6.3 Eigenvalues and characteristic polynomials.

0. *Assumed background.*

   - What has been covered in Topics 1-5.
   - 6.1 *Eigenvalues and eigenvectors for square matrices.*
   - 6.2 *Diagonalization and diagonalizability for square matrices.*

Abstract. We introduce:—

   - the notion of characteristic polynomials of square matrices.

1. Recall the result below, labelled Theorem ($\sharp$), about equivalent formulation for the notion of eigenvalues for a square matrix in terms of invertibility and determinants:—

   Theorem ($\sharp$). (Equivalent formulations for the notion of eigenvalues for square matrices.)

   Suppose A is an $(n \times n)$-square matrix, and λ is a number. Then the statements below are logically equivalent:—

   (1) *λ is an eigenvalue of A.*

   (2) *The homogeneous system $\mathcal{LS}(A - \lambda I_n, \mathbf{0}_n)$ has a non-trivial solution.*

   (3) *$A - \lambda I_n$ is not invertible.*

   (4) *The rank of $A - \lambda I_n$ is at most $n - 1$.*

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

   Now suppose one of (1), (2), (3), (4), (5) *holds. (So all of them hold.)*

   Further suppose $\mathbf{v}$ is a column vector with n entries.

   Then $\mathbf{v}$ is an eigenvector of A with eigenvalue λ if and only if $\mathbf{v}$ is a non-trivial solution of $\mathcal{LS}(A - \lambda I_n, \mathbf{0}_n)$.

2. The logical equivalence between the statements (1), (5) motivates the definition below:

   Definition. (Characteristic polynomial of a matrix.)

   Let A be an $(n \times n)$-square matrix. The (algebraic) expression $\det(A - xI_n)$ *(with indeterminate x) is called the* **characteristic polynomial of the square matrix** *A. It is denoted by $p_A(x)$.*

3. Using mathematical induction, we can prove the result below (which confirms that it makes sense to use the word 'polynomial' in the phrase 'characteristic polynomial of a square matrix').

   Theorem (1).

   Suppose A is an $(n \times n)$-square matrix.

   Then $p_A(x)$ is a degree-n polynomial with indeterminate x, with leading coefficient $(-1)^n$, and with constant coefficient $\det(A)$.

   Remark. The multiple of $(-1)^{n-1}$ with the coefficient of the degree-$(n - 1)$ term in the polynomial $p_A(x)$ is called the **trace of** A, and is denoted by $\mathrm{tr}(A)$.

4. Theorem ($\sharp$) now gives an equivalent formulation for the notion of eigenvalues in terms of polynomials and roots.

   Theorem (2).

   Suppose A is an $(n \times n)$-square matrix, and λ is a number. Then the statements below are logically equivalent:

   (1) *λ is an eigenvalue of A.*

   (2) *λ is a root of $p_A(x)$.*

5. **Example (1). (Characteristic polynomials for square matrices.)**

   (a) Suppose $A = \begin{bmatrix} 13 & 30 \\ -6 & -14 \end{bmatrix}$. Then

$$\begin{aligned} p_A(x) &= \det(A - xI_2) = \det(\begin{bmatrix} 13 - x & 30 \\ -6 & -14 - x \end{bmatrix}) \\ &= (13 - x)(-14 - x) - (-6) \cdot 30 = x^2 + x - 2 = (x - 1)(x + 2). \end{aligned}$$

(b) Suppose $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$. Then

$$\begin{aligned} p_A(x) &= \det(A - xI_3) = \det(\begin{bmatrix} 1-x & 1 & 1 \\ 0 & 2-x & 2 \\ 0 & 0 & 3-x \end{bmatrix}) \\ &= (1-x)(2-x)(3-x) = -(x-1)(x-2)(x-3) = -x^3 + 6x^2 - 11x + 6. \end{aligned}$$

