}_V$ is obviously linear, and the function $f^1 = f$ is linear by assumption. The result now follows from Proposition 4.1.7(c) via an obvious induction.

(b) First of all, by definition, $f^0 = \mathrm{Id}_V$, and so $\mathrm{Ker}(f^0) = \{\mathbf{0}\}$. Now, fix a non-negative integer r. We must show that $\mathrm{Ker}(f^r) \subseteq \mathrm{Ker}(f^{r+1})$. Fix $\mathbf{v} \in \mathrm{Ker}(f^r)$. Then

$$f^{r+1}(\mathbf{v}) \quad = \quad f\big(f^r(\mathbf{v})\big) \quad \overset{(*)}{=} \quad f(\mathbf{0}) \quad \overset{(**)}{=} \quad \mathbf{0},$$

where (*) follows from the fact that $\mathbf{v} \in \mathrm{Ker}(f^r)$, and (**) follows from the fact that f is linear and from Proposition 4.1.6. This proves that $\mathrm{Ker}(f^r) \subseteq \mathrm{Ker}(f^{r+1})$.

(d) Fix a non-negative integer r and a scalar $\lambda \in \mathbb{F}$. We first show that $\mathrm{Ker}(f^r)$ is $(f + \lambda \mathrm{Id}_V)$-invariant. Fix any $\mathbf{v} \in \mathrm{Ker}(f^r)$, so that $f^r(\mathbf{v}) = \mathbf{0}$; we must show that $(f + \lambda \mathrm{Id}_V)(\mathbf{v}) \in \mathrm{Ker}(f^r)$. We compute:

$$\begin{aligned} f^r\Big((f + \lambda \mathrm{Id}_V)(\mathbf{v})\Big) \quad &\overset{(*)}{=} \quad f^{r+1}(\mathbf{v}) + \lambda f^r(\mathbf{v}) \\ &= \quad f\big(f^r(\mathbf{v})\big) + \lambda f^r(\mathbf{v}) \\ &\overset{(**)}{=} \quad \underbrace{f(\mathbf{0})}_{\overset{(***)}{=}\mathbf{0}} + \lambda \mathbf{0} \quad = \quad \mathbf{0}, \end{aligned}$$

where (*) follows from the linearity of f^r, (**) follows from the fact that $\mathbf{v} \in \mathrm{Ker}(f^r)$ (and so $f^r(\mathbf{v}) = \mathbf{0}$), and (***) follows from the linearity of f and from Proposition 4.1.6. Thus, $(f + \lambda \mathrm{Id}_V)(\mathbf{v}) \in \mathrm{Ker}(f^r)$, and it follows that $\mathrm{Ker}(f^r)$ is $(f + \lambda \mathrm{Id}_V)$-invariant.

It remains to show that $\mathrm{Im}(f^r)$ is $(f + \lambda \mathrm{Id}_V)$-invariant. Fix any $\mathbf{v} \in \mathrm{Im}(f^r)$; we must show that $(f + \lambda \mathrm{Id}_V)(\mathbf{v}) \in \mathrm{Im}(f^r)$. Since $\mathbf{v} \in \mathrm{Im}(f^r)$, we know that there

exists some $\mathbf{u} \in V$ such that $\mathbf{v} = f^r(\mathbf{u})$. We now compute:

$$\begin{aligned}(f + \lambda \mathrm{Id}_V)(\mathbf{v}) &= f(\mathbf{v}) + \lambda \mathbf{v} \\ &= f\big(f^r(\mathbf{u})\big) + \lambda f^r(\mathbf{u}) \\ &= f^r\big(f(\mathbf{u})\big) + \lambda f^r(\mathbf{u}) \\ &\overset{(*)}{=} f^r\big(\underbrace{f(\mathbf{u}) + \lambda \mathbf{u}}_{=:\mathbf{w}}\big),\end{aligned}$$

where (*) follows from the linearity of f^r. We have now obtained that $(f+\lambda \mathrm{Id}_V)(\mathbf{v}) = f^r(\mathbf{w})$, and we deduce that $(f + \lambda \mathrm{Id}_V)(\mathbf{v}) \in \mathrm{Im}(f^r)$. This proves that $\mathrm{Im}(f^r)$ is $(f + \lambda \mathrm{Id}_V)$-invariant.

(e) Fix a positive integer r and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V$, and assume that $\mathbf{v}_i \in \mathrm{Ker}(f^i) \setminus \mathrm{Ker}(f^{i-1})$ for all $i \in \{1, \ldots, r\}$. We must show that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent. Fix scalars $\alpha_1, \ldots, \alpha_r \in \mathbb{F}$ such that

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_r \mathbf{v}_r = \mathbf{0}.$$

We must show that $\alpha_1 = \cdots = \alpha_r = 0$. Suppose otherwise, and let $k \in \{1, \ldots, r\}$ be maximal with the property that $\alpha_k \neq 0$. Then

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k = \mathbf{0}.$$

We apply f^{k-1} to both sides of the equation above, and using the linearity of f^{k-1} and Proposition 4.1.6, we obtain

$$\alpha_1 f^{k-1}(\mathbf{v}_1) + \cdots + \alpha_k f^{k-1}(\mathbf{v}_k) = \mathbf{0}.$$

In view of (b), we have that $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1} \in \mathrm{Ker}(f^{k-1})$, and consequently, $f^{k-1}(\mathbf{v}_1) = \cdots = f^{k-1}(\mathbf{v}_{k-1}) = \mathbf{0}$. Thus, $\alpha_k f^{k-1}(\mathbf{v}_k) = \mathbf{0}$. Since $\alpha_k \neq 0$, it follows that $f^{k-1}(\mathbf{v}_k) = \mathbf{0}$, and consequently, $\mathbf{v}_k \in \mathrm{Ker}(f^{k-1})$, a contradiction.

(f) Fix a non-negative integer integer r, an index $i \in \{0, \ldots, r\}$, and a vector $\mathbf{v} \in V$. Then $f^r(\mathbf{v}) = f^i\big(f^{r-i}(\mathbf{v})\big)$, and consequently, we have the following sequence of equivalences:

$$\begin{aligned}\mathbf{v} \in \mathrm{Ker}(f^r) &\iff f^r(\mathbf{v}) = \mathbf{0} \\ &\iff f^i\big(f^{r-i}(\mathbf{v})\big) = \mathbf{0} \\ &\iff f^{r-i}(\mathbf{v}) \in \mathrm{Ker}(f^i).\end{aligned}$$

(g) Fix a positive integer r and a vector $\mathbf{v} \in \text{Ker}(f^r) \setminus \text{Ker}(f^{r-1})$. By (f), we know that $f^{r-i}(\mathbf{v}) \in \text{Ker}(f^i) \setminus \text{Ker}(f^{i-1})$ for all $i \in \{1, \ldots, r\}$.[70] But now by (e), we have that the set $\{f^{r-1}(\mathbf{v}), f^{r-2}(\mathbf{v}), \ldots, f^2(\mathbf{v}), f(\mathbf{v}), \mathbf{v}\}$ is linearly independent.

(c) Fix a non-negative integer r, and assume that $\text{Ker}(f^r) = \text{Ker}(f^{r+1})$. We must show that $\text{Ker}(f^r) = \text{Ker}(f^{r+1}) = \text{Ker}(f^{r+2}) = \text{Ker}(f^{r+3}) = \ldots$. Clearly, it suffices to show that for all non-negative integers j, we have that $\text{Ker}(f^{r+j}) = \text{Ker}(f^{r+j+1})$. Suppose otherwise. In view of (b), this means that there exists some non-negative integer j such that $\text{Ker}(f^{r+j}) \subsetneqq \text{Ker}(f^{r+j+1})$. Fix some $\mathbf{v} \in \text{Ker}(f^{r+j+1}) \setminus \text{Ker}(f^{r+j})$. Then (f) guarantees that $f^j(\mathbf{v}) \in \text{Ker}(f^{r+1}) \setminus \text{Ker}(f^r)$, contrary to the fact that $\text{Ker}(f^r) = \text{Ker}(f^{r+1})$. □

Proposition 8.6.19. *Let V be a non-trivial, finite-dimensional vector space, and let $f : V \to V$ be a linear function. Then there exists a (unique) non-negative integer p such that*

$$\underbrace{Ker(f^0)}_{=\{\mathbf{0}\}} \subsetneqq Ker(f^1) \subsetneqq Ker(f^2) \subsetneqq \ldots \subsetneqq Ker(f^p) = Ker(f^{p+1}) = \ldots$$

Moreover, for this integer p, we have that $p \leq dim\big(Ker(f^p)\big) \leq dim(V)$ and $V = Ker(f^p) \oplus Im(f^p)$.

Remark: It is possible that $p = 0$. In this case, we simply have that $\{\mathbf{0}\} = \text{Ker}(f^0) = \text{Ker}(f^1) = \text{Ker}(f^2) = \ldots$. In view of Theorem 4.2.4, this happens precisely when our linear function f is one-to-one.

Proof. Set $n := \dim(V)$.[71]

Claim 1. $\text{Ker}(f^n) = \text{Ker}(f^{n+1})$.

Proof of Claim 1. Suppose otherwise. Then by Proposition 8.6.18(b-c), we have that $\text{Ker}(f^0) \subsetneqq \text{Ker}(f^1) \subsetneqq \cdots \subsetneqq \text{Ker}(f^n) \subsetneqq \text{Ker}(f^{n+1})$. For each $i \in \{1, \ldots, n+1\}$, we fix some $\mathbf{v}_i \in \text{Ker}(f^i) \setminus \text{Ker}(f^{i-1})$. By Proposition 8.6.18(e), $\{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}\}$ is linearly independent. But this is impossible since by Theorem 3.2.17(a), all linearly independent sets in the n-dimensional vector space V are of size at most n. ♦

Now, using Claim 1, we fix the smallest non-negative integer p such that $\text{Ker}(f^p) = \text{Ker}(f^{p+1})$. By Proposition 8.6.18(b-c), we have that

$$\underbrace{\text{Ker}(f^0)}_{=\{\mathbf{0}\}} \subsetneqq \text{Ker}(f^1) \subsetneqq \text{Ker}(f^2) \subsetneqq \ldots \subsetneqq \text{Ker}(f^p) = \text{Ker}(f^{p+1}) = \ldots.$$

[70] Let us check this in full detail. Fix $i \in \{1, \ldots, r\}$. Since $\mathbf{v} \in \text{Ker}(f^r)$, part (f) applied to r, i, and $\mathbf{v}$ implies that $f^{r-i}(\mathbf{v}) \in \text{Ker}(f^i)$. On the other hand, since $\mathbf{v} \notin \text{Ker}(f^{r-1})$, part (f) applied to $r-1$, $i-1$, and $\mathbf{v}$, implies that $f^{r-i}(\mathbf{v}) = f^{(r-1)-(i-1)}(\mathbf{v}) \notin \text{Ker}(f^{i-1})$.

[71] Since V is non-trivial and finite-dimensional, we have that n is a positive integer.

Claim 2. $p \leq \dim\big(\mathrm{Ker}(f^p)\big) \leq n$.

Proof of Claim 2. Since $\mathrm{Ker}(f^p)$ is a subspace of the n-dimensional subspace V, Theorem 3.2.21(b) guarantees that $\dim\big(\mathrm{Ker}(f^p)\big) \leq n$. It remains to show that $p \leq \dim\big(\mathrm{Ker}(f^p)\big)$. If $p = 0$, then this is immediate. So, we may assume that $p \geq 1$. We now proceed similarly as in Claim 1. By the choice of p, we have that

$$\underbrace{\mathrm{Ker}(f^0)}_{=\{\mathbf{0}\}} \subsetneq \mathrm{Ker}(f^1) \subsetneq \mathrm{Ker}(f^2) \subsetneq \ldots \subsetneq \mathrm{Ker}(f^p).$$

For all $i \in \{1, \ldots, p\}$, we fix $\mathbf{v}_i \in \mathrm{Ker}(f^i) \setminus \mathrm{Ker}(f^{i-1})$. By Proposition 8.6.18(e), vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are linearly independent, and obviously, they all belong to $\mathrm{Ker}(f^p)$. By Theorem 3.2.17(a), all linearly independent sets in $\mathrm{Ker}(f^p)$ are of size at most $\dim\big(\mathrm{Ker}(f^p)\big)$, and we deduce that $p \leq \dim\big(\mathrm{Ker}(f^p)\big)$. ♦

It remains to show that $V = \mathrm{Ker}(f^p) \oplus \mathrm{Im}(f^p)$, that is, that $\mathrm{Ker}(f^p) \cap \mathrm{Im}(f^p) = \{\mathbf{0}\}$ and $V = \mathrm{Ker}(f^p) + \mathrm{Im}(f^p)$.

We first show that $\mathrm{Ker}(f^p) \cap \mathrm{Im}(f^p) = \{\mathbf{0}\}$. Since both $\mathrm{Ker}(f^p)$ and $\mathrm{Im}(f^p)$ are subspaces of V, they both contain $\mathbf{0}$, and so $\mathbf{0} \in \mathrm{Ker}(f^p) \cap \mathrm{Im}(f^p)$. Now, fix any $\mathbf{v} \in \mathrm{Ker}(f^p) \cap \mathrm{Im}(f^p)$; we must show that $\mathbf{v} = \mathbf{0}$. Since $\mathbf{v} \in \mathrm{Im}(f^p)$, we know that there exists some $\mathbf{u} \in V$ such that $f^p(\mathbf{u}) = \mathbf{v}$. We now apply f^p to both sides of the equation, and we obtain $f^{2p}(\mathbf{u}) = f^p(\mathbf{v}) = \mathbf{0}$, where the last equality follows from the fact that $\mathbf{v} \in \mathrm{Ker}(f^p)$. But now $\mathbf{u} \in \mathrm{Ker}(f^{2p}) = \mathrm{Ker}(f^p)$,[72] and we deduce that $f^p(\mathbf{u}) = \mathbf{0}$, i.e. $\mathbf{v} = \mathbf{0}$. This proves that $\mathrm{Ker}(f^p) \cap \mathrm{Im}(f^p) = \{\mathbf{0}\}$.

It remains to show that $V = \mathrm{Ker}(f^p) + \mathrm{Im}(f^p)$. We compute:

$$\begin{aligned}\dim\Big(\mathrm{Ker}(f^p) + \mathrm{Im}(f^p)\Big) &\overset{(*)}{=} \dim\Big(\mathrm{Ker}(f^p) + \mathrm{Im}(f^p)\Big) + \dim\Big(\mathrm{Ker}(f^p) \cap \mathrm{Im}(f^p)\Big) \\ &\overset{(**)}{=} \dim\big(\mathrm{Ker}(f^p)\big) + \underbrace{\dim\big(\mathrm{Im}(f^p)\big)}_{\overset{(***)}{=}\mathrm{rank}(f^p)} \overset{(****)}{=} \dim(V),\end{aligned}$$

where (*) follows from the fact that $\mathrm{Ker}(f^p) \cap \mathrm{Im}(f^p) = \{\mathbf{0}\}$ (proven above), (**) follows from Theorem 3.2.23, (***) follows from the definition of $\mathrm{rank}(f^p)$, and (****) follows from the rank-nullity theorem. Since V is finite-dimensional, Theorem 3.2.21(c) now guarantees that $\mathrm{Ker}(f^p) + \mathrm{Im}(f^p) = V$, and we are done. □

Generalized eigenvectors and generalized eigenspaces. Suppose that V is a vector space over a field $\mathbb{F}$, and that $f : V \to V$ is a linear function, and that $\lambda \in \mathbb{F}$ is a scalar. Then Propositions 4.1.7 and 8.6.18(a) together guarantee that the function

[72] We saw above that $\mathrm{Ker}(f^p) = \mathrm{Ker}(f^{p+1}) = \mathrm{Ker}(f^{p+2}) = \ldots$. In particular, $\mathrm{Ker}(f^p) = \mathrm{Ker}(f^{2p})$.

$(f - \lambda \mathrm{Id}_V)^r : V \to V$ is linear for all linear non-negative integers r, and consequently, we can define the set

$$\begin{aligned} G_\lambda(f) &:= \big\{\mathbf{v} \in V \mid \exists r \in \mathbb{N}_0 \text{ s.t. } (f - \lambda \mathrm{Id}_V)^r(\mathbf{v}) = \mathbf{0}\big\} \\ &= \bigcup_{r=0}^{\infty} \mathrm{Ker}\Big(\big(f - \lambda \mathrm{Id}_V\big)^r\Big). \end{aligned}$$

If λ is an eigenvalue of f, then the set $G_\lambda(f)$ is called the *generalized eigenspace* of f associated with the eigenvalue λ.[73]

Proposition 8.6.20. *Let V be a vector space over a field $\mathbb{F}$, and let $f : V \to V$ be a linear function. Then all the following hold:*

(a) for all scalars $\lambda \in \mathbb{F}$, $G_\lambda(f)$ is an f-invariant subspace of V, and moreover, $E_\lambda(f)$ is a subspace of $G_\lambda(f)$;[74]

(b) for all scalars $\lambda \in \mathbb{F}$, the subspace $G_\lambda(f)$ is non-trivial if and only if λ is an eigenvalue of f;

(c) for all distinct $\lambda_1, \lambda_2 \in \mathbb{F}$, we have that $G_{\lambda_1}(f) \cap G_{\lambda_2}(f) = \{\mathbf{0}\}$.

Proof. To simplify notation, for each scalar $\lambda \in \mathbb{F}$, we set $G_\lambda := G_\lambda(f)$ and $E_\lambda := E_\lambda(f)$. We begin with a sequence of technical claims.

Claim 1. For all scalars $\lambda \in \mathbb{F}$, G_λ is a subspace of V.

Proof of Claim 1. Fix a scalar $\lambda \in \mathbb{F}$, and to simplify notation, set $g := f - \lambda \mathrm{Id}_V$, so that $G_\lambda = \bigcup_{r=0}^{\infty} \mathrm{Ker}(g^r)$. We must show that G_λ is a subspace of V. It suffices to verify that G_λ satisfies the three conditions from Theorem 3.1.7.

First, we note that $g^0(\mathbf{0}) = \mathrm{Id}_V(\mathbf{0}) = \mathbf{0}$, and we deduce that $\mathbf{0} \in \mathrm{Ker}(g^0) \subseteq G_\lambda$.

Next, fix $\mathbf{u}, \mathbf{v} \in G_\lambda$. Then there exist non-negative integers p and q such that $\mathbf{u} \in \mathrm{Ker}(g^p)$ and $\mathbf{v} \in \mathrm{Ker}(g^q)$. By symmetry, we may assume that $p \leq q$. Then by Proposition 8.6.18(b), we have that $\mathrm{Ker}(g^p) \subseteq \mathrm{Ker}(g^q)$, and we deduce that $\mathbf{u}, \mathbf{v} \in \mathrm{Ker}(g^q)$. Since $\mathrm{Ker}(g^q)$ is a subspace of V,[75] we deduce that $\mathbf{u} + \mathbf{v} \in \mathrm{Ker}(g^q) \subseteq G_\lambda$.

Finally, fix $\mathbf{u} \in G_\lambda$ and $\alpha \in \mathbb{F}$. Then there exists a non-negative integer p such that $\mathbf{u} \in \mathrm{Ker}(g^p)$. We once again use the fact that $\mathrm{Ker}(g^p)$ is a subspace of V, and we deduce that $\alpha\mathbf{u} \in \mathrm{Ker}(g^p) \subseteq G_\lambda$.

We have now verified that G_λ satisfies the three conditions from Theorem 3.1.7, and we deduce that G_λ is indeed a subspace of V. ♦

[73] If λ is not an eigenvalue of f, the set $G_\lambda(f)$ is still defined, but we do not refer to it as an eigenspace.

[74] Recall that $E_\lambda(f) = \{\mathbf{v} \in V \mid f(\mathbf{v}) = \lambda\mathbf{v}\}$. If λ is an eigenvalue of f, then we referred to $E_\lambda(f)$ as the eigenspace of f associated with λ.

[75] This follows from Theorem 4.2.3(d) applied to the linear function g^p.

Claim 2. For all scalars $\lambda \in \mathbb{F}$, E_λ is a subspace of G_λ.

Proof of Claim 2. Fix a scalar $\lambda \in \mathbb{F}$. By Proposition 8.1.4, E_λ is a subspace of V, and by Claim 1, G_λ is a subspace of V. So, it suffices to show that $E_\lambda \subseteq G_\lambda$. So, fix any $\mathbf{v} \in E_\lambda$, so that $f(\mathbf{v}) = \lambda\mathbf{v}$. But then $(f - \lambda\mathrm{Id}_V)(\mathbf{v}) = \mathbf{0}$, and it follows that $\mathbf{v} \in \mathrm{Ker}(f - \lambda\mathrm{Id}_V) \subseteq G_\lambda$. ♦

Claim 3. For all scalars $\lambda \in \mathbb{F}$, positive integers p, and vectors $\mathbf{v} \in V$, if $\mathbf{v} \in \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^p\big) \setminus \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^{p-1}\big)$, then the scalar λ is an eigenvalue of f, and $(f - \lambda\mathrm{Id}_V)^{p-1}(\mathbf{v})$ is an eigenvector of f associated with the eigenvalue λ.

Proof of Claim 3. Fix a scalar $\lambda \in \mathbb{F}$, a positive integer p, and a vector $\mathbf{v} \in V$. Assume that $\mathbf{v} \in \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^p\big) \setminus \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^{p-1}\big)$. We must show that $\mathbf{w} := (f - \lambda\mathrm{Id}_V)^{p-1}(\mathbf{v})$ is a non-zero vector in V that satisfies $f(\mathbf{w}) = \lambda\mathbf{w}$.

Set $g := f - \lambda\mathrm{Id}_V$, so that $\mathbf{v} \in \mathrm{Ker}(g^p) \setminus \mathrm{Ker}(g^{p-1})$ and $\mathbf{w} = g^{p-1}(\mathbf{v})$. By Proposition 8.6.18(f), we have that $\mathbf{w} \in \mathrm{Ker}(g^1) \setminus \mathrm{Ker}(g^0) = \mathrm{Ker}(g) \setminus \{\mathbf{0}\}$. In particular, we have that $g(\mathbf{w}) = \mathbf{0}$, which means that $(f - \lambda\mathrm{Id}_V)(\mathbf{w}) = \mathbf{0}$, i.e. $f(\mathbf{w}) = \lambda\mathbf{w}$. Since $\mathbf{w} \neq \mathbf{0}$, we deduce that λ is indeed an eigenvalue of f, and that $\mathbf{w}$ is an eigenvector of f associated with the eigenvalue λ. ♦

Claim 4. For all scalars $\lambda, \lambda' \in \mathbb{F}$, G_λ is an $(f - \lambda'\mathrm{Id}_V)$-invariant subspace of V.

