Matrix Analysis (Lecture 1)

Yikun Zhang[*]

March 29, 2018

Abstract

Linear algebra and matrix analysis have long played a fundamental and indispensable role in fields of science. Many modern research projects in applied science rely heavily on the theory of matrix. Meanwhile, matrix analysis and its ramifications are active fields for research in their own right. In this course, we aim to study some fundamental topics in matrix theory, such as eigen-pairs and equivalence relations of matrices, scrutinize the proofs of essential results, and dabble in some up-to-date applications of matrix theory. Our lecture will maintain a cohesive organization with the main stream of Horn's book[1] and complement some necessary background knowledge omitted by the textbook sporadically.

1 Introduction

We begin our lectures with Chapter 1 of Horn's book[1], which focuses on eigenvalues, eigenvectors, and similarity of matrices. In this lecture, we review the concepts of eigenvalues and eigenvectors with which we are familiar in linear algebra, and investigate their connections with coefficients of the characteristic polynomial. Here we first outline the main concepts in Chapter 1 (Eigenvalues, Eigenvectors, and Similarity).

[*]School of Mathematics, Sun Yat-sen University

1.1 Change of basis and similarity (Page 39-40, 43)

Let V be an n-dimensional vector space over the field $\mathbf{F}$, which can be $\mathbf{R}$, $\mathbf{C}$, or even $\mathbf{Z}(p)$ (the integers modulo a specified prime number p), and let the list $\mathcal{B}_1 = \{v_1, v_2, ..., v_n\}$ be a basis for V. Any vector $x \in V$ can be represented as $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$ because $\mathcal{B}_1$ spans V. This representation of x in $\mathcal{B}_1$ is unique, since if there were some other representation of $x = \beta_1 v_1 + \cdots + \beta_n v_n$ in the same basis, then

$$0 = x - x = (\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_n - \beta_n)v_n$$

from which it follows that $\alpha_i - \beta_i = 0$ because the list $\mathcal{B}_1$ is independent. Thus, given the basis $\mathcal{B}_1$, the linear mapping

$$x \rightarrow [x]_{\mathcal{B}_1} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \text{ in which } x = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

from V to $\mathbf{F}^n$ is well-defined, one-to-one, and onto.
The scalars α_i are called the *coordinates* of x with respect to the basis $\mathcal{B}_1$, and the column vector $[x]_{\mathcal{B}_1}$ is the unique $\mathcal{B}_1$-*coordinate representation of* x.

We now move on to talk about linear transformations on different bases and properties of change-of-basis matrices. Let $T : V \rightarrow V$ be a given linear transformation. The action of T on any $x \in V$ is determined by its action on the basis $\mathcal{B}_1$, i.e., the n vectors $Tv_1, ..., Tv_n$. This is because any $x \in V$ has a unique representation $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and

$$Tx = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 Tv_1 + \cdots + \alpha_n Tv_n$$

by linearity of T. Thus the value of Tx is clear once $[x]_{\mathcal{B}_1}$ is known.

What happen if we change the basis of V? Can we explicitly uncover the representation of T in terms of two given bases?
Let $\mathcal{B}_2 = \{w_1, ..., w_n\}$ also be a basis for V (either different from or the same as $\mathcal{B}_1$) and suppose that the $\mathcal{B}_2$-coordinate representation of Tv_j is

$$[Tv_j]_{\mathcal{B}_2} = \begin{bmatrix} t_{1j} \\ \vdots \\ t_{nj} \end{bmatrix}, \text{ where } j = 1, 2, ..., n.$$

Then, for any $x \in V$, we have

$$[Tx]_{\mathcal{B}_2} = \left[\sum_{j=1}^{n} \alpha_j T v_j\right]_{\mathcal{B}_2} = \sum_{j=1}^{n} \alpha_j [Tv_j]_{\mathcal{B}_2} = \sum_{j=1}^{n} \alpha_j \begin{bmatrix} t_{1j} \\ \vdots \\ t_{nj} \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

What we have shown is that $[Tx]_{\mathcal{B}_2} =_{\mathcal{B}_2} [T]_{\mathcal{B}_1}[x]_{\mathcal{B}_1}$. The n-by-n array $[t_{ij}]$ depends on T and on the choice of the bases $\mathcal{B}_1$ and $\mathcal{B}_2$, but it does not depend on x.

Definition 1.1. *The $\mathcal{B}_1$-$\mathcal{B}_2$* ***basis representation of*** *T is defined to be*

$$_{\mathcal{B}_2}[T]_{\mathcal{B}_1} = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nn} \end{bmatrix} = \begin{bmatrix} [Tv_1]_{\mathcal{B}_2} & \dots & [Tv_n]_{\mathcal{B}_2} \end{bmatrix}.$$

In the important special case where $\mathcal{B}_2 = \mathcal{B}_1$, we have $_{\mathcal{B}_1}[T]_{\mathcal{B}_1}$, which is called the $\mathcal{B}_1$ ***basis representation of*** *T.*

Consider the identity linear transformation $I : V \to V$ defined by $Ix = x$ for all x.

Definition 1.2. *The matrix $_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is called the $\mathcal{B}_1 - \mathcal{B}_2$* ***change of basis matrix****, where I is the identity linear transformation.*

Now we are well-prepared to prove the main result of this section, which is stated as a proposition.

Proposition 1.3. *Every invertible matrix is a change-of-basis matrix, and every change-of-basis matrix is invertible.*

Proof. Given the identity transformation $I : V \to V$,

$$[x]_{\mathcal{B}_2} = [Ix]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,[x]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,[Ix]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}\,[x]_{\mathcal{B}_2}$$

for all $x \in V$. By successively choosing $x = w_1, ..., w_n$ and using the fact that

$$[w_i]_{\mathcal{B}_2} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i, \text{ where } i = 1, 2, ..., n,$$

this identity permits us to identify each column of ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1 \mathcal{B}_1}[I]_{\mathcal{B}_2}$ and shows that

$$ {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\, {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} = I_n $$

Similarly, if we do the same computation starting with $[x]_{\mathcal{B}_1} = [Ix]_{\mathcal{B}_1} = \cdots$, we obtain that

$$ {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}\, {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = I_n $$

