Linear Algebra 2

Lecture #22

Further properties of eigenvalues and eigenvectors. The Cayley-Hamilton theorem

Irena Penev

April 23, 2025

- This lecture has four parts:

- This lecture has four parts:
    1. A brief review of eigenvalues and eigenvectors (last lecture)

- This lecture has four parts:
  1. A brief review of eigenvalues and eigenvectors (last lecture)
  2. Further properties of eigenvalues and eigenvectors, plus the Invertible Matrix Theorem (version 4)

- This lecture has four parts:
  1. A brief review of eigenvalues and eigenvectors (last lecture)
  2. Further properties of eigenvalues and eigenvectors, plus the Invertible Matrix Theorem (version 4)
  3. The relationship between algebraic and geometric multiplicities of eigenvalues

- This lecture has four parts:
  1. A brief review of eigenvalues and eigenvectors (last lecture)
  2. Further properties of eigenvalues and eigenvectors, plus the Invertible Matrix Theorem (version 4)
  3. The relationship between algebraic and geometric multiplicities of eigenvalues
  4. The Cayley-Hamilton theorem

1. A brief review of eigenvalues and eigenvectors (last lecture)

1. A brief review of eigenvalues and eigenvectors (last lecture)

Definition

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 $\mathbf{v}$, s.t.

$$f(\mathbf{v}) \;=\; \lambda\mathbf{v}.$$

Under these circumstances, we also say that $\mathbf{v}$ is an eigenvector of f associated with the eigenvalue λ.

- 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

$$E_\lambda(f) \quad := \quad \{\mathbf{v} \in V \mid f(\mathbf{v}) = \lambda \mathbf{v}\}.$$

  - $E_\lambda(f)$ is a subspace of $\mathbb{F}^n$, and it is non-trivial iff λ is an eigenvalue of f (Proposition 8.1.4).
  - If λ is an eigenvalue of f, then $E_\lambda(f)$ is called the *eigenspace* of f associated with the eigenvalue λ, and the *geometric multiplicity of* λ is $\dim\big(E_\lambda(f)\big)$.

Definition

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 $\mathbf{v}$, s.t.

$$A\mathbf{v} = \lambda \mathbf{v}.$$

Under these circumstances, we also say that $\mathbf{v}$ is an eigenvector of A associated with the eigenvalue λ.

- For a field $\mathbb{F}$, a matrix $A \in \mathbb{F}^{n \times n}$, and a scalar $\lambda \in \mathbb{F}$, we define

$$E_\lambda(A) \quad := \quad \{\mathbf{v} \in \mathbb{F}^n \mid A\mathbf{v} = \lambda\mathbf{v}\}.$$

  - $E_\lambda(A)$ is a subspace of $\mathbb{F}^n$, and it is non-trivial iff λ is an eigenvalue of A (Proposition 8.1.6).
  - If λ is an eigenvalue of A, then $E_\lambda(A)$ is called the *eigenspace* of A associated with the eigenvalue λ, and the *geometric multiplicity of* λ is $\dim\big(E_\lambda(A)\big)$.

Definition

Given a field $\mathbb{F}$ and a matrix $A \in \mathbb{F}^{n \times n}$, the *characteristic polynomial* of A is defined to be

$$p_A(\lambda) \quad := \quad \det(\lambda I_n - A).$$

The *characteristic equation* of A is the equation

$$\det(\lambda I_n - A) \quad = \quad 0.$$

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.

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:

$$\begin{aligned} p_A(\lambda) &= \det(\lambda I_3 - A) = \begin{vmatrix} \lambda - 1 & 2 & -3 \\ 1 & \lambda & -2 \\ -2 & 1 & \lambda + 3 \end{vmatrix} \\ &= \lambda^3 + 2\lambda^2 - 9\lambda - 3. \end{aligned}$$

$\square$

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

$$E_{\lambda_0}(A) \;=\; \operatorname{Nul}\big(\lambda_0 I_n - A\big) \;=\; \operatorname{Nul}\big(A - \lambda_0 I_n\big).$$

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. $\det\big(\lambda_0 I_n - A\big) = 0$.

- By Theorem 8.2.2, the eigenvalues of a square matrix are precisely the roots of its characteristic polynomial.

- By Theorem 8.2.2, the eigenvalues of a square matrix are precisely the roots of its characteristic polynomial.
- 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 $p_A(\lambda)$.

- By Theorem 8.2.2, the eigenvalues of a square matrix are precisely the roots of its characteristic polynomial.
- 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 $p_A(\lambda)$.
- Since $\deg(p_A(\lambda)) = n$, 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.

- By Theorem 8.2.2, the eigenvalues of a square matrix are precisely the roots of its characteristic polynomial.
- 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 $p_A(\lambda)$.
- Since $\deg(p_A(\lambda)) = n$, 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.
- 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.
    - 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.

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: Later!
- Schematically, Theorem 8.2.3 states that for an eigenvalue λ of A:

$$\text{geometric multiplicity of } \lambda \quad \leq \quad \text{algebraic multiplicity of } \lambda.$$

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: Later!
- Schematically, Theorem 8.2.3 states that for an eigenvalue λ of A:

$$\text{geometric multiplicity of } \lambda \quad \leq \quad \text{algebraic multiplicity of } \lambda.$$

- For now, we have only stated Theorem 8.2.3. We will not use this theorem before proving it.

2. Further properties of eigenvalues and eigenvectors, plus the Invertible Matrix Theorem (version 4)

2. Further properties of eigenvalues and eigenvectors, plus the Invertible Matrix Theorem (version 4)

- 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.

2. Further properties of eigenvalues and eigenvectors, plus the Invertible Matrix Theorem (version 4)

- 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

$$\rho(A) = \max\{|1|, |1+i|, |1+i|, |1-i|, |1-i|\} = \sqrt{2},$$

since $|1| = 1$, $|1+i| = \sqrt{2}$, and $|1-i| = \sqrt{2}$.

2. Further properties of eigenvalues and eigenvectors, plus the Invertible Matrix Theorem (version 4)

- 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

$$\rho(A) \;=\; \max\{|1|, |1+i|, |1+i|, |1-i|, |1-i|\} \;=\; \sqrt{2},$$

    since $|1| = 1$, $|1+i| = \sqrt{2}$, and $|1-i| = \sqrt{2}$.
- Reminder:

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 $p(x)$ iff its complex conjugate $\overline{z}$ is a root of $p(x)$.