Proof of Claim 4. Fix scalars $\lambda, \lambda' \in \mathbb{F}$. By Claim 1, G_λ is a subspace of V, and we just need to show that G_λ is an $(f - \lambda'\mathrm{Id}_V)$-invariant. Fix a vector $\mathbf{v} \in G_\lambda$; we must show that $(f - \lambda'\mathrm{Id}_V)(\mathbf{v}) \in G_\lambda$. Since $\mathbf{v} \in G_\lambda$, we know that there exists a non-negative integer r such that $\mathbf{v} \in \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^r\big)$. By Proposition 8.6.18(d), $\mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^r\big)$ is $(f - \lambda'\mathrm{Id}_V)$-invariant,[76] and so $(f - \lambda'\mathrm{Id}_V)(\mathbf{v}) \in \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^r\big) \subseteq G_\lambda$. ♦

We are now ready to prove (a) and (b). Fix a scalar $\lambda \in \mathbb{F}$. By Claims 1 and 4, G_λ is an f-invariant subspace of V.[77] Moreover, by Claim 2, E_λ is a subspace of G_λ. This proves (a). For (b), suppose first that G_λ is non-trivial, and fix some $\mathbf{v} \in G_\lambda \setminus \{\mathbf{0}\}$. Let p be the smallest positive integer such that $\mathbf{v} \in \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^p\big)$. Then $\mathbf{v} \in \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^p\big) \setminus \mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^{p-1}\big)$, and so by Claim 3, λ is an eigenvalue of f. Suppose, conversely, that λ is an eigenvalue of f. Then Proposition 8.1.4 guarantees that E_λ is a non-trivial subspace of V, and so by Claim 2, G_λ is also non-trivial. This proves (b).

It remains to prove (c). Fix distinct scalars $\lambda_1, \lambda_2 \in \mathbb{F}$. We must show that $G_{\lambda_1} \cap G_{\lambda_2} = \{\mathbf{0}\}$. By Claim 1, G_{λ_1} and G_{λ_2} are both subspaces of V, and so they

[76] Indeed, we set $\widetilde{f} := f - \lambda\mathrm{Id}_V$ and $\widetilde{\lambda} := \lambda - \lambda'$. By Proposition 8.6.18(d), we have that $\mathrm{Ker}(\widetilde{f}^r)$ is $(\widetilde{f} + \widetilde{\lambda}\mathrm{Id}_V)$-invariant, i.e. $\mathrm{Ker}\big((f - \lambda\mathrm{Id}_V)^r\big)$ is $(f - \lambda'\mathrm{Id}_V)$-invariant.

[77] We apply Claim 4 to $\lambda' := 0$.

both contain $\mathbf{0}$, i.e. $\mathbf{0} \in G_{\lambda_1} \cap G_{\lambda_2}$. Now, fix any $\mathbf{u} \in G_{\lambda_1} \cap G_{\lambda_2}$; we must show that $\mathbf{u} \neq \mathbf{0}$. Fix the smallest non-negative integer p such that $\mathbf{u} \in \mathrm{Ker}\big((f - \lambda_1 \mathrm{Id}_V)^p\big)$. If $p = 0$, then $\mathbf{u} = \mathbf{0}$,[78] and we are done. We may therefore assume that $p \geq 1$, so that $\mathbf{u} \in \mathrm{Ker}\big((f - \lambda_1 \mathrm{Id}_V)^p\big) \setminus \mathrm{Ker}\big((f - \lambda_1 \mathrm{Id}_V)^{p-1}\big)$. Then by Claim 3, λ_1 is an eigenvalue of f, and $\mathbf{w} := (f - \lambda_1 \mathrm{Id}_V)^{p-1}(\mathbf{u})$ is an eigenvector of f associated with the eigenvalue λ_1. On the other hand, by Claim 4, G_{λ_2} is $(f - \lambda_1 \mathrm{Id}_V)$-invariant. So, since $\mathbf{u} \in G_{\lambda_2}$, we have that $\mathbf{w} \in G_{\lambda_2}$. We will derive a contradiction by showing that $\mathbf{w} = \mathbf{0}$, contrary to the fact that $\mathbf{w}$ is an eigenvector of f.

> **Claim 5.** For all non-negative integers k, we have that $(f - \lambda_2 \mathrm{Id}_V)^k(\mathbf{w}) = (\lambda_1 - \lambda_2)^k \mathbf{w}$.

Proof of Claim 5. We proceed by induction on k. For the base case, we compute:

$$(f - \lambda_2 \mathrm{Id}_V)^0(\mathbf{w}) \;=\; \mathrm{Id}_V(\mathbf{w}) \;=\; \mathbf{w} \;=\; (\lambda_1 - \lambda_2)^0(\mathbf{w}).$$

Now, fix a non-negative integer k, and assume inductively that $(f - \lambda_2 \mathrm{Id}_V)^k(\mathbf{w}) = (\lambda_1 - \lambda_2)^k \mathbf{w}$. We then compute:

$$\begin{aligned}
(f - \lambda_2 \mathrm{Id}_V)^{k+1}(\mathbf{w}) &= (f - \lambda_2 \mathrm{Id}_V)^k\big(f(\mathbf{w}) - \lambda_2 \mathrm{Id}_V(\mathbf{w})\big) \\
&\overset{(*)}{=} (f - \lambda_2 \mathrm{Id}_V)^k(\lambda_1 \mathbf{w} - \lambda_2 \mathbf{w}) \\
&= (f - \lambda_2 \mathrm{Id}_V)^k\big((\lambda_1 - \lambda_2)\mathbf{w}\big) \\
&\overset{(**)}{=} (\lambda_1 - \lambda_2)(f - \lambda_2 \mathrm{Id}_V)^k(\mathbf{w}) \\
&\overset{(***)}{=} (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_2)^k \mathbf{w} \\
&= (\lambda_1 - \lambda_2)^{k+1} \mathbf{w},
\end{aligned}$$

where (*) follows from the fact that $\mathbf{w}$ is an eigenvector of f associated with the eigenvalue λ_1 (and so $f(\mathbf{w}) = \lambda_1 \mathbf{w}$), (**) follows from the linearity of $(f - \lambda_2 \mathrm{Id}_V)^k$, and (***) follows from the induction hypothesis. This completes the induction. ♦

Now, suppose that $\mathbf{w} \in G_{\lambda_2}$. Then there exists a non-negative integer k such that $\mathbf{w} \in \mathrm{Ker}\big((f - \lambda_2 \mathrm{Id}_V)^k\big)$. But now

$$\mathbf{0} \;\overset{(*)}{=}\; (f - \lambda_2 \mathrm{Id}_V)^k(\mathbf{w}) \;\overset{(**)}{=}\; (\lambda_1 - \lambda_2)^k(\mathbf{w}),$$

where (*) follows from the fact that $\mathbf{w} \in \mathrm{Ker}\big((f - \lambda_2 \mathrm{Id}_V)^k\big)$, and (**) follows from

[78] This is because $\mathrm{Ker}\big((f - \lambda_1 \mathrm{Id}_V)^0\big) = \mathrm{Ker}(\mathrm{Id}_V) = \{\mathbf{0}\}$.

Claim 5. Since $\lambda_1 \neq \lambda_2$, it follows that $\mathbf{w} = \mathbf{0}$, contrary to the fact that $\mathbf{w}$ is an eigenvector of f. This completes the proof of (c). $\square$

We now introduce some terminology. Suppose that V is a vector space over a field $\mathbb{F}$, that $f : V \to V$ is a linear function, and that $\lambda \in \mathbb{F}$ is an eigenvalue of f. We then define the following:

- a *generalized eigenvector* of f associated with λ is any **non-zero** vector $\mathbf{v} \in V$ such that for some positive integer r, we have that $(f - \lambda \mathrm{Id}_V)^r(\mathbf{v}) = \mathbf{0}$;[79]
- the *rank* of a generalized eigenvector $\mathbf{v}$ of f associated with λ is the smallest positive integer r such that $(f - \lambda \mathrm{Id}_V)^r(\mathbf{v}) = \mathbf{0}$;[80]
- for a generalized eigenvector $\mathbf{v}$ of f associated with λ and of rank r, the *Jordan chain started by* $\mathbf{v}$ is the set

$$\big\{(f - \lambda \mathrm{Id}_V)^{r-1}(\mathbf{v}), \ldots, (f - \lambda \mathrm{Id}_V)^2(\mathbf{v}), (f - \lambda \mathrm{Id}_V)(\mathbf{v}), \mathbf{v}\big\}.$$

Note that the elements of the generalized eigenspace $G_\lambda(f)$ are precisely the generalized eigenvectors of f associated with λ. Moreover, by Proposition 8.1.4, we have that $E_\lambda(f) = \mathrm{Ker}(f - \lambda \mathrm{Id}_V)$. So, eigenvectors of f associated with λ are precisely the the generalized eigenvectors of rank 1 of f associated with λ.

Remark: In view of Proposition 8.6.20(c), any generalized eigenvector of a linear function $f : V \to V$ (where V is a vector space over a field $\mathbb{F}$) is associated with exactly one eigenvalue of f.

Remark: By Proposition 8.6.18(g), every Jordan chain is a linearly independent set. We note that the basis $\mathcal{B}$ from the proof of Theorem 8.6.4 will turn out to be the union of pairwise disjoint Jordan chains (associated with various eigenvalues).

Theorem 8.6.21. *Let V be a non-trivial, finite-dimensional vector space over an* ***algebraically closed field*** *$\mathbb{F}$, let $f : V \to V$ be a linear function, and let*

$$\big\{\underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{m_k}\big\}$$

be the spectrum of f, where $\lambda_1, \ldots, \lambda_k$ are pairwise distinct eigenvalues of f and $m_1, \ldots, m_k$ are positive integers. Then $V = G_{\lambda_1}(f) \oplus \cdots \oplus G_{\lambda_k}(f)$. Moreover, for each index $i \in \{1, \ldots, k\}$, all the following hold:

- *$G_{\lambda_i}(f)$ is an f-invariant subspace of V of dimension m_i,*

[79] The reason we specify that r is positive (rather than merely non-negative) is simply that $(f - \lambda \mathrm{Id}_V)^0 = \mathrm{Id}_V$, and so there are no non-zero vectors $\mathbf{v} \in V$ such that $(f - \lambda \mathrm{Id}_V)^0(\mathbf{v}) = \mathbf{0}$.

[80] Note that this means that $\mathbf{v} \in \mathrm{Ker}(f^r) \setminus \mathrm{Ker}(f^{r-1})$.

- *there exists a positive integer $p_i \leq m_i$ such that*

$$\underbrace{Ker\Big((f-\lambda_i Id_V)^0\Big)}_{=\{\mathbf{0}\}} \subsetneq Ker\Big((f-\lambda_i Id_V)^1\Big) \subsetneq Ker\Big((f-\lambda_i Id_V)^2\Big) \subsetneq \ldots$$

$$\ldots \subsetneq Ker\Big((f-\lambda_i Id_V)^{p_i}\Big) = Ker\Big((f-\lambda_i Id_V)^{p_i+1}\Big) = \ldots,$$

and in particular, $G_{\lambda_i}(f) = Ker\Big((f-\lambda_i Id_V)^{p_i}\Big)$.

- *λ_i is the only eigenvalue of $f|_{G_{\lambda_i}(f)}$, and the algebraic multiplicity of the eigenvalue λ_i of $f|_{G_{\lambda_i}(f)}$ is m_i.*

Remark: Since the field $\mathbb{F}$ is algebraically closed, we know that $p_f(\lambda)$ can be factored into linear terms, and so $p_f(\lambda) = (\lambda-\lambda_1)^{m_1}\ldots(\lambda-\lambda_k)^{m_k}$ and $m_1+\cdots+m_k = \dim(V)$.

Proof. Set $n := \dim(V)$. To simplify notation, for all indices $i \in \{1,\ldots,k\}$, we set $G_i := G_{\lambda_i}(f)$ and $g_i := f - \lambda_i \mathrm{Id}_V$,[81] so that $G_i = \bigcup_{i=0}^{\infty} \mathrm{Ker}\big(g_i^r\big)$. Now, by Proposition 8.6.19 applied to g_i (for $i \in \{1,\ldots,k\}$), we know that there exists a (unique) non-negative integer p_i such that

$$\mathrm{Ker}(g_i^0) \subsetneq \mathrm{Ker}(g_i^1) \subsetneq \mathrm{Ker}(g_i^2) \subsetneq \ldots \subsetneq \mathrm{Ker}(g_i^{p_i}) = \mathrm{Ker}(g_i^{p_i+1}) = \ldots,$$

and consequently, $G_i = \mathrm{Ker}(g_i^{p_i})$; but in fact, since $G_i \neq \{\mathbf{0}\}$ (because λ_i is an eigenvalue of f, and G_i contains all eigenvectors of f associated with λ_i), we see that $p_i \geq 1$.

> **Claim 1.** For all $i \in \{1,\ldots,k\}$, both G_i and $\mathrm{Im}(g_i^{p_i})$ are f-invariant subspaces of V, and moreover, $V = G_i \oplus \mathrm{Im}(g_i^{p_i})$.

Proof of Claim 1. Fix an index $i \in \{1,\ldots,k\}$. By Proposition 8.6.19 applied g_i, we know that $V = \mathrm{Ker}(g_i^{p_i}) \oplus \mathrm{Im}(g_i^{p_i}) = G_i \oplus \mathrm{Im}(g_i^{p_i})$. The fact that $G_i = \mathrm{Ker}(g_i^{p_i})$ and $\mathrm{Im}(g_i^{p_i})$ are f-invariant now follows from Proposition 8.6.18(d) applied to the linear function g_i and the scalar λ_i.[82] ♦

To simplify notation, for each index $i \in \{1,\ldots,k\}$, we set $f_i := f|_{G_i}$. The fact that the functions $f_1,\ldots,f_k$ are well defined follows from the fact that, by Claim 1, $G_1,\ldots,G_k$ are f-invariant subspaces of V. Moreover, since f is linear, so are $f_1,\ldots,f_k$.

> **Claim 2.** For all indices $i \in \{1,\ldots,k\}$, we have that $p_{f_i}(\lambda) = (\lambda-\lambda_i)^{m_i}$, $\dim(G_i) = m_i$, and $p_i \leq m_i$.

[81] Since f and Id_V are linear, Proposition 4.1.7 guarantees that g_i is linear as well.

[82] Indeed, by Proposition 8.6.18(d), both $\mathrm{Ker}(g_i^{p_i})$ and $\mathrm{Im}(g_i^{p_i})$ are $(g_i+\lambda_i\mathrm{Id}_V)$-invariant, and by construction, $g_i + \lambda_i \mathrm{Id}_V = f$.

Proof of Claim 2. Fix an index $i \in \{1, \ldots, k\}$. By the definition of the characteristic polynomial, we know that the degree of $p_{f_i}(\lambda)$ is equal to $\dim(G_i)$, and by Proposition 8.6.19 applied to g_i, we know that $p_i \leq \dim(G_i)$. So, it suffices to show that $p_{f_i}(\lambda) = (\lambda - \lambda_i)^{m_i}$.

By Claim 1, $\mathrm{Im}(g_i^{p_i})$ is an f-invariant subspace of V, and so $h_i := f|_{\mathrm{Im}(g_i^{p_i})}$ is well defined. Further, by Claim 1 and by Proposition 8.6.11, we know that

$$p_f(\lambda) \;\; = \;\; p_{f_i}(\lambda)\, p_{h_i}(\lambda).$$

Since $p_f(\lambda) = (\lambda - \lambda_1)^{m_1} \ldots (\lambda - \lambda_k)^{m_k}$, it is now enough to show that f_i has no eigenvalues other than λ_i, and that λ_i is not an eigenvalue of h_i. This essentially follows from Proposition 8.6.20, but we give the full details below, as follows.

We first show that f_i has no eigenvalues other than λ_i. Suppose that $\mathbf{u} \in G_i \setminus \{\mathbf{0}\}$ is an eigenvector of f_i associated with an eigenvalue λ_0 of f_i. Then $\mathbf{u}$ is an eigenvector of f associated with λ_0,[83] and so by Proposition 8.6.20(a), $\mathbf{u} \in G_{\lambda_0}$, and consequently, $\mathbf{u} \in G_{\lambda_0}(f) \cap G_{\lambda_i}(f)$. Since $\mathbf{u} \neq \mathbf{0}$, Proposition 8.6.20(c) guarantees that $\lambda_0 = \lambda_i$. This proves that f_i indeed has no eigenvalues other than λ_i.

It remains to show that λ_i is not an eigenvalue of h_i. Fix any $\mathbf{u} \in \mathrm{Im}(g_i^{p_i})$ such that $h_i(\mathbf{u}) = \lambda_i \mathbf{u}$; we must show that $\mathbf{u} = \mathbf{0}$. Now, we have that $f(\mathbf{u}) = h_i(\mathbf{u}) = \lambda_i \mathbf{u}$, and so $\mathbf{u} \in \mathrm{Ker}(f - \lambda_i \mathrm{Id}_V) \subseteq G_i$. Thus, $\mathbf{u} \in G_i \cap \mathrm{Im}(g_i^{p_i})$. But by Claim 1, we have that $V = G_i \oplus \mathrm{Im}(g_i^{p_i})$, which in particular means that $G_i \cap \mathrm{Im}(g_i^{p_i}) = \{\mathbf{0}\}$. Thus, $\mathbf{u} = \mathbf{0}$. This proves that λ_i is not an eigenvalue of h_i. ♦

In view of Claims 1 and 2, it now just remains to show that $V = G_1 \oplus \cdots \oplus G_k$. Set $U := G_1 + \cdots + G_k$. In view of Proposition 8.6.20(c), it follows that $U = G_1 \oplus \cdots \oplus G_k$. But now

$$\dim(U) \;\overset{(*)}{=}\; \dim(G_1) + \cdots + \dim(G_k) \;\overset{(**)}{=}\; m_1 + \cdots + m_k \;\; = \;\; n,$$

where (*) follows from Theorem 3.2.23, and (**) follows from Claim 2. Thus, U is an n-dimensional subspace of the n-dimensional vector space V, and so by Theorem 3.2.21(c), we have that $U = V$, that is, $V = G_1 \oplus \cdots \oplus G_k$. This completes the argument. □

Linear independence over a subspace. Given a vector space V over a field $\mathbb{F}$, a subspace U of V, and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$ $(t \geq 0)$, we say that vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are *linearly independent over* U, or that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_t\}$ is *linearly independent over* U, provided that the following two conditions are satisfied:

- vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent;
- $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U = \{\mathbf{0}\}$.

[83] Indeed, $\mathbf{u}$ is a non-zero vector of V that satisfies $f(\mathbf{u}) = f_i(\mathbf{u}) \overset{(*)}{=} \lambda_0 \mathbf{u}$, where (*) follows from the fact that $\mathbf{u}$ is an eigenvector of f_i associated with the eigenvalue λ_0.

Proposition 8.6.22. *Let V be a vector space over a field $\mathbb{F}$, let U be a subspace of V, and let $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$. Assume that vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over U. Then $\mathbf{v}_1, \ldots, \mathbf{v}_t \notin U$.*

Proof. Suppose otherwise, and fix an index $i \in \{1, \ldots, t\}$ such that $\mathbf{v}_i \in U$. Then $\mathbf{v}_i \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U = \{\mathbf{0}\}$, and consequently, $\mathbf{v}_i = \mathbf{0}$. But this is impossible since vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent. □

Remark: The converse of Proposition 8.6.22 is false, even if we additionally assume that $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent. For example, consider the subspace

$$U \;:=\; \left\{ \begin{bmatrix} x \\ x \end{bmatrix} \mid x \in \mathbb{R} \right\}$$

of $\mathbb{R}^2$, and consider the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ of $\mathbb{R}^2$. Then vectors $\mathbf{e}_1, \mathbf{e}_2$ are linearly independent and do not belong to U, but $\mathrm{Span}(\mathbf{e}_1, \mathbf{e}_2) \cap U = \mathbb{R}^2 \cap U = U \neq \{\mathbf{0}\}$, and so $\mathbf{e}_1, \mathbf{e}_2$ are not linearly independent over U.

Proposition 8.6.23. *Let V be a vector space over a field $\mathbb{F}$, let U be a finite-dimensional subspace of V, and let $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$. Set $m := \dim(U)$. Then the following are equivalent:*

(a) vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over U;

(b) for all scalars $\alpha_1, \ldots, \alpha_t \in \mathbb{F}$, if $\alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t \in U$, then $\alpha_1 = \cdots = \alpha_t = 0$;

(c) for all bases $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ of U, vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m, \mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent;

(d) there exists a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ of U such that vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m, \mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent.

Remark: Some texts take (b) as the **definition** of linear independence over U. By Proposition 8.6.23, and in particular, by the equivalence of (a) and (b), this alternative definition is equivalent to our definition.

Proof. We will prove the implications "(a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) $\Longrightarrow$ (a)." Since U is m-dimensional and therefore has at least one basis of size m, the implication "(c) $\Longrightarrow$ (d)" is immediate. We must prove the remaining three implications.

First, we assume (a), and we prove (b). Fix scalars $\alpha_1, \ldots, \alpha_t \in \mathbb{F}$, and assume that $\alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t \in U$. Then $\alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U$. But by (a), we have that $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U = \{\mathbf{0}\}$, and we deduce that $\alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t = \mathbf{0}$. But once again by (a), vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent. So, $\alpha_1 = \cdots = \alpha_t = 0$. This proves (b).

Next, we assume (b), and we prove (c). Fix any basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ of U. We must show that vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m, \mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent. Fix scalars $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_t \in \mathbb{F}$ such that $\alpha_1\mathbf{b}_1 + \cdots + \alpha_m\mathbf{b}_m + \beta_1\mathbf{v}_1 + \cdots + \beta_t\mathbf{v}_t = \mathbf{0}$. Then $\beta_1\mathbf{v}_1 + \cdots + \beta_t\mathbf{v}_t = -\alpha_1\mathbf{b}_1 - \cdots - \alpha_m\mathbf{b}_m \in \operatorname{Span}(\mathbf{b}_1, \ldots, \mathbf{b}_m) = U$, and so by (b), we have that $\beta_1 = \cdots = \beta_t = 0$. Consequently, $\alpha_1\mathbf{b}_1 + \cdots + \alpha_m\mathbf{b}_m = \mathbf{0}$, and so since $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ is linearly independent (because it is a basis of U), we have that $\alpha_1 = \cdots = \alpha_m = 0$. We have now shown that $\alpha_1 = \cdots = \alpha_t = \beta_1 = \cdots = \beta_m = 0$, and (c) follows.