Thus, every matrix of the form ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is invertible and ${}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ is its inverse. Conversely, every invertible matrix $S = [s_1\, s_2\, \dots s_n] \in M_n(\mathbf{F})$ has the form ${}_{\mathcal{B}_1}[I]_{\mathcal{B}}$ for some basis $\mathcal{B}$. We may take $\mathcal{B}$ to be the vectors $\{\tilde{s}_1, ..., \tilde{s}_n\}$ defined by

$$ [\tilde{s}_i]_{\mathcal{B}_1} = {}_{\mathcal{B}_1}[I]_{\mathcal{B}}\, [\tilde{s}_i]_{\mathcal{B}} = \begin{bmatrix} s_1 & s_2 & \cdots & s_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = s_i, \text{ where } i = 1, 2, ..., n. $$

The list $\mathcal{B}$ is independent, because if $a_1\tilde{s}_1 + \cdots + a_n\tilde{s}_n = 0$, then

$$ 0 = [0]_{\mathcal{B}_1} = [a_1\tilde{s}_1 + \cdots + a_n\tilde{s}_n]_{\mathcal{B}_1} = \begin{bmatrix} s_1 & \cdots & s_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} $$

Due to the fact that S is invertible, we find that $a_i = 0$, $i = 1, 2, ..., n$. $\square$

Notice that

$$ {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = \begin{bmatrix} [Iv_1]_{\mathcal{B}_2} & \cdots & [Iv_n]_{\mathcal{B}_2} \end{bmatrix} = \begin{bmatrix} [v_1]_{\mathcal{B}_2} & \cdots & [v_n]_{\mathcal{B}_2} \end{bmatrix}, $$

so ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ describes how the elements of the basis $\mathcal{B}_1$ are represented in terms of elements of the basis $\mathcal{B}_2$. Now let $x \in V$ and compute

$$ \begin{aligned} {}_{\mathcal{B}_2}[T]_{\mathcal{B}_2}\, [x]_{\mathcal{B}_2} = [Tx]_{\mathcal{B}_2} &= [I(Tx)]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\, [Tx]_{\mathcal{B}_1} \\ &= {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\, {}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\, [x]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\, {}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\, [Ix]_{\mathcal{B}_1} \\ &= {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\, {}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\, {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}\, [x]_{\mathcal{B}_2} \end{aligned} $$

By choosing $x = w_1, ..., w_n$ successively, we conclude that

$$ {}_{\mathcal{B}_2}[T]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\, {}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\, {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} \tag{1} $$

This identity shows how the $\mathcal{B}_1$ basis representation of T changes if the basis is changed to $\mathcal{B}_2$. That is why we define the matrix ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ to be the $\mathcal{B}_1 - \mathcal{B}_2$ change of basis matrix.

Therefore, if $\mathcal{B}_1$ is a given basis of a vector space V, if T is a given linear transformation on V, and if $A = {}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}$ is the $\mathcal{B}_1$ basis representation of T, the set of all possible basis representation of T is

$$\begin{aligned}&\{{}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} : \mathcal{B}_2 \text{ is a basis of } V\}\\ &= \{S^{-1}AS : S \in M_n(\mathbf{F}) \text{ is invertible}\}\end{aligned}$$

This, at the same time, indicates the set of all matrices that are *similar to* the given matrix A.
Remark. Similar but not identical matrices are just different basis representations of a single linear transformation.

1.2 Constrained extrema and eigenvalues (Page 43-44, Appendix E)

Nonzero vectors x such that Ax is a scalar multiple of x play a major role in analyzing the structure of a matrix or linear transformation, but such vectors arise in the more elementary context of maximizing (or minimizing) a real symmetric quadratic form subject to a geometric constraint: For a given real symmetric $A \in M_n(\mathbf{R})$,

$$\text{maximize } x^T Ax, \text{ subject to } x \in \mathbf{R}^n,\ x^T x = 1 \tag{2}$$

Commonly, we introduce a Lagrangian multiplier to convert the constrained optimization problem into an unconstrained one, $L(x) = x^T Ax - \lambda x^T x$, whose necessary conditions for an extremum are

$$0 = \nabla L = 2(Ax - \lambda x) = 0$$

Thus, if a vector $x \in \mathbf{R}^n$ with $x^T x = 1$ (and hence $x \neq 0$) is an extremum of $x^T Ax$, it must satisfy the equation $Ax = \lambda x$.

Definition 1.4. *Let $A \in M_n$ (M_n is the set of n-by-n matrices over $\mathbf{C}$. We will inherit this notation in the rest of lectures). If a scalar λ and a nonzero vector x satisfy the equation*

$$Ax = \lambda x, \quad x \in \mathbf{C}^n, x \neq 0, \lambda \in \mathbf{C} \tag{3}$$

then λ is called an eigenvalue of A and x is called an eigenvector of A associated with λ. The pair λ, x is an eigenpair for A.

Remark. It is a key element of the definition that *an eigenvector can never be the zero vector.*

Theorem 1.5 (Weierstrass)**.** *Let S be a compact subset of a finite-dimensional real or complex vector space V with a given norm $||\cdot||$, and let $f : S \to \mathbf{R}$ be a continuous function. There exists a point $x_{min} \in S$ such that*

$$f(x_{min}) \leq f(x) \text{ for all } x \in S$$

and a point $x_{max} \in S$ such that

$$f(x) \leq f(x_{max}) \text{ for all } x \in S$$

That is, f attains its minimum and maximum values on S.

We omit the proof here. For readers who are interested in its proof, we refer them to Theorem 4.16 (Page 89) of Rudin's book[2].

With Weierstrass's theorem in our mind, we conclude that the constrained extremum problem (2) has a solution, since $f(x) = x^T Ax$ is a continuous function on the compact set $\{x \in \mathbf{R}^n : x^T x = 1\}$. This result, in turn, indicates that every real symmetric matrix has at least one real eigenvalue.

2 The eigenvalue-eigenvector equation (Page 44-48, Appendix C)

We have came across eigenvalues and eigenvectors in the course of linear algebra and proficiently known how to calculate eigenpairs of a specified matrix. Essentially, Equation (3) can be rewritten as $\lambda x - Ax = (\lambda I - A)x = 0$, a square system of homogeneous linear equations. If this system has a nontrivial solution, then λ is an eigenvalue of A and the matrix $\lambda I - A$ is singular. Conversely, if $\lambda \in \mathbf{C}$ and if $\lambda I - A$ is singular, then there exists a nonzero vector x such that $(\lambda I - A)x = 0$, and hence $Ax = \lambda x$, i.e., λ, x is an eigenvalue-eigenvector pair for A.

