Linear Algebra 1: Lecture 4

Irena Penev

Winter 2022/2023

In what follows, $\mathbb{F}$ is a fixed field. For now, you may assume that $\mathbb{F}$ is $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p$ (for some prime p). However, everything that we prove in this lecture also works for general fields $\mathbb{F}$.

1 Linear transformations from $\mathbb{F}^m$ to $\mathbb{F}^n$

A function $f : \mathbb{F}^m \to \mathbb{F}^n$ is a *linear function* (or a *linear transformation*) if it satisfies the following two properties:

1. for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{F}^m$, we have $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$;
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 1.1. *Let $A \in \mathbb{F}^{n\times m}$, and define $f : \mathbb{F}^m \to \mathbb{F}^n$ by setting $f(\mathbf{u}) = A\mathbf{u}$ for all $\mathbf{u} \in \mathbb{F}^m$. Then f is a linear transformation.*

Proof. By Corollary 2.2 from Lecture Notes 3, we have that

(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
$$f(\mathbf{u}+\mathbf{v}) \;=\; A(\mathbf{u}+\mathbf{v}) \;\overset{(i)}{=}\; A\mathbf{u}+A\mathbf{v} \;=\; f(\mathbf{u})+f(\mathbf{v});$$
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. □

Functions given by the formula $f(\mathbf{u}) = A\mathbf{u}$, where A is some matrix, are sometimes called *matrix transformations*. By Proposition 1.1, every matrix transformation is a linear transformation. Interestingly, a reverse of sorts also holds, as we now show.

Theorem 1.2. *Let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear transformation. Then there exists a unique matrix A (called the* standard matrix of f*) such that for all $\mathbf{u} \in \mathbb{F}^m$, we have that $f(\mathbf{u}) = A\mathbf{u}$. Moreover, the standard matrix A of f is given by*

$$A \;=\; \begin{bmatrix} f(\mathbf{e}_1) & \dots & f(\mathbf{e}_m) \end{bmatrix},$$

where $\mathbf{e}_1, \dots, \mathbf{e}_m$ are the standard basis vectors of $\mathbb{F}^m$.

Proof. We first show that $A = \begin{bmatrix} f(\mathbf{e}_1) & \dots & f(\mathbf{e}_m) \end{bmatrix}$ has the property that for all $\mathbf{u} \in \mathbb{F}^m$, we have that $f(\mathbf{u}) = A\mathbf{u}$. Fix any $\mathbf{u} \in \mathbb{F}^m$, and set $\mathbf{u} = \begin{bmatrix} u_1 & \dots & u_m \end{bmatrix}^T$. Then $\mathbf{u} = \sum_{k=1}^{m} u_k \mathbf{e}_k$. We now compute:

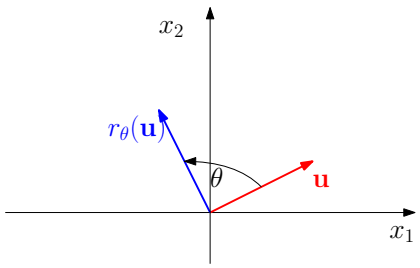
$$\begin{aligned}
A\mathbf{u} &= A\Big(\sum_{k=1}^{m} u_k \mathbf{e}_k\Big) \\
&= \sum_{k=1}^{m} A(u_k \mathbf{e}_k) && \text{because matrix transformations are linear} \\
&= \sum_{k=1}^{m} u_k (A\mathbf{e}_k) && \text{because matrix transformations are linear} \\
&= \sum_{k=1}^{m} u_k f(\mathbf{e}_k) && \text{because } f(\mathbf{e}_k) \text{ is the } k\text{-th column of } A \\
&= \sum_{k=1}^{m} f(u_k \mathbf{e}_k) && \text{because } f \text{ is linear} \\
&= f\Big(\sum_{k=1}^{m} u_k \mathbf{e}_k\Big) && \text{because } f \text{ is linear} \\
&= f(\mathbf{u}).
\end{aligned}$$