(c) Suppose $A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 2 & 1 & 1 & 1 \\ -5 & 2 & 5 & -1 \end{bmatrix}$. Then

$$p_A(x) = \det(A - xI_4) = \cdots = x^4 - 10x^2 + 9 = (x+3)(x+1)(x-1)(x-3).$$

(d) Suppose $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Then

$$\begin{aligned} p_A(x) &= \det(A - xI_3) = \det(\begin{bmatrix} 2-x & 1 & 1 \\ 1 & 2-x & 1 \\ 1 & 1 & 2-x \end{bmatrix}) = \det(\begin{bmatrix} 2-x & 1 & 1 \\ 1 & 2-x & 1 \\ 0 & -1+x & 1-x \end{bmatrix}) \\ &= \det(\begin{bmatrix} 2-x & 2 & 1 \\ 1 & 3-x & 1 \\ 0 & 0 & 1-x \end{bmatrix}) \\ &= (1-x)\det(\begin{bmatrix} 2-x & 2 \\ 1 & 3-x \end{bmatrix}) = (1-x)\det(\begin{bmatrix} 2-x & 2 \\ -1+x & 1-x \end{bmatrix}) = (1-x)\det(\begin{bmatrix} 4-x & 2 \\ 0 & 1-x \end{bmatrix}) \\ &= (1-x)^2(4-x) = -(x-1)^2(x-4) = -x^3 + 6x^2 - 9x + 4. \end{aligned}$$

(e) Suppose $A = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$.Then

$$p_A(x) = \det(A - xI_2) = \det(\begin{bmatrix} 1-x & 4 \\ 0 & 1-x \end{bmatrix}) = (1-x)^2 = (x-1)^2 = x^2 - 2x + 1$$

(f) Suppose $A = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$. Then

$$p_A(x) = \det(A - xI_3) = \det(\begin{bmatrix} 1-x & 4 & 0 \\ 0 & 1-x & 2 \\ 0 & 0 & 1-x \end{bmatrix}) = (1-x)^3 = -(x-1)^3 = -x^3 + 3x^2 - 3x + 1$$

(g) Suppose $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Then

$$\begin{aligned} p_A(x) &= \det(A - xI_2) = \det(\begin{bmatrix} 1-x & -1 \\ 1 & 1-x \end{bmatrix}) \\ &= (1-x)^2 + 1 = x^2 - 2x + 2 = [x - (1+i)][x - (1-i)] \end{aligned}$$

(h) Suppose $A = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$. Then

$$\begin{aligned} p_A(x) &= \det(A - xI_4) \\ &= \cdots \\ &= (1-x)^4 + 1 = x^4 - 4x^3 + 6x^2 - 4x + 2 \\ &= [(x-1)^2 - i][(x-1)^2 + i] \\ &= \left[x - \left(1 + \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)\right]\left[x - \left(1 - \frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}\right)\right]\left[x - \left(1 - \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)\right]\left[x - \left(1 + \frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}\right)\right]. \end{aligned}$$

6. We now state a famous result first discovered and established by Gauss, whose proof is beyond the scope of this course:—

Fundamental Theorem of Algebra.

Suppose $f(x)$ is a non-constant polynomial whose coefficients are complex numbers. Then $f(x)$ has a root amongst complex numbers.

7. Using the Fundamental Theorem of Algebra, and using Factor Theorem, we can deduce the result below from Theorem (2). Example (1) has already provided some illustrations for its content.

Theorem (3).

Suppose A is an $(n \times n)$-square matrix. Then there are some pairwise distinct numbers $\lambda_1, \lambda_2, \cdots, \lambda_s$, and some positive integers $m_1, m_2, \cdots, m_s$, such that:—

(1) $p_A(x) = (-1)^n (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_s)^{m_s}$ *as polynomials,*

(2) $\lambda_1, \lambda_2, \cdots, \lambda_s$ *are all the (pairwise distinct) eigenvalues of A, and*

(3) $m_1 + m_2 + \cdots + m_s = n$.