Therefore, we often compute eigenvalues of A by solve its characteristic equation $\det(\lambda I - A) = 0$.

Definition 2.1. *The spectrum of $A \in M_n$ is the set of all $\lambda \in \mathbf{C}$ that are eigenvalues of A; we denote this set by $\sigma(A)$.*

If x is an eigenvector of $A \in M_n$ associated with λ, then for any $c \in \mathbf{C}, c \neq 0$, we find that $A(cx) = c(Ax) = c(\lambda x) = \lambda(cx)$, that is, any nonzero scalar multiple of x is an eigenvector of A associated with λ. Thus, it is often convenient to normalize an eigenvector x to form a unit vector $\xi = x/||x||_2$, which is still an eigenvector of A associated with λ.
Remark. However, normalization does not select a unique eigenvector associated with λ: $\lambda, e^{i\theta}\xi$ is an eigenvalue-eigenvector pair for A for all $\theta \in \mathbf{R}$.

Example 2.2. Consider the matrix

$$A = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix} \in M_2$$

Then $3 \in \sigma(A)$ and $[1\ 2]^T$ is an eigenpair, while $5 \in \sigma(A)$ and $[1\ 1]^T$ is the other.

Example 2.3. Let J_n be the n-by-n matrix whose entries are all equal to 1. Consider the n-vector e whose entries are all equal to 1, and let $x_k = e - ne_k$, in which $\{e_1, ..., e_n\}$ is the standard basis for $\mathbf{C}^n$. Then $\{e, x_1, ..., x_{n-1}\}$ are linearly independent, since the matrix whose columns are $\{e, x_1, ..., x_{n-1}\}$ can be reduced to I_n via elementary row operations (or its determinant is n^{n-1}). Furthermore, $\{e, x_1, ..., x_{n-1}\}$ are eigenvectors of J_n associated with eigenvalues $\{n, 0, ..., 0\}$, respectively. This can be checked by direct computations.
Consider a matrix

$$A = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$$

Then $Ae = (4I - J3)e = e$ and $Ax_1 = (4I - J_3)x_1 = 4x_1$. Hence 1 and 4 are eigenvalues of A.

Proposition 2.4. *Let $A \in M_n$. Then $\sigma(\bar{A}) = \overline{\sigma(A)}$. If $A \in M_n(\mathbf{R})$ and $\lambda \in \sigma(A)$, then $\bar{\lambda} \in \sigma(A)$ as well.*

Proof. If $Ax = \lambda x, x \neq 0$, then $\bar{A}\bar{x} = \bar{\lambda}\bar{x}$, showing that $\bar{\lambda}$ is an eigenvalue of $\bar{A}$ and thus $\overline{\sigma(A)} \subset \sigma(\bar{A})$.
Conversely, if $\lambda' \in \sigma(\bar{A})$, namely, $\bar{A}x = \lambda' x$, then $A\bar{x} = \bar{\lambda}'\bar{x}$, i.e., $\bar{\lambda}' \in \sigma(A)$.

Whenever $A \in M_n(\mathbf{R})$ and $\lambda \in \sigma(A)$, $\overline{\sigma(A)} = \sigma(\bar{A}) = \sigma(A)$, yielding that $\bar{\lambda} \in \sigma(A)$. □

For a given $A \in M_n$, we are not sure at this point whether $\sigma(A)$ is empty, or, if it is not empty, whether it contains finitely or infinitely many complex numbers. Fortunately, we can resort to the *fundamental theorem of algebra*[1]to tackle this problem.

Theorem 2.5 (fundamental theorem of algebra)**.** *Any polynomial p with complex coefficients and of degree at least 1 has at least one zero in $\mathbf{C}$.*

Using synthetic divisions and inductive argument, we can derive that

Corollary 2.6. *A polynomial of degree $n \geq 1$ with complex coefficients has, counting multiplicities, exactly n zeroes among the complex numbers.*

Evaluation of a polynomial of degree k

$$p(t) = a_k t^k + a_{k-1} t^{k-1} + \cdots + a_1 t + a_0,\ a_k \neq 0 \tag{4}$$

with real or complex coefficients at a matrix $A \in M_n$ is well-defined since we may form linear combinations of integral powers of a given square matrix. We define

$$p(A) = a_k A^k + a_{k-1} A^{k-1} + \cdots + a_1 A + a_0 I \tag{5}$$

in which we observe the universal convention that $A^0 = I$. A polynomial (4) of degree k is said to be *monic* if $a_k = 1$; since $a_k \neq 0$, $a_k^{-1} p(t)$ is always monic.

Remark. A monic polynomial cannot be the zero polynomial.

There is an alternative way to represent $p(A)$ that has very important consequences. The *fundamental theorem of algebra* ensures that any monic polynomial of degree $k \geq 1$ can be represented as a product of exactly k complex of real linear factors:

$$p(t) = (t - \alpha_1) \cdots (t - \alpha_k) \tag{6}$$

[1]Mathematicians have announced miscellaneous approaches to prove this well-known theorem. Some of them addressed it with elementary algebra, while others utilized topological tools to conjure up proofs. The method via results in Complex Analysis (Liouville's theorem) is one of the most elegant proofs, which can be found in Chapter 2 of Stein's book[3].

The representation of $p(t)$ is unique up to permutation of its factors. It tells us that $p(\alpha_i) = 0$ for each $j = 1, ..., k$, so that each α_i is a *root* of the equation $p(t) = 0$ or a *zero* of $p(t)$. Conversely, if β is a complex number such that $p(\beta) = 0$, then $\beta \in \{\alpha_1, ..., \alpha_k\}$, so a polynomial of degree $k \geq 1$ has at most k distinct zeroes. The number of times a factor $(t - \alpha_j)$ is repeated is the *multiplicity* of α_j as a zero of $p(t)$. The factorization (6) gives a factorization of $p(A)$:

$$p(A) = (A - \alpha_1 I) \cdots (A - \alpha_k I) \tag{7}$$

The eigenvalues of $p(A)$ are linked to the eigenvalues of A in a simple way.

