Guangliang Chen

Introduction to Matrix-based Data Science: Mathematics, Computing and Data

Volume 1: Mathematical foundations, data plotting and visualization, and dimension reduction

January 18, 2022

Springer Nature

Chapter 1
Matrix Algebra and Computing

Abstract We review some basic linear algebra in this chapter. We assume that the reader has already taken a first course in linear algebra. We will also use the chance to introduce notation and mention some matrix operations frequently needed in computing, such as matrix reshaping, matrix row or column summing, matrix-vector multiplication, and the Hardmard product. We will focus on the computational and applied aspects of linear algebra rather than the abstract theory, as the former is much more useful for learning basic machine learning concepts and theory. We will also use many illustrations to demonstrate the challenging matrix concepts and operations, as we believe the ability to visualize matrices and matrix operations is tremendously important for mastering linear algbera.

1.1 Euclidean vectors

A *Euclidean vector*, or simply a *vector*, is a geometric object that has both magnitude and direction. Algebraically, it is typically represented as an ordered list of numbers in column form, e.g., $\mathbf{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. For the sake of saving space, we often write $\mathbf{a} = (1, 2, 3)^T$ instead. The ith element of a vector $\mathbf{a}$ is written as a_i, or in some cases, $\mathbf{a}(i)$.

In this book we denote vectors by **boldface** lowercase letters (such as $\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}$). In contrast, scalars are denoted in plain, lowercase letters (such as x, y, λ, μ). Sometimes, we have to deal with vectors in row form and we will denote them by plain, uppercase letters, e.g., $A = (1, 2, 3)$. (Note, however, that plain, uppercase letters may be used later in the book to denote other mathematical objects such as sets and random variables as well.)

When vectors start from the origin, they are identified with points in Euclidean spaces:

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, \ldots, x_n)^T \mid x_1, \ldots, x_n \in \mathbb{R}\}, \tag{1.1}$$

where $\mathbb{R}$ is the set of real numbers, and n is a positive integer, called the *dimension* of the Euclidean space $\mathbb{R}^n$. To indicate the dimension of a vector, we use notation like $\mathbf{a} \in \mathbb{R}^3$, which reads $\mathbf{a}$ lies in (or belongs to) $\mathbb{R}^3$.

Below are some notable constant vectors:

- **The n-dimensional zero vector**

$$\mathbf{0}_n = (0, 0, \ldots, 0)^T \in \mathbb{R}^n \ . \tag{1.2}$$

- **The n-dimensional vector of ones**

$$\mathbf{1}_n = (1, 1, \ldots, 1)^T \in \mathbb{R}^n \ . \tag{1.3}$$

- **The canonical basis vectors of $\mathbb{R}^n$**

$$\mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{i\text{th entry}}, 0, \ldots, 0)^T \in \mathbb{R}^n, \quad 1 \leq i \leq n. \tag{1.4}$$

In many cases, their dimension are not directly specified but rather should be inferred from the context. For example, in the sum expression $\mathbf{a} + \mathbf{1}$, where $\mathbf{a} = (1, 2, 3)^T \in \mathbb{R}^3$, the vector of ones must be three dimensional: $\mathbf{1} = (1, 1, 1)^T \in \mathbb{R}^3$.

The *dot product* of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i . \tag{1.5}$$

This operation satisfies the following properties:

- $\mathbf{x} \cdot \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$;
- $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$;
- $(k\mathbf{x}) \cdot \mathbf{y} = k(\mathbf{x} \cdot \mathbf{y})$ and $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$ for any $k \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$.

Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are said to be *orthogonal* if their dot product is zero: $\mathbf{x} \cdot \mathbf{y} = 0$. Geometrically, two orthogonal vectors always have an angle of 90 degrees.

A *norm* on $\mathbb{R}^n$ is a function

$$\| \cdot \| \ : \ \mathbb{R}^n \longrightarrow \mathbb{R} \tag{1.6}$$

that satisfies the following three conditions:

- $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
- $\|k\mathbf{x}\| = |k| \|\mathbf{x}\|$ for any scalar $k \in \mathbb{R}$ and vector $\mathbf{x} \in \mathbb{R}^n$.
- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Note that $\|\mathbf{x}\|$ can be thought of as the length or magnitude of the vector $\mathbf{x}$.