Finally, we assume (d), and we prove (a). Using (d), we fix a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ of U such that vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m, \mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent. It is clear that $\{\mathbf{v}_1, \ldots, \mathbf{v}_t\}$ is linearly independent, and we just need to show that $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U = \{\mathbf{0}\}$. Since $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t)$ and U are both subspaces of V, they both contain $\mathbf{0}$, and so $\mathbf{0} \in \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U$. Now, fix any $\mathbf{u} \in \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U$; we must show that $\mathbf{u} = \mathbf{0}$. Since $\mathbf{u} \in \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t)$, we know that there exist scalars $\alpha_1, \ldots, \alpha_t \in \mathbb{F}$ such that $\mathbf{u} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t$. On the other hand, since $\mathbf{u} \in U$, and since $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ is a basis of U, we know that there exist scalars $\beta_1, \ldots, \beta_m \in \mathbb{F}$ such that $\mathbf{u} = \beta_1\mathbf{b}_1 + \cdots + \beta_m\mathbf{b}_m$. But now $\alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t = \beta_1\mathbf{b}_1 + \cdots + \beta_m\mathbf{b}_m$, and consequently, $\beta_1\mathbf{b}_1 + \cdots + \beta_m\mathbf{b}_m - \alpha_1\mathbf{v}_1 - \cdots - \alpha_t\mathbf{v}_t = \mathbf{0}$. Since $\{\mathbf{b}_1, \ldots, \mathbf{b}_m, \mathbf{v}_1, \ldots, \mathbf{v}_t\}$ is linearly independent, we deduce that $\beta_1 = \cdots = \beta_m = -\alpha_1 = \cdots = -\alpha_t = 0$. So, $\mathbf{u} = \mathbf{0}$. This proves that $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_t) \cap U = \{\mathbf{0}\}$, and (a) follows. □

Proposition 8.6.24. *Let V be a vector space over a field $\mathbb{F}$, let r be a positive integer, and let $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$ ($t \geq 0$). If vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over $\operatorname{Ker}(f^r)$, then vectors $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_t)$ are linearly independent over $\operatorname{Ker}(f^{r-1})$.*

Proof. Assume that vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over $\operatorname{Ker}(f^r)$. We must show that vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over $\operatorname{Ker}(f^{r-1})$. We use the equivalence of (a) and (b) from Proposition 8.6.23. Fix scalars $\alpha_1, \ldots, \alpha_t \in \mathbb{F}$ such that $\mathbf{v} := \alpha_1 f(\mathbf{v}_1) + \cdots + \alpha_t f(\mathbf{v}_t) \in \operatorname{Ker}(f^{r-1})$; we must show that $\alpha_1 = \cdots = \alpha_t = 0$.[84] By the linearity of f, we have that $\mathbf{v} = f(\alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t)$. Since $\mathbf{v} \in \operatorname{Ker}(f^{r-1})$, Proposition 8.6.18(f) guarantees that $\alpha_1\mathbf{v}_1 + \cdots + \alpha_t\mathbf{v}_t \in \operatorname{Ker}(f^r)$. But since vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over $\operatorname{Ker}(f^r)$, Proposition 8.6.23 guarantees that $\alpha_1 = \cdots = \alpha_t = 0$. □

Nilpotent linear functions. A linear function $f : V \to V$ (where V is a vector space over a field $\mathbb{F}$) is said to be *nilpotent* if there exists a positive integer p such that f^p is the zero function,[85] i.e. $\operatorname{Ker}(f^p) = V$. Nilpotent matrices can be defined in an analogous way: a square matrix $A \in \mathbb{F}^{n \times n}$ (where $\mathbb{F}$ is a field) is *nilpotent* if there exists a positive integer p such that $A^p = O_{n \times n}$.

[84] By the equivalence of (a) and (b) from Proposition 8.6.23, this will immediately imply that vectors $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_t)$ are linearly independent over $\operatorname{Ker}(f^{r-1})$, which is what we need to show.

[85] This simply means that $f^p(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$.

Proposition 8.6.25. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then the following are equivalent:*

(a) A is nilpotent;

(b) $p_A(\lambda) = \lambda^n$;

(c) A has only one eigenvalue, namely 0, and the algebraic multiplicity of this eigenvalue is n.

Remark: Note that Proposition 8.6.25 immediately implies that the only eigenvalue of a nilpotent $n \times n$ matrix is 0, and moreover, the algebraic multiplicity of this eigenvalue is n.

Proof. Obviously, (b) and (c) are equivalent. We will complete the proof by showing that (a) and (b) are equivalent, that is, that A is nilpotent if and only if $p_A(\lambda) = \lambda^n$.

If $p_A(\lambda) = \lambda^n$, then the Cayley-Hamilton theorem (see section 8.3) guarantees that $A^n = O_{n\times n}$, and consequently, A is nilpotent.

For the reverse implication, we assume that A is nilpotent, and we prove that $p_A(\lambda) = \lambda^n$. Using the fact that A is nilpotent, we fix a positive integer p such that $A^p = O_{n\times n}$. Now, recall the following factoring formula:

$$x^p - y^p \quad = \quad (x-y)\Big(\sum_{i=0}^{p-1} x^{p-i-1}y^i\Big).$$

If we plug in $x := \lambda I_n$ and $y := A$ into the formula above, we get

$$\lambda^p I_n - A^p \quad = \quad (\lambda I_n - A)\Big(\sum_{i=0}^{p-1} \lambda^{p-i-1}A^i\Big).$$

Taking the determinant of both sides, and keeping the multiplicative property of determinants in mind (see Theorem 7.5.2), we get that

$$\det(\lambda^p I_n - A^p) \quad = \quad \det(\lambda I_n - A)\,\det\Big(\sum_{i=0}^{p-1} \lambda^{p-i-1}A^i\Big).$$

Since $A^p = O_{n\times n}$, we have that $\det(\lambda^p I_n - A^p) = \det(\lambda^p I_n) = \lambda^{pn}$. On the other hand, by definition, we have that $p_A(\lambda) = \det(\lambda I_n - A)$. So,

$$\lambda^{pn} \quad = \quad p_A(\lambda)\,\det\Big(\sum_{i=0}^{p-1} \lambda^{p-i-1}A^i\Big),$$

and it follows that $p_A(\lambda) \mid \lambda^{pn}$. Since $p_A(\lambda)$ is a polynomial of degree n with leading coefficient 1, we deduce that $p_A(\lambda) = \lambda^n$. □

Proposition 8.6.26. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $f : V \to V$ be a linear function. Set $n := dim(V)$. Then the following are equivalent:*

(a) f is nilpotent;

(b) $p_f(\lambda) = \lambda^n$;

(c) f has only one eigenvalue, namely 0, and the algebraic multiplicity of this eigenvalue is n.

Proof. Obviously, (b) and (c) are equivalent, and we just need to prove that (a) and (b) are equivalent. For this, we simply "translate" Proposition 8.6.25 into the language of linear functions, as follows. Let $\mathcal{B}$ be any basis of V, and set $B := {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$; by Proposition 8.2.12, we have that $p_f(\lambda) = p_B(\lambda)$. We note that Theorem 4.5.3(c) and an easy induction on p imply that for all non-negative integers p, we have that $B^p = {}_{\mathcal{B}}\big[\ f^p\ \big]_{\mathcal{B}}$. We now have the following sequence of equivalent statements:

$$\begin{array}{lcl} f \text{ is nilpotent} & \iff & \exists p \in \mathbb{N} \text{ s.t. } f^p \text{ is the zero function} \\ & \iff & \exists p \in \mathbb{N} \text{ s.t. } {}_{\mathcal{B}}\big[\ f^p\ \big]_{\mathcal{B}} = O_{n\times n} \\ & \iff & \exists p \in \mathbb{N} \text{ s.t. } B^p = O_{n\times n}. \\ & \iff & B \text{ is nilpotent} \\ & \overset{(*)}{\iff} & p_B(\lambda) = \lambda^n \\ & \iff & p_f(\lambda) = \lambda^n, \end{array}$$

where (*) follows from Proposition 8.6.25. □

We now introduce some terminology and notation. Suppose that $f : V \to V$ is a nilpotent linear function, where V is a non-trivial, finite-dimensional vector space over some field $\mathbb{F}$. For a vector $\mathbf{v} \in V \setminus \{\mathbf{0}\}$, we define

$$\mathcal{J}_f(\mathbf{v}) \ := \ \big\{f^{r-1}(\mathbf{v}), f^{r-2}(\mathbf{v}), \ldots, f^2(\mathbf{v}), f(\mathbf{v}), \mathbf{v}\big\},$$

where r is the smallest positive integer such that $f^r(\mathbf{v}) = \mathbf{0}$.[86] Here, we have that $\mathbf{v} \in \mathrm{Ker}(f^r) \setminus \mathrm{Ker}(f^{r-1})$, i.e. $\mathbf{v}$ is a generalized eignvector of rank r of the nilpotent function f and associated with the eigenvalue 0, and $\mathcal{J}_f(\mathbf{v})$ is the Jordan chain started by $\mathbf{v}$. By Proposition 8.6.18(g), $\mathcal{J}_f(\mathbf{v})$ is a linearly independent set. Now, a basis $\mathcal{B}$ of V is *canonical* with respect to the nilpotent linear function f if it is of the form

$$\mathcal{B} \ = \ \mathcal{J}_f(\mathbf{u}_1) \cup \cdots \cup \mathcal{J}_f(\mathbf{u}_k),$$

[86] Such an r exists because f is nilpotent.

where $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are some non-zero vectors in V, and $\mathcal{J}_f(\mathbf{u}_1), \ldots, \mathcal{J}_f(\mathbf{u}_k)$ are pairwise disjoint. So, a canonical basis of V associated with the nilpotent linear function f is a basis that is the union of pairwise disjoint Jordan chains. Our goal is to prove Theorem 8.6.28, which states that such a basis always exists. We begin with a technical proposition, which readily implies Theorem 8.6.28.

Proposition 8.6.27. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, let $f : V \to V$ be a nilpotent linear function, and let p be a positive integer such that f^p is the zero function, i.e. $Ker(f^p) = V$. Then for all vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$ $(t \geq 0)$ that are linearly independent over $Ker(f^{p-1})$, there exist non-zero vectors $\mathbf{v}_{t+1}, \ldots, \mathbf{v}_{t+s} \in V$ $(s \geq 0)$ such that $\mathcal{J}_f(\mathbf{v}_1), \ldots, \mathcal{J}_f(\mathbf{v}_{t+s})$ are pairwise disjoint and $\mathcal{B} = \mathcal{J}_f(\mathbf{v}_1), \ldots, \mathcal{J}_f(\mathbf{v}_{t+s})$ is a basis of V.*

Remark: Note that the basis $\mathcal{B}$ is canonical with respect to f.

Proof. We may assume inductively that the proposition is true for non-trivial, finite-dimensional vector spaces of dimension strictly smaller than $\dim(V)$.

To simplify notation, for each non-negative integer r, we set $K_r := \mathrm{Ker}(f^r)$. So, $K_p = V$. Let us first explain why we may assume that $K_{p-1} \subsetneqq V$. Let q be the smallest non-negative integer such that $K_q = V$ (so, $q \leq p$). Since $K_0 = \{\mathbf{0}\}$ and V is non-trivial, we know that $K_0 \subsetneqq V$, and in particular, $q \geq 1$. Now, suppose that $q < p$. By Proposition 8.6.19, we then have that

$$\underbrace{K_0}_{=\{\mathbf{0}\}} \subsetneqq K_1 \subsetneqq K_2 \subsetneqq \ldots \subsetneqq \underbrace{K_q}_{=V} = K_{q+1} = \ldots,$$

and in particular, $K_{p-1} = K_p = V$. So, by Proposition 8.6.22, if vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$ are linearly independent over K_{p-1}, then $t = 0$, that is, $\mathbf{v}_1, \ldots, \mathbf{v}_t$ is in fact an empty list of vectors. Therefore, we just need to show that there exists a basis of V that is canonical with respect to f. Thus, we may simply prove the proposition with q instead of p, since any canonical basis of V that works for q also works for p.

In view of the discussion above, we assume from now on that $K_{p-1} \subsetneqq V$. By Proposition 8.6.19, we have that

$$\underbrace{K_0}_{=\{\mathbf{0}\}} \subsetneqq K_1 \subsetneqq K_2 \subsetneqq \ldots \subsetneqq \underbrace{K_p}_{=V} = K_{p+1} = \ldots,$$

Now, fix vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t \in V$ $(t \geq 0)$ that are linearly independent over K_{p-1}. In view of Proposition 8.6.22, we have that $\mathbf{v}_1, \ldots, \mathbf{v}_t \in K_p \setminus K_{p-1}$.[87] Fix any basis $\mathcal{N}_{p-1}$ of K_{p-1}. By Proposition 8.6.23, $\mathcal{N}_{p-1} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_t\}$ is linearly independent, and so by Theorem 3.2.19, it can be extended to a basis $\mathcal{N}_{p-1} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_t, \mathbf{z}_1, \ldots, \mathbf{z}_\ell\}$ of V. Once again by Proposition 8.6.23, vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t, \mathbf{z}_1, \ldots, \mathbf{z}_\ell$ are linearly

[87] Indeed, $\mathbf{v}_1, \ldots, \mathbf{v}_t \in K_p$ because $V = K_p$. On the other hand, since $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over K_{p-1}, Proposition 8.6.22 guarantees that $\mathbf{v}_1, \ldots, \mathbf{v}_t \notin K_{p-1}$.

independent over K_{p-1}. So, we may now assume that $\ell = 0$, for otherwise, we simply consider vectors $\mathbf{v}_1, \dots, \mathbf{v}_t, \mathbf{z}_1, \dots, \mathbf{z}_\ell$ instead of $\mathbf{v}_1, \dots, \mathbf{v}_t$. With this set-up, $\mathcal{N}_{p-1} \cup \{\mathbf{v}_1, \dots, \mathbf{v}_t\}$ is a basis of V. Note that this implies that $t = \dim(V) - \dim(K_{p-1})$.

Suppose first that $p = 1$, so that $\{\mathbf{0}\} = K_0 \subsetneqq \mathrm{Ker}(f) = V$.[88] Then $\mathcal{J}_f(\mathbf{v}_i) = \{\mathbf{v}_i\}$ for all $i \in \{1, \dots, t\}$. Moreover, we have that $\mathcal{N}_{p-1} = \emptyset$,[89] and therefore, $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_t\}$ is a basis of V.[90] But note that $\mathcal{B} = \mathcal{J}_f(\mathbf{v}_1) \cup \dots \cup \mathcal{J}_f(\mathbf{v}_t)$, and obviously, the Jordan chains $\mathcal{J}_f(\mathbf{v}_1), \dots, \mathcal{J}_f(\mathbf{v}_t)$ are pairwise disjoint.[91] So, $\mathcal{B}$ satisfies the requirements from the statement of the proposition, and we are done.

From now on, we may assume that $p \geq 2$. Note that this implies that $\{\mathbf{0}\} = K_0 \subsetneqq K_{p-1} \subsetneqq K_p = V$, and in particular, by Theorem 3.2.21, $0 < \dim(K_{p-1}) < \dim(V)$. So, we will be able to apply the induction hypothesis to the vector space K_{p-1}.

By Proposition 8.6.18(d), K_{p-1} is f-invariant, and so $g := f|_{K_{p-1}}$ is well defined and obviously linear (because f is linear). Clearly, g^{p-1} is the zero function,[92] and in particular, g is nilpotent. Moreover, it is clear that for all $i \in \{0, \dots, p-1\}$, we have that $\mathrm{Ker}(g^i) = K_i$.[93]

By Proposition 8.6.18(f), vectors $f(\mathbf{v}_1), \dots, f(\mathbf{v}_t)$ belong to K_{p-1}, and by Proposition 8.6.24, they are linearly independent over K_{p-2}. We now apply the induction hypothesis applied to K_{p-1}, the nilpotent linear function g, and the vectors $f(\mathbf{v}_1), \dots, f(\mathbf{v}_t)$, and we deduce that there exist non-zero vectors $\mathbf{v}_{t+1}, \dots, \mathbf{v}_{t+s} \in K_{p-1}$ $(s \geq 0)$ such that the Jordan chains

$$\mathcal{J}_g\big(f(\mathbf{v}_1)\big), \dots, \mathcal{J}_g\big(f(\mathbf{v}_t)\big), \mathcal{J}_g(\mathbf{v}_{t+1}), \dots, \mathcal{J}_g(\mathbf{v}_{t+s})$$

are pairwise disjoint and such that

$$\mathcal{C} \;:=\; \mathcal{J}_g\big(f(\mathbf{v}_1)\big) \cup \dots \cup \mathcal{J}_g\big(f(\mathbf{v}_t)\big) \cup \mathcal{J}_g(\mathbf{v}_{t+1}) \cup \dots \cup \mathcal{J}_g(\mathbf{v}_{t+s})$$

is a basis of K_{p-1}. We will show that

$$\mathcal{B} \;:=\; \mathcal{J}_f(\mathbf{v}_1) \cup \dots \cup \mathcal{J}_f(\mathbf{v}_t) \cup \mathcal{J}_f(\mathbf{v}_{t+1}) \cup \dots \cup \mathcal{J}_f(\mathbf{v}_{t+s})$$

is the basis of V that we need.

[88] Note that this means that f itself is a zero function.

[89] This is because $\mathcal{N}_{p-1}$ is now a basis of $K_{p-1} = K_0 = \mathrm{Ker}(f^0) = \mathrm{Ker}(\mathrm{Id}_V) = \{\mathbf{0}\}$.

[90] This is because $\mathcal{N}_{p-1} \cup \{\mathbf{v}_1, \dots, \mathbf{v}_t\}$ is a basis of V, and $\mathcal{N}_{p-1} = \emptyset$.

[91] Indeed, $\{\mathbf{v}_1, \dots, \mathbf{v}_t\}$ is a basis of V, and in particular, $\mathbf{v}_1, \dots, \mathbf{v}_t$ are pairwise distinct. So, the Jordan chains $\mathcal{J}_f(\mathbf{v}_i) = \{\mathbf{v}_i\}$ $(i \in \{1, \dots, t\})$ are pairwise disjoint.

[92] Indeed, for any $\mathbf{u} \in K_{p-1}$, we have that

$$g^{p-1}(\mathbf{u}) \;\overset{(*)}{=}\; f^{p-1}(\mathbf{u}) \;\overset{(**)}{=}\; \mathbf{0},$$

where (*) follows from the fact that $g = f|_{K_{p-1}}$, and (**) follows from the fact that $\mathbf{u} \in K_{p-1}$.

[93] This readily follows from the fact that $g = f|_{K_{p-1}}$, but here is a formal proof. Fix an index $i \in \{0, \dots, p-1\}$. Clearly, for all $\mathbf{u} \in \mathrm{Ker}(g^i)$, we have that $f^i(\mathbf{u}) = g^i(\mathbf{u}) = \mathbf{0}$, and so $\mathbf{u} \in K_i$. This proves that $\mathrm{Ker}(g^i) \subseteq K_i$. For the reverse inclusion, fix any $\mathbf{u} \in K_i$. Since $K_i \subseteq K_{p-1}$, we know that $\mathbf{u} \in K_{p-1}$, and in particular, $g^i(\mathbf{u})$ is defined. But now $g^i(\mathbf{u}) = f^i(\mathbf{u}) = \mathbf{0}$, and we deduce that $\mathbf{u} \in \mathrm{Ker}(g^i)$. This proves that $K_i \subseteq \mathrm{Ker}(g^i)$.

First of all, since $g = f|_{K_{p-1}}$, it is clear that for any $\mathbf{u} \in K_{p-1}$, we have that $\mathcal{J}_g(\mathbf{u}) = \mathcal{J}_f(\mathbf{u})$. Therefore,

$$\mathcal{C} \;=\; \mathcal{J}_f\big(f(\mathbf{v}_1)\big) \cup \cdots \cup \mathcal{J}_f\big(f(\mathbf{v}_t)\big) \cup \mathcal{J}_f(\mathbf{v}_{t+1}) \cup \cdots \cup \mathcal{J}_f(\mathbf{v}_{t+s}),$$

and the Jordan chains $\mathcal{J}_f\big(f(\mathbf{v}_1)\big), \ldots, \mathcal{J}_f\big(f(\mathbf{v}_t)\big), \mathcal{J}_f(\mathbf{v}_{t+1}), \ldots, \mathcal{J}_f(\mathbf{v}_{t+s})$ are pairwise disjoint. Moreover, for all $i \in \{1, \ldots, t\}$, we have that $\mathcal{J}_f(\mathbf{v}_i) = \mathcal{J}_f\big(f(\mathbf{v}_i)\big) \cup \{\mathbf{v}_i\}$. Since vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are pairwise distinct (because they are linearly independent) and do not belong to K_{p-1}, whereas all vectors of $\mathcal{C}$ do belong to K_{p-1}, we see that the Jordan chains $\mathcal{J}_f(\mathbf{v}_1), \ldots, \mathcal{J}_f(\mathbf{v}_t), \mathcal{J}_f(\mathbf{v}_{t+1}), \ldots, \mathcal{J}_f(\mathbf{v}_{t+s})$ are pairwise disjoint.