Theorem 2.7. *Let $p(t)$ be a given polynomial of degree k. If λ, x is an eigenvalue-eigenvector pair of $A \in M_n$, then $p(\lambda), x$ is an eigenvalue-eigenvector pair of $p(A)$. Conversely, if $k \geq 1$ and if μ is an eigenvalue of $p(A)$, then there is some eigenvalue λ of A such that $\mu = p(\lambda)$.*

Proof. If $Ax = \lambda x$, then we have

$$p(A)x = a_k A^k x + a_{k-1} A^{k-1} x + \cdots + a_1 Ax + a_0 x, \quad a_k \neq 0$$

and $A^j x = A^{j-1} \lambda x = \cdots = \lambda^j x$ by repeated application of the eigenvalue-eigenvector equation. Thus,

$$p(A) = a_k \lambda^k x + \cdots + a_0 x = (a_k \lambda^k + \cdots + a_0)x = p(\lambda)x.$$

Conversely, if μ is an eigenvalue of $p(A)$, then $p(A) - \mu I$ is singular. Since $p(t)$ has degree $k \geq 1$, the polynomial $q(t) = p(t) - \mu$ has degree $k \geq 1$, and we can factor it as $q(t) = (t - \beta_1) \cdots (t - \beta_k)$ for some complex or real $\beta_1, ..., \beta_k$. Since $p(A) - \mu I = q(A) = (A - \beta_1 I) \cdots (A - \beta_k I)$ is singular, some factor $A - \beta_j I$ is singular, which means that β_j is an eigenvalue of A. But $0 = q(\beta_j) = p(\beta_j) - \mu$, so $\mu = p(\beta_j)$, as claimed. $\square$

Remark. In the "converse" part of the preceding theorem, we can say nothing about the relation between the eigenvector of $p(A)$ and the eigenvector of A. Consider a singular matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. e_1 is an eigenvector of both A and A^2, associated with the eigenvalue $\lambda = 0$. However, e_2 is an eigenvector of A^2 but *not* of A, since $Ae_2 = e_1$ while $A^2 e_2 = Ae_1 = 0$.

Example 2.8. Suppose that $A \in M_n$. If $\sigma(A) = \{-1, 1\}$, then $\sigma(A^2) = \{1\}$. This is because by Theorem 2.7, if $\lambda \in \sigma(A)$, then $\lambda^2 \in \sigma(A^2)$, so $1 \in \sigma(A^2)$.

Furthermore, the solutions of $p(t) = t^2 - 1 = 0$ does not go beyond the set $\{-1, 1\}$. Therefore, 1 is the only point in $\sigma(A^2)$.

Before we prove the main result of this section, that is, *every complex matrix has a nonempty spectrum*, we make some nontrivial observations.

Observation 2.9. *A matrix $A \in M_n$ is singular if and only if $0 \in \sigma(A)$.*

Proof. The matrix A is singular if and only if $Ax = 0$ for some $x \neq 0$. This happens if and only if $Ax = 0x$ for some $x \neq 0$, that is, if and only if $\lambda = 0$ is an eigenvalue of A. □

Observation 2.10. *Let $A \in M_n$ and $\lambda, \mu \in \mathbf{C}$ be given. Then $\lambda \in \sigma(A)$ if and only if $\lambda + \mu \in \sigma(A + \mu I)$.*

Proof. If $\lambda \in \sigma(A)$, there exists a nonzero vector x such that $Ax = \lambda x$ and hence $(A + \mu I)x = Ax + \mu x = (\lambda + \mu)x$. Thus, $\lambda + \mu \in \sigma(A + \mu I)$. Conversely, if $\lambda + \mu \in \sigma(A + \mu I)$, there exists a nonzero vector y such that $Ay + \mu y = (A + \mu I)y = (\lambda + \mu)y = \lambda y + \mu y$. Thus, $Ay = \lambda y$ and $\lambda \in \sigma(A)$. □

With these observations, we are about to answer a question aroused earlier, that is, for each $A \in M_n$, there exist some scalar $\lambda \in \mathbf{C}$ and some nonzero $x \in \mathbf{C}^n$ such that $Ax = \lambda x$. We formulate this result in a sightly different way.

Theorem 2.11. *Let $A \in M_n$ be given. Then A has an eigenvalue. In fact, for each given nonzero $y \in C^n$, there is a polynomial $g(t)$ of degree at most $n - 1$ such that $g(A)y$ is an eigenvector of A.*

Proof. Let m be the *least* integer k such that the vectors $y, Ay, A^2y, ..., A^ky$ are linearly dependent. Then $m \geq 1$ since $y \neq 0$, and $m \leq n$ since any $n+1$ vectors in $\mathbf{C}^n$ are linearly dependent. Let $a_0, a_1, ..., a_m$ be scalars, not all zero, such that

$$a_m A^m y + a_{m-1} A^{m-1} y + \cdots + a_1 Ay + a_0 y = 0. \tag{8}$$

We claim that $a_m \neq 0$, for, otherwise, if $a_m = 0$, then (8) implies that the vectors $y, Ay, ..., A^{m-1}y$ are linearly dependent, contradicting the minimality of m.
We may consider the polynomial $p(t) = t^m + (a_{m-1}/a_m)t^{m-1} + \cdots + (a_1/a_m)t +$

(a_0/a_m). The identity (8) ensures that $p(A)y = 0$, so $0, y$ is an eigenvalue-eigenvector pair for $p(A)$. Theorem 2.7 assures that one of the m zeroes of $p(t)$ is an eigenvalue of A.
Suppose that λ is a zero if $p(t)$ that is an eigenvalue of A and factor $p(t) = (t-\lambda)g(t)$, in which $g(t)$ is a polynomial of degree $m-1$. If $g(A)y = 0$, the minimality of m would be contradicted again (in Identity (8)), so $g(A)y \neq 0$. But $0 = p(A)y = (A-\lambda I)(g(A)y)$, so the nonzero vector $g(A)y$ is an eigenvector of A associated with the eigenvalue λ. □

Remark. The preceding argument shows that for a given $A \in M_n$ we can find a polynomial of degree at most n such that *at least one* of its zeroes is an eigenvalue of A. In the next section, we will introduce the characteristic polynomial $p_A(t)$ of degree exactly n, whose zeroes are eigenvalues of A, and vice versa, that is, $p_A(\lambda) = 0$ if and only if $\lambda \in \sigma(A)$.