For any fixed number $p \geq 1$, the ℓ_p *norm*, or simply the *p-norm*, on $\mathbb{R}^n$ is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \quad \mathbf{x} \in \mathbb{R}^n. \tag{1.7}$$

It is a rich family of norms on Euclidean spaces, including the following three:

- **2-norm (Euclidean norm)**:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\mathbf{x}\cdot\mathbf{x}}. \tag{1.8}$$

- **1-norm (Manhattan norm)**:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|. \tag{1.9}$$

- **∞-norm (maximum norm)**:

$$\|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i|. \tag{1.10}$$

See Figure 1.1 for illustrations of these three specific norms.

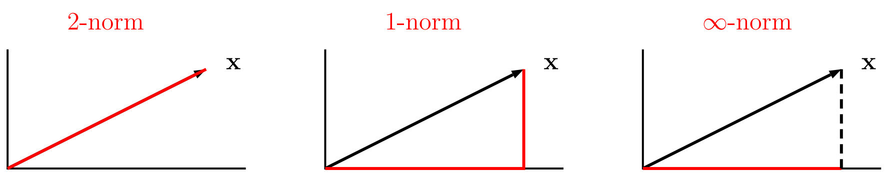

Fig. 1.1 Three particular p-norms on $\mathbb{R}^2$. In each case, the total length of the line segment(s) in red is the corresponding norm of the vector $\mathbf{x}$.

Remark 1.1 When the vector p-norm $\|\cdot\|$ has an unspecified subscript, it is understood as the ℓ_2 norm. This is often for the purpose of simplifying notation.

Given two nonzero vectors $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, let θ be the angle between them. Then

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\cdot\|\mathbf{y}\|} \tag{1.11}$$

Because $|\cos\theta|\le 1$, we have the so-called *Cauchy-Schwartz Inequality*:

$$|\mathbf{x}\cdot\mathbf{y}| \le \|\mathbf{x}\|\cdot\|\mathbf{y}\|, \quad \text{for all } \mathbf{x},\mathbf{y}\in\mathbb{R}^n, \tag{1.12}$$

where the equality holds true if and only if the two vectors $\mathbf{x},\mathbf{y}$ are parallel to each other, i.e., $\mathbf{y} = k\mathbf{x}$ for some constant $k\in\mathbb{R}$. An alternative, yet equivalent form of the above inequality is

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \le \left(\sum_{i=1}^{n} x_i^2\right)\cdot\left(\sum_{i=1}^{n} y_i^2\right). \tag{1.13}$$

Given any norm $\|\cdot\|$ on the Euclidean space $\mathbb{R}^n$, the set of all vectors in $\mathbb{R}^n$ that have a unit norm is called a *unit circle* (under the particular norm):

$$\{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\| = 1\}. \tag{1.14}$$

Figure 1.2 shows three different unit circles in $\mathbb{R}^2$ under three different p-norms.

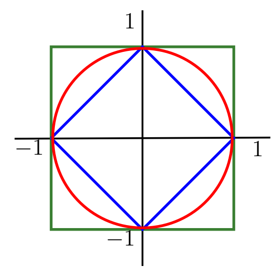

Fig. 1.2 Three different kinds of unit circles in $\mathbb{R}^2$ corresponding to different p-norms: blue (1-norm), red (2-norm) and green (∞-norm).

Any norm $\|\cdot\|$ on $\mathbb{R}^n$ can be used to measure the distance between two vectors:

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|, \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{1.15}$$

For example, the Euclidean norm defines the Euclidean distance:

$$\text{dist}_E(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{1.16}$$

Let $V \subseteq \mathbb{R}^n$ be a set of vectors in $\mathbb{R}^n$. It is called a *subspace* of $\mathbb{R}^n$ if it contains the zero vector and is closed under scalar multiplication and vector addition. That is,

- $\mathbf{0} \in V$;
- For any vector $\mathbf{x} \in V$ and scalar $k \in \mathbb{R}$, we have $k\mathbf{x} \in V$;
- For any two vectors $\mathbf{x}, \mathbf{y} \in V$, we have $\mathbf{x} + \mathbf{y} \in V$.

Given several vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, a *linear combination* of them is a vector of the form

$$\mathbf{v} = c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k, \tag{1.17}$$

for some scalars $c_1, \ldots, c_k$. The *linear span*, or in short *span*, of those given vectors is the set consisting of all of their linear combinations, i.e.,

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k \mid c_1, \ldots, c_k \in \mathbb{R}\}. \tag{1.18}$$

The span of any set of vectors is a subspace of $\mathbb{R}^n$.