It remains to show that A is the only matrix with the desired property. So, fix any matrix B such that for all $\mathbf{u} \in \mathbb{F}^m$, we have that $f(\mathbf{u}) = B\mathbf{u}$. We must show that $A = B$. Since $f : \mathbb{F}^m \to \mathbb{F}^m$, it is clear that $B \in \mathbb{F}^{n \times m}$, and so A and B are of the same size. So, it is enough to show that the corresponding columns of A and B are the same. Set $B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_m \end{bmatrix}$. Now for any $j \in \{1, \dots, m\}$, we have that

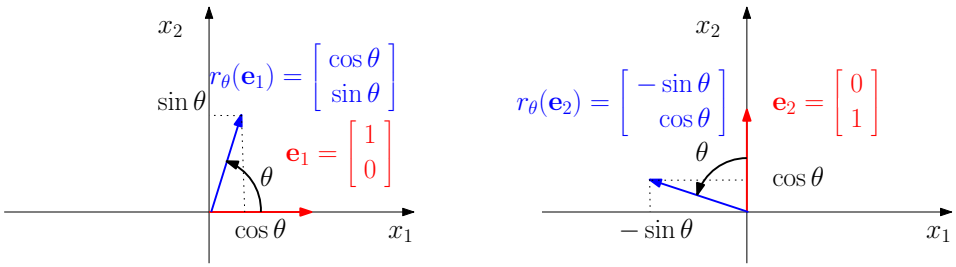
$$\mathbf{b}_j \;=\; B\mathbf{e}_j \;=\; f(\mathbf{e}_j),$$

and by construction, $f(\mathbf{e}_j)$ is the j-th column of A. So, the corresponding columns of A and B are the same, and it follows that $A = B$. □

Example 1.3. *Find the standard matrix of the linear transformation r_θ that rotates vectors $\mathbb{R}^2$ by the angle θ (counterclockwise). You may assume this function is linear.*



Solution. Note that $r_\theta(\mathbf{e}_1) = r_\theta\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and $r_\theta(\mathbf{e}_2) = r_\theta\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ (see the picture below).



So, by Theorem 1.2, the standard matrix of r_θ is

$$\begin{bmatrix} r_\theta(\mathbf{e}_1) & r_\theta(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

□

Proposition 1.4. *Let $f : \mathbb{F}^m \to \mathbb{F}^n$ be a linear transformation. Then $f(\mathbf{0}) = \mathbf{0}$.*[1]

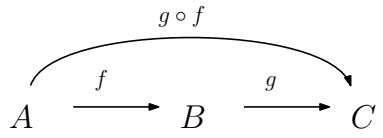
Proof. We observe that

$$f(\mathbf{0}) = f(0 \cdot \mathbf{0}) \overset{(*)}{=} 0f(\mathbf{0}) = \mathbf{0},$$

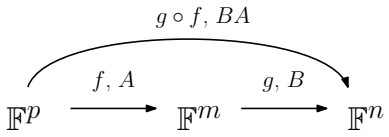
where (*) follows from the fact that f is linear. □

Given functions $f : A \to B$ and $g : B \to C$, we define the *composition of functions* f and g to be the function $h : A \to C$ given by $h(a) = g(f(a))$ for all $a \in A$. The function h is denoted by $g \circ f$ (see the diagram below).

[1] Note that the $\mathbf{0}$ from $f(\mathbf{0})$ is the zero vector in $\mathbb{F}^m$, whereas the $\mathbf{0}$ on the right-hand-side of the equation $f(\mathbf{0}) = \mathbf{0}$ is the zero vector in $\mathbb{F}^n$. So, the two zero vectors aren't actually the same (unless $m = n$). Furthermore, 0 (from the proof of Proposition 1.4) is the zero element of the field $\mathbb{F}$.