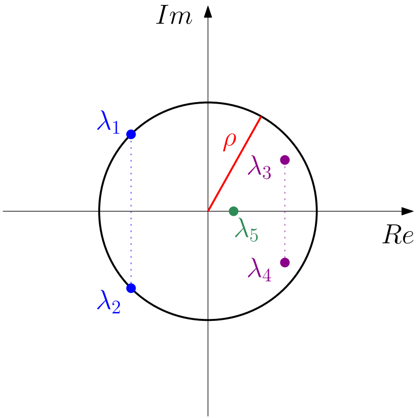
- In view of Theorems 0.3.6 and 8.2.2., we can visualize the complex eigenvalues of an $n \times n$ matrix A with **real** entries (however, we consider A to be a matrix in the vector space $\mathbb{C}^{n \times n}$, so that it can have complex eigenvalues).

- 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.

- 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 (each real root is its own conjugate pair), and moreover, by Theorem 8.2.2, those roots are precisely the eigenvalues of A.

- 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 (each real root is its own conjugate pair), 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.

- 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 (each real root is its own conjugate pair), 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 on the next slide (the conjugate pairs are color coded for emphasis).


Im
Re
ρ
λ_1
λ_2
λ_3
λ_4
λ_5

Definition

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 $P \in \mathbb{F}^{n \times n}$ s.t. $B = P^{-1}AP$.

Definition

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 $P \in \mathbb{F}^{n \times n}$ s.t. $B = P^{-1}AP$.

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 $f : V \to V$ s.t. $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$, there exists a basis $\mathcal{C}$ of V s.t. $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$;
c. for all bases $\mathcal{C}$ of V and linear functions $f : V \to V$ s.t. $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$, there exists a basis $\mathcal{B}$ of V s.t. $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$;
d. there exist bases $\mathcal{B}$ and $\mathcal{C}$ of V and a linear function $f : V \to V$ s.t. $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$ and $C = {}_{\mathcal{C}}[\ f\ ]_{\mathcal{C}}$.

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:

$$E_\lambda(A) \;\;\cancel{\asymp}\;\; E_\lambda(B)$$

- Reminder:

Proposition 8.1.7

Let V be a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$, let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ 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

$$E_\lambda\Big({}_{\mathcal{B}}\Big[\ f\ \Big]_{\mathcal{B}} \Big) \quad = \quad \Big\{ \Big[\ \mathbf{v}\ \Big]_{\mathcal{B}} \mid \mathbf{v} \in E_\lambda(f) \Big\}.$$

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.

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**.

Proof.

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**.

Proof. Let us first show that A and B have the same eigenvalues with the same corresponding geometric multiplicities.

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**.

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 $\mathbb{F}^n$ s.t. $A = {}_{\mathcal{A}}[\ f\ ]_{\mathcal{A}}$ and $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.

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**.

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 $\mathbb{F}^n$ s.t. $A = {}_{\mathcal{A}}[\ f\ ]_{\mathcal{A}}$ and $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.
But then by Proposition 8.1.7, the linear function f and the matrix $A = {}_{\mathcal{A}}[\ f\ ]_{\mathcal{A}}$ have exactly the same eigenvalues, with exactly the same corresponding geometric multiplicities, and the same holds for f and the matrix $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.

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**.

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 $\mathbb{F}^n$ s.t. $A = {}_{\mathcal{A}}[\ f\ ]_{\mathcal{A}}$ and $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.

But then by Proposition 8.1.7, the linear function f and the matrix $A = {}_{\mathcal{A}}[\ f\ ]_{\mathcal{A}}$ have exactly the same eigenvalues, with exactly the same corresponding geometric multiplicities, and the same holds for f and the matrix $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$. So, A and B have exactly the same eigenvalues with exactly the same corresponding geometric multiplicities.

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**.

Proof (continued). 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.

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**.

Proof (continued). 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 $P \in \mathbb{F}^{n \times n}$ s.t. $B = P^{-1}AP$.

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**.

Proof (continued). 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 $P \in \mathbb{F}^{n \times n}$ s.t. $B = P^{-1}AP$. We now compute (next slide):

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**.

Proof (continued).

$$\begin{aligned}
p_B(\lambda) &= \det\big(\lambda I_n - B\big) \\
&= \det\big(\lambda I_n - P^{-1}AP\big) \\
&= \det\big(P^{-1}(\lambda I_n - A)P\big) \\
&= \det(P^{-1})\det\big(\lambda I_n - A\big)\det(P) && \text{by Theorem 7.5.2} \\
&= \tfrac{1}{\det(P)}\det\big(\lambda I_n - A\big)\det(P) && \text{by Corollary 7.5.3} \\
&= \det\big(\lambda I_n - A\big) \\
&= p_A(\lambda).
\end{aligned}$$

$\square$

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**.

- **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."

Definition

The *trace* of a square matrix $A = \left[\ a_{i,j}\ \right]_{n\times n}$ with entries in some field $\mathbb{F}$ is defined to be $\operatorname{trace}(A) := \sum_{i=1}^{n} a_{i,i}$, 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 $\operatorname{trace}(A) = 1 + 5 + 9 = 15$.

Theorem 8.2.10

Let $\mathbb{F}$ be a field, let $A = \left[\ a_{i,j} \ \right]_{n \times n}$ be a matrix in $\mathbb{F}^{n \times n}$, and assume that $\{\lambda_1, \dots, \lambda_n\}$ is the spectrum of A. Then

(a) $\det(A) = \lambda_1 \dots \lambda_n$;

(b) $\text{trace}(A) = \lambda_1 + \dots + \lambda_n$.

Proof (outline).

Theorem 8.2.10

Let $\mathbb{F}$ be a field, let $A = \left[\ a_{i,j}\ \right]_{n \times n}$ be a matrix in $\mathbb{F}^{n \times n}$, and assume that $\{\lambda_1, \dots, \lambda_n\}$ is the spectrum of A. Then

(a) $\det(A) = \lambda_1 \dots \lambda_n$;

(b) $\text{trace}(A) = \lambda_1 + \dots + \lambda_n$.