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$ is said to be *linearly dependent*, if there exist scalars $c_1, \ldots, c_k \in \mathbb{R}$ that are not all zero such that

$$c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k = \mathbf{0}. \tag{1.19}$$

On the other hand, if all the coefficients $\{c_i\}_{1\le i\le k}$ have to be set to zero in order for (1.19) to hold true, then the vectors are said to be *linearly independent*.

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$ form a *basis* for a subspace $V \subseteq \mathbb{R}^n$ if the vectors span V, and meanwhile, they are linearly independent. The number of vectors in the basis is then called the *dimension* of the subspace (It can be shown that different bases of the same subspace must have the same cardinality).

A basis for a subspace $V \subseteq \mathbb{R}^n$ is further called an *orthonormal basis* if the basis vectors are orthogonal to each other and all have unit norm. An example would be the set of canonical basis vectors, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Lastly, we mention the concepts of affine and convex combinations of several vectors. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$, and $\mathbf{v} \in \mathbb{R}^n$ a linear combination of the vectors in S, i.e.,

$$\mathbf{v} = c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k \tag{1.20}$$

for some scalars $c_1, \ldots, c_k \in \mathbb{R}$. We call $\mathbf{v}$ an *affine combination* of the vectors in S if the coefficients satisfy

$$c_1 + \cdots + c_k = 1 \tag{1.21}$$

The set of all affine combination of the vectors in S is called the *affine span* or *affine hull* of S:

$$\text{aSpan}(S) = \{c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k \mid c_1, \ldots, c_k \in \mathbb{R},\ c_1 + \cdots + c_k = 1\}. \tag{1.22}$$

An affine combination of the vectors in S, i.e., $\mathbf{v} = c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k$ where $c_1 + \cdots + c_k = 1$, is further called a *convex combination* of the vectors in S if the coefficients are all nonnegative, i.e., $c_1 \ge 0, \ldots, c_k \ge 0$. The set of all convex combination of the vectors in S is called the *convex hull* of S:

$$\text{cSpan}(S) = \{c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k \mid c_1 \ge 0, \ldots, c_k \ge 0,\ c_1 + \cdots + c_k = 1\}. \tag{1.23}$$

Note that for any given set S of vectors in $\mathbb{R}^n$, we always have

$$\text{cSpan}(S) \subseteq \text{aSpan}(S) \subseteq \text{span}(S) \subseteq \mathbb{R}^n. \tag{1.24}$$

Example 1.1 For two linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, their span is a two-dimensional linear subspace of $\mathbb{R}^n$ containing them. In contrast, the affine span of them is the line that goes through the two points corresponding to $\mathbf{v}_1, \mathbf{v}_2$:

$$\begin{aligned}\text{aSpan}\{\mathbf{v}_1, \mathbf{v}_2\} &= \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \mid c_1, c_2 \in \mathbb{R},\ c_1 + c_2 = 1\} \\ &= \{\mathbf{v}_1 + c_2(\mathbf{v}_2 - \mathbf{v}_1) \mid c_2 \in \mathbb{R}\}\end{aligned}$$

and the convex hull of them is the line segment connecting the two points $\mathbf{v}_1, \mathbf{v}_2$ (it is a part of the affine span). See Figure 1.3, left for a demonstration.

For three linearly dependent vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^n$ which span a two-dimensional subspace, their affine span coincides with the linear span:

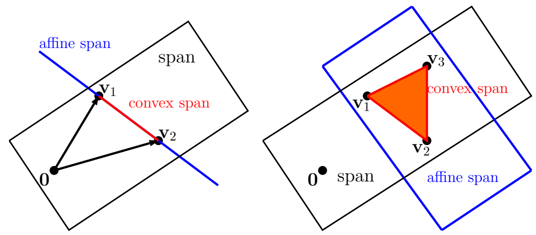

Fig. 1.3 Illustration of the linear, affine and convex spans of vectors in space. Left: two linearly independent vectors; right: three linearly dependent vectors that have a two dimensional linear span.

$$\begin{aligned}\text{aSpan}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} &= \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 \mid c_1, c_2, c_3 \in \mathbb{R},\ c_1 + c_2 + c_3 = 1\} \\ &= \{\mathbf{v}_1 + c_2(\mathbf{v}_2 - \mathbf{v}_1) + c_3(\mathbf{v}_3 - \mathbf{v}_1) \mid c_2, c_3 \in \mathbb{R}\}\end{aligned}$$

and the convex hull of them is the triangular region determined by the three points (including the interior). See Figure 1.3, right for a demonstration.