Proposition 1.5. *Let $f : \mathbb{F}^p \to \mathbb{F}^m$ and $g : \mathbb{F}^m \to \mathbb{F}^n$ be linear transformations, and let A and B, respectively, be their standard matrices. Then $g \circ f : \mathbb{F}^p \to \mathbb{F}^n$ is a linear transformation, and its standard matrix is BA. (See the diagram below.)*



Proof. For any $\mathbf{u} \in \mathbb{F}^p$, we have that

$$(g \circ f)(\mathbf{u}) \;=\; g(f(\mathbf{u})) \;\overset{(*)}{=}\; g(A\mathbf{u}) \;\overset{(**)}{=}\; B(A\mathbf{u}) \;=\; (BA)\mathbf{u}.$$

where (*) follows from the fact that A is the standard matrix of f, and (**) follows from the fact that B is the standard matrix of g. We have now shown that $g \circ f$ is a matrix transformation, and so (by Proposition 1.1) it is linear. Moreover, since (by the calculation above) we have that $(g \circ f)(\mathbf{u}) = (BA)\mathbf{u}$ for all vectors $\mathbf{u} \in \mathbb{F}^n$, we see that BA is the standard matrix of $g \circ f$. □

1.1 Bijections and inverse functions

For a set X, we define the function $Id_X : X \to X$ by setting $Id_X(x) = x$ for all $x \in X$. Id_X is called the *identity function* on X.

A function $f : A \to B$ is

- *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 $f(a_1) \neq f(a_2)$;[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 1.6. *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 $f \circ g = Id_B$.

[2]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 $a_1 = a_2$.

Proof. Suppose first that (a) holds. Then for all $b \in B$, there exists a unique $a \in A$ such that $f(a) = b$.[3] We now define $g : B \to A$ by letting, for each $b \in B$, $g(b)$ be the unique $a \in A$ such that $f(a) = b$. Then clearly, $g \circ f = Id_A$ and $f \circ g = Id_B$.[4]

Suppose now that (b) holds, and fix a function $g : B \to A$ such that $g \circ f = Id_A$ and $f \circ g = Id_B$. We first show that A is one-to-one. Fix distinct $a_1, a_2 \in A$, and suppose that $f(a_1) = f(a_2)$. Then

$$\begin{aligned}
a_1 &= Id_A(a_1) \\
&= (g \circ f)(a_1) && \text{because } g \circ f = Id_A \\
&= g(f(a_1)) \\
&= g(f(a_2)) && \text{because } f(a_1) = f(a_2) \\
&= (g \circ f)(a_2) \\
&= Id_A(a_2) && \text{because } g \circ f = Id_A \\
&= a_2.
\end{aligned}$$

So, f is one-to-one. We now show that f is onto. Fix $b \in B$, and set $a = g(b)$. Then

$$f(a) \;=\; f(g(b)) \;=\; (f \circ g)(b) \;=\; Id_B(b) \;=\; b.$$

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 1.7. *Let $f : A \to B$ be a bijection. Then there exists a unique function $f : B \to A$ such that $g \circ f = Id_A$ and $f \circ g = Id_B$.*

Proof. The existence of g follows immediately from Proposition 1.6. It remains to prove uniqueness. So, suppose that functions $g_1, g_2 : B \to A$ satisfy

- $g_1 \circ f = Id_A$ and $f \circ g_1 = Id_B$;
- $g_2 \circ f = Id_A$ and $f \circ g_2 = Id_B$.

[3] 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.

[4] 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 $(f \circ g)(b) = f(g(b)) = f(a) = b$.