Proof (outline). (a) Compute $p_A(0)$ in two different ways. (b) Compute the coefficient in front of λ^{n-1} in $p_A(\lambda)$ in two different ways. (Details: Lecture Notes.) $\square$

Theorem 8.2.10

Let $\mathbb{F}$ be a field, let $A = \left[\ a_{i,j}\ \right]_{n \times n}$ be a matrix in $\mathbb{F}^{n \times n}$, and assume that $\{\lambda_1, \dots, \lambda_n\}$ is the spectrum of A. Then

(a) $\det(A) = \lambda_1 \dots \lambda_n$;

(b) $\text{trace}(A) = \lambda_1 + \dots + \lambda_n$.

Proof (outline). (a) Compute $p_A(0)$ in two different ways.
(b) Compute the coefficient in front of λ^{n-1} in $p_A(\lambda)$ in two different ways. (Details: Lecture Notes.) $\square$

- **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.

Proposition 8.2.11

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$. Then A is invertible iff 0 is **not** an eigenvalue of A.

Proof.

Proposition 8.2.11

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$. Then A is invertible iff 0 is **not** an eigenvalue of A.

Proof. It suffices to show that 0 is an eigenvalue of A iff A is not invertible.

Proposition 8.2.11

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$. Then A is invertible iff 0 is **not** an eigenvalue of A.

Proof. It suffices to show that 0 is an eigenvalue of A iff A is not invertible. We have the following sequence of equivalent statements:

$$
\begin{aligned}
0 \text{ is eigenvalue of } A \quad & \overset{\text{Thm. 8.2.2}}{\iff} \quad \det(0I_n - A) = 0 \\
& \iff \quad \det(-A) = 0 \\
& \overset{\text{Prop. 7.2.3}}{\iff} \quad (-1)^n \det(A) = 0 \\
& \iff \quad \det(A) = 0 \\
& \overset{\text{Thm. 7.4.1}}{\iff} \quad A \text{ is not invertible}
\end{aligned}
$$

$\square$

Proposition 8.2.11

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n \times n}$. Then A is invertible iff 0 is **not** an eigenvalue of A.

- We now add the eigenvalue condition from Proposition 8.2.11 to our previous version of the Invertible Matrix Theorem to obtain the fourth and final version of that theorem (next three slides).
    - It uses 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 $\mathbf{x} \in \mathbb{F}^n$.[a] Then the following are equivalent:

- (a) A is invertible (i.e. A has an inverse);
- (b) A^T is invertible;
- (c) $\mathrm{RREF}(A) = I_n$;
- (d) $\mathrm{RREF}(\left[\begin{array}{c|c} A & I_n \end{array}\right]) = \left[\begin{array}{c|c} I_n & B \end{array}\right]$ for some matrix $B \in \mathbb{F}^{n\times n}$;
- (e) $\mathrm{rank}(A) = n$;
- (f) $\mathrm{rank}(A^T) = n$;
- (g) is a product of elementary matrices;

[a] Since f is a matrix transformation, Proposition 1.10.4 guarantees that f is linear. Moreover, A is the standard matrix of f.

The Invertible Matrix Theorem (version 4, continued)

- (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 $\mathbf{b} \in \mathbb{F}^n$ s.t. 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;

The Invertible Matrix Theorem (version 4, continued)

(p) there exists a matrix $B \in \mathbb{F}^{n\times n}$ s.t. $BA = I_n$ (i.e. A has a left inverse);

(q) there exists a matrix $C \in \mathbb{F}^{n\times n}$ s.t. $AC = I_n$ (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. $\mathrm{Col}(A) = \mathbb{F}^n$);

(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. $\mathrm{Row}(A) = \mathbb{F}^{1\times n}$);

(w) the rows of A form a basis of $\mathbb{F}^{1\times n}$;

(x) $\mathrm{Nul}(A) = \{\mathbf{0}\}$ (i.e. $\dim\big(\mathrm{Nul}(A)\big) = 0$);

(y) $\det(A) \neq 0$;

(z) 0 is not an eigenvalue of A.

- Reminder:
  - 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

$$\det(f) \quad := \quad \det\Big({}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}} \Big),$$

where $\mathcal{B}$ is any basis of V.

- Reminder:
  - 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

$$\det(f) \quad := \quad \det\Big({}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \Big),$$

    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 ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ are similar, and therefore (by Corollary 7.5.4) have the same determinant.

Definition

Let V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$. The *characteristic polynomial* of a linear function $f : V \to V$ is defined to be the polynomial

$$p_f(\lambda) \quad := \quad \det(\lambda \mathrm{Id}_V - f) \quad = \quad \det\Big({}_{\mathcal{B}}\Big[\ \lambda \mathrm{Id}_V - f \ \Big]_{\mathcal{B}} \Big),$$

where $\mathcal{B}$ is **any** basis of V.[a]

---

[a] 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 $\mathbf{v} \in V$.

Definition

Let V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$. The *characteristic polynomial* of a linear function $f : V \to V$ is defined to be the polynomial

$$p_f(\lambda) \quad := \quad \det\big(\lambda \mathrm{Id}_V - f\big) \quad = \quad \det\Big({}_{\mathcal{B}}\Big[\ \lambda \mathrm{Id}_V - f\ \Big]_{\mathcal{B}} \Big),$$

where $\mathcal{B}$ is **any** basis of V.[a]

[a] 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 $\mathbf{v} \in V$.

- As per our discussion above, the polynomial $p_f(\lambda)$ depends only on f, and not on the particular choice of the basis $\mathcal{B}$.

Definition

Let V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$. The *characteristic polynomial* of a linear function $f : V \to V$ is defined to be the polynomial

$$p_f(\lambda) \;:=\; \det\big(\lambda \mathrm{Id}_V - f\big) \;=\; \det\Big({}_{\mathcal{B}}\big[\ \lambda \mathrm{Id}_V - f\ \big]_{\mathcal{B}}\Big),$$

where $\mathcal{B}$ is **any** basis of V.[a]

[a] 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 $\mathbf{v} \in V$.

- 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.

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.

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 $B := {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$. Then $p_f(\lambda) = p_B(\lambda)$.