1.2 Review of matrix algebra

A *matrix* is a two-dimensional, rectangular array of real numbers arranged in rows and columns.[1] For any two positive integers m, n, the space of all real valued matrices that have m rows and n columns is denoted by $\mathbb{R}^{m\times n}$. We say that the *size* of the matrices in $\mathbb{R}^{m\times n}$ is $m \times n$.

Matrices are denoted by **boldface** UPPERCASE letters (such as $\mathbf{A}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \mathbf{\Lambda}, \mathbf{\Sigma}$). We write $\mathbf{A} \in \mathbb{R}^{m\times n}$ to indicate the size of a matrix $\mathbf{A}$ with m rows and n columns. If either m or n is equal to 1, then the matrix $\mathbf{A}$ reduces to a row/column vector. The (i, j) entry of $\mathbf{A}$ is typically denoted by a_{ij}. Sometimes, alternative notation such as $\mathbf{A}(i, j)$ and $\mathbf{A}_{ij}$ may be used instead of a_{ij}. The ith row of $\mathbf{A}$ is denoted by A_i (or sometimes $\mathbf{A}(i, :)$), while the jth column of $\mathbf{A}$ is written as $\mathbf{a}_j$ (or sometimes $\mathbf{A}(:, j)$). See Figure 1.4 for an illustration.

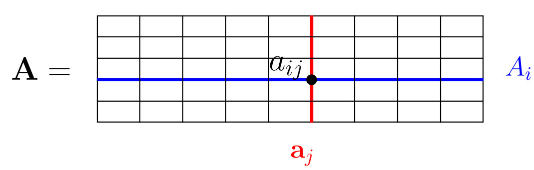

Fig. 1.4 Notation for a single row, or column, or element of a matrix **A**.

In many cases, there is a need to convert a matrix to a vector by preserving all its entries (this is called *vectorizing* a matrix), or to reshape a vector to a matrix with the same elements (this is called *matricizing* a vector). One typically perform those operations in a column fashion. For example, to vectorize a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$,

[1] Note that there could be complex-valued matrices in general, but in this book, we only consider matrices with real-value entries, as they are sufficient for our purposes.

we can take the columns of the matrix and stack them vertically following the same order as they appear in the matrix:

$$\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{m\times n} \qquad \longrightarrow \qquad \operatorname{vec}(\mathbf{A}) = \begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\\ \vdots\\ \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{mn}. \tag{1.25}$$

Similarly, to matricize a vector $\mathbf{a} = (a_i) \in \mathbb{R}^{\ell}$ to have m rows, where m must be a divisor of ℓ, we take the entries of $\mathbf{a}$ in the original order and complete one column of the matrix at a time:

$$\mathbf{a} = (a_i) \in \mathbb{R}^{\ell} \qquad \longrightarrow \qquad \operatorname{mat}(\mathbf{a}) = \begin{pmatrix} a_1 & a_{m+1} & \cdots & a_{\ell-m+1}\\ a_2 & a_{m+2} & \cdots & a_{\ell-m+2}\\ \vdots & \vdots & \ddots & \vdots\\ a_m & a_{2m} & \cdots & a_{\ell}\end{pmatrix}. \tag{1.26}$$

Below are some notable constant matrices:

- *The $n \times n$ zero matrix*: $\mathbf{O}_n$, where $\mathbf{O}_n(i, j) = 0$ for all i, j.
- *The $n \times n$ identity matrix*:

$$\mathbf{I}_n = (\delta_{ij}), \qquad \text{where} \quad \delta_{ij} = \begin{cases} 1, & i = j\\ 0, & i \neq j \end{cases} \tag{1.27}$$

- *The $n \times n$ matrix of ones*: $\mathbf{J}_n$, where $\mathbf{J}_n(i, j) = 1$ for all i, j.

That is,

$$\mathbf{O}_n = \begin{pmatrix} 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix}, \qquad \mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}, \qquad \mathbf{J}_n = \begin{pmatrix} 1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 1\end{pmatrix}. \tag{1.28}$$

Similarly, we will not directly specify the size of any such matrix when it is clear based on the context.

A matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is said to be *positive* if all its entries are positive, i.e., $a_{ij} > 0$ for all i, j. Similarly, we say that $\mathbf{A}$ is *nonnegative* if $a_{ij} \geq 0$ for all i, j. For example, $\mathbf{I}_n$ is a nonnegative matrix while $\mathbf{J}_n$ is a positive matrix.

If a matrix has mostly zero entries and relatively few nonzero entries, then we say that the matrix is *sparse* and often leave the zero entries blank when writing it out. For example,

$$\mathbf{I} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}, \quad \mathbf{E} = \begin{pmatrix} 1 & & 3 \\ & 1 & \\ & & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3}, \quad \mathbf{P} = \begin{pmatrix} & 1 & \\ 1 & & \\ & & \\ & & \end{pmatrix} \in \mathbb{R}^{4\times 3}. \tag{1.29}$$

Sparse matrices are easier and more efficient to handle in computing tasks, as we shall see later in the book.

In some cases we need to characterize a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ based on its shape. We say that $\mathbf{A}$ is a

- *square* matrix, if $m = n$;
- *long* matrix, if $m < n$;
- *tall* matrix, if $m > n$.

The last two families of matrices can also be collectively referred to as *rectangular* matrices. See Figure 1.5 for an illustration.

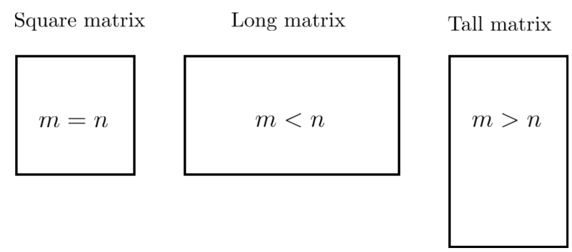

Fig. 1.5 Characterization of a matrix based on its shape.

The notion of a *diagonal* matrix usually refers to a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ whose off diagonal entries are all zero ($a_{ij} = 0$ for all $i \neq j$):

$$\mathbf{A} = \begin{pmatrix} a_{11} & & \\ & \ddots & \\ & & a_{nn} \end{pmatrix}. \tag{1.30}$$

Sometimes, a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is also said to be diagonal if all of its nonzeros are in the locations $\{(i,i), 1 \leq i \leq \min(m,n)\}$, e.g.,

$$\mathbf{B} = \begin{pmatrix} 1 & & & \\ & 2 & & \\ & & 3 & \\ & & & \\ & & & \end{pmatrix} \in \mathbb{R}^{5\times 4}, \quad \mathbf{C} = \begin{pmatrix} 0 & & & & \\ & 2 & & & \\ & & 4 & & \end{pmatrix} \mathbb{R}^{3\times 5} \tag{1.31}$$

A (square) diagonal matrix is uniquely determined by the vector that contains all the diagonal entries, and there is a simple way to denote it. For example,

$$\mathbf{A} = \begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix} \quad \longrightarrow \quad \mathbf{A} = \mathrm{diag}(\mathbf{a}), \quad \mathbf{a} = (1,2,3)^T \ .$$

Alternatively, we may simply write

$$\mathbf{A} = \mathrm{diag}(1,2,3).$$

1.2.1 Matrix multiplication

Given two matrices with compatible sizes, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$, their matrix product is another matrix $\mathbf{C} \in \mathbb{R}^{m\times p}$ with entries

$$\mathbf{C} = (c_{ij}), \quad c_{ij} = A_i \mathbf{b}_j = \sum_{k=1}^{n} a_{ik} b_{kj} \ . \tag{1.32}$$

See Figure 1.6 for a demonstration.

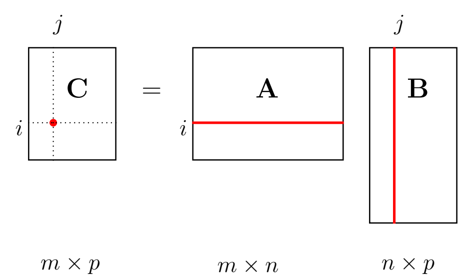

Fig. 1.6 Matrix multiplication (in an entrywise fashion).