3 The characteristic polynomial and algebraic multiplicity (Page 49-55, Page 25-26, Page 28-29)

Recall that we can rewrite the eigenvalue-eigenvector equation (3) into a square system of homogeneous linear equations,

$$(\lambda I - A)x = 0, \quad x \neq 0. \tag{9}$$

Thus, $\lambda \in \sigma(A)$ if and only if $\lambda I - A$ is singular, that is, if and only if

$$\det(\lambda I - A) = 0. \tag{10}$$

This identity, in terms of λ, is in effect a polynomial of degree n when $A \in M_n$.

Definition 3.1. *Thought of as a formal polynomial in t, the* characteristic polynomial *of $A \in M_n$ is*

$$p_A(t) = \det(tI - A).$$

We refer to the equation $p_A(t) = 0$ as the characteristic equation *of A.*

Observation 3.2. *The characteristic polynomial of each $A = [a_{ij}] \in M_n$ has degree n and $p_A(t) = t^n - (\mathrm{tr}A)t^{n-1} + \cdots + (-1)^n \det(A)$. Moreover, $p_A(\lambda) = 0$ if and only if $\lambda \in \sigma(A)$, so $\sigma(A)$ contains at most n complex numbers.*

Proof. We define a *permutation* of $\{1, ..., n\}$ to be a one-to-one function $\sigma : \{1, ..., n\} \to \{1, ..., n\}$ and the determinant of A can be presented as

$$\det A = \sum_{\sigma} (sgn\, \sigma \prod_{i=1}^{n} a_{i\sigma(i)}). \tag{11}$$

Each summand in this presentation of $\det(tI - A)$ is a product of exactly n entries of $tI - A$, each from a different row and column, so each summand is a polynomial in t of degree at most n. The degree of a summand can be n if and only if every factor in the product involves t, which happens only for the summand

$$(t - a_{11}) \cdots (t - a_{nn}) = t^n - (a_{11} + \cdots + a_{nn})t^{n-1} + \cdots \tag{12}$$

which is just the product of the diagonal entries.
By the presentation (11), any other summand must contain a factor $-a_{ij}$ with $i \neq j$, so the diagonal entries $(t - a_{ii})$ (in the same row as a_{ij}) and $(t - a_{jj})$ (in the same column as a_{ij}) cannot also be factors; this summand therefore cannot have degree larger than $n - 2$.
Thus, the coefficients of t^n and t^{n-1} in the polynomial $p_A(t)$ arise only from the summand (12).
The constant term in $p_A(t)$ is just $p_A(0) = \det(0I - A) = \det(-A) = (-1)^n \det A$. The remaining assertion is the equivalence of (9) and (10), together with the fact that a polynomial of degree $n \geq 1$ has at most n distinct zeroes. □

Remark. The characteristic polynomial could alternatively be defined as $\det(A - tI) = (-1)^n \det(tI - A) = (-1)^n p_A(t)$. Conventionally, we would choose the coefficient of t^n in the characteristic polynomial to be $+1$.

In fact, the coefficients of $t^i, i = 0, ..., n-1$ in $p_A(t)$ can also be expressed in terms of eigenvalues of A. Consider a matrix $A \in M_n$ with $n > 1$ and factor its characteristic polynomial as $p_A(t) = (t - \alpha_1) \cdots (t - \alpha_n)$. We know that each zero α_i of $p_A(t)$ (regardless of its multiplicity) is an eigenvalue of A. A computation reveals that

$$p_A(t) = t^n - (\alpha_1 + \cdots + \alpha_n)t^{n-1} + \cdots + (-1)^n \alpha_1 \cdots \alpha_n \tag{13}$$

A comparison of Observation 3.2 and Identity (13) tells us that the sum of the zeroes of $p_A(t)$ is $tr\, A$, and the product of the zeroes of $p_A(t)$ is $\det A$. If each zero of $p_A(t)$ has multiplicity 1, that is, if $\alpha_i \neq \alpha_j$ whenever $i \neq j$, then

$\sigma(A) = \{\alpha_1, ..., \alpha_n\}$, so $tr\,A$ is the sum of the eigenvalues of A and $\det A$ is the product of the eigenvalues of A.

What if an eigenvalue λ of a matrix $A \in M_n$ with $n > 1$ is a multiple zero of $p_A(t)$ (equivalently, a multiple root of its characteristic equation)? For instance, the characteristic polynomial of $I \in M_n$ is

$$p_I(t) = \det(tI - I) = \det((t-1)I) = (t-1)^n \det I = (t-1)^n$$

so the eigenvalue $\lambda = 1$ has multiplicity n as a zero of $p_I(t)$.

Definition 3.3. *Let $A \in M_n$. The multiplicity of an eigenvalue λ of A is its multiplicity as a zero of the characteristic polynomial $p_A(t)$. For clarity, we sometimes refer to the multiplicity of an eigenvalue as its* algebraic multiplicity.

Henceforth, the eigenvalues of $A \in M_n$ will always mean the eigenvalues together with their respective (algebraic) multiplicities. Thus, the zeroes of the characteristic polynomial of A (including their multiplicities) are the same as the eigenvalues of A (including their multiplicities):

$$p_A(t) = (t - \lambda_1) \cdots (t - \lambda_n) \tag{14}$$

in which $\lambda_1, ..., \lambda_n$ are the n eigenvalues of A.

We can now say without qualification that *each matrix $A \in M_n$ has exactly n eigenvalues among the complex numbers*; the trace and determinant of A are the sum and product, respectively, of its eigenvalues.
Remark. The elements of the set $\sigma(A)$ are the *distinct eigenvalues* of A.

Since we now know that each n-by-n complex matrix has finitely many eigenvalues, we may unambiguously make the following definition.

Definition 3.4. *Let $A \in M_n$. The spectral radius of A is $\rho(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}$.*

Thanks to the finiteness of eigenvalues of $A \in M_n$, we conclude that all its eigenvalues lie in the closed bounded disk $\{z : z \in \mathbf{C} \text{ and } |z| \leq \rho(A)\}$ in the complex plane.