We must show that $g_1 = g_2$. Fix $b \in B$. Since f is one-to-one, there exists a unique $a \in A$ such that $f(a) = b$. We now have that

$$\begin{aligned} g_1(b) &= g_1(f(a)) && \text{because } f(a) = b \\ &= (g_1 \circ f)(a) \\ &= Id_A(a) && \text{because } g_1 \circ f = Id_A \\ &= (g_2 \circ f)(a) && \text{because } g_2 \circ f = Id_A \\ &= g_2(f(a)) \\ &= g_2(b) && \text{because } f(a) = b. \end{aligned}$$

So, $g_1 = g_2$. □

If $f : A \to B$ is a bijection, then the unique function $g : B \to A$ that satisfies $g \circ f = Id_A$ and $f \circ g = Id_B$,[5] is called the *inverse* of f and is denoted by f^{-1}. Note that this means that:

- $f^{-1} \circ f = Id_A$;
- $f \circ f^{-1} = Id_B$;
- for all $a \in A$ and $b \in B$, we have that $b = f(a)$ if and only if $a = f^{-1}(b)$.

Proposition 1.8. *Let $f : \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation. If f is a bijection, then its inverse $f^{-1} : \mathbb{F}^n \to \mathbb{F}^n$ is also linear.*

Proof. First, fix $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{F}^n$. We must show that $f^{-1}(\mathbf{v}_1 + \mathbf{v}_2) = f^{-1}(\mathbf{v}_1) + f^{-1}(\mathbf{v}_2)$. Fix $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{F}^n$ such that $\mathbf{v}_1 = f(\mathbf{u}_1)$ and $\mathbf{v}_2 = f(\mathbf{u}_2)$. Then

$$\begin{aligned} f^{-1}(\mathbf{v}_1 + \mathbf{v}_2) &= f^{-1}(f(\mathbf{u}_1) + f(\mathbf{u}_2)) \\ &\overset{(*)}{=} f^{-1}(f(\mathbf{u}_1 + \mathbf{u}_2)) \\ &= \mathbf{u}_1 + \mathbf{u}_2 \\ &= f^{-1}(\mathbf{v}_1) + f^{-1}(\mathbf{v}_2), \end{aligned}$$

where (*) follows from the fact that f is linear.

[5] So, g is the function from Proposition 1.7.

Next, fix $\mathbf{v} \in \mathbb{F}^n$ and $\alpha \in \mathbb{F}$. We must show that $f^{-1}(\alpha\mathbf{v}) = \alpha f(\mathbf{v})$. Fix $\mathbf{u} \in \mathbb{F}^n$ such that $\mathbf{v} = f(\mathbf{u})$. Then

$$\begin{aligned} f^{-1}(\alpha\mathbf{v}) &= f^{-1}(\alpha f(\mathbf{u})) \\ &\overset{(*)}{=} f^{-1}(f(\alpha\mathbf{u})) \\ &= \alpha\mathbf{u} \\ &= \alpha f^{-1}(\mathbf{v}), \end{aligned}$$

where (*) follows from the fact that f is linear.

We have now proven that f^{-1} linear. □

2 Invertible and non-invertible matrices

A matrix $A \in \mathbb{F}^{n\times n}$ is *invertible* if there exists a matrix $B \in \mathbb{F}^{n\times n}$ such that $AB = BA = I_n$. If such a matrix B exist, then it is unique (see Proposition 2.1(b) below), and it is denoted by A^{-1}. A square matrix that is not invertible is called *non-invertible*.

Proposition 2.1. *Let $A, B, C \in \mathbb{F}^{n\times n}$.*

(a) If $AB = I_n$ and $CA = I_n$, then $B = C$.[6]

(b) If $AB = BA = I_n$ and $AC = CA = I_n$, then $B = C$.

Proof. Obviously, (a) implies (b). For (a), we assume that $AB = I_n$ and $CA = I_n$, and we observe that

$$C = CI_n = C(AB) = (CA)B = I_nB = B,$$

which is what we needed. □

Note that Proposition 2.1 guarantees that if $A, B, C \in \mathbb{F}^{n\times n}$ are such that $AB = I_n$ and $CA = I_n$, then A is invertible, and $A^{-1} = B = C$.