Remarks on terminologies.

(a) For each $j = 1, 2, \cdots, s$, the positive integer m_j is called the **algebraic multiplicity of the eigenvalue λ_j of A.**

(b) The set $\{\lambda_1, \lambda_2, \cdots, \lambda_s\}$ is called the **spectrum of A**.

Further remark. Theorem (3) is stated only for reference; we are not going to rely on it in the rest of material.

8. Examples on non-diagonalizable upper-triangular matrices suggest that the main obstacle against diagonalizability is a lack of 'sufficiently many' distinct eigenvalues. So it is natural for us to ask whether a square matrix which has 'sufficiently many' distinct eigenvalues will be guaranteed to be diagonalizable.

This turns out to be the case. But at the same time, we will know that the same square matrix cannot have 'too many' eigenvalues.

Theorem (4).

Suppose A is an $(n \times n)$-square matrix. Then the statements below hold:—

(1) *A possesses at most n (pairwise distinct) eigenvalues.*

(2) *If A possesses exactly n (pairwise distinct) eigenvalues, then A is diagonalizable.*

Remarks.

(a) The statement (1) is already (implicitly) implied by Theorem (3). This is a consequence of three 'facts':—
   - The degree of the characteristic polynomial of a square matrix is the same as the size of that matrix.
   - There are at most as many roots for a polynomial as its degree.
   - A number is an eigenvalue of a square matrix exactly when it is a root of the characteristic polynomial of that square matrix.

   Nonetheless, we will give an argument for the statement (1) which does not rely on Theorem (3).

(b) The statement (2) only provide a sufficient condition for diagonalizability. Its converse is false. It can happen that a square matrix has just one eigenvalue and yet it is diagonalizable: two trivial examples are the zero square-matrix and the identity matrix.

9. We will give an argument for Theorem (4) with the help of Theorem (6). The latter relies on the 'technical result' which is Lemma (5) below. The proofs of Lemma (5) and Theorem (6) will be given later.

Lemma (5).

Let A be an $(n \times n)$-square matrix, and $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p, \mathbf{v}$ are non-zero column vectors with n entries.

Suppose (1), (2), (3), (4) *hold:—*

(1) *$\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p$ are linearly independent.*

(2) *$\mathbf{v}$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p$.*

(3) *$\mathbf{v}$ is not a linear combination of (at most) $p - 1$ column vectors amongst $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p$.*

(4) *$\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p, \mathbf{v}$ are eigenvectors of A, with corresponding eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p, \mu$ respectively.*

Then $\lambda_1 = \lambda_2 = \cdots = \lambda_p = \mu$.

10. **Theorem (6). (Linear independence of eigenvectors of distinct eigenvalues.)**

Let A be a square matrix.

Suppose $\kappa_1, \kappa_2, \cdots, \kappa_q$ are pairwise distinct eigenvalues of A.

Further suppose $\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_q$ are eigenvalues of A, with corresponding eigenvalues $\kappa_1, \kappa_2, \cdots, \kappa_q$ respectively.

Then $\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_q$ are linearly independent.

11. For the moment, we take for granted the validity of Lemma (5) and Theorem (6), and apply the latter to give an argument for Theorem (4).

Proof of Theorem (4).

Suppose A is an $(n \times n)$-square matrix.

(1) We verify that A possesses at most n (pairwise distinct) eigenvalues, with the proof-by-contradiction method.
Suppose it were true that A possessed $n+1$ (pairwise distinct) eigenvalues, say, $\lambda_1, \lambda_2, \cdots, \lambda_n, \lambda_{n+1}$, (and perhaps other eigenvalues as well).
For each $j = 1, 2, \cdots, n, n+1$, there is some eigenvector $\mathbf{u}_j$ of A corresponding to the eigenvalue λ_j.
Now by Theorem (5), $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n, \mathbf{u}_{n+1}$ would be linearly independent column vectors, each with n entries. Contradiction arises.