If $A = [a_{ij}] \in M_n(\mathbf{R})$, some or all of its eigenvalues might not be real. Suppose that A has an eigenvalue λ that is not real. We claim that $\bar{\lambda}$ is also an eigenvalue of A. This is because by the representation (11) of $\det(tI - A)$, the coefficients of $p_A(t)$ are products of a_{ij} or the sums of products of a_{ij}, which are all real. Since $p_A(\lambda) = 0$, we find that

$$0 = \bar{0} = \overline{p_A(\lambda)} = p_A(\bar{\lambda}),$$

which means that $\bar{\lambda}$ is an eigenvalue, as claimed. Furthermore, rewriting $p_A(t) = (t - \lambda)(t - \bar{\lambda})g(t)$, together with inductive arguments, we can deduce that the algebraic multiplicities of λ and $\bar{\lambda}$ are the same. In addition, if x, λ is an eigenpair for A, we immediately know that $\bar{x}, \bar{\lambda}$ is also an eigenpair.

Before conducting more in-depth investigations on the coefficients of characteristic polynomials, we digress for a moment to discuss eigenvalues of some special matrices. Here we introduce some preliminary knowledge about submatrices and block matrices.

Let $A \in M_{m,n}(\mathbf{F})$. For index sets $\alpha \subseteq \{1, ..., m\}$ and $\beta \subseteq \{1, ..., n\}$, we denote by $A[\alpha, \beta]$ the (sub)matrix of entries that lie in the rows of A indexed by α and the columns indexed by β. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \left[\{1,3\}, \{1,2,3\}\right] = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

If $\alpha = \beta$, the submatrix $A[\alpha] = A[\alpha, \alpha]$ is a *principal submatrix* of A.

The determinant of an r-by-r submatrix of A is called a *minor*; if we wish to indicate the size of the submatrix, we call its determinant a *minor of size* r. If the r-by-r submatrix is a principal submatrix, then its determinant is a *principal minor (of size r).*

Remark. By convention, the empty principal minor is 1, i.e., $\det A[\emptyset] = 1$.

If a matrix is partitioned by sequential partitions of its rows and columns, the resulting partitioned matrix is called a *block matrix*. For example, if the rows and columns of $A \in M_n(\mathbf{F})$ are partitioned by the same sequential partition $\alpha_1 = \{1, ..., k\}, \alpha_2\{k+1, ..., n\}$, the resulting block matrix is

$$A = \begin{bmatrix} A[\alpha_1, \alpha_1] & A[\alpha_1, \alpha_2] \\ A[\alpha_2, \alpha_1] & A[\alpha_2, \alpha_2] \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

in which the *blocks* are $A_{ij} = A[\alpha_i, \alpha_j]$.

Remark. Computations with block matrices are employed throughout the book; 2-by-2 block matrices are the most important and useful.

To compute the eigenvalues and determinant of $I + xy^*$, where $x, y \in \mathbf{C}^n$, we introduce an important concept in the field of both matrix theory and convex optimization, *Schur complements*. Let $A \in M_n(\mathbf{F})$ be partitioned into a 2-by-2 block matrix,

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

in which A_{11} is nonsingular. Suppose that we want to compute $\det A$ as well as the inverse of A. It would be convenient if A has a block diagonal form. Thus, we utilize *block Gaussian elimination* to transform A as

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \tag{15}$$

The special matrix

$$S = A/A_{11} = A_{22} - A_{21}A_{11}^{-1}A_{12} \tag{16}$$

is called the *Schur complement of* A_{11} *in* A. Therefore, we have the following determinantal formula for A,

$$\det A = \det A_{11} \det(A_{22} - A_{21}A_{11}^{-1}A_{12}) \tag{17}$$

Meanwhile, a useful expression for the corresponding partitioned presentation of A^{-1} is

$$\begin{aligned} A^{-1} &= \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & S^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} A_{11} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{bmatrix} \end{aligned} \tag{18}$$

When A_{22} consists of a single element, the Schur complement of A_{11} in A is a scalar and A can be rewritten as

$$A = \begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix}$$

with $a \in \mathbf{F}, x, y \in \mathbf{F}^{n-1}$, and $\tilde{F} \in M_{n-1}(\mathbf{F})$. Then (17) reduces to the identity

$$\begin{aligned} \det A = \det \begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix} &= (\det \tilde{A})\,(a - y^T\tilde{A}^{-1}x) \\ &= a \det \tilde{A} - y^T(adj\,\tilde{A})x \end{aligned} \tag{19}$$

where we use the fact that the adjugate adj $\tilde{A} = (\det \tilde{A})\tilde{A}^{-1}$. (The definition of the *adjugate* of A is adj $A = \left[(-1)^{i+j} \det A[\{j\}^c, \{i\}]\right]$, which is also called the *classical adjoint* of A.) The expression (19) is called the *Cauchy expansion* of the determinant of a bordered matrix, which is valid even if A_{11} is singular. If $a \neq 0$, we can use the Schur complement of $[a]$ in A to express

$$\det \begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix} = a \det(\tilde{A} - a^{-1} x y^T)$$

Equating the right-hand side of this identity to that of (19) and setting $a = -1$ gives *Cauchy's formula for the determinant of a rank-one perturbation*

$$\det(\tilde{A} + xy^T) = \det(\tilde{A}) + y^T (adj\, \tilde{A}) x \tag{20}$$

Now we are ready to compute the determinants of some particular matrices. One of them is the so-called Brauer's theorem.

Example 3.5. Let $x, y \in \mathbf{C}^n$. We now compute the eigenvalues and determinant of $I + xy^*$.
Using (20) and the fact that $adj(\alpha I) = \alpha^{n-1} I$, we calculate

$$\begin{aligned} p_{I+xy^*}(t) &= \det(tI - (I + xy^*)) = \det((t-1)I - xy^*) \\ &= \det((t-1)I) - y^* adj((t-1)I)x \\ &= (t-1)^n - (t-1)^{n-1} y^* x = (t-1)^{n-1}(t - (1 + y^* x)) \end{aligned}$$

Thus, the eigenvalues of $I + xy^*$ are $1 + y^*x$ and 1 (with multiplicity $n-1$), so $\det(I + xy^*) = (1 + y^*x)(1)^{n-1} = 1 + y^*x$.