Now, $\mathcal{C}$ is a basis of K_{p-1}, and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent over K_{p-1}. So, by Proposition 8.6.23, $\mathcal{C} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_t\}$ is linearly independent. But $\mathcal{B}$ was obtained from $\mathcal{C} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_t\}$ by simply rearranging the elements of the ordered set $\mathcal{C} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_t\}$ (and placing $\mathbf{v}_1, \ldots, \mathbf{v}_t$ in the appropriate places). So, $\mathcal{B}$ is linearly independent. Moreover, $|\mathcal{B}| = |\mathcal{C}| + t$. Since $t = \dim(V) - \dim(K_{p-1})$, and since $|\mathcal{C}| = \dim(K_{p-1})$ (because $\mathcal{C}$ is a basis of K_{p-1}), we have that $|\mathcal{B}| = \dim(V)$. Corollary 3.2.20(a) now guarantees that $\mathcal{B}$ is a basis of V. This completes the argument. □

Theorem 8.6.28. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and let $f : V \to V$ be a nilpotent linear function. Then V has a basis that is canonical with respect to f.*

Proof. This follows immediately from Proposition 8.6.27, as follows. Since f is nilpotent, there exists a positive integer p such that f^p is the zero function. We now apply Proposition 8.6.27 to f, p, and the empty list of vectors (i.e. $t = 0$), and we deduce that there exist vectors $\mathbf{v}_1, \ldots, \mathbf{v}_s \in V$ such that the Jordan chains $\mathcal{J}_f(\mathbf{v}_1), \ldots, \mathcal{J}_f(\mathbf{v}_s)$ are pairwise disjoint, and such that $\mathcal{B} = \mathcal{J}_f(\mathbf{v}_1) \cup \cdots \cup \mathcal{J}_f(\mathbf{v}_s)$ is a basis of V. By definition, this basis $\mathcal{B}$ is canonical with respect to f. □

Theorem 8.6.29. *Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, and set $n := \dim(V)$. Let $f : V \to V$ be a nilpotent linear function. Then there exists a basis $\mathcal{B}$ of V such that ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is a Jordan matrix.*

Proof. Using Theorem 8.6.28, we fix a canonical basis $\mathcal{B}$ of V with respect to f. By the definition of a canonical basis, there exist non-zero vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$ such that the Jordan chains $\mathcal{J}_f(\mathbf{u}_1), \ldots, \mathcal{J}_f(\mathbf{u}_k)$ are pairwise disjoint and such that

$$\mathcal{B} \;=\; \mathcal{J}_f(\mathbf{u}_1) \cup \cdots \cup \mathcal{J}_f(\mathbf{u}_k).$$

Our goal is to show that ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is a Jordan matrix.

For each $i \in \{1, \ldots, k\}$, let a_i be the positive integer for which $\mathbf{u}_i \in \mathrm{Ker}(f^{a_i}) \setminus \mathrm{Ker}(f^{a_i-1})$, so that $\mathcal{J}_f(\mathbf{u}_i) = \big\{f^{a_i-1}(\mathbf{u}_i), f^{a_i-2}(\mathbf{u}_i), \ldots, f^2(\mathbf{u}_i), f(\mathbf{u}_i), \mathbf{u}_i\big\}$, and set $U_i := \mathrm{Span}\big(\mathcal{J}_f(\mathbf{u}_i)\big)$, so that $\mathcal{J}_i(\mathbf{u}_i)$ is a basis of U_i. We will show that ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} = J_{a_1}(0) \oplus \cdots \oplus J_{a_k}(0)$.

Claim. $V = U_1 \oplus \cdots \oplus U_k$. Moreover, subspaces $U_1, \ldots, U_k$ of V are f-invariant.

Proof of the Claim. The fact that $V = U_1 \oplus \cdots \oplus U_k$ follows immediately from Proposition 8.6.10. It remains to show that $U_1, \ldots, U_k$ are f-invariant. Fix an index $i \in \{1, \ldots, k\}$ and a vector $\mathbf{u} \in U_i$. Then there exist scalars $\alpha_0, \ldots, \alpha_{a_i - 1} \in \mathbb{F}$ such that

$$\mathbf{u} \;=\; \sum_{j=0}^{a_i - 1} \alpha_j f^j(\mathbf{u}_i).$$

By applying f to both sides of the equation, and by using the linearity of f, we obtain

$$f(\mathbf{u}) \;=\; \sum_{j=0}^{a_i - 1} \alpha_j f^{j+1}(\mathbf{u}_i).$$

But note that $f^{(a_i-1)+1}(\mathbf{u}_i) = f^{a_i}(\mathbf{u}_i) = \mathbf{0}$, and so we in fact have that

$$f(\mathbf{u}) \;=\; \sum_{j=0}^{a_i - 2} \alpha_j f^{j+1}(\mathbf{u}_i) \;=\; \sum_{j=1}^{a_i - 1} \alpha_{j-1} f^j(\mathbf{u}_i).$$

So, $f(\mathbf{u}) \in \mathrm{Span}\big(f^{a_1-1}(\mathbf{u}_1), \ldots, f^2(\mathbf{u}_1), f(\mathbf{u}_1)\big) \subseteq \mathrm{Span}\big(\mathcal{J}_f(\mathbf{a}_i)\big) = U_i$. ♦

In view of the Claim, for each $i \in \{1, \ldots, k\}$, we may define $f_i := f|_{U_i}$. By the Claim and Proposition 8.6.11, we have that

$${}_{\mathcal{B}}\big[\; f \;\big]_{\mathcal{B}} \;=\; {}_{\mathcal{J}_f(\mathbf{u}_1)}\big[\; f_1 \;\big]_{\mathcal{J}_f(\mathbf{u}_1)} \oplus \cdots \oplus {}_{\mathcal{J}_f(\mathbf{u}_k)}\big[\; f_k \;\big]_{\mathcal{J}_f(\mathbf{u}_k)}.$$

On the other hand, for all indices $i \in \{1, \ldots, k\}$, we have the following:

$$\begin{aligned}
& {}_{\mathcal{J}_f(\mathbf{u}_i)}\big[\; f_i \;\big]_{\mathcal{J}_f(\mathbf{u}_i)} \\
\overset{(*)}{=} \;& \Big[\; \big[\; f_i\big(f^{a_i-1}(\mathbf{u}_i)\big) \;\big]_{\mathcal{J}_f(\mathbf{u}_i)} \;\; \big[\; f_i\big(f^{a_i-2}(\mathbf{u}_i)\big) \;\big]_{\mathcal{J}_f(\mathbf{u}_i)} \;\; \cdots \;\; \big[\; f_i\big(\mathbf{u}_i\big) \;\big]_{\mathcal{J}_f(\mathbf{u}_i)} \;\Big] \\
= \;& \Big[\; \big[\; f^{a_i}(\mathbf{u}_i) \;\big]_{\mathcal{J}_f(\mathbf{u}_i)} \;\; \big[\; f^{a_i-1}(\mathbf{u}_i) \;\big]_{\mathcal{J}_f(\mathbf{u}_i)} \;\; \cdots \;\; \big[\; f(\mathbf{u}_i) \;\big]_{\mathcal{J}_f(\mathbf{u}_i)} \;\Big] \\
\overset{(**)}{=} \;& \big[\; \mathbf{0} \;\; \mathbf{e}_1 \;\; \ldots \;\; \mathbf{e}_{a_i-1} \;\big] \;=\; J_{a_i}(0),
\end{aligned}$$

where (*) follows from Theorem 4.5.1, and where in (**), $\mathbf{e}_1, \ldots, \mathbf{e}_{a_i}$ are the standard basis vectors of $\mathbb{F}^{a_i}$. It now follows that

$$\begin{aligned}
{}_{\mathcal{B}}\big[\; f \;\big]_{\mathcal{B}} \;&=\; {}_{\mathcal{J}_f(\mathbf{u}_1)}\big[\; f_1 \;\big]_{\mathcal{J}_f(\mathbf{u}_1)} \oplus \cdots \oplus {}_{\mathcal{J}_f(\mathbf{u}_k)}\big[\; f_k \;\big]_{\mathcal{J}_f(\mathbf{u}_k)} \\
&=\; J_{a_1}(0) \oplus \cdots \oplus J_{a_k}(0),
\end{aligned}$$

which completes the argument. □

The existence part of Theorems 8.6.2 and 8.6.4. We are now ready to prove the main result of this section. We first prove the existence part of Theorem 8.6.4 (see Theorem 8.6.30 below); as we shall see, it readily follows from Theorems 8.6.21 and 8.6.29. Corollary 8.6.31 readily follows from Theorem 8.6.30, and it constitutes the existence part of Theorem 8.6.2.

Theorem 8.6.30. *Let V be a non-trivial, finite-dimensional vector space over an* ***algebraically closed field*** *$\mathbb{F}$, and let $f : V \to V$ be a linear function. Then there exists a basis $\mathcal{B}$ of V such that ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is a Jordan matrix.*

Proof. Set $n := \dim(V)$, and let

$$\big\{\underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{m_k}\big\}$$

be the spectrum of f, where $\lambda_1, \ldots, \lambda_k$ are pairwise distinct eigenvalues of f and $m_1, \ldots, m_k$ are positive integers. Since the field $\mathbb{F}$ is algebraically closed, we know that $m_1 + \cdots + m_k = n$.

By Theorem 8.6.21, we know that $V = G_{\lambda_1}(f) \oplus \cdots \oplus G_{\lambda_k}(f)$, and that for all indices $i \in \{1, \ldots, k\}$, the generalized eigenspace $G_{\lambda_i}(f)$ is f-invariant and satisfies $\dim\big(G_{\lambda_i}(f)\big) = m_i$. To simplify notation, for each index $i \in \{1, \ldots, k\}$, we set $G_i := G_{\lambda_i}(f)$ and $f_i := f|_{G_i}$. Moreover, for each $i \in \{1, \ldots, k\}$, we let p_i be the smallest positive integer such that

$$G_i = \mathrm{Ker}\big((f - \lambda_i \mathrm{Id}_V)^{p_i}\big) = \mathrm{Ker}\big((f_i - \lambda_i \mathrm{Id}_{G_i})^{p_i}\big).$$

Now, for each index $i \in \{1, \ldots, k\}$, we proceed as follows. Obviously, the linear function $f_i - \lambda_i \mathrm{Id}_{G_i}$ is nilpotent.[94] Using Theorem 8.6.29, for each $i \in \{1, \ldots, k\}$, we fix a basis $\mathcal{B}_i$ such that $J_i := {}_{\mathcal{B}_i}\big[\ f_i - \lambda_i \mathrm{Id}_{G_i}\ \big]_{\mathcal{B}_i}$ is a matrix in Jordan normal form; by Theorem 8.6.29, we know that all the Jordan blocks of the Jordan matrix J_i are of the form $J_t(0)$ for some positive integer t. But now

$${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \overset{(*)}{=} {}_{\mathcal{B}}\big[\ f - \lambda_i \mathrm{Id}_{G_i}\ \big]_{\mathcal{B}} + \lambda_i\, {}_{\mathcal{B}}\big[\ \mathrm{Id}_{G_i}\ \big]_{\mathcal{B}} \overset{(**)}{=} J_i + \lambda_i I_{m_i},$$

where (*) follows from Theorem 4.5.3, and (**) follows from the definition of J_i and from formula from Theorem 4.5.1. Since the matrix J_i is in Jordan normal form, so is the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$, and moreover, we see that each Jordan block $J_t(0)$ of the former corresponds to a Jordan block $J_t(\lambda_i)$ of the latter.

Now, set $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$. Then by Proposition 8.6.11, we have that

$${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \;=\; {}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1} \oplus \cdots \oplus {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}.$$

Since ${}_{\mathcal{B}_1}\big[\ f_1\ \big]_{\mathcal{B}_1}, \ldots, {}_{\mathcal{B}_k}\big[\ f_k\ \big]_{\mathcal{B}_k}$ are Jordan matrices, so is ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. □

[94] Indeed, for all $\mathbf{v} \in G_i$, we have that $(f_i - \lambda_i \mathrm{Id}_{G_i})^{p_i}(\mathbf{v}) = (f - \lambda_i \mathrm{Id}_V)^{p_i}(\mathbf{v}) = \mathbf{0}$, and it follows that $(f_i - \lambda_i \mathrm{Id}_{G_i})^{p_i}$ is a zero function. So, $f_i - \lambda_i \mathrm{Id}_{G_i}$ is nilpotent.

Corollary 8.6.31. *Let $\mathbb{F}$ be an **algebraically closed field**, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then A is similar to a matrix in Jordan normal form.*

Proof. Define $f_A : \mathbb{F}^n \to \mathbb{F}^n$ by setting $f_A(\mathbf{u}) = A\mathbf{u}$ for all $\mathbf{u} \in \mathbb{F}^n$; then f_A is linear (by Proposition 1.10.4), and clearly, its standard matrix is A, i.e. $A = {}_{\mathcal{E}_n}\big[\ f_A\ \big]_{\mathcal{E}_n}$, where $\mathcal{E}_n$ is the standard basis of $\mathbb{F}^n$. By Theorem 8.6.30, there exists a basis $\mathcal{B}$ of $\mathbb{F}^n$ such that the matrix $J := {}_{\mathcal{B}}\big[\ f_A\ \big]_{\mathcal{B}}$ is a Jordan matrix. Finally, by Theorem 4.5.16, matrices $A = {}_{\mathcal{E}_n}\big[\ f_A\ \big]_{\mathcal{E}_n}$ and $J = {}_{\mathcal{B}}\big[\ f_A\ \big]_{\mathcal{B}}$ are similar. This completes the argument. □

8.6.6 The proof of Theorems 8.6.2, 8.6.4, and 8.6.6

We are finally ready to prove Theorems 8.6.2, 8.6.4, and 8.6.6, restated below for the reader's convenience.

Theorem 8.6.2. *Assume that $\mathbb{F}$ is an **algebraically closed field**, and let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then A is similar to a matrix J in Jordan normal form. Moreover, this matrix J is unique up to a reordering of the Jordan blocks.*

Proof. The existence part follows from Corollary 8.6.31. Uniqueness follows from Theorem 8.6.1, since any two Jordan matrices that are similar to A are (by Proposition 4.5.13) similar to each other. □

Theorem 8.6.4. *Let V be a non-trivial, finite-dimensional vector space over an **algebraically closed field** $\mathbb{F}$, and let $f : V \to V$ be a linear function. Then there exists a basis $\mathcal{B}$ such that the matrix ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ is in Jordan normal form. Moreover, this matrix is unique in the following sense: if $\mathcal{B}_1$ and $\mathcal{B}_2$ are bases of V such that both ${}_{\mathcal{B}_1}\big[\ f\ \big]_{\mathcal{B}_1}$ and ${}_{\mathcal{B}_2}\big[\ f\ \big]_{\mathcal{B}_2}$ are in Jordan normal form, then these two matrices are the same up to a reordering of the Jordan blocks.*

Proof. The existence part follows immediately from Theorem 8.6.30. The uniqueness part is a consequence of Theorem 8.6.1, as we now explain. Suppose that $\mathcal{B}_1$ and $\mathcal{B}_2$ are bases of V such that both ${}_{\mathcal{B}_1}\big[\ f\ \big]_{\mathcal{B}_1}$ and ${}_{\mathcal{B}_2}\big[\ f\ \big]_{\mathcal{B}_2}$ are in Jordan normal form. By Theorem 4.5.16, these two matrices are similar. But now Theorem 8.6.1 guarantees that these two matrices have exactly the same Jordan blocks (counting repetitions). □

Theorem 8.6.6. *Let $\mathbb{F}$ be an **algebraically closed field**, let $A \in \mathbb{F}^{n\times n}$, and let*

$$\big\{\underbrace{\lambda_1,\dots,\lambda_1}_{m_1},\dots,\underbrace{\lambda_k,\dots,\lambda_k}_{m_k}\big\}$$

be the spectrum of A, where $\lambda_1,\dots,\lambda_k$ are pairwise distinct eigenvalues of f and $m_1,\dots,m_k$ are positive integers.[95] Then A is similar to a matrix $J \in \mathbb{F}^{n\times n}$ in Jordan normal form that has the following properties:

[95] Since $\mathbb{F}$ is algebraically closed, we know that $m_1 + \cdots + m_k = n$.

(i) each Jordan block of the Jordan matrix J is of the form $J_t(\lambda_i)$ for some $i \in \{1, \dots, k\}$ and $t \in \{1, \dots, m_i\}$;

(ii) for each $i \in \{1, \dots, k\}$ and each positive integer r, the Jordan matrix J has exactly

$$rank\big((A - \lambda_i I_n)^{r-1}\big) - rank\big((A - \lambda_i I_n)^r\big)$$

many Jordan blocks $J_t(\lambda_i)$ satisfying $t \geq r$.

Moreover, A is similar to any Jordan matrix in $\mathbb{F}^{n\times n}$ that satisfies conditions (i) and (ii) above.

Proof. By Theorem 8.6.2, A is similar to a matrix J in Jordan normal form. The fact that the matrix J satisfies (i) and (ii) follows immediately from Proposition 8.6.17. On the other hand, it is clear that any Jordan matrix in $\mathbb{F}^{n\times n}$ that satisfies (i) and (ii) has exactly the same Jordan blocks as our Jordan matrix J, and is therefore (by Theorem 8.6.1) similar to J. Since matrix similarity in $\mathbb{F}^{n\times n}$ is an equivalence relation (by Proposition 4.5.13), it follows that any Jordan matrix in $\mathbb{F}^{n\times n}$ that satisfies (i) and (ii) is indeed similar to A. □

8.6.7 Computing the Jordan normal form of a square matrix (once more): computing both J and P

Generalied eigenvectors and generalized eigenspaces of square matrices. In what follows, we will need to adapt to matrices some of the terminology and notation that we originally introduced for linear functions. For a field $\mathbb{F}$, a square matrix $A \in \mathbb{F}^{n\times n}$, and an eigenvalue λ of A, we define the following:

- a *generalized eigenvector* of A associated with λ is any **non-zero** vector $\mathbf{v} \in \mathbb{F}^n$ such that for some positive integer r, we have that $(A - \lambda I_n)^r \mathbf{v} = \mathbf{0}$;[96]
- the *rank* of a generalized eigenvector $\mathbf{v}$ of A associated with λ is the smallest positive integer r such that $(A - \lambda I_n)^r \mathbf{v} = \mathbf{0}$;[97]
- for a generalized eigenvector $\mathbf{v}$ of A associated with λ and of rank r, the *Jordan chain started by* $\mathbf{v}$ is the set

$$\big\{(A - \lambda I_n)^{r-1}\mathbf{v}, \dots, (A - \lambda I_n)^2\mathbf{v}, (A - \lambda I_n)\mathbf{v}, \mathbf{v}\big\};$$

- the *generalized eigenspace* of A associated with λ is the set

$$\begin{aligned} G_\lambda(A) &:= \big\{\mathbf{v} \in V \mid \exists r \in \mathbb{N}_0 \text{ s.t. } (A - \lambda I_n)^r \mathbf{v} = \mathbf{0}\big\} \\ &= \bigcup_{r=0}^{\infty} \mathrm{Nul}\Big(\big(A - \lambda I_n\big)^r\Big), \end{aligned}$$

[96] The reason we specify that r is positive (rather than merely non-negative) is simply that $(A - \lambda I_n)^0 = I_n$, and so there are no non-zero vectors $\mathbf{v} \in \mathbb{F}^n$ such that $(A - \lambda I_n)^0 \mathbf{v} = \mathbf{0}$.

[97] Note that this means that $\mathbf{v} \in \mathrm{Nul}(A^r) \setminus \mathrm{Nul}(A^{r-1})$.

i.e. the set whose elements are precisely the generalized eigenvectors of A associated with λ, plus the vector $\mathbf{0}$.

Let $f_A : \mathbb{F}^n \to \mathbb{F}^n$ given by $f_A(\mathbf{v}) = A\mathbf{v}$, so that f_A is the linear function whose standard matrix is A. So, all the results from subsection 8.6.5 that apply to the linear function f_A can be translated into results for the matrix A (we note, however, that some of those results only hold when $\mathbb{F}$ is an algebraically closed field). Importantly, if $\mathbb{F}$ is an algebraically closed field, then Theorem 8.6.21 guarantees that $\mathbb{F}^n$ is the direct sum of the generalized eigenspaces of A. Moreover, by Proposition 8.6.18(g), all Jordan chains are linearly independent.

Computing J and P. Suppose that $\mathbb{F}$ is an algebraically closed field, and that $A \in \mathbb{F}^{n\times n}$ is a square matrix. We saw in subsection 8.6.2 how Theorem 8.6.6 can be used to compute the Jordan normal form of A. Suppose we would like to do more: we would like to compute both a Jordan matrix $J \in \mathbb{F}^{n\times n}$ and an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $J = P^{-1}AP$. Here, we give a recipe for doing precisely that. We will not give a fully formal proof of the correctness of our recipe, except to note that it essentially follows from our proof of Theorems 8.6.2 and 8.6.4 (see subsections 8.6.4, 8.6.5, and 8.6.6). The steps are as follows.

1. Compute the characteristic polynomial $p_A(\lambda)$ of A, and factor this polynomial into linear terms.[98]

2. Compute the spectrum

$$\{\underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{m_k}\}$$

of A, where $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are pairwise distinct and $m_1, \ldots, m_k$ are positive integers.

3. For each index $i \in \{1, \ldots, k\}$, we process the eigenvalue λ_i as follows.

   (a) We form the matrix $A_i := A - \lambda_i I_n$, and we compute the matrices A_i^r and their ranks for non-negative integers r. We keep computing until the rank becomes $n - m_i$, that is, until we get $\operatorname{rank}\left(A_i^{p_i}\right) = n - m_i$ for some positive integer p_i. For this p_i, we will have

$$\underbrace{\operatorname{Nul}(A_i^0)}_{=\{\mathbf{0}\}} \subsetneq \operatorname{Nul}(A_i^1) \subsetneq \ldots \subsetneq \underbrace{\operatorname{Nul}(A_i^{p_i})}_{=G_{\lambda_i}(A)} = \operatorname{Nul}(A_i^{p_i+1}) = \ldots,$$

[98] The fact that $p_A(\lambda)$ can be factored into linear terms follows from the fact that the field $\mathbb{F}$ is algebraically closed. Of course, we only know that such a factorization exists and do not have an actual formula/algorithm for finding it. If we get stuck factoring $p_A(\lambda)$, then we are well and truly stuck: we will not be able to compute the matrices J and P that we need.

and we will further have $\operatorname{rank}\left(A_i^0\right) > \operatorname{rank}\left(A_i^1\right) > \cdots > \operatorname{rank}\left(A_i^{p_i-1}\right) > \operatorname{rank}\left(A_i^{p_i}\right) = \operatorname{rank}\left(A_i^{p_i+1}\right) = \operatorname{rank}\left(A_i^{p_i+2}\right) = \ldots$,[99] and in particular, $\operatorname{rank}\left(A_i^{p_i}\right) = \operatorname{rank}\left(A_i^{p_i+1}\right)$ will be the first instance of "repeating rank."[100]

- The largest Jordan block of J associated with the eigenvalue λ_i will be $J_{p_i}(\lambda_i)$, and there may possibly be more than one copy of this block in J.