(2) Suppose A possesses exactly n (pairwise distinct) eigenvalues, say, $\lambda_1, \lambda_2, \cdots, \lambda_n$.
For each $j = 1, 2, \cdots, n$, there is some eigenvector $\mathbf{u}_j$ of A corresponding to the eigenvalue λ_j.
Now by Theorem (5), $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are linearly independent column vectors, each with n entries.
Then A is diagonalizable, with a diagonalization given by $U^{-1}AU = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, in which U is the invertible matrix given by $U = [\ \mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n\ ]$.

12. We can in fact give a necessary and sufficient condition for diagonalizability with focus put on eigenvalues.

For simplicity we restrict ourselves to real numbers, and to matrices and vectors with real entries:—

Theorem (7). (Necessary and sufficient conditions for diagonalizability, in terms dimensions and bases for eigenspaces of various eigenvalues.)

Let A is an $(n \times n)$-square matrix with real entries.

Suppose $\lambda_1, \lambda_2, \cdots, \lambda_s$ are all the eigenvalues of A, pairwise distinct. Further suppose all of $\lambda_1, \lambda_2, \cdots, \lambda_s$ are real numbers.
Then the statements below are logically equivalent:—

(1) *The equality $\dim(\mathcal{N}(A - \lambda_1 I_n)) + \dim(\mathcal{N}(A - \lambda_2 I_n)) + \cdots + \dim(\mathcal{N}(A - \lambda_s I_n)) = n$ holds.*

(2) *A is diagonalizable.*

Now suppose either of (1), (2) holds. (So both of them hold.)

For each $k = 1, 2, \cdots, p$, write $\dim(\mathcal{N}(A - \lambda_k)) = n_k$, and suppose that $\mathbf{v}_{k,1}, \mathbf{v}_{k,2}, \cdots, \mathbf{v}_{k,n_k}$ constitute a basis for $\mathcal{N}(A - \lambda_k)$.
Then a basis for $\mathbb{R}^n$, in which all n column vectors are eigenvectors of A, is constituted by

$$\begin{array}{llll} \mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \cdots, \mathbf{v}_{1,n_1}, & & & \\ & \mathbf{v}_{2,1}, \mathbf{v}_{2,2}, \cdots, \mathbf{v}_{2,n_2}, & & \\ & & \ddots & \\ & & & \mathbf{v}_{s,1}, \mathbf{v}_{s,2}, \cdots, \mathbf{v}_{s,n_s}. \end{array}$$

Proof of Theorem (7). Omitted. (It is not difficult, but it is a tedious exercise in book-keeping.)

13. **Example (2). (Illustration of the content of Theorem (7).)**

Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, and $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$.

$\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are eigenvectors of A with respective eigenvalues $4, 1, 1$.

Write $\lambda_1 = 4$, $\lambda_2 = 1$.

The only eigenspaces of A are $\mathcal{N}(A - \lambda_1 I_3), \mathcal{N}(A - \lambda_2 I_3)$.

Note that $\dim(\mathcal{N}(A - \lambda_1 I_3))$. A basis for $\mathcal{N}(A - \lambda_1 I_3)$ given by $\mathbf{u}_1$.

Note that $\dim(\mathcal{N}(A - \lambda_2 I_3)) = 2$. A basis for $\mathcal{N}(A - \lambda_2 I_3)$ is given by $\mathbf{u}_2, \mathbf{u}_3$.

The respective dimensions of $\mathcal{N}(A - \lambda_1 I_3), \mathcal{N}(A - \lambda_2 I_3)$ add up to give 3.

Coincidentally, A is diagonalizable (as expected from theory), with a diagonalization of A given by $U^{-1}AU = \mathrm{diag}(4, 1, 1)$, in which $U = [\ \mathbf{u}_1 \mid \mathbf{u}_2 \mid \mathbf{u}_3\ ]$.