Theorem 3.6 (Brauer)**.** *Let $x, y \in \mathbf{C}^n, x \neq 0$, and $A \in M_n$. Suppose that $Ax = \lambda x$ and let the eigenvalues of A be $\lambda, \lambda_2, ..., \lambda_n$. Then the eigenvalues of $A + xy^*$ are $\lambda + y^*x, \lambda_2, ..., \lambda_n$.*

Proof. First, $(t - \lambda)x = (tI - A)x$ implies that $(t - \lambda) adj(tI - A)x = adj(tI - A)(tI - A)x = \det(tI - A)x$, that is,

$$(t - \lambda) adj(tI - A)x = p_A(t)x \tag{21}$$

Apply (20) to compute

$$\begin{aligned} p_{A+xy^*}(t) &= \det(tI - (A + xy^*)) = \det((tI - A) - xy^*) \\ &= \det(tI - A) - y^* adj(tI - A)x \end{aligned}$$

Multiply both sides by $(t-\lambda)$, use (21), and obtain

$$\begin{aligned}(t-\lambda)p_{A+xy^*}(t) &= (t-\lambda)\det(tI-A) - y^*(t-\lambda)adj(tI-A)x \\ &= (t-\lambda)p_A(t) - p_A(t)y^*x\end{aligned}$$

which is the polynomial identity

$$(t-\lambda)p_{A+xy^*}(t) = (t-(\lambda+y^*x))p_A(t)$$

The zeroes of the left-hand polynomial are λ together with the n eigenvalues of $A+xy^*$, while the zeroes of the right-hand polynomial are $\lambda+y^*x, \lambda, \lambda_2, ..., \lambda_n$. It follows that the eigenvalues of $A+xy^*$ are simply $\lambda+y^*x, \lambda_2, ..., \lambda_n$. □

We now return to the main stream of this section, and express the coefficients of the characteristic polynomial in terms of the sum of *principle minors*.

Definition 3.7. *Let $A \in M_n$. The sum of its principle minors of size k (there are $\binom{n}{k}$ of them) is denoted by $E_k(A)$.*

We have already encountered principle minor sums as two coefficients of the characteristic polynomial in Observation (3.2)

$$p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_2t^2 + a_1t + a_0 \tag{22}$$

If $k=1$, then $\binom{n}{k} = n$ and $E_1(A) = a_{11} + \cdots + a_{nn} = tr\,A = -a_{n-1}$; if $k=n$, then $\binom{n}{k} = 1$ and $E_n(A) = \det A = (-1)^n a_0$. The border connection between coefficients and principal minor sums is a consequence of the fact that the coefficients are explicit functions of certain derivatives of $p_A(t)$ at $t=0$:

$$a_k = \frac{1}{k!}p_A^{(k)}(0), \ \ k=0,...,n-1 \tag{23}$$

We need to deduce the derivative of the determinant.

Proposition 3.8. *Let $A(t) = [a_1(t) \ \cdots \ a_n(t)] = [a_{ij}(t)]$ be an n-by-n complex matrix whose entries are differentiable functions of t and define $A'(t) = [a'_{ij}(t)]$. The derivative of $\det A(t)$ is $\frac{d}{dt}\det A(t) = \mathrm{tr}((\mathrm{adj}\,A(t))A'(t))$. Moreover, $\frac{d}{dt}\det(tI-A) = \mathrm{tr}(\mathrm{adj}(tI-A))$.*

Proof. The first formula follows from multilinearity of the determinant (11) and the fact that $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij} = \operatorname{tr}(A^T B)$,

$$\begin{aligned}\frac{d}{dt}\det A(t) &= \sum_{j=1}^{n}\det(A(t) \underset{j}{\leftarrow} a_j'(t)) = \sum_{i=1}^{n}\sum_{j=1}^{n}((\operatorname{adj} A(t))^T)_{ij}a_{ij}'(t) \\ &= \operatorname{tr}((\operatorname{adj} A(t))A'(t)),\end{aligned} \tag{24}$$

where the notation $(A(t) \underset{j}{\leftarrow} a_j'(t))$ denotes the matrix whose jth column is $a_j'(t)$ and whose remaining columns coincide with those of $A(t)$, i.e., $(A(t) \underset{j}{\leftarrow} a_j'(t)) = [a_1(t) \cdots a_{j-1}(t)\ a_j'(t)\ a_{j+1}(t) \cdots a_n(t)]$.

If $A \in M_n$ and $A(t) = tI - A$, then $A'(t) = I$ and

$$\frac{d}{dt}\det(tI - A) = \operatorname{tr}((\operatorname{adj} A(t))I) = (\operatorname{tr}(\operatorname{adj}(tI - A)) \tag{25}$$

as claimed. □

Proposition 3.8 tells us that $p_A'(t) = \operatorname{tr}\operatorname{adj}(tI - A)$. Observe that $\operatorname{tr}(\operatorname{adj} A)$ is the sum of the principal minors of A of size $n-1$, i.e., $\operatorname{tr}(\operatorname{adj} A) = E_{n-1}(A)$. Then

$$\begin{aligned} a_1 &= p_A'(t)|_{t=0} = \operatorname{tr}(\operatorname{adj}(tI - A))|_{t=0} = \operatorname{tr}(\operatorname{adj}(-A)) \\ &= (-1)^{n-1}\operatorname{tr}(\operatorname{adj} A) = (-1)^{n-1}E_{n-1}(A)\end{aligned}$$

Again, observe that $\operatorname{tr}(\operatorname{adj}(tI - A)) = E_{n-1}(tI - A) = \sum_{i=1}^{n} p_{A_{(i)}}(t)$ is the sum of the characteristic polynomials of the n principal submatrices of A of size $n-1$, which we denote by $A_{(1)}, ..., A_{(n)}$. Use the second result of Proposition 3.8 again to evaluate

$$p_A''(t) = \frac{d}{dt}\operatorname{tr}(\operatorname{adj}(tI - A)) = \sum_{i=1}^{n}\frac{d}{dt}p_{A(i)}(t) = \sum_{i=1}^{n}\operatorname{tr}(\operatorname{adj}(tI - A_{(i)})) \tag{26}$$