(b) For each $r \in \{0, 1, \ldots, p_i\}$, we compute a basis $\mathcal{N}_{i,r}$ of $\operatorname{Nul}(A_i^r)$.[101]

(c) We compute the Jordan blocks $J_t(\lambda_i)$ of the Jordan matrix J using the formula from Theorem 8.6.6, as in subsection 8.6.2.[102]

(d) First, we deal with the Jordan blocks of the form $J_{p_i}(\lambda_i)$ of J. Each of these blocks will correspond to a Jordan chain $\left\{A_i^{p_i-1}\mathbf{u}, \ldots, A_i\mathbf{u}, \mathbf{u}\right\}$ associated with A and λ_i; the goal is to find the vectors $\mathbf{u}$ that start these Jordan chains (one vector $\mathbf{u}$ per Jordan block $J_{p_i}(\lambda_i)$ in J). We proceed as follows. We extend the basis $\mathcal{N}_{i,p_i-1}$ of $\operatorname{Nul}(A_i^{p_i-1})$ to a basis of $\operatorname{Nul}(A_i^{p_i})$. For this, we can use Proposition 3.3.19 (the needed matrix is formed by the vectors of $\mathcal{N}_{i,p_i-1}$ and of $\mathcal{N}_{i,p_i}$, with a vertical dotted line placed between them). Each vector $\mathbf{u}$ that we added to $\mathcal{N}_{i,p_i-1}$ to form our basis of $\operatorname{Nul}(A_i^{p_i})$ starts a Jordan chain $\left\{A_i^{p_i-1}\mathbf{u}, \ldots, A_i\mathbf{u}, \mathbf{u}\right\}$ corresponding to one of the Jordan blocks $J_{p_i}(\lambda_i)$.

(e) If J contains no Jordan blocks associated with λ_i other than blocks $J_{p_i}(\lambda_i)$, then we are done processing the eigenvalue λ_i. Otherwise, we proceed as follows.

(f) Suppose that $t < p_i$, that J has at least one Jordan block $J_t(\lambda_i)$, and that we have already dealt with Jordan blocks associated with λ_i and of size

[99] Let us justify this (and in particular, explain why such a p_i exists). By Proposition 8.6.21, we have that $\dim\left(G_{\lambda_i}(A)\right) = m_i$. Moreover, Proposition 8.6.21 guarantees that there exists a positive integer p_i such that

$$\underbrace{\operatorname{Nul}(A_i^0)}_{=\{\mathbf{0}\}} \subsetneq \operatorname{Nul}(A_i^1) \subsetneq \ldots \subsetneq \underbrace{\operatorname{Nul}(A_i^{p_i})}_{=G_{\lambda_i}(A)} = \operatorname{Nul}(A_i^{p_i+1}) = \ldots$$

For this p_i, we have that $\dim\Big(\operatorname{Nul}\left(A_i^{p_i}\right)\Big) = \dim\Big(G_{\lambda_i}(A)\Big) = m_i$, and so by the rank-nullity theorem, $\operatorname{rank}\left(A_i^{p_i}\right) = n - m_i$. On the other hand, Theorem 3.2.21 guarantees that $0 = \dim\Big(\operatorname{Nul}\left(A_i^0\right)\Big) < \cdots < \dim\Big(\operatorname{Nul}\left(A_i^{p_i}\right)\Big)$, and consequently (by the rank-nullity theorem), $n = \operatorname{rank}\left(A_i^0\right) > \cdots > \operatorname{rank}\left(A_i^{p_i}\right) = p_i - m_i$. For integers $r > p_i$, we have that $\operatorname{Nul}\left(A_i^r\right) = \operatorname{Nul}\left(A_i^{p_i}\right)$, and consequently (by the rank-nullity theorem), $\operatorname{rank}\left(A_i^r\right) = \operatorname{rank}\left(A_i^{p_i}\right) = n - m_i$.

[100] In subsection 8.6.2, our recipe said that we should keep computing until we get "repeating rank" for the first time. Here, we give a bit of a shortcut: we keep computing until we get rank $n - m_i$ for the first time, and we are guaranteed that the rank will start repeating itself after that.

[101] Note that $A_i^0 = I_n$, and consequently, $\operatorname{Nul}(A_i^0) = \operatorname{Nul}(I_n) = \{\mathbf{0}\}$ and $\mathcal{N}_{i,0} = \emptyset$.

[102] Since we are currently processing the eigenvalue λ_i, we are only interested in the Jordan blocks associated with this particular eigenvalue. The Jordan blocks associated with the other eigenvalues are computed when we process those other eigenvalues.

greater than $t \times t$. We now take the basis $\mathcal{N}_{i,t-1}$ of $\mathrm{Nul}\big(A_i^{t-1}\big)$, and we add to it all the generalized eigenvectors of rank t of A associated with λ_i that come from the Jordan chains that we have already generated.[103] Let us call the resulting set $\mathcal{N}_{i,t}^-$. This is a linearly independent set of vectors in $\mathrm{Nul}\big(A_i^t\big)$. Using Proposition 8.6.19, we extend $\mathcal{N}_{i,t}^-$ to a basis of $\mathrm{Nul}(A_i^t)$ (the columns of the needed matrix are the vectors of $\mathcal{N}_{i,t}^-$ and the vectors of $\mathcal{N}_{i,t}$, with a vertical dotted line placed between the vectors of $\mathcal{N}_{i,t}^-$ and the vectors of $\mathcal{N}_{i,t}$). Each vector $\mathbf{u}$ that we added to $\mathcal{N}_{i,t}^-$ to create a basis of $\mathrm{Nul}(A_i^t)$ starts a Jordan chain $\big\{A_i^{t-1}\mathbf{u}, \ldots, A_i\mathbf{u}, \mathbf{u}\big\}$ that corresponds to one of our Jordan blocks $J_t(\lambda_i)$. We continue this process until we have created a Jordan chain for each Jordan block of J associated with λ_i.[104]

4. We form J as the direct sum of the Jordan blocks that we have computed

[103] Let us be more precise. Suppose we have already generated a Jordan chain $A_i^{s-1}\mathbf{u}, \ldots, A_i\mathbf{u}, \mathbf{u}$ for some $s > t$. Then $A_i^{s-t}\mathbf{u}$ is the unique generalized eigenvector of rank t of A associated with the eigenvalue λ_i that comes from this Jordan chain. We select such a generalized eigenvector of rank t out of each Jordan chain associated with A and λ_i that we have already generated.

[104] Let us consider an example. Suppose that we have determined (using Theorem 8.6.6) that the Jordan blocks of J associated with the eigenvalue λ_i are precisely

$$J_{18}(\lambda_i), J_9(\lambda_i), J_9(\lambda_i), J_9(\lambda_i), J_5(\lambda_i), J_5(\lambda_i), J_3(\lambda_i), J_1(\lambda_i), J_1(\lambda_i),$$

counting repetitions. Suppose that we have already generated the needed Jordan chains that correspond to the Jordan blocks $J_{18}(\lambda_i), J_9(\lambda_i), J_9(\lambda_i), J_9(\lambda_i)$, and that we are currently trying to generate the Jordan chains that correspond to the Jordan blocks $J_5(\lambda_i), J_5(\lambda_i)$. Suppose that the Jordan chains that we have already generated are the following:

- the Jordan chain $\{A_i^{17}\mathbf{u}_1, A_i^{16}\mathbf{u}_1, \ldots, A_i\mathbf{u}_1, \mathbf{u}_1\}$ corresponding to the Jordan block $J_{18}(\lambda_i)$,
- the Jordan chain $\{A_i^{8}\mathbf{w}_1, A_i^{7}\mathbf{w}_1, \ldots, A_i\mathbf{w}_1, \mathbf{w}_1\}$ corresponding to the first Jordan block $J_9(\lambda_i)$,
- the Jordan chain $\{A_i^{8}\mathbf{w}_2, A_i^{7}\mathbf{w}_2, \ldots, A_i\mathbf{w}_2, \mathbf{w}_2\}$ corresponding to the second Jordan block $J_9(\lambda_i)$,
- the Jordan chain $\{A_i^{8}\mathbf{w}_3, A_i^{7}\mathbf{w}_3, \ldots, A_i\mathbf{w}_3, \mathbf{w}_3\}$ corresponding to the third Jordan block $J_9(\lambda_i)$.

Since we are now processing the Jordan blocks $J_5(\lambda_i), J_5(\lambda_i)$, we first to need to identify the generalized eigenvectors of rank 5 of the matrix A associated with the eigenvalue λ_i from the four Jordan chains above. These are the vectors $A_i^{13}\mathbf{u}_1, A_i^4\mathbf{w}_1, A_i^4\mathbf{w}_2, A_i^4\mathbf{w}_3$. Now we form the set

$$\mathcal{N}_{i,5}^- \quad := \quad \mathcal{N}_{i,4} \cup \big\{A_i^{13}\mathbf{u}_1, A_i^4\mathbf{w}_1, A_i^4\mathbf{w}_2, A_i^4\mathbf{w}_3\big\}.$$

The set $\mathcal{N}_{i,5}^-$ is linearly independent in $\mathrm{Nul}(A_i^5)$, and so using Proposition 3.3.19, we can extend it to a basis of $\mathrm{Nul}(A_i^5)$ (the columns of the needed matrix are the vectors of $\mathcal{N}_{i,5}^-$, followed by the vectors of $\mathcal{N}_{i,5}$, with a vertical dotted line between the vectors of $\mathcal{N}_{i,5}^-$ and the vectors of $\mathcal{N}_{i,5}$). Because we have two Jordan blocks $J_5(\lambda_i), J_5(\lambda_i)$, we will need to add exactly two vectors, say $\mathbf{z}_1, \mathbf{z}_2$, to $\mathcal{N}_{i,5}^-$ in order to obtain a basis of $\mathrm{Nul}(A_i^5)$. (Vectors $\mathbf{z}_1, \mathbf{z}_2$ will be the vectors of $\mathcal{N}_{i,5}$ that Proposition 3.3.19 gives us. More precisely, they will be the the pivot columns to the right of the vertical dotted line of the matrix that we formed when we used Proposition 3.3.19.) Now $\mathbf{z}_1, \mathbf{z}_2$ are generalized eigenvectors of rank 5 of A associated with the eigenvalue λ_i, and they will start the Jordan chains of length five corresponding to our two Jordan blocks $J_5(\lambda_i), J_5(\lambda_i)$. So, the Jordan chains that correspond to our two Jordan blocks $J_5(\lambda_i), J_5(\lambda_i)$ are $\{A_i^4\mathbf{z}_1, A_i^3\mathbf{z}_1, A_i^2\mathbf{z}_1, A_i\mathbf{z}_1, \mathbf{z}_1\}$ and $\{A_i^4\mathbf{z}_2, A_i^3\mathbf{z}_2, A_i^2\mathbf{z}_2, A_i\mathbf{z}_2, \mathbf{z}_2\}$.

(associated with all the eigenvalues of A), and we form P using the corresponding Jordan chains. We make sure that the Jordan blocks of J and the Jordan chains in P are placed in a corresponding order.

Some numerical examples. We now revisit Examples 8.6.8 and 8.6.9 from subsection 8.6.2. In each case, we are given a square matrix A with entries in $\mathbb{C}$. In Examples 8.6.8 and 8.6.9, we computed the Jordan normal form of the matrix in question. We will now compute both a Jordan matrix J and an invertible matrix P such that $J = P^{-1}AP$.

Example 8.6.32. *Consider the following matrix in* $\mathbb{C}^{10\times 10}$*:*

$$A := \begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -3 & 1 & 5 & 2 & -2 & -4 & -7 & 4 & -1 & 3 \\ 0 & 1 & 3 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -2 & -1 & 3 & 4 & -1 & -2 & -3 & 2 & -1 & 2 \\ -1 & 0 & 2 & 1 & 2 & -2 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 2 & -1 & 0 & 0 & 1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 7 & -2 & 1 & -2 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 3 & -1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & 0 \end{bmatrix}.$$

Compute a Jordan matrix J and an invertible matrix P, both in $\mathbb{C}^{10\times 10}$, such that $J = P^{-1}AP$.

Remark: This is the matrix from Example 8.6.8. In that example, we computed the matrix J. Here, we will see how to compute the matrix P.

Solution. We first compute the characteristic polynomial of A, and we factor it into linear terms:

$$p_A(\lambda) = \det\big(\lambda I_{10} - A\big) = (\lambda - 3)^8(\lambda - 2)^2$$

The eigenvalues of A are $\lambda_1 = 3$ (with algebraic multiplicity 8) and $\lambda_2 = 2$ (with algebraic multiplicity 2). We handle the two eigenvalues separately.

The eigenvalue $\lambda_1 = 3$. To simplify notation, we write

$$A_1 \;:=\; A - \lambda_1 I_{10} \;=\; \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -3 & -2 & 5 & 2 & -2 & -4 & -7 & 4 & -1 & 3 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -2 & -1 & 3 & 1 & -1 & -2 & -3 & 2 & -1 & 2 \\ -1 & 0 & 2 & 1 & -1 & -2 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 4 & -2 & 1 & -2 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & -3 \end{bmatrix}.$$

We proceed as follows. We compute the matrices A_1^r for $r = 0, 1, 2, 3, \ldots$, and we simultaneously compute their ranks. Our matrix A is of size 10×10, and the algebraic multiplicity of the eigenvalue λ_1 is 8. So, we will keep computing until we reach the first positive integer p for which $\operatorname{rank}\big(A_1^p\big) = 10 - 8 = 2$. For this p, we will have that $\operatorname{Nul}\big(A_1^p\big) = \operatorname{Nul}\big(A_1^{p+1}\big) = \operatorname{Nul}\big(A_1^{p+2}\big) = \ldots$, and consequently, $\operatorname{rank}\big(A_1^p\big) = \operatorname{rank}\big(A_1^{p+1}\big) = \operatorname{rank}\big(A_1^{p+2}\big) = \ldots$. For $r = 0, 1, \ldots, p$, we also compute a basis $\mathcal{N}_{1,r}$ of $\operatorname{Nul}\big(A_1^r\big)$. Here is our computation.

$r = 0$: $A_1^0 = I_{10}$, $\operatorname{rank}\big(A_1^0\big) = 10$. A basis of $\operatorname{Nul}\big(A_1^0\big)$ is $\mathcal{N}_{1,0} := \emptyset$.

$r = 1$:

$$A_1^1 \;=\; \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -3 & -2 & 5 & 2 & -2 & -4 & -7 & 4 & -1 & 3 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -2 & -1 & 3 & 1 & -1 & -2 & -3 & 2 & -1 & 2 \\ -1 & 0 & 2 & 1 & -1 & -2 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 4 & -2 & 1 & -2 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & -3 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}(A_1^1) = \begin{bmatrix} 1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\mathrm{rank}\big(A_1^1\big) = 7$. The following is a basis of $\mathrm{Nul}\big(A_1^1\big)$:

$$\mathcal{N}_{1,1} := \left\{ \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

$r = 2$:

$$A_1^2 = \begin{bmatrix} -1 & -1 & 2 & 1 & -1 & -1 & -2 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 1 & -1 & -1 & -2 & 2 & -1 & 0 \\ 1 & 1 & -2 & -1 & 1 & 2 & 4 & -2 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1 \\ 1 & 0 & -1 & 0 & 0 & 1 & 2 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 & 2 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

By row reducing, we obtain

$$\mathrm{RREF}(A_1^2) = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 & 0 & -2 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\mathrm{rank}\big(A_1^2\big) = 4$. The following is a basis of $\mathrm{Nul}\big(A_1^1\big)$:

$$\mathcal{N}_{1,2} := \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

$r = 3$:

$$A_1^3 = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & -1 & -3 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -1 & -3 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -1 & -2 & 0 & 0 & 2 \\ -1 & 0 & 1 & 0 & 0 & -1 & -3 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -1 & -3 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}$$

By row reducing, we obtain:

$$\text{RREF}(A_1^3) = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\text{rank}(A_1^3) = 2$. The following is basis of $\text{Nul}\big(A_1^3\big)$:

$$\mathcal{N}_{1,3} := \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Since $\text{rank}\big(A_1^3\big) = 2$,[105] we know that $\text{rank}\big(A_1^3\big) = \text{rank}\big(A_1^4\big) = \text{rank}\big(A_1^5\big) = \ldots$, and in particular, we can now stop computing powers of A_1.

Next, we compute the Jordan blocks of our Jordan matrix J that correspond to the eigenvalue λ_1. This is done precisely as in Example 8.6.8 (using Theorem 8.6.6). In Example 8.6.8, we saw that J contains two Jordan blocks $J_3(\lambda_1) = J_3(3)$ and one Jordan block $J_2(\lambda_1) = J_2(3)$, and that it contains no other Jordan blocks associated with the eigenvalue $\lambda_1 = 3$. We now need to generate the Jordan chains corresponding to these Jordan blocks.

We first generate the Jordan chains corresponding to the two Jordan blocks $J_3(\lambda_1)$. (Note that each of these Jordan chains will contain three vectors and will be of the form $\{A_1^2\mathbf{u}, A_1\mathbf{u}, \mathbf{u}\}$.) We extend our basis $\mathcal{N}_{1,2}$ of $\text{Nul}\big(A_1^2\big)$ to a basis of $\text{Nul}\big(A_1^3\big)$. We use Proposition 3.3.19. We form the matrix whose columns are the vectors of $\mathcal{N}_{1,2}$, followed by the vectors of $\mathcal{N}_{1,3}$, with a vertical dotted line between

[105] Recall that A is of size 10×10, whereas the eigenvalue λ_1 is of algebraic multiplicity 8. So, we needed to compute powers of A_1 until we got rank $10 - 8 = 2$.

the vectors of $\mathcal{N}_{1,2}$ and the vectors of $\mathcal{N}_{1,3}$. The matrix we obtain is the following:

$$\left[\begin{array}{cccccc:cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 2 & -1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

By row reducing this matrix, we see that its pivot columns are all its columns to the left of the vertical dotted line, plus the first and fifth column to the right of the vertical dotted line. So, $\mathcal{N}_{1,2}$ can be extended to a basis of $\mathrm{Nul}\big(A_1^3\big)$ by adding the first and fifth vector of the basis $\mathcal{N}_{1,3}$ to it. The first and fifth vector of $\mathcal{N}_{1,3}$ are the following vectors:

- $\mathbf{u}_1 := \big[\begin{array}{cccccccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\big]^T$;
- $\mathbf{u}_2 := \big[\begin{array}{cccccccccc} -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array}\big]^T$.

These two vectors will start our first two Jordan chains. More precisely, one of our Jordan blocks $J_3(\lambda_1)$ corresponds to the Jordan chain $\{A_1^2\mathbf{u}_1, A_1\mathbf{u}_1, \mathbf{u}_1\}$, whereas the other Jordan block $J_3(\lambda_1)$ corresponds to the Jordan chain $\{A_1^2\mathbf{u}_2, A_1\mathbf{u}_2, \mathbf{u}_2\}$. If we compute the values of all these vectors and arrange them into a matrix, we obtain the following (the vertical dotted line separates the two Jordan chains):

$$\left[\begin{array}{ccc:ccc} A_1^2\mathbf{u}_1 & A_1\mathbf{u}_1 & \mathbf{u}_1 & A_1^2\mathbf{u}_2 & A_1\mathbf{u}_2 & \mathbf{u}_2 \end{array}\right] = \left[\begin{array}{ccc:ccc} -1 & 1 & 0 & 0 & 0 & -1 \\ 0 & -2 & 1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{array}\right].$$

It remains to compute a Jordan chain that corresponds to the Jordan block $J_2(\lambda_1)$. (This Jordan chain will contain two vectors, and it will be of the form $\{A_1\mathbf{u}, \mathbf{u}\}$.) We first identify the generalized eigenvectors of rank 2 of A associated with the eigenvalue λ_1 inside the Jordan chains (associated with λ_1) that we have already

created. These are the vectors $A_1\mathbf{u}_1$ and $A_1\mathbf{u}_2$. Now, $\mathcal{N}_{1,2}^- := \mathcal{N}_{1,1} \cup \{A_1\mathbf{u}_1, A_1\mathbf{u}_2\}$ is a linearly independent set of vectors in $\mathrm{Nul}\big(A_1^2\big)$, and we would like to extend it to a basis of $\mathrm{Nul}\big(A_1^2\big)$. We once again use Proposition 3.3.19. We form the matrix whose columns are the vectors of the linearly independent set $\mathcal{N}_{1,2}^- = \mathcal{N}_{1,1} \cup \{A_1\mathbf{u}_1, A_1\mathbf{u}_2\}$, followed by the vectors of $\mathcal{N}_{1,2}$, with a vertical dotted line between the vectors of the two sets. The matrix that we obtain is the following:

$$\left[\begin{array}{rrrrr:rrrrrr} -1 & 1 & -1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & -1 & 1 & 1 & -1 & 2 & -1 & -2 \\ -1 & 1 & -1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

By row reducing, we see that the pivot columns of this matrix are all the columns to the left of the vertical dotted line, plus the fifth column to the right of the vertical dotted line. So, $\mathcal{N}_{1,2}^- = \mathcal{N}_{1,1} \cup \{A_1\mathbf{u}_1, A_1\mathbf{u}_2\}$ can be extended to a basis of $\mathrm{Nul}\big(A_1^2\big)$ by adding the fifth vector of $\mathcal{N}_{1,2}$ to it. The fifth vector of $\mathcal{N}_{1,2}$ is the following:

- $\mathbf{u}_3 = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$.

This vector starts our Jordan chain. More precisely, our Jordan block $J_2(\lambda_1)$ corresponds to the Jordan chain $\{A_1\mathbf{u}_3, \mathbf{u}_3\}$. By computing and placing this chain into a matrix, we obtain:

$$\begin{bmatrix} A_1\mathbf{u}_3 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

We are now done processing the eigenvalue $\lambda_1 = 3$.

The eigenvalue $\lambda_2 = 2$. To simplify notation, we write

$$A_2 \;:=\; A - \lambda_2 I_{10} \;=\; \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -3 & -1 & 5 & 2 & -2 & -4 & -7 & 4 & -1 & 3 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -2 & -1 & 3 & 2 & -1 & -2 & -3 & 2 & -1 & 2 \\ -1 & 0 & 2 & 1 & 0 & -2 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & -2 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & -2 \end{bmatrix}.$$

We now proceed as we did in the case of the eigenvalue $\lambda_1 = 3$. We compute the matrices A_2^r for $r = 0, 1, 2, 3, \ldots$, and we simultaneously compute their ranks. Our matrix A is of size 10×10, and the algebraic multiplicity of the eigenvalue λ_2 is 2. So, we will keep computing until we reach the first positive integer p for which $\operatorname{rank}\big(A_2^p\big) = 10 - 2 = 8$. For this p, we will have that $\operatorname{Nul}\big(A_2^p\big) = \operatorname{Nul}\big(A_2^{p+1}\big) = \operatorname{Nul}\big(A_2^{p+2}\big) = \ldots$, and consequently, $\operatorname{rank}\big(A_2^p\big) = \operatorname{rank}\big(A_2^{p+1}\big) = \operatorname{rank}\big(A_2^{p+2}\big) = \ldots$ For $r = 0, 1, \ldots, p$, we also compute a basis $\mathcal{N}_{2,r}$ of $\operatorname{Nul}\big(A_2^r\big)$. Here is our computation.