14. Before moving onto something else, we first complete the arguments leading towards Theorem (6) by providing a proof for Lemma (4) and a proof for Theorem (5).

Proof of Lemma (4).

Let A be an $(n \times n)$-square matrix, and $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p, \mathbf{v}$ are non-zero column vectors with n entries.

Suppose (1), (2), (3), (4) hold:—

(1) $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p$ are linearly independent.

(2) $\mathbf{v}$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p$.

(3) $\mathbf{v}$ is not a linear combination of (at most) $p-1$ column vectors amongst $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p$.

(4) $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p, \mathbf{v}$ are eigenvectors of A, with corresponding eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p, \mu$ respectively.

By (2), there exist some numbers $\alpha_1, \alpha_2, \cdots, \alpha_p$ such that $\mathbf{v} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_p\mathbf{u}_p$.

By (4), we have

$$\begin{aligned}\mu\mathbf{v} &= A\mathbf{v} \\ &= A(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_p\mathbf{u}_p) \\ &= \alpha_1 A\mathbf{u}_1 + \alpha_2 A\mathbf{u}_2 + \cdots + \alpha_p A\mathbf{u}_p \\ &= \alpha_1\lambda_1\mathbf{u}_1 + \alpha_2\lambda_2\mathbf{u}_2 + \cdots + \alpha_p\lambda_p\mathbf{u}_p \quad \text{——} \ (\sharp)\end{aligned}$$

Note that we also have

$$\begin{aligned}\mu\mathbf{v} &= \mu(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_p\mathbf{u}_p) \\ &= \alpha_1\mu\mathbf{u}_1 + \alpha_2\mu\mathbf{u}_2 + \cdots + \alpha_p\mu\mathbf{u}_p \quad \text{——} \ (\natural)\end{aligned}$$

Now by $(\sharp), (\natural)$, we obtain

$$\alpha_1(\lambda_1 - \mu)\mathbf{u}_1 + \alpha_2(\lambda_2 - \mu)\mathbf{u}_2 + \cdots + \alpha_p(\lambda_p - \mu)\mathbf{u}_p = \cdots = A\mathbf{v} - \mu\mathbf{v} = \mathbf{0}_n$$

By (1), $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p$ are linearly independent. Then by the definition of linear independence, we have

$$\alpha_j(\lambda_j - \mu) = 0 \quad \text{for each } j = 1, 2, \cdots, p.$$

By (3), each of $\alpha_1, \alpha_2, \cdots, \alpha_p$ is non-zero. Then $\lambda_1 = \lambda_2 = \cdots = \lambda_p = \mu$.

15. **Proof of Theorem (5).** [We apply the proof-by-contradiction method.]

Let A be a square matrix. Suppose $\kappa_1, \kappa_2, \cdots, \kappa_q$ are pairwise distinct eigenvalues of A.

Further suppose $\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_q$ are eigenvalues of A, with corresponding eigenvalues $\kappa_1, \kappa_2, \cdots, \kappa_q$ respectively.

Also suppose it were true that $\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_q$ were linearly dependent.

Without loss of generality, we might assume that $\mathbf{t}_1$ was a linear combination of $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_q$.

(a) Note that $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_q$ are linearly independent.
By assumption, each of $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_q, \mathbf{t}_1$ is not the zero column vector.
Also by assumption, '$\kappa_2 = \kappa_3 = \cdots = \kappa_q = \kappa_1$' fails to hold.
Then, by Lemma (4) (and by logic), $\mathbf{t}_1$ is a linear combination of at most $q-2$ column vectors amongst $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_q$. Without loss of generality, we may assume they are $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_{q-1}$.

(b) Repeating the above argument, we deduce that $\mathbf{t}_1$ is a linear combination of at most $q-3$ column vectors amongst $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_{q-1}$. Without loss of generality, we may assume they are $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_{q-2}$.