Each summand $\operatorname{tr}(\operatorname{adj}(tI - A_{(i)}))$ is the sum of the $n-1$ principal minors of size $n-2$ of a principal minor of $tI - A$, so each summand is a sum of certain principal minors of $tI - A$ of size $n-2$. Each of the $\binom{n}{n-2}$ principal minors of $tI - A$ of size $n-2$ appears twice in (26): *the principal minor with* *rows and columns k and ℓ omitted appears when $i = k$ as well as when $i = \ell$.* Thus,

$$\begin{aligned} a_2 &= \frac{1}{2}p_A''(t)|_{t=0} = \frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}(\mathrm{adj}(tI - A_{(i)}))|_{t=0} = \frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}(\mathrm{adj}(-A_{(i)})) \\ &= \frac{1}{2}(-1)^{n-2}\mathrm{tr}(\mathrm{adj}\, A_{(i)}) = \frac{1}{2}(-1)^{n-2}(2E_{n-2}(A)) \\ &= (-1)^{n-2}E_{n-2}(A). \end{aligned}$$

Repeating this argument reveals that $p_A^{(k)}(0) = k!(-1)^{n-k}E_{n-k}(A), k = 0, 1, ..., n-1$, so the coefficients of the characteristic polynomial (22) are

$$a_k = \frac{1}{k!}p_A^{(k)}(0) = (-1)^{n-k}E_{n-k}(A),\ \ k = 0, 1, ..., n-1$$

and hence

$$p_A(t) = t^n - E_1(A)t^{n-1} + \cdots + (-1)^{n-1}E_{n-1}(A)t + (-1)^n E_n(A) \tag{27}$$

In order to express the coefficients of the characteristic function $p_A(t)$ (or $E_i(A)$) in terms of eigenvalues of A, we make the following definition:

Definition 3.9. *The kth* elementary symmetric function *of n complex numbers $\lambda_1, ..., \lambda_n, k \le n$, is*

$$S_k(\lambda_1, ..., \lambda_n) = \sum_{1\le i_1<\cdots<i_k\le n} \prod_{j=1}^{k} \lambda_{i_j}$$

Notice that the sum has $\binom{n}{k}$ summands. If $A \in M_n$ and $\lambda_1, ..., \lambda_n$ are its eigenvalues, we define $S_k(A) = S_k(\lambda_1, ..., \lambda_n)$.

A calculation with (14) reveals that

$$p_A(t) = t^n - S_1(A)t^{n-1} + \cdots + (-1)^{n-1}S_{n-1}(A)t + (-1)^n S_n(A) \tag{28}$$

Comparison of (27) and (28) yields the following identities between elementary symmetric functions of eigenvalues of a matrix and sums of its principal minors.

Theorem 3.10. *Let $A \in M_n$. Then $S_k(A) = E_k(A)$ for each $k = 1, ..., n$.*

Singular matrices are annoying but ubiquitous under real-world scenarios. The next theorem ensures that a singular matrix can always be shifted slightly to become nonsingular and therefore we can compute its inverse. Also, it permits us to use continuity arguments to deduce results about singular matrices from properties of nonsingular matrices.

Theorem 3.11. *Let $A \in M_n$. There is some $\delta > 0$ such that $A + \epsilon I$ is nonsingular whenever $\epsilon \in \mathbf{C}$ and $0 < |\epsilon| < \delta$*

Proof. Observation 2.10 shows that $\lambda \in \sigma(A)$ if and only if $\lambda + \epsilon \in \sigma(A + \epsilon I)$. Thus, $0 \in \sigma(A + \epsilon I)$ if and only if $\lambda + \epsilon = 0$ for some $\lambda \in \sigma(A)$, that is, if and only if $\epsilon = -\lambda$ for some $\lambda \in \sigma(A)$.
If all the eigenvalues of A are zero, take $\delta = 1$.
If some eigenvalues of A is nonzero, let $\delta = \min\{|\lambda| : \lambda \in \sigma(A) \text{ and } \lambda \neq 0\}$.
If we choose any ϵ such that $0 < |\epsilon| < \delta$, we are assured that $-\epsilon \notin \sigma(A)$, so $0 \notin \sigma(A + \epsilon I)$ and $A + \epsilon I$ is nonsingular. $\square$

There is a useful connection between the derivatives of a polynomial $p(t)$ and the multiplicity of its zeroes: α is a zero of $p(t)$ with multiplicity $k \geq 1$ if and only if we can write $p(t)$ in the form

$$p(t) = (t - \alpha)^k q(t)$$

where $q(t)$ is a polynomial such that $q(\alpha) \neq 0$. Differentiating this identity with respect to t repeatedly, we know that α is a zero of $p(t)$ of multiplicity k if and only if $p(\alpha) = p'(\alpha) = \cdots = p^{(k-1)}(\alpha) = 0$ and $p^{(k)}(\alpha) \neq 0$.

Theorem 3.12. *Let $A \in M_n$ and suppose that $\lambda \in \sigma(A)$ has algebraic multiplicity k. Then* $\operatorname{rank}(A - \lambda I) \geq n - k$ *with equality for $k = 1$.*

Proof. Apply the previous observation to the characteristic polynomial $p_A(t)$ of a matrix $A \in M_n$ that has an eigenvalue λ with multiplicity $k \geq 1$. Let $B = A - \lambda I$. Obviously, 0 is an eigenvalue of B with multiplicity k and hence $p_B^{(k)}(0) \neq 0$.
Since $p_B^{(k)}(0) = k!(-1)^{n-k} E_{n-k}(B)$, we find that $E_{n-k}(B) \neq 0$. In particular, *some* principal minor of $B = A - \lambda I$ of size $n - k$ is nonzero, so $\operatorname{rank}(A - \lambda I) \geq n - k$.
If $k = 1$, we can say more: $A - \lambda I$ is singular, so $n > \operatorname{rank}(A - \lambda I) \geq n - 1$, which means that $\operatorname{rank}(A - \lambda I) = n - 1$ if the eigenvalue λ has algebraic multiplicity 1. $\square$

References

[1] Roger A. Horn, Charles R. Johnson (2012) *Matrix Analysis, Second Edition.* Cambridge University Press.

[2] Walter Rudin (1976) *Principles of Mathematical Analysis, Third Edition.* McGraw-Hill Companies, Inc.

[3] Elias M. Stein, Rami Shakarchi (2003) *Complex Analysis.* Princeton University Press.