The following theorem describes a procedure for determining whether a square matrix is invertible, and if so, finding its inverse.

Theorem 2.2. *Let $A \in \mathbb{F}^{n\times n}$, and let $\left[\begin{array}{c|c} U & B \end{array}\right] = RREF(\left[\begin{array}{c|c} A & I_n \end{array}\right])$, 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.

[6] This means that if A has a "right inverse" B and a "left inverse" C, then $B = C$.

Proof. Later! □

Example 2.3. *Consider the following matrices.*

(a) $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, *with entries understood to be in* $\mathbb{R}$*;*

(b) $B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, *with entries understood to be in* $\mathbb{Z}_2$*;*

(c) $C = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 2 & 0 & 1 \end{bmatrix}$, *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) The reduced row echelon form of the matrix

$$\left[\begin{array}{c|c} A & I_2 \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right]$$

(with entries in $\mathbb{R}$) is

$$\left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right]$$

The submatrix to the left of the dotted line is I_2. So, A is invertible, and

$$A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}.$$

(b) The reduced row echelon form of the matrix

$$\left[\begin{array}{c|c} B & I_3 \end{array}\right] = \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

(with entries in $\mathbb{Z}_2$) is

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

The submatrix to the left of the dotted line is I_3. So, B is invertible, and

$$B^{-1} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

(c) The reduced row echelon form of the matrix

$$\left[\begin{array}{c:c} C & I_3 \end{array}\right] = \left[\begin{array}{ccc:ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

(with entries in $\mathbb{Z}_3$) is

$$\left[\begin{array}{ccc:ccc} 1 & 0 & 2 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 2 \end{array}\right].$$

The submatrix to the left of the dotted line is not I_3. So, C is **not** invertible. □

Theorem 2.4. *Let $f : \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation, and let A be its standard matrix. Then f is a bijection if and only if A is invertible. Moreover, if f is a bijection, then $f^{-1} : \mathbb{F}^n \to \mathbb{F}^n$ is linear, and the standard matrix of f^{-1} is A^{-1}.*

Proof. Suppose first that f is a bijection. By Proposition 1.8, f^{-1} is linear. Let B be the standard matrix of f^{-1}. Since f and f^{-1} are linear, Proposition 1.5 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 $f^{-1} \circ f = f \circ f^{-1} = Id_{\mathbb{F}^n}$, and clearly, the standard matrix of $Id_{\mathbb{F}^n}$ is I_n. So, $AB = BA = I_n$. But now A is invertible and $B = A^{-1}$.

Suppose now that A is invertible. Define $g : \mathbb{F}^n \to \mathbb{F}^n$ by setting $g(\mathbf{u}) = A^{-1}\mathbf{u}$ for all $\mathbf{u} \in \mathbb{F}^n$. Our goal is to show that $f \circ g = g \circ f = Id_{\mathbb{F}^n}$. In view of Proposition 1.6, this will imply that f is a bijection. For any $\mathbf{u} \in \mathbb{F}^n$, we have that

- $(f \circ g)(\mathbf{u}) = f(g(\mathbf{u})) = A(A^{-1}\mathbf{u}) = (AA^{-1})\mathbf{u} = I_n\mathbf{u} = \mathbf{u}$;
- $(g \circ f)(\mathbf{u}) = g(f(\mathbf{u})) = A^{-1}(A\mathbf{u}) = (A^{-1}A)\mathbf{u} = I_n\mathbf{u} = \mathbf{u}$.

This proves that $f \circ g = g \circ f = Id_{\mathbb{F}^n}$, and it follows that f is a bijection. □

Proposition 2.5. *Let $A \in \mathbb{F}^n$ be an invertible matrix. Then A^{-1} is invertible and $(A^{-1})^{-1} = A$.*

Proof. Since $AA^{-1} = A^{-1}A = I_n$, we see that A^{-1} is invertible that that its inverse is A. □

Proposition 2.6. *Let $A, B \in \mathbb{F}^n$ be invertible matrices. Then AB is invertible. Moreover, $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. 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):

- $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$;
- $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n$.