(c) Further repeating the above arguments (for finitely many times), we eventually deduce that $\mathbf{t}_1$ is a linear combination of exactly one of $\mathbf{t}_2, \mathbf{t}_3, \cdots, \mathbf{t}_q$, say, $\mathbf{t}_2$.

But by assumption, $\kappa_1 \neq \kappa_2$. Contradiction arises.

16. We now proceed to deduce a relation between the positive integral powers of a diagonalizable matrix and the coefficients of the characteristic polynomial of the same matrix.

Lemma (8).

Suppose A is a diagonalizable $(n \times n)$-square matrix, with a diagonalization given by

$$U^{-1}AU = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n),$$

in which U is some invertible $(n \times n)$-square matrix.

Then $p_A(x) = (-1)^n (x - \lambda_1)(x - \lambda_2) \cdot \ldots \cdot (x - \lambda_n)$ as polynomials.

Remark. We can further deduce from the basic properties of diagonalizable matrices:—

(a) For each positive integer q,

$$p_{A^q}(x) = (-1)^n (x - \lambda_1{}^q)(x - \lambda_2{}^q) \cdot \ldots \cdot (x - \lambda_n{}^q) \quad \text{as polynomials.}$$

(b) When A is invertible,

$$p_{A^{-1}}(x) = (-1)^n (x - 1/\lambda_1)(x - 1/\lambda_2) \cdot \ldots \cdot (x - 1/\lambda_n) \quad \text{as polynomials.}$$

17. **Proof of Lemma (8).**

Suppose A is a diagonalizable $(n \times n)$-square matrix, with a diagonalization given by

$$U^{-1}AU = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n),$$

in which U is some invertible $(n \times n)$-square matrix.

Write $D = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

Note that $A - xI_n = UDU^{-1} - U(xI_n)U^{-1} = U(D - xI_n)U^{-1}$.

Also note that $D - xI_n = \mathrm{diag}(\lambda_1 - x, \lambda_2 - x, \cdots, \lambda_n - x)$.

Then, as polynomials,

$$\begin{aligned} p_A(x) &= \det(A - xI_n) = \det(U(D - xI_n)U^{-1}) \\ &= \det(U) \cdot \det(D - xI_n) \cdot \det(U^{-1}) \\ &= \det(U) \cdot \det(D - xI_n) \cdot (\det(U))^{-1} \\ &= (\lambda_1 - x)(\lambda_2 - x) \cdot \ldots \cdot (\lambda_n - x) = (-1)^n (x - \lambda_1)(x - \lambda_2) \cdot \ldots \cdot (x - \lambda_n) \end{aligned}$$

18. Lemma (8) will be used in deducing the result below.

Theorem (9). (A special case of the Cayley-Hamilton Theorem, for diagonalizable square matrices.)

Suppose A is a diagonalizable $(n \times n)$-square matrix.

For each j, denote the coefficient of the j-th power term of $p_A(x)$ is c_j.

(So $p_A(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1}x^{n-1} + c_n x^n$ as polynomials.)

Then $c_0 I_n + c_1 A + c_2 A^2 + \cdots + c_{n-1}A^{n-1} + c_n A^n = \mathcal{O}_{n \times n}$.

Remark. The conclusion in Theorem (9) is often presented as $p_A(A) = \mathcal{O}_{n \times n}$.

Further note that $c_n = (-1)^n$. As a consequence,

$$A^n = (-1)^{n+1} c_0 I_n + (-1)^{n+1} c_1 A + (-1)^{n+1} c_2 A^2 + \cdots + (-1)^{n+1} c_{n-1} A^{n-1}.$$

So A^n is a 'linear combination' of $I_0, A, A^2, \cdots, A^{n-1}$.