Proof. We compute:

$$
\begin{aligned}
p_f(\lambda) &= \det\big(\lambda \mathrm{Id}_V - f\big) && \text{by definition} \\
&= \det\Big({}_{\mathcal{B}}\big[\ \lambda \mathrm{Id}_V - f\ \big]_{\mathcal{B}}\Big) && \text{by definition} \\
&= \det\Big(\lambda\, {}_{\mathcal{B}}\big[\ \mathrm{Id}_V\ \big]_{\mathcal{B}} - {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}\Big) && \text{by Theorem 4.5.3} \\
&= \det\big(\lambda I_n - B\big) && \\
&= p_B(\lambda) && \text{by definition.}
\end{aligned}
$$

□

- Reminder:

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

$$E_{\lambda_0}(A) \quad = \quad \mathrm{Nul}(\lambda_0 I_n - A) \quad = \quad \mathrm{Nul}(A - \lambda_0 I_n).$$

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. $\det(\lambda_0 I_n - A) = 0$.

- Reminder:

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

$$E_{\lambda_0}(A) \quad = \quad \mathrm{Nul}(\lambda_0 I_n - A) \quad = \quad \mathrm{Nul}(A - \lambda_0 I_n).$$

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. $\det(\lambda_0 I_n - A) = 0$.

- Analogously, we have the following (next slide):

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

$$E_{\lambda_0}(f) \;=\; \mathrm{Ker}(\lambda_0 \mathrm{Id}_V - f) \;=\; \mathrm{Ker}(f - \lambda_0 \mathrm{Id}_V).$$

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. $\det(\lambda_0 \mathrm{Id}_V - f) = 0$.

- Proof: Lecture Notes. (Similar to the proof of Theorem 8.2.2.)

- Suppose that $f : V \to V$ is a linear function, where V is a non-trivial, finite-dimensional vector space over a field $\mathbb{F}$.

- 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 $p_f(\lambda)$.

- 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.

- 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.
  - **Reminder:** The *geometric multiplicity* of an eigenvalue λ_0 of f is $\dim\big(E_f(\lambda_0)\big)$.

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 ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ have the same **characteristic polynomial**, and the same **spectrum**. Moreover, f and ${}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$ have exactly the same **eigenvalues**, with exactly the same corresponding **geometric multiplicities**, and exactly the same corresponding **algebraic multiplicities**.

Proof.

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.

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.

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**. $\square$

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 ${}_{\mathcal{B}}\left[\ f\ \right]_{\mathcal{B}}$ have the same **characteristic polynomial**, and the same **spectrum**. Moreover, f and ${}_{\mathcal{B}}\left[\ f\ \right]_{\mathcal{B}}$ have exactly the same **eigenvalues**, with exactly the same corresponding **geometric multiplicities**, and exactly the same corresponding **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 (next slide).

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.

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 $A = {}_{\mathcal{E}_n}\left[\ f\ \right]_{\mathcal{E}_n}$, where $\mathcal{E}_n$ is the standard basis of $\mathbb{F}^n$.

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 $A = {}_{\mathcal{E}_n}\left[\ f\ \right]_{\mathcal{E}_n}$, 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. $\square$

3. The relationship between algebraic and geometric multiplicities of eigenvalues

3. The relationship between algebraic and geometric multiplicities of eigenvalues

- Let's now prove Theorem 8.2.3!

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.

- Schematically, Theorem 8.2.3 states that for an eigenvalue λ of A:

$$\text{geometric multiplicity of } \lambda \quad \leq \quad \text{algebraic multiplicity of } \lambda.$$

3. The relationship between algebraic and geometric multiplicities of eigenvalues

- Let's now prove Theorem 8.2.3!

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.

- Schematically, Theorem 8.2.3 states that for an eigenvalue λ of A:

$$\text{geometric multiplicity of } \lambda \quad \leq \quad \text{algebraic multiplicity of } \lambda.$$

- In fact, it will be a bit more convenient to prove this theorem for linear functions first (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.

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.

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 $(\lambda - \lambda_0)^k \mid p_f(\lambda)$.

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, $p_f(\lambda) = p_B(\lambda)$.

Proof (continued). Reminder: λ_0 is an eigenvalue of f; WTS there exists a basis $\mathcal{B}$ of V s.t. $(\lambda - \lambda_0)^k \mid p_B(\lambda)$, where $B = {}_{\mathcal{B}}\left[\ f\ \right]_{\mathcal{B}}$.

Proof (continued). Reminder: λ_0 is an eigenvalue of f; WTS there exists a basis $\mathcal{B}$ of V s.t. $(\lambda - \lambda_0)^k \mid p_B(\lambda)$, where $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.

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 $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$.

Proof (continued). Reminder: λ_0 is an eigenvalue of f; WTS there exists a basis $\mathcal{B}$ of V s.t. $(\lambda - \lambda_0)^k \mid p_B(\lambda)$, where $B = {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$.

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 $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$. In particular, $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ is a linearly independent set of vectors in V, and so by Theorem 3.2.19, it can be extended to a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_k, \mathbf{b}_{k+1}, \dots, \mathbf{b}_n\}$ of V.

Proof (continued). Reminder: λ_0 is an eigenvalue of f; WTS there exists a basis $\mathcal{B}$ of V s.t. $(\lambda - \lambda_0)^k \mid p_B(\lambda)$, where $B = {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}}$.

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 $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$. In particular, $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ is a linearly independent set of vectors in V, and so by Theorem 3.2.19, it can be extended to a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_k, \mathbf{b}_{k+1}, \dots, \mathbf{b}_n\}$ of V. We now compute:

$$\begin{array}{rcl}
B & := & {}_{\mathcal{B}}\big[\ f\ \big]_{\mathcal{B}} \\
& \overset{(*)}{=} & \Big[\ \big[\ f(\mathbf{b}_1)\ \big]_{\mathcal{B}}\ \ \dots\ \ \big[\ f(\mathbf{b}_k)\ \big]_{\mathcal{B}}\ \ \big[\ f(\mathbf{b}_{k+1})\ \big]_{\mathcal{B}}\ \ \dots\ \ \big[\ f(\mathbf{b}_n)\ \big]_{\mathcal{B}}\ \Big] \\
& \overset{(**)}{=} & \Big[\ \big[\ \lambda_0\mathbf{b}_1\ \big]_{\mathcal{B}}\ \ \dots\ \ \big[\ \lambda_0\mathbf{b}_k\ \big]_{\mathcal{B}}\ \ \big[\ f(\mathbf{b}_{k+1})\ \big]_{\mathcal{B}}\ \ \dots\ \ \big[\ f(\mathbf{b}_n)\ \big]_{\mathcal{B}}\ \Big] \\
& = & \Big[\ \lambda_0\mathbf{e}_1^n\ \ \dots\ \ \lambda_0\mathbf{e}_k^n\ \ \big[\ f(\mathbf{b}_{k+1})\ \big]_{\mathcal{B}}\ \ \dots\ \ \big[\ f(\mathbf{b}_n)\ \big]_{\mathcal{B}}\ \Big] \\
& = & \left[\begin{array}{c|ccc} \lambda_0 I_k & & & \\ \hline O_{(n-k)\times k} & \big[\ f(\mathbf{b}_{k+1})\ \big]_{\mathcal{B}} & \dots & \big[\ f(\mathbf{b}_n)\ \big]_{\mathcal{B}} \end{array}\right],
\end{array}$$