□

3 Elementary matrices

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 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 3.1. *Let R be some elementary row operation performed on a matrix with n rows, and let $E \in \mathbb{F}^{n\times n}$ be the elementary matrix that corresponds to it. Then for any matrix $A \in \mathbb{F}^{n\times m}$, the matrix obtained by performing the row operation R on A is precisely the matrix EA.*

Proof. Define $f_R : \mathbb{F}^n \to \mathbb{F}^n$ by letting, for each $\mathbf{u} \in \mathbb{F}^n$, $f(\mathbf{u})$ be the vector obtained by performing the elementary row operation R on $\mathbf{u}$. It is easy

to see that f_R is linear.[7] So, f_R has a standard matrix. But clearly, the standard matrix of f_R is precisely the matrix E.

Now, fix any matrix $A \in \mathbb{F}^{n\times m}$, and set $A = \left[\begin{array}{ccc} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{array}\right]$. Then $EA = \left[\begin{array}{ccc} E\mathbf{a}_1 & \dots & E\mathbf{a}_m \end{array}\right]$. Since E is the standard matrix of f_R, we see that, for each index $i \in \{1, \dots, m\}$, $E\mathbf{a}_i = f(\mathbf{a}_i)$ is the vector obtained by performing the elementary row operation R on the column $\mathbf{a}_i$. So, EA is the matrix obtained by performing the elementary row operation R on A. □

Proposition 3.2.

(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) The product of elementary matrices in $\mathbb{F}^{n\times n}$ is invertible.

Proof. By Proposition 2.6, the product of invertible matrices is invertible. So, (c) follows from (a). It remains to prove (a) and (b). Let R be an elementary row operation performed on a matrix with n rows, and let E be the elementary matrix corresponding to R. Let R' be the elementary row operation that "undoes" R.[8] Let E' be the elementary matrix that corresponds to R'. But now Proposition 3.1 guarantees that $EE' = E'E = I_n$.[9] This proves that E is invertible, and that its inverse is the elementary matrix E'. □

We complete this section by proving a strengthening of part (a) of Theorem 2.2, stated below.

Proposition 3.3. *Let $A \in \mathbb{F}^{n\times n}$, and assume that $\left[\begin{array}{c:c} I_n & B \end{array}\right]$ is the reduced row echelon form of the matrix $\left[\begin{array}{c:c} A & I_n \end{array}\right]$. Then all the following hold:*

- *A is invertible;*
- *$A^{-1} = B$;*
- *A is the product of elementary matrices.*

Proof. By Proposition 3.1, there exists a sequence $E_1, \dots, E_k$ of elementary matrices such that $E_k \dots E_1 \left[\begin{array}{c:c} A & I_n \end{array}\right] = \left[\begin{array}{c:c} I_n & B \end{array}\right]$. Then $E_k \dots E_1 A = I_n$ and $E_k \dots E_1 I_n = B$. Set $E := E_k \dots E_1$, so that $EA = I_n$ and $EI_n = B$ (and so $B = E$). Now, by Proposition 3.2, we know that $B = E$ is invertible and has an inverse B^{-1}.[10] We now have that $B^{-1}B = I_n$ and $BA = I_n$. So,

[7] Check this!

[8] If R swaps rows i and j, then R' also swaps rows i and j. If R scales row i by a non-zero scalar α, then R' scales row i by α^{-1}. If R adds α times row i to row j ($j \neq i$), then R' adds $-\alpha$ times row i to row j.

[9] Details?