$r = 0$: $A_2^0 = I_{10}$, $\operatorname{rank}\big(A_2^0\big) = 10$. A basis of $\operatorname{Nul}\big(A_2^0\big)$ is $\mathcal{N}_{2,0} := \emptyset$.

$r = 1$:

$$A_2^1 \;=\; \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -3 & -1 & 5 & 2 & -2 & -4 & -7 & 4 & -1 & 3 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ -2 & -1 & 3 & 2 & -1 & -2 & -3 & 2 & -1 & 2 \\ -1 & 0 & 2 & 1 & 0 & -2 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & -2 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 1 & 1 & -2 & -1 & 1 & 2 & 5 & -2 & 1 & -2 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}(A_2^1) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\mathrm{rank}\big(A_2^1\big) = 9$. The following is a basis of $\mathrm{Nul}\big(A_2^1\big)$:

$$\mathcal{N}_{2,1} := \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

$r = 2$:

$$A_2^2 = \begin{bmatrix} 0 & 1 & 2 & 1 & -1 & -1 & -2 & 0 & -1 & 2 \\ -6 & -3 & 10 & 4 & -4 & -8 & -14 & 8 & -2 & 6 \\ -1 & 1 & 3 & 1 & -1 & -1 & -2 & 0 & -1 & 2 \\ -3 & -1 & 4 & 2 & -1 & -2 & -2 & 2 & -1 & 2 \\ -2 & 0 & 4 & 2 & -1 & -3 & -1 & 2 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & -4 & -2 & 2 & 4 & 9 & -4 & 2 & -4 \\ -1 & 0 & 1 & 0 & 0 & -1 & 2 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 2 & 2 & -4 & -2 & 2 & 4 & 9 & -4 & 2 & -4 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}(A_2^2) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\mathrm{rank}\big(A_2^2\big) = 8$. The following is a basis of $\mathrm{Nul}\big(A_2^2\big)$:

$$\mathcal{N}_{2,2} := \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Since $\mathrm{rank}\big(A_2^2\big) = 8$,[106] we know that $\mathrm{rank}\big(A_2^2\big) = \mathrm{rank}\big(A_2^3\big) = \mathrm{rank}\big(A_2^4\big) = \ldots$, and in particular, we can now stop computing powers of A_2.

Next, we compute the Jordan blocks of our Jordan matrix J that correspond to the eigenvalue λ_2. We already did this in Example 8.6.8, and we saw that J contains one Jordan block $J_2(\lambda_2) = J_2(2)$. We must compute a Jordan chain that corresponds to this Jordan block. (This Jordan chain will contain two vectors, and it will be of the form $\{A_2\mathbf{u}, \mathbf{u}\}$.) We proceed as follows. We first extend our basis $\mathcal{N}_{2,1}$ of $\mathrm{Nul}\big(A_2^1\big)$ to a basis of $\mathrm{Nul}\big(A_2^2\big)$. We use Proposition 3.3.19. We form the matrix whose columns are the vectors of $\mathcal{N}_{2,1}$, followed by the vectors of $\mathcal{N}_{2,2}$, with a vertical dotted line between the vectors of $\mathcal{N}_{2,1}$ and the vectors of $\mathcal{N}_{2,2}$. The matrix

[106] Recall that A is of size 10×10, whereas the eigenvalue λ_2 is of algebraic multiplicity 2. So, we needed to compute powers of A_2 until we got rank $10 - 2 = 8$.

that we obtain is the following:

$$\left[\begin{array}{c|cc} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}\right].$$

By row reducing, we see that the pivot columns of this matrix are its first column (the only column to the left of the vertical dotted line), plus the second column after the vertical dotted line. So, $\mathcal{N}_{2,1}$ can be extended to a basis of $\mathrm{Nul}\big(A_2^2\big)$ by adding the second vector of the basis $\mathcal{N}_{2,2}$ to it. The second vector of $\mathcal{N}_{2,2}$ is the following vector:

- $\mathbf{w}_1 := \left[\begin{array}{cccccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{array}\right]^T$.

The Jordan chain corresponding to the Jordan block $J_2(\lambda_2)$ will therefore be the chain $\{A_2\mathbf{w}_1, \mathbf{w}_1\}$. In a matrix form, we get

$$\left[\begin{array}{cc} A_2\mathbf{w}_1 & \mathbf{w}_1 \end{array}\right] = \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{array}\right].$$

We are now done processing the eigenvalue $\lambda_2 = 2$.

The matrices J and P. We now put everything together. Our Jordan matrix is

$$\begin{aligned} J &:= J_3(\lambda_1) \oplus J_3(\lambda_1) \oplus J_2(\lambda_1) \oplus J_2(\lambda_2) \\ &= J_3(3) \oplus J_3(3) \oplus J_2(3) \oplus J_2(2) \\ &= \begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}. \end{aligned}$$

Our matrix P consists of the Jordan chains that we have created, where we make sure that we place the Jordan chains in the order that corresponds to the order in which we placed our Jordan blocks in J. Our matrix P is the following (with vertical dotted lines places between different Jordan chains to facilitate reading):

$$\begin{aligned} P &:= \left[\begin{array}{ccc:ccc:ccc:cc} A_1^2\mathbf{u}_1 & A_1\mathbf{u}_1 & \mathbf{u}_1 & A_1^2\mathbf{u}_2 & A_1\mathbf{u}_2 & \mathbf{u}_2 & A_1\mathbf{u}_3 & \mathbf{u}_3 & A_2\mathbf{w}_1 & \mathbf{w}_1 \end{array} \right] \\ &= \left[\begin{array}{ccc:ccc:ccc:cc} -1 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & -2 & 1 & 0 & -1 & 0 & 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{array} \right]. \end{aligned}$$

For this choice of J and P, we have that $J = P^{-1}AP$.

Optional: Because our computation is so long and complicated, it is very easy to miscompute (even if we use a calculator throughout). Therefore, it is a good idea to check our answer. We ask a calculator to check whether our matrix P is invertible,

and if so, to compute its inverse. Indeed, we obtain

$$P^{-1} = \begin{bmatrix} 3 & 0 & -5 & -1 & 1 & 3 & 2 & -2 & 0 & -1 \\ 1 & 0 & -2 & -1 & 1 & 1 & 2 & -1 & 0 & -1 \\ 2 & 1 & -3 & -1 & 1 & 2 & 4 & -2 & 1 & -2 \\ -2 & 0 & 3 & 1 & 0 & -2 & 0 & 1 & 0 & 0 \\ -1 & 0 & 2 & 1 & -1 & -1 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}.$$

Next, we ask the calculator to compute the product $P^{-1}AP$. We are in luck: we do indeed get the Jordan matrix J that we computed above. So, our answer is correct. □

Example 8.6.33. *Consider the following matrix in* $\mathbb{C}^{13\times 13}$*:*

$$A := \begin{bmatrix} 4 & 0 & 0 & 3 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 5 & 1 & -3 & 0 & 1 & -2 & 1 & 1 & 0 & 1 & -2 & 0 \\ -1 & -2 & 3 & 4 & 0 & -6 & 3 & 1 & -2 & 2 & -3 & 4 & -2 \\ 0 & -3 & 0 & 10 & -1 & -2 & 4 & 0 & -2 & -1 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 & 4 & 3 & -1 & 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & -12 & 1 & 2 & -4 & 0 & 2 & 1 & -2 & -6 & 0 \\ 1 & 1 & 1 & -1 & 0 & 6 & -1 & 3 & 1 & -2 & 3 & -2 & 2 \\ 0 & -1 & 0 & 3 & 0 & -1 & 2 & 0 & 3 & 0 & 0 & 2 & 0 \\ -1 & -1 & -1 & 4 & 0 & -4 & 3 & -1 & -1 & 4 & -2 & 3 & -2 \\ 0 & 1 & 0 & -3 & 0 & 1 & -2 & 0 & 1 & 0 & 4 & -2 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Compute a Jordan matrix J and an invertible matrix P, both in $\mathbb{C}^{5\times 5}$, such that $J = P^{-1}AP$.

Remark: This is the matrix from Example 8.6.9. In that example, we computed the matrix J. Here, we will see how to compute the matrix P.

Solution. We first compute the characteristic polynomial of A, and we factor it into linear terms:

$$p_A(\lambda) = \det\big(\lambda I_{13} - A\big) = (\lambda-4)^{10}(\lambda-2)^3.$$

The eigenvalues of A are $\lambda_1 = 4$ (with algebraic multiplicity 10) and $\lambda_2 = 2$ (with algebraic multiplicity 3). We handle the two eigenvalues separately.

The eigenvalue $\lambda_1 = 4$. To simplify notation, we write

$$A_1 := A - \lambda_1 I_{13}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 3 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 1 & 1 & -3 & 0 & 1 & -2 & 1 & 1 & 0 & 1 & -2 & 0 \\ -1 & -2 & -1 & 4 & 0 & -6 & 3 & 1 & -2 & 2 & -3 & 4 & -2 \\ 0 & -3 & 0 & 6 & -1 & -2 & 4 & 0 & -2 & -1 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 & 0 & 3 & -1 & 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & -12 & 1 & 2 & -8 & 0 & 2 & 1 & -2 & -6 & 0 \\ 1 & 1 & 1 & -1 & 0 & 6 & -1 & -1 & 1 & -2 & 3 & -2 & 2 \\ 0 & -1 & 0 & 3 & 0 & -1 & 2 & 0 & -1 & 0 & 0 & 2 & 0 \\ -1 & -1 & -1 & 4 & 0 & -4 & 3 & -1 & -1 & 0 & -2 & 3 & -2 \\ 0 & 1 & 0 & -3 & 0 & 1 & -2 & 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \end{bmatrix}$$

We proceed as follows. We compute the matrices A_1^r for $r = 0, 1, 2, 3, \ldots$, and we simultaneously compute their ranks. Our matrix A is of size 13×13, and the algebraic multiplicity of the eigenvalue λ_1 is 10. So, we will keep computing until we reach the first positive integer p for which $\operatorname{rank}(A_1^p) = 13 - 10 = 3$. For this p, we will have $\operatorname{Nul}(A_1^p) = \operatorname{Nul}(A_1^{p+1}) = \operatorname{Nul}(A_1^{p+2}) = \ldots$, and consequently, $\operatorname{rank}(A_1^p) = \operatorname{rank}(A_1^{p+1}) = \operatorname{rank}(A_1^{p+1}) = \ldots$. For $r = 0, 1, \ldots, p$, we also compute a basis $\mathcal{N}_{1,r}$ of $\operatorname{Nul}(A_1^r)$. Here is our computation.

$r = 0$: $A_1^0 = I_{13}$, $\operatorname{rank}(A_1^0) = 13$. A basis of $\operatorname{Nul}(A_1^0)$ is $\mathcal{N}_{1,0} := \emptyset$.

$r = 1$:

$$A_1^1 = \begin{bmatrix} 0 & 0 & 0 & 3 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 1 & 1 & -3 & 0 & 1 & -2 & 1 & 1 & 0 & 1 & -2 & 0 \\ -1 & -2 & -1 & 4 & 0 & -6 & 3 & 1 & -2 & 2 & -3 & 4 & -2 \\ 0 & -3 & 0 & 6 & -1 & -2 & 4 & 0 & -2 & -1 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 & 0 & 3 & -1 & 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & -12 & 1 & 2 & -8 & 0 & 2 & 1 & -2 & -6 & 0 \\ 1 & 1 & 1 & -1 & 0 & 6 & -1 & -1 & 1 & -2 & 3 & -2 & 2 \\ 0 & -1 & 0 & 3 & 0 & -1 & 2 & 0 & -1 & 0 & 0 & 2 & 0 \\ -1 & -1 & -1 & 4 & 0 & -4 & 3 & -1 & -1 & 0 & -2 & 3 & -2 \\ 0 & 1 & 0 & -3 & 0 & 1 & -2 & 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}(A_1^1) = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\mathrm{rank}\big(A_1^1\big) = 9$. The following is a basis of $\mathrm{Nul}\big(A_1^1\big)$:

$$\mathcal{N}_{1,1} := \left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

$r = 2$:

$$A_1^2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 3 & -1 & 0 & 2 & 1 & 0 & -1 & 1 & 2 & 0 \\ -2 & 3 & -2 & 9 & 3 & 8 & 6 & -6 & 0 & -1 & 2 & 2 & 4 \\ -2 & 2 & -2 & 0 & 2 & 0 & 0 & -2 & 0 & 2 & -2 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & -4 & 0 & 0 & 0 & -1 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 2 & 12 & -2 & 0 & 8 & 2 & 0 & -2 & 6 & 4 & 0 \\ 1 & -2 & 1 & -9 & -2 & -8 & -6 & 5 & 0 & 2 & -3 & -2 & -4 \\ -1 & 1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & -3 & 2 & 4 & -2 & -1 & 0 & 2 & -1 & -2 & 4 \\ 1 & -1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & -12 & -1 & 0 & -8 & 1 & 0 & -1 & -3 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}(A_1^2) = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2/3 & 0 & 0 & 0 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\operatorname{rank}(A_1^2) = 6$. The following is a basis of $\operatorname{Nul}(A_1^2)$:

$$\mathcal{N}_{1,2} := \left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

$r = 3$:

$$A_1^3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 & 1 & 0 & 2 & 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & -1 & 0 & -42 & -1 & -16 & -28 & 8 & 0 & 7 & -12 & -16 & -8 \\ 0 & 0 & 0 & -6 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -18 & 0 & 0 & -12 & 0 & 0 & 0 & -8 & -4 & 0 \\ 0 & 1 & 0 & 39 & 1 & 16 & 26 & -8 & 0 & -7 & 12 & 14 & 8 \\ 0 & 0 & 0 & -3 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & -1 & 0 & -3 & -1 & -8 & -2 & 0 & 0 & -1 & 0 & -2 & -8 \\ 0 & 0 & 0 & 3 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 27 & 0 & 0 & 18 & 0 & 0 & 0 & 8 & 10 & 0 \\ 0 & 0 & 0 & 0 & 0 & -8 & 0 & 0 & 0 & 0 & 0 & 0 & -8 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}(A_1^3) = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2/3 & 0 & 0 & 0 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\mathrm{rank}(A_1^3) = 5$. The following is a basis of $\mathrm{Nul}(A_1^3)$:

$$\mathcal{N}_{1,3} := \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

$r = 4$:

$$A_1^4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 96 & -2 & 32 & 64 & -16 & 0 & -18 & 32 & 32 & 16 \\ 0 & -2 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -16 & 0 & 0 & 0 & 0 & 0 & 0 & -16 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 48 & 2 & 0 & 32 & 0 & 0 & 2 & 16 & 16 & 0 \\ 0 & 1 & 0 & -96 & 1 & -32 & -64 & 16 & 0 & 17 & -32 & -32 & -16 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 16 & 0 & 0 & 0 & -1 & 0 & 0 & 16 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -48 & 1 & 0 & -32 & 0 & 0 & 1 & -16 & -16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 16 & 0 & 0 & 0 & 0 & 0 & 0 & 16 \end{bmatrix}$$

By row reducing, we obtain:

$$\mathrm{RREF}(A_1^4) = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2/3 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\operatorname{rank}\big(A_1^4\big) = 4$. The following is a basis of $\operatorname{Nul}\big(A_1^4\big)$:

$$\mathcal{N}_{1,4} := \left\{ \begin{bmatrix} 1\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\-1\\0\\0\\1\\0\\0\\0\\0\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\-2\\0\\0\\3\\0\\0\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\0\\0\\0\\0\\1\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\-1\\0\\0\\0\\0\\0\\-1\\0\\1\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\-1\\0\\0\\0\\0\\0\\0\\3\\0\\0 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0\\0\\0\\-1\\0\\0\\0\\0\\0\\0\\0\\3\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\0\\-1\\0\\-1\\0\\0\\0\\0\\1 \end{bmatrix} \right\}.$$

$r = 5$:

$$A_1^5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -240 & 0 & -64 & -160 & 32 & 0 & 32 & -80 & -80 & -32 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 32 & 0 & 0 & 0 & 0 & 0 & 0 & 32 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -96 & 0 & 0 & -64 & 0 & 0 & 0 & -32 & -32 & 0 \\ 0 & 0 & 0 & 240 & 0 & 64 & 160 & -32 & 0 & -32 & 80 & 80 & 32 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -32 & 0 & 0 & 0 & 0 & 0 & 0 & -32 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 96 & 0 & 0 & 64 & 0 & 0 & 0 & 32 & 32 & 0 \\ 0 & 0 & 0 & 0 & 0 & -32 & 0 & 0 & 0 & 0 & 0 & 0 & -32 \end{bmatrix}.$$

By row reducing, we obtain:

$$\mathrm{RREF}(A_1^5) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 2/3 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\operatorname{rank}\left(A_1^5\right) = 3$. The following is a basis of $\operatorname{Nul}\left(A_1^5\right)$:

$$\mathcal{N}_{1,5} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Since $\operatorname{rank}\left(A_1^5\right) = 3$,[107] we know that $\operatorname{rank}\left(A_1^5\right) = \operatorname{rank}\left(A_1^6\right) = \operatorname{rank}\left(A_1^7\right) = \ldots$, and in particular, we can now stop computing powers of A_1.

Now, we compute the Jordan blocks of our Jordan matrix J associated with the eigenvalue $\lambda_1 = 4$. This is done precisely as in Example 8.6.9 (using Theorem 8.6.6). In Example 8.6.9, we saw that J contains one Jordan block $J_5(\lambda_1) = J_5(4)$, two Jordan blocks $J_2(\lambda_1) = J_2(4)$, and one Jordan block $J_1(\lambda_1) = J_1(4)$. The Jordan normal form of A contains no other Jordan blocks of the form $J_t(\lambda_1) = J_t(4)$. We now need to generate the Jordan chains corresponding to these Jordan blocks.

We first generate the Jordan chain corresponding to the Jordan block $J_5(\lambda_1)$. (Note that this Jordan chain will contain five vectors and will be of the form

[107] Recall that A is of size 13×13, whereas the algebraic multiplicity of λ_1 is 10. So, we needed to compute powers of A_1 until we got rank $13 - 10 = 3$.

$\{A_1^4\mathbf{u}, A_1^3\mathbf{u}, A_1^2\mathbf{u}, A_1\mathbf{u}, \mathbf{u}\}$.) We extend our basis $\mathcal{N}_{1,4}$ to a basis of $\mathrm{Nul}\big(A_1^5\big)$. We use Proposition 3.3.19. We form the matrix whose columns are the vectors of $\mathcal{N}_{1,4}$, followed by the vectors of $\mathcal{N}_{1,5}$, with a vertical dotted line between the vectors of $\mathcal{N}_{1,4}$ and the vectors of $\mathcal{N}_{1,5}$. The matrix that we obtain is the following:

$$\left[\begin{array}{ccccccccc:cccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -2 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & -1 & -1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right].$$

By row reducing this matrix, we see that its pivot columns are all its columns to the left of the vertical dotted line, plus the second column to the right of the vertical dotted line. So, $\mathcal{N}_{1,4}$ can be extended to a basis of $\mathrm{Nul}\big(A_1^5\big)$ by adding the second vector of the basis $\mathcal{N}_{1,5}$ to it. The second vector of $\mathcal{N}_{1,5}$ is the following vector:

- $\mathbf{u}_1 := \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$.

The vector $\mathbf{u}_1$ will start the Jordan chain that corresponds to the Jordan block $J_5(\lambda_1)$. This Jordan chain is $\{A_1^4\mathbf{u}_1, A_1^3\mathbf{u}_1, A_1^2\mathbf{u}_1, A_1\mathbf{u}_1, \mathbf{u}_1\}$. If we compute the values of all these vectors and arrange them into a matrix, we obtain the following:

$$\begin{bmatrix} A_1^4\mathbf{u}_1 & A_1^3\mathbf{u}_1 & A_1^2\mathbf{u}_1 & A_1\mathbf{u}_1 & \mathbf{u}_1 \end{bmatrix} = \begin{bmatrix}
0 & 0 & 1 & 0 & 0 \\
1 & 1 & -1 & 1 & 1 \\
-2 & -1 & 3 & -2 & 0 \\
-2 & 0 & 2 & -3 & 0 \\
0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
2 & 0 & -2 & 3 & 0 \\
1 & 1 & -2 & 1 & 0 \\
-1 & 0 & 1 & -1 & 0 \\
-1 & -1 & 2 & -1 & 0 \\
1 & 0 & -1 & 1 & 0 \\
1 & 0 & -1 & 2 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}.$$

Next, we generate the Jordan chains corresponding to the two Jordan chains corresponding to the two Jordan blocks $J_2(\lambda_1)$. Each of these Jordan chains will contain two vectors and will be of the form $\{A_1\mathbf{u}, \mathbf{u}\}$. We first identify the generalized eigenvectors of rank 2 of A associated with the eigenvalue λ_1 inside the Jordan chains (associated with λ_1) that we have already created. There is exactly on such vector, namely, $A_1^3\mathbf{u}_1$. Now, $\mathcal{N}_{1,2}^- := \mathcal{N}_{1,1} \cup \{A_1^3\mathbf{u}_1\}$ is a linearly independent set of vectors in $\text{Nul}\big(A_1^2\big)$, and we would like to extend it to a basis of $\text{Nul}\big(A_1^2\big)$. We once again use Proposition 3.3.19. We form the matrix whose columns are the vectors of the linearly independent set $\mathcal{N}_{1,2}^- = \mathcal{N}_{1,1} \cup \{A_1^3\mathbf{u}_1\}$, followed by the vectors of $\mathcal{N}_{1,2}$, with a vertical dotted line between the vectors of the two sets. The matrix that we obtain is the following:

$$\left[\begin{array}{ccccc:ccccccc}
-1 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & -3 & 1 \\
0 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -2 & 0 & 0 & 0 & -2 & 0 & 0 & -2 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]$$

By row reducing this matrix, we see that its pivot columns are all its columns to the left of the vertical dotted line, plus the third and seventh column to the right of the vertical dotted line. So, $\mathcal{N}_{1,2}^-$ can be extended to a basis of $\text{Nul}\big(A_1^2\big)$ by adding the third and the seventh vector of $\mathcal{N}_{1,2}$ to it. The third and the seventh vector of $\mathcal{N}_{1,2}$ are the following:

- $\mathbf{u}_2 := \begin{bmatrix} 0 & 0 & 0 & -2 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$;
- $\mathbf{u}_3 := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$.