where (*) follows from Theorem 4.5.1, and (**) follows from the fact that $\mathbf{b}_1, \dots, \mathbf{b}_k \in E_{\lambda_0}(f)$.

Proof (continued). Reminder: λ_0 is an eigenvalue of f; WTS there exists a basis $\mathcal{B}$ of V s.t. $(\lambda - \lambda_0)^k \mid p_B(\lambda)$, where $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.

We showed:

$$B \quad := \quad {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}} \quad = \quad \left[\begin{array}{c:cccc} \lambda_0 I_k & \multirow{2}{*}{$[\ f(\mathbf{b}_{k+1})\ ]_{\mathcal{B}}$} & \multirow{2}{*}{$\dots$} & \multirow{2}{*}{$[\ f(\mathbf{b}_n)\ ]_{\mathcal{B}}$} \\ \hdashline O_{(n-k)\times k} & & & \end{array}\right].$$

Proof (continued). Reminder: λ_0 is an eigenvalue of f; WTS there exists a basis $\mathcal{B}$ of V s.t. $(\lambda - \lambda_0)^k \mid p_B(\lambda)$, where $B = {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}}$.

We showed:

$$B \ := \ {}_{\mathcal{B}}[\ f\ ]_{\mathcal{B}} \ = \ \left[\begin{array}{c:ccc} \lambda_0 I_k & & & \\ \hdashline O_{(n-k)\times k} & [\ f(\mathbf{b}_{k+1})\ ]_{\mathcal{B}} & \dots & [\ f(\mathbf{b}_n)\ ]_{\mathcal{B}} \end{array}\right].$$

Thus, $p_B(\lambda)$ is of the form

$$p_B(\lambda) \ = \ \left|\begin{array}{cccc:cccc} \lambda - \lambda_0 & 0 & \dots & 0 & * & * & \dots & * \\ 0 & \lambda - \lambda_0 & \dots & 0 & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda - \lambda_0 & * & * & \dots & * \\ \hdashline 0 & 0 & \dots & 0 & * & * & \dots & * \\ 0 & 0 & \dots & 0 & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & * & * & \dots & * \end{array}\right|,$$

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 $(\lambda - \lambda_0)^k$. $\square$

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.

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.

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.

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.

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 Prop. 1.10.4), and its standard matrix is A.

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.

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 Prop. 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. $\square$

4. The Cayley-Hamilton theorem

4. 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 $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$ be the characteristic polynomial of A. Then

$$A^n + a_{n-1}A^{n-1} + \dots + a_1 A + a_0 I_n \quad = \quad O_{n\times n}.$$

- The 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

$$\begin{aligned} A^2 - 5A - 2I_2 &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^2 - 5\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix} - \begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

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

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n \quad = \quad O_{n \times n}.$$

- The proof of the Cayley-Hamilton theorem relies on the adjugate matrix and a theorem that we proved about it (namely, Theorem 7.8.2).
    - Reminder: Next slide.

Definition

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 $\begin{bmatrix} C_{i,j} \end{bmatrix}_{n\times n}$. The *adjugate matrix* (also called the *classical adjoint*) of A, denoted by $\text{adj}(A)$, is the transponse of the cofactor matrix of A, i.e.

$$\text{adj}(A) \quad := \quad \left(\begin{bmatrix} C_{i,j} \end{bmatrix}_{n\times n} \right)^T.$$

So, the i, j-th entry of $\text{adj}(A)$ is the cofactor $C_{j,i}$ (note the swapping of the indices).

Definition

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 $i, j \in \{1, \dots, n\}$), the *cofactor matrix* of A is the matrix $\begin{bmatrix} C_{i,j} \end{bmatrix}_{n\times n}$. 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.

$$\mathrm{adj}(A) \quad := \quad \left(\begin{bmatrix} C_{i,j} \end{bmatrix}_{n\times n} \right)^T.$$

So, the i, j-th entry of $\mathrm{adj}(A)$ is the cofactor $C_{j,i}$ (note the swapping of the indices).

Theorem 7.8.2

Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ ($n \geq 2$). Then

$$\mathrm{adj}(A)\, A \quad = \quad A\, \mathrm{adj}(A) \quad = \quad \det(A) I_n.$$

Consequently, if A is invertible, then $A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)$.

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

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n \quad = \quad O_{n \times n}.$$

Proof.

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

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n \quad = \quad O_{n\times n}.$$

Proof. If $n = 1$, then the result is immediate.

- Indeed, suppose that $n = 1$, and consider any matrix $A = \left[\ a_{1,1}\ \right]$ in $\mathbb{F}^{1\times 1}$.
- Then $p_A(\lambda) = \det(\lambda I_1 - A) = \det(\left[\ \lambda - a_{1,1}\ \right]) = \lambda - a_{1,1}$, and we see that $A - a_{1,1}I_1 = O_{1\times 1}$.

Proof (continued). From now on, we assume that $n \geq 2$.

Proof (continued). From now on, we 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

$$(\lambda I_n - A)\,\mathrm{adj}(\lambda I_n - A) \quad = \quad \det(\lambda I_n - A) I_n.$$

Proof (continued). From now on, we 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

$$(\lambda I_n - A) \operatorname{adj}(\lambda I_n - A) \quad = \quad \det(\lambda I_n - A) I_n.$$

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}.

Proof (continued). From now on, we 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

$$(\lambda I_n - A)\,\text{adj}(\lambda I_n - A) \;=\; \det(\lambda I_n - A) I_n.$$

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 $\text{adj}(\lambda I_n - A)$ are precisely the cofactors of $\lambda I_n - A$, it follows that each entry of $\text{adj}(\lambda I_n - A)$ is a polynomial (in the variable λ) of degree at most $n - 1$.