[10] Note that by Proposition 2.6, we have that $B^{-1} = E_1^{-1} \dots E_k^{-1}$.

by Proposition 2.1, we have that $A = B^{-1}$. This proves that A is invertible, and by Proposition 2.5, $A^{-1} = (B^{-1})^{-1} = B$. But now by Proposition 2.6, we have that $A = B^{-1} = (E_k \dots E_1)^{-1} = E_1^{-1} \dots E_k^{-1}$. By Proposition 3.2, $E_1^{-1}, \dots, E_k^{-1}$ are elementary matrices, and it follows that A is the product of elementary matrices. □

4 Singular and non-singular matrices

A *homogeneous matrix-vector equation* is any equation of the form $A\mathbf{x} = \mathbf{0}$, where A is a matrix, $\mathbf{0}$ is the zero vector of the appropriate size, and the vector $\mathbf{x}$ is unknown. Note that any homogeneous matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has at least one solution, namely, $\mathbf{x} = \mathbf{0}$; this solution is called the *trivial solution*. A *non-trivial solution* of $A\mathbf{x} = \mathbf{0}$ is any solution of this equation other than $\mathbf{0}$.

A matrix $A \in \mathbb{F}^{n \times n}$ is *singular* if the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution; otherwise, A is *non-singular*.[11]

Proposition 4.1. *A square matrix is invertible if and only if it is non-singular.*

Proof. Fix a matrix $A \in \mathbb{F}^{n \times n}$.

Suppose first that A is invertible. Fix any solution $\mathbf{x}_0$ of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. Then we get the following sequence of implications.

$$\begin{aligned} & A\mathbf{x}_0 = \mathbf{0} \\ \implies \quad & A^{-1}(A\mathbf{x}_0) = A^{-1}\mathbf{0} \\ \implies \quad & (A^{-1}A)\mathbf{x}_0 = \mathbf{0} \\ \implies \quad & I_n\mathbf{x}_0 = \mathbf{0} \\ \implies \quad & \mathbf{x}_0 = \mathbf{0} \end{aligned}$$

Thus, the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, i.e. A is non-singular.

Suppose now that A is non-singular, so that the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. The augmented matrix of the equation $A\mathbf{x} = \mathbf{0}$ is $\left[\begin{array}{c|c} A & \mathbf{0} \end{array}\right]$. Let $\left[\begin{array}{c|c} U & \mathbf{0} \end{array}\right] = RREF(\left[\begin{array}{c|c} A & \mathbf{0} \end{array}\right])$. Since $A\mathbf{x} = \mathbf{0}$ has only one solution (namely, the trivial solution), the same is true for $U\mathbf{x} = \mathbf{0}$. So, the system corresponding to the augmented matrix $\left[\begin{array}{c|c} U & \mathbf{0} \end{array}\right]$ has no free variables, i.e. all columns of $\left[\begin{array}{c|c} U & \mathbf{0} \end{array}\right]$ other than the last column (the

[11] So, A is *non-singular* if the matrix-vector equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. the solution $\mathbf{x} = \mathbf{0}$).

one to the right of the dotted line) are pivot columns. Since $\left[\begin{array}{c:c} U & \mathbf{0} \end{array}\right]$ is in reduced row echelon form, and since $U \in \mathbb{F}^{n\times n}$, it follows that $U = I_n$. So, some sequence of elementary row operations transforms the matrix $\left[\begin{array}{c:c} A & \mathbf{0} \end{array}\right]$ into the matrix $\left[\begin{array}{c:c} I_n & \mathbf{0} \end{array}\right]$. Now we apply that same sequence of elementary row operations to the matrix $\left[\begin{array}{c:c} A & I_n \end{array}\right]$; this produces a matrix $\left[\begin{array}{c:c} I_n & B \end{array}\right]$, which is clearly in reduced row echelon form. So, by Proposition 3.3, A is invertible. □

5 Proof of Theorem 2.2

We are now ready to prove Theorem 2.2, restated below.