These two vectors will start our two Jordan chains that correspond to the two Jordan blocks $J_2(\lambda_1)$. These two Jordan chains are $\{A_1\mathbf{u}_2, \mathbf{u}_2\}$ and $\{A_1\mathbf{u}_3, \mathbf{u}_3\}$. If we compute the values of these vectors and arrange them into a matrix, we obtain

the following (the vertical dotted line separates the two Jordan chains):

$$\left[\begin{array}{cc:cc} A_1\mathbf{u}_2 & \mathbf{u}_2 & A_1\mathbf{u}_3 & \mathbf{u}_3 \end{array}\right] = \left[\begin{array}{cc:cc} 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & -2 & 2 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 3 & -2 & 0 \\ -1 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

Finally, we generate the Jordan chain that corresponds to the Jordan block $J_1(\lambda_1)$. This Jordan chain will contain only one vector, which we find as follows. We first identify the generalized eigenvectors of rank 1 of A associated with the eigenvalue λ_1 inside the Jordan chains (associated with λ_1) that we have already generated. There are exactly three such vectors, namely, $A_1^4\mathbf{u}_1, A_1\mathbf{u}_2, A_1\mathbf{u}_3$. Now, $\mathcal{N}_{1,1}^- := \mathcal{N}_{1,0} \cup \{A_1^4\mathbf{u}_1, A_1\mathbf{u}_2, A_1\mathbf{u}_3\} = \{A_1^4\mathbf{u}_1, A_1\mathbf{u}_2, A_1\mathbf{u}_3\}$ is a linearly independent set in $\mathrm{Nul}\big(A_1^1\big)$, and we would like to extend it to a basis of $\mathrm{Nul}\big(A_1^1\big)$. We once again use Proposition 3.3.19. We form the matrix whose columns are the vectors of the linearly independent set $\mathcal{N}_{1,1}^- = \{A_1^4\mathbf{u}_1, A_1\mathbf{u}_2, A_1\mathbf{u}_3\}$, followed by the vectors of $\mathcal{N}_{1,1}$, with a vertical dotted line between the vectors of $\mathcal{N}_{1,1}^-$ and the vectors of $\mathcal{N}_{1,1}$. The matrix that we obtain is the following:

$$\left[\begin{array}{ccc:cccc} 0 & 0 & 1 & -1 & 0 & 1 & -1 \\ 1 & 0 & -1 & 0 & -1 & 0 & 0 \\ -2 & 1 & 2 & 1 & 0 & 0 & 0 \\ -2 & 0 & 2 & 0 & 0 & 0 & -2 \\ 0 & -1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & -2 & 0 & 0 & 0 & 2 \\ 1 & -1 & -2 & 0 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

By row reducing this matrix, we see that its pivot columns are all its columns to the left of the vertical dotted line, plus the second column to the right of the vertical

dotted line. So, $\mathcal{N}_{1,1}^{-}$ can be extended to a basis of $\mathrm{Nul}\big(A_1^1\big)$ by adding the second vector of $\mathcal{N}_{1,1}$ to it. The second vector of $\mathcal{N}_{1,1}$ is the following:

- $\mathbf{u}_4 := \begin{bmatrix} 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^T$.

This vector forms the Jordan chain $\{\mathbf{u}_4\}$ that corresponds to the Jordan block $J_1(\lambda_1)$.

We are now done processing the eigenvalue $\lambda_1 = 4$.

The eigenvalue $\lambda_2 = 2$. To simplify notation, we write

$$A_2 := A - \lambda_2 I_{13} = \begin{bmatrix} 2 & 0 & 0 & 3 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 3 & 1 & -3 & 0 & 1 & -2 & 1 & 1 & 0 & 1 & -2 & 0 \\ -1 & -2 & 1 & 4 & 0 & -6 & 3 & 1 & -2 & 2 & -3 & 4 & -2 \\ 0 & -3 & 0 & 8 & -1 & -2 & 4 & 0 & -2 & -1 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 & 2 & 3 & -1 & 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & -12 & 1 & 2 & -6 & 0 & 2 & 1 & -2 & -6 & 0 \\ 1 & 1 & 1 & -1 & 0 & 6 & -1 & 1 & 1 & -2 & 3 & -2 & 2 \\ 0 & -1 & 0 & 3 & 0 & -1 & 2 & 0 & 1 & 0 & 0 & 2 & 0 \\ -1 & -1 & -1 & 4 & 0 & -4 & 3 & -1 & -1 & 2 & -2 & 3 & -2 \\ 0 & 1 & 0 & -3 & 0 & 1 & -2 & 0 & 1 & 0 & 2 & -2 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We proceed similarly as in the case of the eigenvalue λ_1. We compute matrices A_2^r for $r = 0, 1, 2, 3, \ldots$, and we simultaneously compute their ranks. Our matrix A is of size 13×13, and the algebraic multiplicity of the eigenvalue λ_2 is 3. So, we will keep computing until we reach the first positive integer p for which $\mathrm{rank}\big(A_2^p\big) = 13 - 3 = 10$. For this p, we will have that $\mathrm{Nul}\big(A_2^p\big) = \mathrm{Nul}\big(A_2^{p+1}\big) = \mathrm{Nul}\big(A_2^{p+2}\big) = \ldots$, and consequently, $\mathrm{rank}\big(A_2^p\big) = \mathrm{rank}\big(A_2^{p+1}\big) = \mathrm{rank}\big(A_2^{p+2}\big) = \ldots$. For $r = 0, 1, \ldots, p$, we also compute a basis $\mathcal{N}_{2,r}$ of $\mathrm{Nul}\big(A_2^r\big)$. Here is our computation.

$r = 0$: $A_2^0 = I_{13}$, $\mathrm{rank}\big(A_2^0\big) = 13$. A basis of $\mathrm{Nul}\big(A_2^0\big)$ is $\mathcal{N}_{2,0} := \emptyset$.

$r = 1$:

$$A_2^1 = \begin{bmatrix} 2 & 0 & 0 & 3 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 3 & 1 & -3 & 0 & 1 & -2 & 1 & 1 & 0 & 1 & -2 & 0 \\ -1 & -2 & 1 & 4 & 0 & -6 & 3 & 1 & -2 & 2 & -3 & 4 & -2 \\ 0 & -3 & 0 & 8 & -1 & -2 & 4 & 0 & -2 & -1 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 & 2 & 3 & -1 & 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & -12 & 1 & 2 & -6 & 0 & 2 & 1 & -2 & -6 & 0 \\ 1 & 1 & 1 & -1 & 0 & 6 & -1 & 1 & 1 & -2 & 3 & -2 & 2 \\ 0 & -1 & 0 & 3 & 0 & -1 & 2 & 0 & 1 & 0 & 0 & 2 & 0 \\ -1 & -1 & -1 & 4 & 0 & -4 & 3 & -1 & -1 & 2 & -2 & 3 & -2 \\ 0 & 1 & 0 & -3 & 0 & 1 & -2 & 0 & 1 & 0 & 2 & -2 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By row reducing, we obtain

$$\mathrm{RREF}(A_2^1) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\operatorname{rank}\big(A_2^1\big) = 11$. A basis of $\operatorname{Nul}\big(A_2^1\big)$ is

$$\mathcal{N}_{2,1} := \left\{ \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

$r = 2$:

$$A_2^2 = \begin{bmatrix} 4 & 1 & 0 & 12 & 1 & -4 & 8 & 0 & 0 & 1 & 0 & 8 & 0 \\ 5 & 7 & 5 & -9 & -1 & 4 & -6 & 5 & 4 & -1 & 5 & -6 & 0 \\ -6 & -5 & -2 & 25 & 3 & -16 & 18 & -2 & -8 & 7 & -10 & 18 & -4 \\ -2 & -10 & -2 & 28 & -2 & -8 & 16 & -2 & -8 & -2 & -2 & 16 & 0 \\ 0 & -1 & 0 & -4 & 3 & 8 & -4 & 0 & 0 & -1 & 4 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 10 & 2 & -36 & 2 & 8 & -20 & 2 & 8 & 2 & -2 & -20 & 0 \\ 5 & 2 & 5 & -13 & -2 & 16 & -10 & 5 & 4 & -6 & 9 & -10 & 4 \\ -1 & -3 & -1 & 12 & 1 & -4 & 8 & -1 & 0 & 1 & -1 & 8 & 0 \\ -5 & -2 & -5 & 13 & 2 & -12 & 10 & -5 & -4 & 6 & -9 & 10 & -4 \\ 1 & 3 & 1 & -12 & -1 & 4 & -8 & 1 & 4 & -1 & 5 & -8 & 0 \\ 1 & 7 & 1 & 0 & 3 & 4 & 0 & 1 & 4 & 3 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By row reducing, we obtain:

$$\mathrm{RREF}(A_2^2) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and we see that $\mathrm{rank}\big(A_2^2\big) = 10$. A basis of $\mathrm{Nul}\big(A_2^2\big)$ is

$$\mathcal{N}_{2,2} := \left\{ \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Since $\mathrm{rank}\big(A_2^2\big) = 10$,[108] we know that $\mathrm{rank}\big(A_2^2\big) = \mathrm{rank}\big(A_2^3\big) = \mathrm{rank}(A_2^4) = \ldots$, and in particular, we can now stop computing powers of A_2.

We now compute the Jordan blocks of the Jordan matrix J associated with the eigenvalue $\lambda_2 = 2$. This is done precisely as in Example 8.6.9 (using Theorem 8.6.6). In Example 8.6.9, we saw that J contains one Jordan block $J_2(\lambda_2) = J_2(2)$, one Jordan block $J_1(\lambda_2) = J_1(2)$, and no other Jordan blocks of the form $J_t(\lambda_2) = J_t(2)$.

We first generate the Jordan chain corresponding to the Jordan block $J_2(\lambda_2)$. (Note that this Jordan chain will contain two vectors and will be of the form $\{A_2\mathbf{u}, \mathbf{u}\}$.)

[108] Recall that A is of size 13×13, whereas the eigenvalue λ_2 is of algebraic multiplicity 3. So, we need to compute powers of A_2 until we got rank $13 - 3 = 10$.

We extend our basis $\mathcal{N}_{2,1}$ to a basis of $\mathrm{Nul}\big(A_2^2\big)$. We use Proposition 3.3.19. We form the matrix whose columns are the vectors of $\mathcal{N}_{2,1}$, followed by the vectors of $\mathcal{N}_{2,2}$, with a vertical dotted line between the vectors of $\mathcal{N}_{2,1}$ and the vectors of $\mathcal{N}_{2,2}$. The matrix that we obtain is the following:

$$\left[\begin{array}{cc:ccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right].$$

By row reducing, we see that the pivot columns of the matrix above are both columns to the left of the vertical dotted line, plus the second column to the right of the vertical dotted line. So, $\mathcal{N}_{2,1}$ can be extended to a basis of $\mathrm{Nul}\big(A_2^2\big)$ by adding the second vector of $\mathcal{N}_{2,2}$ to it. The second vector of $\mathcal{N}_{2,1}$ is the following vector:

- $\mathbf{w}_1 := \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$.

The vector $\mathbf{w}_1$ will starts the Jordan chain $\{A_2\mathbf{w}_1, \mathbf{w}_1\}$ that corresponds to the Jordan block $J_2(\lambda_2)$. If we place this Jordan chain into a matri, we obtain the following:

$$\begin{bmatrix} A_2\mathbf{w}_1 & \mathbf{w}_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & -1 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

It remains to generate a Jordan chain that corresponds to the Jordan block $J_1(\lambda_2)$ of the Jordan matrix J. This Jordan chain will contain only one vector,

which we find as follows. We first identify the generalized eigenvectors of rank 1 of A associated with the eigenvalue λ_2 inside the Jordan chains (associated with λ_2) that we have already generated. There is exactly one such vector, namely, $A_2\mathbf{w}_1$. Now, $\mathcal{N}_{2,1}^- := \mathcal{N}_{2,0} \cup \{A_2\mathbf{w}_1\} = \{A_2\mathbf{w}_1\}$ is a linearly independent set in $\text{Nul}\big(A_2^1\big)$, and we would like to extend it to a basis of $\text{Nul}\big(A_2^1\big)$.

We once again use Proposition 3.3.19. We form the matrix whose columns are the unique vector of the linearly independent set $\mathcal{N}_{2,1}^- = \{A_2\mathbf{w}_1\}$, followed by the vectors of $\mathcal{N}_{2,1}$, with a vertical dotted line between the vector of $\mathcal{N}_{2,1}^-$ and the vectors of $\mathcal{N}_{2,1}$. The matrix that we obtain is the following:

$$\left[\begin{array}{c:cc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

By row reducing this matrix, we see that its pivot columns are its one column to the left of the vertical dotted line, plus the second column to the right of the vertical dotted line. So, $\mathcal{N}_{2,1}^-$ can be extended to a basis of $\text{Nul}\big(A_2^1\big)$ by adding the second vector of $\mathcal{N}_{2,1}$ to it. The second vector of $\mathcal{N}_{2,1}$ is the following:

- $\mathbf{w}_2 := \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}^T$.

This vector forms the Jordan chain $\{\mathbf{w}_2\}$ that corresponds to the Jordan block $J_1(\lambda_2)$.

We are now done processing the eigenvalue $\lambda_2 = 2$.

The matrices J and P. We now put everything together. Our Jordan matrix is

$$\begin{aligned} J &:= J_5(\lambda_1) \oplus J_2(\lambda_1) \oplus J_2(\lambda_1) \oplus J_1(\lambda_1) \oplus J_2(\lambda_2) \oplus J_1(\lambda_2) \\ &= J_5(4) \oplus J_2(4) \oplus J_2(4) \oplus J_1(4) \oplus J_2(2) \oplus J_1(2) \\ &= \begin{bmatrix} 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}. \end{aligned}$$

Our matrix P consists of the Jordan chains that we have generated, where we make sure that we place the Jordan chain in the order that correspoinds to the order in which we places our Jordan blocks in J. Our matrix P is the following (with vertical dotted lines between different Jordan chains to facilitate reading):

$$\begin{aligned} P &:= \left[\begin{array}{ccccc:cc:cc:c:cc:c} A_1^4\mathbf{u}_1 & A_1^3\mathbf{u}_1 & A_1^2\mathbf{u}_1 & A_1\mathbf{u}_1 & \mathbf{u}_1 & A_1\mathbf{u}_2 & \mathbf{u}_2 & A_1\mathbf{u}_3 & \mathbf{u}_3 & \mathbf{u}_4 & A_2\mathbf{w}_1 & \mathbf{w}_1 & \mathbf{w}_2 \end{array}\right] \\ &= \left[\begin{array}{ccccc:cc:cc:c:cc:c} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ -2 & -1 & 3 & -2 & 0 & 1 & 0 & 2 & 0 & 0 & 1 & 0 & 0 \\ -2 & 0 & 2 & -3 & 0 & 0 & -2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & -2 & 3 & 0 & 0 & 3 & -2 & 0 & 0 & 0 & -1 & 0 \\ 1 & 1 & -2 & 1 & 0 & -1 & 0 & -2 & -1 & 0 & -1 & 0 & 0 \\ -1 & 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ -1 & -1 & 2 & -1 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \end{aligned}$$

Optional: Let us check our anger. We first check that our matrix P is invertible,

and we compute its inverse. Indeed, the calculator tells us that

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 & -3 & 0 & 1 & -2 & 0 & 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & -2 & 0 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & -2 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & -1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -3 & 0 & 0 & -2 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Next, we ask the calculator to compute the product $P^{-1}AP$, and indeed, we obtain our matrix J. So, our answer is correct. □

8.7 Symmetric matrices and orthogonal diagonalization

As usual, the complex conjugate of a complex number z is denoted by $\overline{z}$. Recall from section 6.1 that the *standard scalar product* in $\mathbb{R}^n$ or $\mathbb{C}^{n\times n}$, denoted by $\cdot$, was defined as follows:

- for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$:

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k y_k;$$

- for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{C}^n$:

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k \overline{y_k}.$$

Throughout this section, we shall denote by $||\cdot||$ the norm induced by the standard scalar product $\cdot$ in $\mathbb{R}^n$ or $\mathbb{C}^n$ (as appropriate). In particular, orthogonality and orthonormality will always be assumed to be with respect to the standard scalar product and the induced norm.

8.7.1 Symmetric and Hermitian matrices

For any field $\mathbb{F}$, a matrix $A \in \mathbb{F}^{n\times n}$ is *symmetric* if $A^T = A$. If $\mathbb{F} = \mathbb{C}$, then it turns out that symmetric matrices are less interesting than the so-called "Hermitian

matrices." For a matrix $A = \begin{bmatrix} a_{i,j} \end{bmatrix}_{n\times m}$ in $\mathbb{C}^{n\times m}$, we set $\overline{A} = \begin{bmatrix} \overline{a_{i,j}} \end{bmatrix}_{n\times m}$, i.e. $\overline{A}$ is an $n \times m$ matrix such that for all indices $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, the i, j-th entry of $\overline{A}$ is $\overline{a_{i,j}}$ (the complex conjugate of $a_{i,j}$). The *Hermitian transpose* of a A is the matrix $A^* = (\overline{A})^T$. For example, for

$$A := \begin{bmatrix} -1+i & 3 & 2i \\ 1+2i & 4-2i & 3 \end{bmatrix},$$

we have the following:

$$\overline{A} = \begin{bmatrix} -1-i & 3 & -2i \\ 1-2i & 4+2i & 3 \end{bmatrix}, \qquad A^* = \begin{bmatrix} -1-i & 1-2i \\ 3 & 4+2i \\ -2i & 3 \end{bmatrix}.$$

A square matrix $A \in \mathbb{C}^{n\times n}$ is *Hermitian* if $A^* = A$. For example, the matrix

$$\begin{bmatrix} -1 & 1+i & 2-i \\ 1-i & 2 & -3+i \\ 2+i & -3-i & 0 \end{bmatrix}$$

is Hermitian. Note that all entries on the main diagonal of a Hermitian matrix are real. Note also that if all entries of a matrix in $\mathbb{C}^{n\times n}$ happen to be real, then that matrix is Hermitian if and only if it is symmetric.

Proposition 8.7.1. *For all $\mathbf{x} \in \mathbb{C}^n$, we have that $\mathbf{x}^*\mathbf{x} = ||\mathbf{x}||^2$.*

Proof. For any vector $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$ in $\mathbb{C}^n$, we have that

$$\mathbf{x}^*\mathbf{x} = \begin{bmatrix} \overline{x_1} & \ldots & \overline{x_n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{k=1}^{n} \overline{x_k} x_k = \mathbf{x} \cdot \mathbf{x} = ||\mathbf{x}||^2,$$

which is what we needed. □

The basic algebraic properties of the Hermitian transpose are essentially the same as those of the ordinary transpose, as we can see by comparing Proposition 1.8.1 (which deals with the ordinary transpose) and Proposition 8.7.2 (below).

Proposition 8.7.2. *For all matrices $A, B \in \mathbb{C}^{n\times m}$ and scalars $\alpha \in \mathbb{C}$, the following hold:*

(a) $(A^*)^* = A$;

(b) $(A + B)^* = A^* + B^*$;

(c) $(\alpha A)^* = \overline{\alpha} A^*$;

(d) $(AB)^* = B^* A^*$.

Proof. This follows from the appropriate definitions. The details are left as an easy exercise. □

Theorem 8.7.3. *All eigenvalues of a Hermetrian matrix are real.*

Remark: Recall that the field $\mathbb{C}$ is algebraically closed, and consequently, every matrix in $\mathbb{C}^{n\times n}$ has n complex eigenvalues (with algebraic multiplicities taken into account). So, Theorem 8.7.3 states that if A is a Hermitian matrix in $\mathbb{C}^{n\times n}$, then all n eigenvalues of A (with algebraic multiplicities taken into account) are real.

Proof. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, let λ be any eigenvalue of A, and let $\mathbf{x}$ be an associated eigenvector of A. After possibly normalizing the eigenvector $\mathbf{x}$ (i.e. replacing $\mathbf{x}$ by $\frac{\mathbf{x}}{||\mathbf{x}||}$), we may assume that $\mathbf{x}$ is a unit vector, i.e. that it satisfies $||\mathbf{x}|| = 1$. Then $A\mathbf{x} = \lambda\mathbf{x}$, and we compute:

$$\begin{aligned}
\mathbf{x}^*A\mathbf{x} &= \mathbf{x}^*(\lambda\mathbf{x}) && \text{because } A\mathbf{x} = \lambda\mathbf{x}\\
&= \lambda(\mathbf{x}^*\mathbf{x})\\
&= \lambda||\mathbf{x}||^2 && \text{by Proposition 8.7.1}\\
&= \lambda && \text{because } ||\mathbf{x}|| = 1.
\end{aligned}$$

But now we have the following:

$$\begin{aligned}
\lambda &= \mathbf{x}^*A\mathbf{x}\\
&= \mathbf{x}^*A^*\mathbf{x} && \text{because } A \text{ is Hermitian}\\
&= \mathbf{x}^*A^*(\mathbf{x}^*)^* && \text{by Proposition 8.7.2(a)}\\
&= (\mathbf{x}^*A\mathbf{x})^* && \text{by Proposition 8.7.2(d)}\\
&= \lambda^* && \text{where we consider } \lambda \text{ as a } 1\times 1 \text{ complex matrix}\\
&= \overline{\lambda} && \text{where we consider } \lambda \text{ as a complex number.}
\end{aligned}$$

We have now shown that $\lambda = \overline{\lambda}$, and it follows that λ is a real number. □

Corollary 8.7.4. *Every symmetric matrix in $\mathbb{R}^{n\times n}$ has n real eigenvalues (with algebraic multiplicities taken into account). In other words, for every symmetric matrix $A \in \mathbb{R}^{n\times n}$, the sum of algebraic multiplicities of its distinct (real) eigenvalues is n.*

Proof. Consider any symmetric matrix $A \in \mathbb{R}^{n\times n}$. If we consider A as a matrix in $\mathbb{C}^{n\times n}$, then A is Hermitian, and so Theorem 8.7.3 guarantees that all complex eigenvalues of A are in fact real. Finally, the fact that A has n complex eigenvalues follows from the fact that $\mathbb{C}$ is algebraically closed. □

8.7.2 Orthogonal diagonalizability

Recall from section 6.8 that a matrix $Q \in \mathbb{R}^{n\times n}$ is *orthogonal* if $Q^TQ = I_n$. By Theorem 6.8.1, the following are equivalent for any matrix Q in $\mathbb{R}^{n\times n}$:

- Q is orthogonal;
- Q is invertible and satisfies $Q^{-1} = Q^T$;
- the columns of Q form an an orthonormal basis of $\mathbb{R}^n$.