Proof (continued). From now on, we 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

$$(\lambda I_n - A)\,\mathrm{adj}(\lambda I_n - A) \;\; = \;\; \det(\lambda I_n - A) I_n.$$

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 $\mathrm{adj}(\lambda I_n - A)$ are precisely the cofactors of $\lambda I_n - A$, it follows that each entry of $\mathrm{adj}(\lambda I_n - A)$ is a polynomial (in the variable λ) of degree at most $n-1$. So, the matrix $\mathrm{adj}(\lambda I_n - A)$ can be expressed in the form

$$\mathrm{adj}(\lambda I_n - A) \;\; = \;\; \lambda^{n-1} B_{n-1} + \lambda^{n-2} B_{n-2} + \cdots + \lambda B_1 + B_0,$$

for some matrices $B_0, B_1, \ldots, B_{n-1} \in \mathbb{F}^{n \times n}$.

Proof (continued). From now on, we 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

$$(\lambda I_n - A)\,\mathrm{adj}(\lambda I_n - A) \;\; = \;\; \det(\lambda I_n - A) I_n.$$

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 $\mathrm{adj}(\lambda I_n - A)$ are precisely the cofactors of $\lambda I_n - A$, it follows that each entry of $\mathrm{adj}(\lambda I_n - A)$ is a polynomial (in the variable λ) of degree at most $n-1$. So, the matrix $\mathrm{adj}(\lambda I_n - A)$ can be expressed in the form

$$\mathrm{adj}(\lambda I_n - A) \;\; = \;\; \lambda^{n-1} B_{n-1} + \lambda^{n-2} B_{n-2} + \cdots + \lambda B_1 + B_0,$$

for some matrices $B_0, B_1, \ldots, B_{n-1} \in \mathbb{F}^{n \times n}$. Consequently,

$$\underbrace{(\lambda I_n - A)(\underbrace{\lambda^{n-1} B_{n-1} + \lambda^{n-2} B_{n-2} + \cdots + \lambda B_1 + B_0}_{=\mathrm{adj}(\lambda I_n - A)})}_{:=\mathrm{LHS}} \;\; = \;\; \underbrace{\det(\lambda I_n - A) I_n}_{:=\mathrm{RHS}}.$$

Proof (continued). Reminder: $n \geq 2$,

$$\underbrace{(\lambda I_n - A)(\underbrace{\lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2} + \cdots + \lambda B_1 + B_0}_{=\mathrm{adj}(\lambda I_n - A)})}_{:=\mathrm{LHS}} \quad = \quad \underbrace{\det(\lambda I_n - A) I_n}_{:=\mathrm{RHS}}.$$

Proof (continued). Reminder: $n \geq 2$,

$$\underbrace{(\lambda I_n - A)(\underbrace{\lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2} + \cdots + \lambda B_1 + B_0}_{=\mathrm{adj}(\lambda I_n - A)})}_{:=\mathrm{LHS}} \quad = \quad \underbrace{\det(\lambda I_n - A)I_n}_{:=\mathrm{RHS}}.$$

For the left-hand-side, we have

$$\begin{aligned} \mathrm{LHS} &= (\lambda I_n - A)(\lambda^{n-1}B_{n-1} + \cdots + \lambda B_1 + B_0) \\ &= \lambda^n B_{n-1} + \lambda^{n-1}(B_{n-2} - AB_{n-1}) + \lambda^{n-2}(B_{n-3} - AB_{n-2}) + \\ &\quad + \cdots + \lambda(B_0 - AB_1) - AB_0. \end{aligned}$$

Proof (continued). Reminder: $n \geq 2$,

$$\underbrace{(\lambda I_n - A)(\underbrace{\lambda^{n-1} B_{n-1} + \lambda^{n-2} B_{n-2} + \cdots + \lambda B_1 + B_0}_{=\mathrm{adj}(\lambda I_n - A)})}_{:=\mathrm{LHS}} \quad = \quad \underbrace{\det(\lambda I_n - A) I_n}_{:=\mathrm{RHS}}.$$

For the left-hand-side, we have

$$\begin{aligned} \mathrm{LHS} &= (\lambda I_n - A)(\lambda^{n-1} B_{n-1} + \cdots + \lambda B_1 + B_0) \\ &= \lambda^n B_{n-1} + \lambda^{n-1}(B_{n-2} - AB_{n-1}) + \lambda^{n-2}(B_{n-3} - AB_{n-2}) + \\ &\quad + \cdots + \lambda(B_0 - AB_1) - AB_0. \end{aligned}$$

For the right-hand-side, we have

$$\begin{aligned} \mathrm{RHS} &= \det(\lambda I_n - A) I_n \quad = \quad p_A(\lambda) I_n \\ &= (\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-1} + \cdots + a_1\lambda + a_0) I_n \\ &= \lambda^n I_n + \lambda^{n-1} a_{n-1} I_n + \lambda^{n-2} a_{n-2} I_n + \cdots + \lambda a_1 I_n + a_0 I_n. \end{aligned}$$

Proof (continued). Reminder: $n \geq 2$,

$$\underbrace{(\lambda I_n - A)(\underbrace{\lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2} + \cdots + \lambda B_1 + B_0}_{=\mathrm{adj}(\lambda I_n - A)})}_{:=\mathrm{LHS}} \quad = \quad \underbrace{\det(\lambda I_n - A) I_n}_{:=\mathrm{RHS}}.$$

For the left-hand-side, we have

$$\begin{aligned} \mathrm{LHS} &= (\lambda I_n - A)(\lambda^{n-1}B_{n-1} + \cdots + \lambda B_1 + B_0) \\ &= \lambda^n B_{n-1} + \lambda^{n-1}(B_{n-2} - AB_{n-1}) + \lambda^{n-2}(B_{n-3} - AB_{n-2}) + \\ &\quad + \cdots + \lambda(B_0 - AB_1) - AB_0. \end{aligned}$$

For the right-hand-side, we have

$$\begin{aligned} \mathrm{RHS} &= \det(\lambda I_n - A) I_n \quad = \quad p_A(\lambda) I_n \\ &= (\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-1} + \cdots + a_1\lambda + a_0) I_n \\ &= \lambda^n I_n + \lambda^{n-1} a_{n-1} I_n + \lambda^{n-2} a_{n-2} I_n + \cdots + \lambda a_1 I_n + a_0 I_n. \end{aligned}$$

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).