19. **Theorem (10). (Corollary to Theorem (9).)**

Suppose A is a diagonalizable $(n \times n)$-square matrix. Then the statements below hold:—

(1) *For each positive number q, there exist some numbers $g_0(q), g_1(q), g_2(q), \cdots, g_{n-1}(q)$ such that*

$$A^q = g_0(q)I_n + g_1(q)A + g_2(q)A^2 + \cdots + g_{n-1}(q)A^{n-1}.$$

(2) *Suppose A is invertible.*

Then, for each positive integer q, there exist some numbers $g_0(-q), g_1(-q), g_2(-q), \cdots, g_{n-1}(-q)$ such that

$$A^{-q} = g_0(-q)I_n + g_1(-q)A + g_2(-q)A^2 + \cdots + g_{n-1}(-q)A^{n-1}.$$

Remark. In plain words, every positive integral power of A is a 'linear combination' of the 'low positive powers' of A given by $I_0, A, A^2, \cdots, A^{n-1}$. The same can be said of every negative integral power of A under the assumption of A being invertible in the first place.

20. **Proof of Theorem (9).**

Suppose A is a diagaonalizable $(n \times n)$-square matrix, with some diagonalization given by $U^{-1}AU = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, for some invertible $(n \times n)$-square matrix U.

For each $k = 1, 2, \cdots, n$, the number λ_k is an eigenvalue of A. Then $p_A(\lambda_k) = 0$.

Note that for each positive integer q,

$$U^{-1}A^qU = (U^{-1}AU)^q = (\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n))^q = \mathrm{diag}({\lambda_1}^q, {\lambda_2}^q, \cdots, {\lambda_n}^q).$$

For each j, denote the coefficient of the j-th power term of $p_A(x)$ is c_j.

(So $p_A(x) = c_0 + c_1x + c_2x^2 + \cdots + c_{n-1}x^{n-1} + c_nx^n$ as polynomials.)

We have

$$\begin{aligned}
& U^{-1}(c_0I_n + c_1A + c_2A^2 + \cdots + c_{n-1}A^{n-1} + c_nA^n)U \\
= \; & c_0I_n + c_1U^{-1}AU + c_2U^{-1}A^2U + \cdots + c_{n-1}U^{-1}A^{n-1}U + c_nU^{-1}A^nU \\
= \; & c_0I_n + c_1\,\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) + c_2\,\mathrm{diag}({\lambda_1}^2, {\lambda_2}^2, \cdots, {\lambda_n}^2) \\
& + \cdots + c_{n-1}\,\mathrm{diag}({\lambda_1}^{n-1}, {\lambda_2}^{n-1}, \cdots, {\lambda_n}^{n-1}) + c_n\,\mathrm{diag}({\lambda_1}^n, {\lambda_2}^n, \cdots, {\lambda_n}^n) \\
= \; & \mathrm{diag}(p_A(\lambda_1), p_A(\lambda_2), \cdots, p_A(\lambda_n)) = \mathrm{diag}(0, 0, \cdots, 0) = \mathcal{O}_{n\times n}
\end{aligned}$$

Then $c_0I_n + c_1A + c_2A^2 + \cdots + c_{n-1}A^{n-1} + c_nA^n = U\mathcal{O}_{n\times n}U^{-1} = \mathcal{O}_{n\times n}$.

21. Theorem (9) is a special case of the result below, known as the Cayley-Hamilton Theorem. Its proof is omitted here. (It can be given with the use of adjoints of square matrices. It is accessible at the level of this course, and can be found in standard textbooks.) Theorem (10) can be generalized accordingly, with the dropping of the assumption on diagonalizability.

Cayley-Hamilton Theorem.

Suppose A is an $(n \times n)$-square matrix.

For each j, denote the coefficient of the j-th power term of $p_A(x)$ is c_j.

(So $p_A(x) = c_0 + c_1x + c_2x^2 + \cdots + c_{n-1}x^{n-1} + c_nx^n$ as polynomials.)

Then $c_0I_n + c_1A + c_2A^2 + \cdots + c_{n-1}A^{n-1} + c_nA^n = \mathcal{O}_{n\times n}$.