In what follows, we will repeatedly use the fact that the three statements above are equivalent, without explicitly mentioning Theorem 6.8.1.

Let us say that a matrix $A \in \mathbb{R}^{n\times n}$ is *orthogonally diagonalizable* if there exists a diagonal matrix D and an orthogonal matrix Q, both in $\mathbb{R}^{n\times n}$, such that $D = Q^TAQ$. Since orthogonal matrices Q are invertible and satisfy $Q^T = Q^{-1}$, we see that orthogonally diagonalizable matrices are, in particular, diagonalizable in the usual sense. The main result of this subsection is Theorem 8.7.6, which states that a matrix in $\mathbb{R}^{n\times n}$ is orthogonally diagonalizable if and only if it is symmetric. The proof proceeds by induction on n, and in the induction step, it will be convenient to reduce the problem to the case when the matrix has an eigenvalue 0. To this end, we will use the following technical proposition.

Proposition 8.7.5. *Let $A \in \mathbb{R}^{n\times n}$ and $\lambda_0 \in \mathbb{R}$. Then all the following hold:*

(a) λ_0 is an eigenvalue of A if and only if 0 is an eigenvalue of $A - \lambda_0 I_n$, and moreover, $E_{\lambda_0}(A) = E_0(A - \lambda_0 I_n)$;[109]

(b) A is symmetric if and only if $A - \lambda_0 I_n$ is symmetric;

(c) A is diagonalizable if and only if $A - \lambda_0$ is digonalizable;

(d) A is orthogonally disagonalizable if and only if $A - \lambda_0 I_n$ is orthogonally diagonalizable.

Proof. (a) For all $\mathbf{v} \in \mathbb{R}^n$, we have that $A\mathbf{v} = \lambda_0\mathbf{v}$ if and only if $(A-\lambda_0 I_n)\mathbf{v} = \mathbf{0} = 0\mathbf{v}$, and we deduce that $\mathbf{v} \in E_{\lambda_0}(A)$ if and only if $\mathbf{v} \in E_0(A - \lambda_0 I_n)$. Thus, $E_{\lambda_0}(A) = E_0(A - \lambda_0 I_n)$. In particular, $E_{\lambda_0}(A)$ is non-trivial if and only if $E_0(A - \lambda_0 I_n)$ is

[109] Here, $E_{\lambda_0}(A) = E_0(A - \lambda_0 I_n)$ holds even if λ_0 is not an eigenvalue of A. In that case, we simply have that $E_{\lambda_0}(A) = E_0(A - \lambda_0 I_n) = \{\mathbf{0}\}$.

non-trivial, and consequently (by definition, or alternatively, by Proposition 8.1.6(a)), λ_0 is an eigenvalue of A if and only if 0 is an eigenvalue of $A - \lambda_0 I_n$.

(b) First, we note that $(A - \lambda_0 I_n)^T = A^T - \lambda_0 I_n^T = A^T - \lambda_0 I_n$; consequently, $(A - \lambda_0 I_n)^T = A - \lambda_0 I_n$ if and only if $A^T = A$, i.e. $A - \lambda_0 I_n$ is symmetric if and only if A is symmetric.

(c) Suppose first that A is diagonalizable. Then there exist a diagonal matrix D and an invertible matrix P, both in $\mathbb{R}^{n\times n}$, such that $D = P^{-1}AP$. But then

$$\begin{aligned} P^{-1}(A - \lambda_0 I_n)P &= P^{-1}AP - P^{-1}(\lambda_0 I_n)P \\ &= \underbrace{P^{-1}AP}_{=D} - \lambda_0 \underbrace{P^{-1}P}_{=I_n} \\ &= D - \lambda_0 I_n, \end{aligned}$$

and obviously, $D - \lambda_0 I_n$ is diagonal. So, $A - \lambda_0 I_n$ is diagonalizable. The proof of the converse is analogous.[110]

(d) This is completely analogous to the proof of (c), except that instead of P and P^{-1} (where $P \in \mathbb{R}^{n\times n}$ is an invertible matrix), we have Q and Q^T (where $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix). □

Theorem 8.7.6. *A matrix in $\mathbb{R}^{n\times n}$ is orthogonally diagonalizable if and only if it is symmetric.*

Proof. Let us first show that orthogonally diagonalizable matrices are symmeteric. Fix any orthogonally diagonalizable matrix $A \in \mathbb{R}^{n\times n}$. Let D be a diagonal and Q an orthogonal matrix, both in $\mathbb{R}^{n\times n}$, such that $D = Q^T AQ$. Then $A = QDQ^T$, and we see that

$$A^T = (QDQ^T)^T = (Q^T)^T D^T Q^T \overset{(*)}{=} QDQ^T = A,$$

where in (*), we used the fact that $D^T = D$, since D is diagonal. Thus, A is symmetric.

It remains to prove the reverse implication: symmetric matrices in $\mathbb{R}^{n\times n}$ are orthogonally diagonalizable. We proceed by induction on n.

For $n = 1$, the result is immediate: indeed, if $A \in \mathbb{R}^{1\times 1}$, then A is diagonal, and we can take $D := A$ and $Q := I_1$ to obtain $D = Q^T AQ$.[111]

Now, fix a positive integer n, and assume inductively that every symmetric matrix in $\mathbb{R}^{n\times n}$ is orthogonally diagonalizable. Fix any symmetric matrix $A \in \mathbb{R}^{(n+1)\times(n+1)}$;

[110] Indeed, set $A' := A - \lambda_0 I_n$ and $\lambda_0' := -\lambda_0$. Then an argument completely analogous to the above shows that if $A' = A - \lambda_0 I_n$ is diagonalizable, then so is $A' - \lambda_0' I_n = A$.

[111] Obviously, I_1 is orthogonal.

we must show that A is orthogonally diagonalizable. By Corollary 8.7.4, A has $n+1$ real eigenvalues (with algebraic multiplicities taken into account). Let $\lambda_0 \in \mathbb{R}$ be an eigenvalue of A. In view of Proposition 8.7.5, we may assume that $\lambda_0 = 0$, for otherwise, we simply consider $A - \lambda_0 I_n$ instead of A.[112] Let $\mathbf{x}_0 \in \mathbb{R}^n$ be an eigenvector of A associated with the eigenvalue 0, so that $A\mathbf{x}_0 = \mathbf{0}$. After possibly normalizing the eigenvector $\mathbf{x}_0$ (i.e. replacing $\mathbf{x}_0$ by $\frac{\mathbf{x}_0}{||\mathbf{x}_0||}$), we may assume that $||\mathbf{x}_0|| = 1$. Now, using Corollary 6.3.11(d), we let $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be an orthonormal basis of $\mathbb{R}^{n+1}$.[113] Set $S := \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \ldots & \mathbf{x}_n \end{bmatrix}$; then S is an orthogonal matrix. Now, since A is symmetric, so is $S^T AS$; indeed,

$$\left(S^T AS\right)^T = S^T A^T S \overset{(*)}{=} S^T AS,$$

where in (*), we used the fact that $A^T = A$ (since A is symmetric). Moreover, it is easy to see that the first (i.e. leftmost) column of $S^T AS$ is $\mathbf{0}$; indeed:

$$\begin{aligned} S^T AS &= S^T A \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \ldots & \mathbf{x}_n \end{bmatrix} \\ &= \begin{bmatrix} S^T A\mathbf{x}_0 & S^T A\mathbf{x}_1 & \ldots & S^T A\mathbf{x}_n \end{bmatrix} && \text{by the definition of matrix multiplication} \\ &= \begin{bmatrix} S^T \mathbf{0} & S^T A\mathbf{x}_1 & \ldots & S^T A\mathbf{x}_n \end{bmatrix} && \text{because } A\mathbf{x}_0 = \mathbf{0} \\ &= \begin{bmatrix} \mathbf{0} & S^T A\mathbf{x}_1 & \ldots & S^T A\mathbf{x}_n \end{bmatrix}. \end{aligned}$$

We now know that $S^T AS \in \mathbb{R}^{(n+1)\times(n+1)}$ is a symmetric matrix, and that its leftmost column is $\mathbf{0}$. So, there exists a symmetric matrix $A_0 \in \mathbb{R}^{n\times n}$ such that

$$S^T AS = \left[\begin{array}{c|c} 0 & \mathbf{0}^T \\ \hline \mathbf{0} & A_0 \end{array}\right].$$

By the induction hypothesis, A_0 is orthogonally diagonalizable, i.e. there exist a diagonal matrix D_0 and an orthogonal matrix Q_0, both in $\mathbb{R}^{n\times n}$, such that $D_0 = Q_0^T A_0 Q_0$. Now, set

- $D := \left[\begin{array}{c|c} 0 & \mathbf{0}^T \\ \hline \mathbf{0} & D_0 \end{array}\right]_{(n+1)\times(n+1)}$;
- $R := \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & Q_0 \end{array}\right]_{(n+1)\times(n+1)}$.

[112] Let us explain this in more detail. By Proposition 8.7.5(a-b), $A - \lambda_0 I_n$ is symmetric and has eigenvalue 0. On the other hand, if we can show that $A - \lambda_0 I_n$ is orthogonally diagonalizable, then Proposition 8.7.5(d) will guarantee that A is also orthogonally diagonalizable. So, we may consider $A - \lambda_0 I_n$ and 0 instead of A and λ_0, respectively.

[113] Indeed, $\{\mathbf{x}_0\}$ is an orthonormal basis of the subspace $U := \text{Span}(\mathbf{x}_0)$ of $\mathbb{R}^{n+1}$, and so by Corollary 6.3.11(d), $\{\mathbf{x}_0\}$ can be extended to an orthonormal basis of $\mathbb{R}^{n+1}$.

Clearly, D is diagonal (because D_0 is diagonal), and R is orthogonal (because Q_0 is orthogonal).[114] Since R and S are orthogonal, Proposition 6.8.3 guarantees that $Q := SR$ is also orthogonal. Finally, we compute:

$$\begin{aligned} Q^T A Q &= (SR)^T A (SR) \\ &= R^T (S^T A S) R \\ &= \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & Q_0^T \end{array}\right] \left[\begin{array}{c|c} 0 & \mathbf{0}^T \\ \hline \mathbf{0} & A_0 \end{array}\right] \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & Q_0 \end{array}\right] \\ &= \left[\begin{array}{c|c} 0 & \mathbf{0}^T \\ \hline \mathbf{0} & Q_0^T A_0 Q_0 \end{array}\right] \\ &= \left[\begin{array}{c|c} 0 & \mathbf{0}^T \\ \hline \mathbf{0} & D_0 \end{array}\right] \\ &= D, \end{aligned}$$

and we are done. □

8.7.3 The spectral theorem for symmetric matrices

For a matrix $A \in \mathbb{R}^{n \times n}$, an *orthogonal eigenbasis* (resp. *orthonormal eigenbasis*) of $\mathbb{R}^n$ associated with A is an orthogonal (resp. orthonormal) basis of $\mathbb{R}^n$, all of whose vectors are eigenvectors of A. Theorem 8.7.6, combined with what we know about diagonalization and orthogonal matrices, readily yields the following theorem.

The spectral theorem for symmetric matrices. *For every matrix $A \in \mathbb{R}^{n \times n}$, the following are equivalent:*

(a) A is symmetric;

(b) A is orthogonally diagonalizable;

(c) $\mathbb{R}^n$ has an orthonormal eigenbasis associated with A;

(d) $\mathbb{R}^n$ has an orthogonal eigenbasis associated with A;

(e) $\mathbb{R}^n$ has an eigenbasis associated with A, and the eigenspaces of A are pairwise orthogonal;

[114] Indeed,

$$R^T R = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & Q_0^T \end{array}\right] \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & Q_0 \end{array}\right] = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & Q^T Q_0 \end{array}\right] \overset{(*)}{=} \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & I_n \end{array}\right] = I_{n+1},$$

where (*) follows from the fact that Q is orthogonal.

(f) A has n pairwise orthogonal eigenvectors.[115]

Proof. Fix a matrix A. To simplify notation, for each eigenvalue λ of A, we set $E_\lambda := E_\lambda(A)$. To prove the theorem, it is enough to prove the implications shown in the diagram below.

$$\begin{array}{ccccc} (a) & \Longleftrightarrow & (b) & \Longleftarrow & (c) \\ & & \Downarrow & & \Uparrow \\ & & (e) & \Longrightarrow & (d) \end{array} \quad \Longleftrightarrow \quad (f)$$

The fact that (a) and (b) are equivalent follows from Theorem 8.7.6. Let us prove that (d) and (f) are equivalent. Since since any basis of $\mathbb{R}^n$ contains exactly n vectors, it is clear that (d) implies (f). Let us now assume (f) and prove (d). Using (f), we fix pairwise orthogonal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of A. By the definition of an eigenvector, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are all non-zero. So, by Proposition 6.3.4(a), $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal basis of $\mathbb{R}^n$. But now $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal eigenbasis of $\mathbb{R}^n$ associated with A, and we see that (d) holds.

Next, we assume (b) and prove (e). Using (b), we fix a diagonal matrix D and an orthogonal matrix Q, both in $\mathbb{R}^{n\times n}$, such that $D = Q^T AQ = Q^{-1}AQ$. By Proposition 8.5.12, the columns of Q form an eigenbasis of $\mathbb{R}^n$ associated with A, and since Q is orthogonal, this basis is orthonormal. It remains to show that the eigenspaces of A are pairwise orthogonal. So, suppose that λ_1 and λ_2 are distinct eigenvalues of A; we must show that $E_{\lambda_1} \perp E_{\lambda_2}$. By Proposition 8.5.12, the eigenvalues of A are precisely the entries on the main diagonal of D, and in particular, both λ_1 and λ_2 appear on the main diagonal of D. Now, suppose the eigenvalue λ_1 appears (precisely) in entries $i_1, \ldots, i_{k_1}$ of the main diagonal of D; then by Proposition 8.5.12, columns number $i_1, \ldots, i_{k_1}$ of Q form a basis $\mathcal{B}_1$ of E_{λ_1}. Similarly, suppose that the eigenvalue λ_2 appears (precisely) in entries $j_1, \ldots, j_{k_2}$ of the main diagonal of D; then by Proposition 8.5.12, columns number $j_1, \ldots, j_{k_2}$ of Q form a basis $\mathcal{B}_2$ of E_{λ_2}. But since Q is orthogonal, we know that its columns form an orthonormal basis of $\mathbb{R}^n$. In particular, $\mathcal{B}_1 \perp \mathcal{B}_2$. Proposition 6.1.5 then implies that $\mathrm{Span}(\mathcal{B}_1) \perp \mathrm{Span}(\mathcal{B}_2)$, that is, $E_{\lambda_1} \perp E_{\lambda_2}$. This proves (e).

Next, we assume (e) and prove (d). Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of A, and for all $i \in \{1, \ldots, k\}$, let $\mathcal{B}_i$ be a basis of E_{λ_i}; after possibly applying the Gram-Schmidt orthogonalization procedure to $\mathcal{B}_i$, we may assume that $\mathcal{B}_i$ is orthogonal. By (e), $\mathbb{R}^n$ has an eigenbasis associated with A, and so by Proposition 8.4.5(c), the sum of geometric multiplicities of A is n, and moreover, $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ is an eigenbasis of $\mathbb{R}^n$ associated with A. Since the eigenspaces of A are pairwise orthogonal, we see that $\mathcal{B}_1, \ldots, \mathcal{B}_k$ are orthogonal to each other. Since $\mathcal{B}_1, \ldots, \mathcal{B}_k$

[115] This means that some n eigenvectors of A are pairwise orthogonal. It does **not** mean that A has exactly n eigenvectors (which happen to be orthogonal).

are orthognal sets of vectors, we deduce that the eigenbasis $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ is orthogonal, i.e. (d) holds.

Next, we assume (d) and prove (c). Using (d), we fix an orthogonal eigenbasis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ associated with A. For all $i \in \{1, \ldots, n\}$, set $\mathbf{u}_i := \frac{\mathbf{v}_i}{||\mathbf{v}_i||}$; since $\mathbf{v}_i$ is an eigenvector of A associated with the eigenvalue λ_i, so is $\mathbf{u}_i$. In view of Proposition 6.3.3(c), we now deduce that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal eigenbasis of $\mathbb{R}^n$ associated with A. This proves (c).

Finally, we assume (c) and prove (b). Using (c), we fix an orthonormal eigenbasis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ associated with A. For each $i \in \{1, \ldots, n\}$, we let λ_i be the eigenvalue of A associated with the eigenvector $\mathbf{v}_i$. Set $D := D(\lambda_1, \ldots, \lambda_n)$ and $Q := \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{bmatrix}$. By Theorem 8.5.6, A is diagonalizable and satisfies $D = Q^{-1}AQ$. But since the columns of Q form an orthonormal basis of $\mathbb{R}^n$, we see that the matrix Q is orthogonal and therefore satisfies $Q^{-1} = Q^T$. Thus, $D = Q^T AQ$, and (b) follows. □

8.7.4 Diagonally orthogonalizing a symmetric matrix

By Theorem 8.7.6, every symmetric matrix in $\mathbb{R}^{n\times n}$ can be orthogonally diagonalized, and in fact, the proof of Theorem 8.7.6 gives us a recipe of sorts for orthogonally diagonalizing such a matrix. However, that recipe is not particularly practical, and we are better off using the spectral theorem from subsection 8.7.3 instead. Suppose we are given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, which we wish to orthogonally diagonalize. So, our goal is to construct a diagonal matrix D and an orthogonal matrix Q, both in $\mathbb{R}^{n\times n}$, such that $D = Q^T AQ$. We proceed as follows. First, we compute the characteristic polynomial of A, we factor it, and we find all the (real) eigenvalues of A along with their algebraic multiplicities. Since A is orthogonally diagonalizable (and in particular, diagonalizable), Theorems 8.4.5(d) and 8.5.6 together guarantee that $\mathbb{R}^n$ has an eigenbasis associated with A, and moreover, that the sum of algebraic multiplicities of the eigenvalues of A is n, and that the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity. Next, for each eigenvalue λ of A, we compute a basis $\mathcal{B}_\lambda$ of the eigenspace $E_\lambda(A)$, and then we apply the Gram-Schmidt orthogonalization process to $\mathcal{B}_\lambda$ in order to obtain an orthonormal basis $\mathcal{C}_\lambda$ of $E_\lambda(A)$. In view of the spectral theorem, we see that the union $\mathcal{C}$ of the $\mathcal{C}_\lambda$'s is an orthonormal eigenbasis of $\mathbb{R}^n$ associated with A.[116] We now form the diagonal matrix D by placing the eigenvalues of A on the main diagonal of D (while

[116]Let us justify this in more detail. Since A is diagonalizable, Theorems 8.4.5(d) and 8.5.6 together guarantee that $\mathcal{C}$ is an eigenbasis of $\mathbb{R}^n$ associated with A. Since $\mathcal{C}$ is the union of the $\mathcal{C}_\lambda$'s, and since each $\mathcal{C}_\lambda$ is an orthonormal set of vectors, it only remains to explain why the $\mathcal{C}_\lambda$'s are orthogonal to each other. But this follows immediately from the spectral theorem (see subsection 8.7.3): since A is symmetric, its eigenspaces are orthogonal to each other, and in particular, the bases $\mathcal{C}_\lambda$ of those eigenspaces are orthogonal to each other.

respecting the algebraic/geometric multiplicity of each eigenvalue),[117] and we form Q by arranging the vectors of our orthonormal eigenbasis $\mathcal{C}$ into a matrix (while respecting the order from D).[118] Since the columns of Q form an orthonormal basis of $\mathbb{R}^n$, we see that Q is orthogonal, and so $Q^{-1} = Q^T$. But now Theorem 8.5.6 guarantees that $D = Q^{-1}AQ = Q^TAQ$.

Example 8.7.7. *Orthogonally diagonalize the following symmetric matrix in* $\mathbb{R}^{3\times 3}$*:*

$$A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}.$$

Proof. First, we compute the characteristic polynomial of A:

$$\begin{aligned} p_A(\lambda) &= \det(\lambda I_3 - A) \\ &= \begin{vmatrix} \lambda-3 & 2 & -4 \\ 2 & \lambda-6 & -2 \\ -4 & -2 & \lambda-3 \end{vmatrix} \\ &= \lambda^3 - 12\lambda^2 + 21\lambda + 98 \\ &= (\lambda+2)(\lambda-7)^2. \end{aligned}$$

Thus, A has two eigenvectors: $\lambda_1 = -2$ (with algebraic multiplicity 1) and $\lambda_2 = 7$ (with algebraic multiplicity 2). We now compute a basis $\mathcal{B}_1 = \{\begin{bmatrix} -2 & -1 & 2 \end{bmatrix}^T\}$ of $E_{\lambda_1}(A)$ and a basis $\mathcal{B}_2 = \{\begin{bmatrix} -1 & 2 & 0 \end{bmatrix}^T, \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T\}$ of $E_{\lambda_2}(A)$. Next, we apply the Gram-Schmidt orthogonalization process to $\mathcal{B}_1$ and $\mathcal{B}_2$. This yields an orthonormal basis $\mathcal{C}_1 = \{\begin{bmatrix} -\frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}^T\}$ of E_{λ_1}, and an orthonormal basis $\mathcal{C}_2 = \{\begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0 \end{bmatrix}^T, \begin{bmatrix} \frac{4}{3\sqrt{5}} & \frac{2}{3\sqrt{5}} & \frac{5}{3\sqrt{5}} \end{bmatrix}^T\}$ of E_{λ_2}. We now set

$$D := \begin{bmatrix} -2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{bmatrix} \quad \text{and} \quad Q := \begin{bmatrix} -2/3 & -1/\sqrt{5} & 4/(3\sqrt{5}) \\ -1/3 & 2/\sqrt{5} & 2/(3\sqrt{5}) \\ 2/3 & 0 & 5/(3\sqrt{5}) \end{bmatrix}.$$

Now D is diagonal, Q is orthogonal, and $D = Q^TAQ$. □

[117] In other words, if $\{\lambda_1, \ldots, \lambda_n\}$ is the spectrum of A, then we set $D := D(\lambda_1, \ldots, \lambda_n)$. Recall that the geometric multiplicity of any eigenvalue of A is equal to the algebraic multiplicity of that eigenvalue.

[118] So, if an eigenvalue λ of A has algebraic/geometric multiplicity k and appears (precisely) in entries $i_1, \ldots, i_k$ of the main diagonal of D, then columns number $i_1, \ldots, i_k$ of Q should be the vectors of the orthonormal basis $\mathcal{C}_\lambda$ of the eigenspace $E_\lambda(A)$.