In the special case when $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a square matrix, one can multiply $\mathbf{A}$ and itself to get the matrix square

$$\mathbf{A}^2 = \mathbf{A} \cdot \mathbf{A}. \tag{1.33}$$

More generally, for any positive integer k, the kth power of the square matrix $\mathbf{A}$ is defined as

$$\mathbf{A}^k = \underbrace{\mathbf{A} \cdot \cdots \cdot \mathbf{A}}_{k \text{ copies of } \mathbf{A}} \tag{1.34}$$

Besides the *entry-wise* multiplication fashion scheme in (1.32), there are at least three other ways to multiply $\mathbf{A}$ and $\mathbf{B}$ together. First, one may carry out the matrix multiplication in a *rowwise* fashion, that is to take the rows of $\mathbf{A}$ to multiply the matrix $\mathbf{B}$ separately to obtain full rows of $\mathbf{C}$, one at a time (see Figure 1.7):

$$C_i = A_i \cdot \mathbf{B} \quad \text{for each } i = 1, \ldots, m \ . \tag{1.35}$$

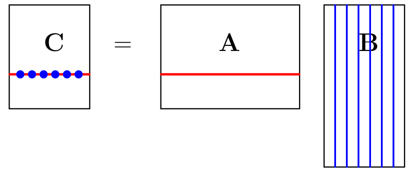

Fig. 1.7 Matrix rowwise multiplication.

Consequently,

$$\mathbf{C} = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} \cdot \mathbf{B} = \begin{bmatrix} A_1 \cdot \mathbf{B} \\ \vdots \\ A_m \cdot \mathbf{B} \end{bmatrix} . \tag{1.36}$$

Secondly, one may carry out the matrix multiplication in a *columnwise* fashion, that is to take the matrix **A** to multiply columns of **B** separately to obtain full columns of **C**, one at a time (see Figure 1.8):

$$\mathbf{c}_j = \mathbf{A} \cdot \mathbf{b}_j, \quad \text{for each } j = 1, \ldots, p . \tag{1.37}$$

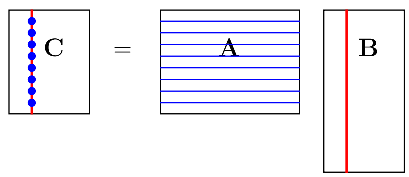

Fig. 1.8 Columnwise matrix multiplication

As a result,

$$\mathbf{C} = \mathbf{A} \cdot \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} \mathbf{A} \cdot \mathbf{b}_1 & \cdots & \mathbf{A} \cdot \mathbf{b}_p \end{bmatrix} . \tag{1.38}$$

Lastly, one may take each column of **A** to multiply the corresponding row of **B**, yielding a set of rank-1 matrices, whose sum is exactly the product matrix **C**:

$$\mathbf{C} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} = \sum_{k=1}^{n} \mathbf{a}_k \cdot B_k . \tag{1.39}$$

See Figure 1.9 for an illustration.

We use the following example to demonstrate the different multiplication schemes.

Example 1.2 Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \qquad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

We compute their matrix product, $\mathbf{C} = \mathbf{AB}$, in four different ways:

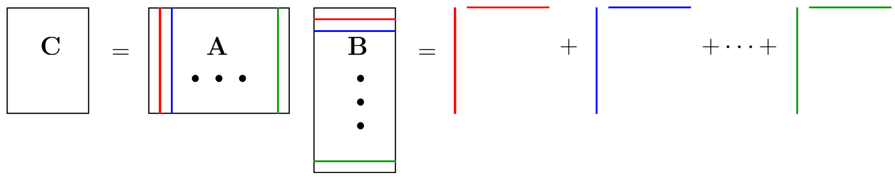

Fig. 1.9 Matrix product through column-row multiplications.

- **Entrywise multiplication**:

$$c_{11} = A_1 \cdot \mathbf{b}_1 = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 6,$$

$$c_{12} = A_1 \cdot \mathbf{b}_2 = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 2,$$

$$c_{21} = A_2 \cdot \mathbf{b}_1 = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 16,$$

$$c_{22} = A_2 \cdot \mathbf{b}_2 = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 2.$$

Putting everything together, we have

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ 15 & 2 \end{pmatrix}.$$

- **Columnwise multiplication**:

$$\mathbf{c}_1 = \mathbf{A} \cdot \mathbf{b}_1 = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 15 \end{pmatrix},$$

$$\mathbf{c}_2 = \mathbf{A} \cdot \mathbf{b}_2 = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Therefore,

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix} = \begin{pmatrix} 6 & 2 \\ 15 & 2 \end{pmatrix}$$

- **Rowwise multiplication**:

$$C_1 = A_1 \cdot \mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 2 \end{pmatrix},$$

$$C_2 = A_2 \cdot \mathbf{B} = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 15 & 2 \end{pmatrix}.$$