Proof (continued). Reminder: $n \geq 2$,

$$\underbrace{(\lambda I_n - A)\underbrace{(\lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2} + \cdots + \lambda B_1 + B_0)}_{=\mathrm{adj}(\lambda I_n - A)}}_{:=\mathrm{LHS}} \quad = \quad \underbrace{\det(\lambda I_n - A) I_n}_{:=\mathrm{RHS}}.$$

For the left-hand-side, we have

$$\begin{aligned} \mathrm{LHS} &= (\lambda I_n - A)(\lambda^{n-1}B_{n-1} + \cdots + \lambda B_1 + B_0) \\ &= \lambda^n B_{n-1} + \lambda^{n-1}(B_{n-2} - AB_{n-1}) + \lambda^{n-2}(B_{n-3} - AB_{n-2}) + \\ &\quad + \cdots + \lambda(B_0 - AB_1) - AB_0. \end{aligned}$$

For the right-hand-side, we have

$$\begin{aligned} \mathrm{RHS} &= \det(\lambda I_n - A) I_n \quad = \quad p_A(\lambda) I_n \\ &= (\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-1} + \cdots + a_1\lambda + a_0) I_n \\ &= \lambda^n I_n + \lambda^{n-1} a_{n-1} I_n + \lambda^{n-2} a_{n-2} I_n + \cdots + \lambda a_1 I_n + a_0 I_n. \end{aligned}$$

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 (next slide).

Proof (continued).

$$\begin{aligned}
B_{n-1} &= I_n \\
B_{n-2} - AB_{n-1} &= a_{n-1} I_n \\
B_{n-3} - AB_{n-2} &= a_{n-2} I_n \\
&\vdots \\
B_0 - AB_1 &= a_1 I_n \\
-AB_0 &= a_0 I_n
\end{aligned}$$

Proof (continued).

$$\begin{aligned}
B_{n-1} &= I_n \\
B_{n-2} - AB_{n-1} &= a_{n-1} I_n \\
B_{n-3} - AB_{n-2} &= a_{n-2} I_n \\
&\vdots \\
B_0 - AB_1 &= a_1 I_n \\
-AB_0 &= a_0 I_n
\end{aligned}$$

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.

Proof (continued).

$$\begin{array}{rcl} B_{n-1} & = & I_n \\ B_{n-2} - AB_{n-1} & = & a_{n-1} I_n \\ B_{n-3} - AB_{n-2} & = & a_{n-2} I_n \\ & \vdots & \\ B_0 - AB_1 & = & a_1 I_n \\ -AB_0 & = & a_0 I_n \end{array}$$

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. This yields the following.

$$\begin{array}{rcl} A^n B_{n-1} & = & A^n \\ A^{n-1} B_{n-2} - A^n B_{n-1} & = & a_{n-1} A^{n-1} \\ A^{n-2} B_{n-3} - A^{n-1} B_{n-2} & = & a_{n-2} A^{n-2} \\ & \vdots & \\ AB_0 - A^2 B_1 & = & a_1 A \\ -AB_0 & = & a_0 I_n \end{array}$$

Proof (continued). Reminder:

$$\begin{aligned} A^n B_{n-1} &= A^n \\ A^{n-1}B_{n-2} - A^n B_{n-1} &= a_{n-1}A^{n-1} \\ A^{n-2}B_{n-3} - A^{n-1}B_{n-2} &= a_{n-2}A^{n-2} \\ &\vdots \\ AB_0 - A^2 B_1 &= a_1 A \\ -AB_0 &= a_0 I_n \end{aligned}$$

Proof (continued). Reminder:

$$\begin{array}{rcl} A^n B_{n-1} & = & A^n \\ A^{n-1} B_{n-2} - A^n B_{n-1} & = & a_{n-1} A^{n-1} \\ A^{n-2} B_{n-3} - A^{n-1} B_{n-2} & = & a_{n-2} A^{n-2} \\ & \vdots & \\ AB_0 - A^2 B_1 & = & a_1 A \\ -AB_0 & = & a_0 I_n \end{array}$$

We now add up the equations that we obtained.

Proof (continued). Reminder:

$$\begin{array}{rcl} A^n B_{n-1} & = & A^n \\ A^{n-1} B_{n-2} - A^n B_{n-1} & = & a_{n-1} A^{n-1} \\ A^{n-2} B_{n-3} - A^{n-1} B_{n-2} & = & a_{n-2} A^{n-2} \\ & \vdots & \\ AB_0 - A^2 B_1 & = & a_1 A \\ -AB_0 & = & a_0 I_n \end{array}$$

We now add up the equations that we obtained.

On the left-hand-side, the sum is obviously $O_{n\times n}$.

Proof (continued). Reminder:

$$\begin{array}{rcl} A^n B_{n-1} & = & A^n \\ A^{n-1} B_{n-2} - A^n B_{n-1} & = & a_{n-1} A^{n-1} \\ A^{n-2} B_{n-3} - A^{n-1} B_{n-2} & = & a_{n-2} A^{n-2} \\ & \vdots & \\ AB_0 - A^2 B_1 & = & a_1 A \\ -AB_0 & = & a_0 I_n \end{array}$$

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.

$$A^n + a_{n-1} A^{n-1} + a_{n-2} A^{n-2} + \cdots + a_1 A + a_0 I_n \quad = \quad O_{n \times n}.$$

But this is precisely what we needed to show. □

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

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n \quad = \quad O_{n\times n}.$$

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

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n \quad = \quad O_{n\times n}.$$

Corollary 8.3.1

Let $\mathbb{F}$ be a field. For all matrices $A \in \mathbb{F}^{n\times n}$:

a. $A^n \in \mathrm{Span}(I_n, A, A^2, \ldots, A^{n-1})$, i.e. A^n is a linear combination of $I_n, A, A^2, \ldots, A^{n-1}$;
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 $I_n, A, A^2, \ldots, A^{n-1}$.

Proof.