Theorem 2.2. *Let $A \in \mathbb{F}^{n\times n}$, and let $\left[\begin{array}{c:c} U & B \end{array}\right] = RREF(\left[\begin{array}{c:c} A & I_n \end{array}\right])$, 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. Part (a) follows immediately from Proposition 3.3. Let us prove part (b). Assume that $U \neq I_n$. Clearly, $RREF(A) = U$ and $RREF(\left[\begin{array}{c:c} A & \mathbf{0} \end{array}\right]) = \left[\begin{array}{c:c} U & \mathbf{0} \end{array}\right]$. So, $A\mathbf{x} = \mathbf{0}$ and $U\mathbf{x} = \mathbf{0}$ have exactly the same solutions. The homogeneous equation $U\mathbf{x} = \mathbf{0}$ is consistent (because $\mathbf{x} = \mathbf{0}$ is a solution, i.e. $U\mathbf{0} = \mathbf{0}$). Moreover, since U is an $n \times n$ matrix in reduced row echelon form, but $U \neq I_n$, we see that U has at least one non-pivot column. So, the (consistent) system corresponding to the augmented matrix $\left[\begin{array}{c:c} U & \mathbf{0} \end{array}\right]$ has at least one free variable, and it follows that the equation $U\mathbf{x} = \mathbf{0}$ has more than one solution. Therefore, the equation $A\mathbf{x} = \mathbf{0}$ also has more than one solution, and so A is singular. But now by Proposition 4.1, we see that A is non-invertible. □

Corollary 5.1. *Let $A \in \mathbb{F}^{n\times n}$. Then the following are equivalent:*

(a) A is invertible (i.e. A has an inverse);

(b) A is non-singular (i.e. the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution);

(c) A is the product of elementary matrices;

(d) $RREF(A) = I_n$;

(e) $RREF(\left[\begin{array}{c:c} A & I_n \end{array}\right]) = \left[\begin{array}{c:c} I_n & B \end{array}\right]$ for some matrix $B \in \mathbb{F}^{n\times n}$.

Proof. By Proposition 4.1, (a) and (b) are equivalent. By Theorem 2.2, (a) and (e) are equivalent. Further, it is obvious that (d) and (e) are equivalent.[12]

[12] If a sequence of elementary row operations transforms A into I_n, then that same sequence of elementary row operations will transform $\left[\begin{array}{c:c} A & I_n \end{array}\right]$ into $\left[\begin{array}{c:c} I_n & B \end{array}\right]$ for some matrix B. The reverse also holds.

So far, we have proven that (a), (b), (d), and (e) are equivalent. Now, by Proposition 3.2, (c) implies (a), and by Proposition 3.3, (e) implies (c). This completes the argument. □

Proposition 5.2. *Let $A \in \mathbb{F}^{n\times n}$ be an invertible matrix, and let $\mathbf{b} \in \mathbb{F}^n$ be a vector. Then the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ has a unique solution, namely $A^{-1}\mathbf{b}$.*

Proof. First, $A^{-1}\mathbf{b}$ is a solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ because

$$A(A^{-1}\mathbf{b}) \;=\; (AA^{-1})\mathbf{b} \;=\; I_n\mathbf{b} \;=\; \mathbf{b}.$$

It remains to show that this solution is unique. So, fix any solution $\mathbf{x}_0$ of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. Then $A\mathbf{x}_0 = \mathbf{b}$. By multiplying both sides on the right by A^{-1}, we get that $A^{-1}(A\mathbf{x}_0) = A^{-1}\mathbf{b}$. We now compute

$$A^{-1}\mathbf{b} \;=\; A^{-1}(A\mathbf{x}_0) \;=\; (A^{-1}A)\mathbf{x}_0 \;=\; I_n\mathbf{x}_0 \;=\; \mathbf{x}_0.$$

So, $\mathbf{x}_0 = A^{-1}\mathbf{b}$, i.e. the solution $A^{-1}\mathbf{b}$ of $A\mathbf{x} = \mathbf{b}$ is unique. □