Therefore,

$$\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{pmatrix} 6 & 2 \\ 15 & 2 \end{pmatrix}.$$

- **Column-row multiplication**:

$$\begin{aligned} \mathbf{C} &= \mathbf{a}_1 \cdot B_1 + \mathbf{a}_2 \cdot B_2 + \mathbf{a}_3 \cdot B_3 \\ &= \begin{pmatrix} 1 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} + \begin{pmatrix} 2 \\ 5 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix} 3 \\ 6 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -1 \\ 4 & -4 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 5 & 0 \end{pmatrix} + \begin{pmatrix} 3 & 3 \\ 6 & 6 \end{pmatrix} \\ &= \begin{pmatrix} 6 & 2 \\ 15 & 2 \end{pmatrix}. \end{aligned}$$

When one of the matrices $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}$ is actually a vector, we can obtain their product $\mathbf{AB}$ as follows:

- If $A = (a_1, \ldots, a_n)$ is a row vector (i.e., $m = 1$), then

$$A\mathbf{B} = \begin{pmatrix} a_1 & \ldots & a_n \end{pmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} = \sum_{i=1}^{n} a_i B_i \ . \tag{1.40}$$

  That is, $A\mathbf{B}$ is a linear combination of the rows of $\mathbf{B}$; see Figure 1.10.
- If $\mathbf{b} = (b_1, \ldots, b_n)^T$ is a column vector ($i.e., p = 1$), then

$$\mathbf{Ab} = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{j=1}^{n} b_j \mathbf{a}_j \ . \tag{1.41}$$

  That is, $\mathbf{Ab}$ is linear combination of columns of $\mathbf{A}$; see Figure 1.11.

This new interpretation implies that for a general matrix product $\mathbf{C} = \mathbf{AB}$,

- Each row of $\mathbf{C}$, $C_i = A_i\mathbf{B}$, is a linear combination of the rows of $\mathbf{B}$, and
- Each column of $\mathbf{C}$, $\mathbf{c}_j = \mathbf{Ab}_j$, is a linear combination of the columns of $\mathbf{A}$.

Example 1.3

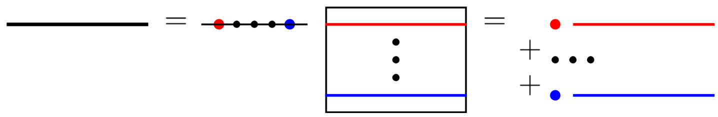

Fig. 1.10 Demonstration of multiplication of a row vector and a matrix. In the diagram, colored dots represent entries of the row vector while colored lines represented rows of the matrix. The end product is also a vector.

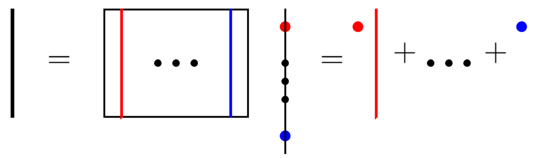

Fig. 1.11 Demonstration of multiplication of a matrix and a column vector.

$$\begin{pmatrix}-1 & 0 & 1\end{pmatrix}\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix} = -1\cdot\begin{pmatrix}1&2&3\end{pmatrix} + 0\cdot\begin{pmatrix}4&5&6\end{pmatrix} + 1\cdot\begin{pmatrix}7&8&9\end{pmatrix} = \begin{pmatrix}6&6&6\end{pmatrix}\ .$$

$$\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}\begin{pmatrix}-1\\0\\1\end{pmatrix} = -1\cdot\begin{pmatrix}1\\4\\7\end{pmatrix} + 0\cdot\begin{pmatrix}2\\5\\8\end{pmatrix} + 1\cdot\begin{pmatrix}3\\6\\9\end{pmatrix} = \begin{pmatrix}2\\2\\2\end{pmatrix}\ .$$

Example 1.4 Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be an arbitrary matrix. To extract a single column of $\mathbf{A}$, we can multiply $\mathbf{A}$ by a canonical basis vector from the right hand side:

$$\mathbf{a}_j = \mathbf{A}\mathbf{e}_j = \begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_j & \dots & \mathbf{a}_n\end{bmatrix}\begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix}, \quad \text{for each } j = 1,\dots,n \tag{1.42}$$