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

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n \quad = \quad O_{n \times n}.$$

Corollary 8.3.1

Let $\mathbb{F}$ be a field. For all matrices $A \in \mathbb{F}^{n \times n}$:

(a) $A^n \in \mathrm{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^n is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$;

(b) if A is invertible, then $A^{-1} \in \mathrm{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^{-1} is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$.

Proof. Fix a matrix $A \in \mathbb{F}^{n \times n}$, and consider its characteristic polynomial $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

Corollary 8.3.1

(a) $A^n \in \mathrm{Span}(I_n, A, A^2, \ldots, A^{n-1})$, i.e. A^n is a linear combination of $I_n, A, A^2, \ldots, A^{n-1}$;

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0.$

Proof of (a).

Corollary 8.3.1

(a) $A^n \in \operatorname{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^n is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$;

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0.$

Proof of (a). By the Cayley-Hamilton theorem, we have that

$$A^n + a_{n-1}A^{n-1} + \cdots + a_a A^2 + a_1 A + a_0 I_n \quad = \quad O_{n\times n}.$$

Corollary 8.3.1

(a) $A^n \in \text{Span}(I_n, A, A^2, \ldots, A^{n-1})$, i.e. A^n is a linear combination of $I_n, A, A^2, \ldots, A^{n-1}$;

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

Proof of (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,

$$A^n \;\; = \;\; -a_0 I_n - a_1 A - a_2 A^2 - \cdots - a_{n-1}A^{n-1}.$$

Corollary 8.3.1

(a) $A^n \in \text{Span}(I_n, A, A^2, \ldots, A^{n-1})$, i.e. A^n is a linear combination of $I_n, A, A^2, \ldots, A^{n-1}$;

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

Proof of (a). By the Cayley-Hamilton theorem, we have that

$$A^n + a_{n-1}A^{n-1} + \cdots + a_a A^2 + a_1 A + a_0 I_n \quad = \quad O_{n\times n}.$$

Consequently,

$$A^n \quad = \quad -a_0 I_n - a_1 A - a_2 A^2 - \cdots - a_{n-1}A^{n-1}.$$

Thus, A^n is a linear combination of the matrices $I_n, A, A^2, \ldots, A^{n-1}$.

Corollary 8.3.1

(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 $I_n, A, A^2, \ldots, A^{n-1}$.

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

Proof of (b). Assume that A is invertible.

Corollary 8.3.1

(b) if A is invertible, then $A^{-1} \in \text{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^{-1} is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$.

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

Proof of (b). Assume that A is invertible. Proposition 8.2.11 then guarantees that 0 is not an eigenvalue of A.

Corollary 8.3.1

(b) if A is invertible, then $A^{-1} \in \text{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^{-1} is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$.

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

Proof of (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 $a_0 \neq 0$.

Corollary 8.3.1

(b) if A is invertible, then $A^{-1} \in \text{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^{-1} is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$.

- Reminder:
  $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + a_0$.

Proof of (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 $a_0 \neq 0$.

Now, by the Cayley-Hamilton theorem, we have that

$$A^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I_n \quad = \quad O_{n\times n}.$$

Corollary 8.3.1

(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 $I_n, A, A^2, \ldots, A^{n-1}$.

Proof of (b) (continued). Reminder: $a_0 \neq 0$,

$$A^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I_n = O_{n\times n}.$$

Corollary 8.3.1

(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 $I_n, A, A^2, \ldots, A^{n-1}$.

Proof of (b) (continued). Reminder: $a_0 \neq 0$,

$$A^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I_n \quad = \quad O_{n\times n}.$$

We multiply both sides of the equation by A^{-1} on the right, and we obtain

$$A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1I_n + a_0A^{-1} \quad = \quad O_{n\times n},$$

Corollary 8.3.1

(b) if A is invertible, then $A^{-1} \in \text{Span}(I_n, A, A^2, \ldots, A^{n-1})$, i.e. A^{-1} is a linear combination of $I_n, A, A^2, \ldots, A^{n-1}$.

Proof of (b) (continued). Reminder: $a_0 \neq 0$,

$$A^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I_n \;\; = \;\; O_{n\times n}.$$

We multiply both sides of the equation by A^{-1} on the right, and we obtain

$$A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1I_n + a_0A^{-1} \;\; = \;\; O_{n\times n},$$

and consequently,

$$a_0A^{-1} \;\; = \;\; -a_1I_n - a_2A - \cdots - a_{n-1}A^{n-2} - A^{n-1}.$$

Corollary 8.3.1

(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 $I_n, A, A^2, \ldots, A^{n-1}$.

Proof of (b) (continued). Reminder: $a_0 \neq 0$,

$$A^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I_n \quad = \quad O_{n\times n}.$$

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

$$A^{-1} \quad = \quad -\tfrac{a_1}{a_0}I_n - \tfrac{a_2}{a_0}A - \cdots - \tfrac{a_{n-1}}{a_0}A^{n-2} - \tfrac{1}{a_0}A^{n-1}.$$

Corollary 8.3.1

(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 $I_n, A, A^2, \ldots, A^{n-1}$.

Proof of (b) (continued). Reminder: $a_0 \neq 0$,

$$A^n + a_{n-1}A^{n-1} + \cdots + a_2A^2 + a_1A + a_0I_n = O_{n\times n}.$$

We multiply both sides of the equation by A^{-1} on the right, and we obtain

$$A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1I_n + a_0A^{-1} = O_{n\times n},$$

and consequently,

$$a_0A^{-1} = -a_1I_n - a_2A - \cdots - a_{n-1}A^{n-2} - A^{n-1}.$$

Since $a_0 \neq 0$, this implies that

$$A^{-1} = -\frac{a_1}{a_0}I_n - \frac{a_2}{a_0}A - \cdots - \frac{a_{n-1}}{a_0}A^{n-2} - \frac{1}{a_0}A^{n-1}.$$

So, A^{-1} is a linear combination of $I_n, A, A^2, \ldots, A^{n-1}$. $\square$

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

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n = O_{n\times n}.$$

Corollary 8.3.1

Let $\mathbb{F}$ be a field. For all matrices $A \in \mathbb{F}^{n\times n}$:

(a) $A^n \in \mathrm{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^n is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$;

(b) if A is invertible, then $A^{-1} \in \mathrm{Span}(I_n, A, A^2, \dots, A^{n-1})$, i.e. A^{-1} is a linear combination of $I_n, A, A^2, \dots, A^{n-1}$.