Similarly, to extract a single row of $\mathbf{A}$, we can multiply $\mathbf{A}$ by a canonical basis vector (in row form) from the left hand side:

$$A_i = \mathbf{e}_i^T\mathbf{A} = \begin{pmatrix}0 & \dots & 1 & \dots & 0\end{pmatrix}\begin{bmatrix}A_1\\ \vdots\\ A_i\\ \vdots\\ A_m\end{bmatrix}, \quad \text{for each } i = 1,\dots,m \tag{1.43}$$

Lastly, we mention a few identities involving the constant vector of ones. First, for the vector $\mathbf{1} \in \mathbb{R}^n$,

$$\mathbf{1}^T\mathbf{1} = \begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} = n\,, \tag{1.44}$$

$$\mathbf{1}\mathbf{1}^T = \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}\begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix} = \begin{pmatrix}1 & 1 & \dots & 1\\1 & 1 & \dots & 1\\\vdots & \vdots & \ddots & \vdots\\1 & 1 & \dots & 1\end{pmatrix} = \mathbf{J}\,. \tag{1.45}$$

Second, for any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$,

$$\mathbf{A}\mathbf{1} = \begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n\end{bmatrix}\begin{pmatrix}1\\\vdots\\1\end{pmatrix} = \sum_{j=1}^{n}\mathbf{a}_j\,, \tag{1.46}$$

$$\mathbf{1}^T\mathbf{A} = \begin{pmatrix}1 & \dots & 1\end{pmatrix}\begin{bmatrix}A_1\\\vdots\\A_m\end{bmatrix} = \sum_{i=1}^{m}A_i\,, \tag{1.47}$$

$$\mathbf{1}^T\mathbf{A}\mathbf{1} = \sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\,. \tag{1.48}$$

Important

The three expressions $\mathbf{A1}, \mathbf{1}^T\mathbf{A}, \mathbf{1}^T\mathbf{A1}$ respectively represent the row sums, column sums and overall sum of the matrix $\mathbf{A}$, and are very useful and convenient notation in a lot of practical applications.

Example 1.5 Verify the last three identities using the matrix

$$\mathbf{A} = \begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\end{pmatrix}$$

Solution 1.1 By direct calculation,

$$\mathbf{A1} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 15 \end{pmatrix}, \qquad \text{(vector of row sums)}$$

$$\mathbf{1}^T\mathbf{A} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 5 & 7 & 9 \end{pmatrix}, \qquad \text{(row vector of column sums)}$$

$$\mathbf{1}^T\mathbf{A1} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 21. \qquad \text{(overall sum of all entries)}$$

1.2.1.1 Computational complexity

In many applications it is important to estimate the amount of calculations needed before carrying out the calculations. Here we assume that addition, subtraction, multiplication and division among real numbers take (roughly) the same amount of time, and thus define the computational complexity of a linear algebra operation as the total number of arithmetic operations. For convenience, we focus on determining the limiting magnitude of the total count, and will disregard any multiplicative factor (as well as lower order terms).

Using such a definition, we have the following facts.

Theorem 1.1 *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *and* $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$. *Assume that* m, n, p *are all large. Then*

1. *Summing up the entries of* $\mathbf{x}$ *requires* $O(n)$ *operations.*
2. *Computing the dot product* $\mathbf{x}\cdot\mathbf{y}$ *also requires* $O(n)$ *operations. This implies that computing the norm of* $\mathbf{x}$, $\|\mathbf{x}\|^2 = \mathbf{x}\cdot\mathbf{x}$, *also has* $O(n)$ *complexity.*
3. *Multiplying a matrix and a vector* $\mathbf{Ax}$ *requires* $O(mn)$ *operations.*
4. *Multiplying two matrices* $\mathbf{AB}$ *requires* $O(mnp)$ *operations. In particular, computing* $\mathbf{C}^2$ *for a square matrix* $\mathbf{C} \in \mathbb{R}^{n\times n}$ *takes* $O(n^3)$ *operations.*

Proof We have:

1. Adding up the n entries of $\mathbf{x}$ takes exactly $n-1$ additions.
2. Computing the dot product $\mathbf{x}^T\mathbf{y}$ requires n multiplications between corresponding entries of the two vecotrs and $n-1$ additions for the products. In total, this is $2n-1$ operations.
3. Multiplying a matrix and a vector $\mathbf{Ax}$ requires m dot product operations between the rows of $\mathbf{A}$ and the vector $\mathbf{x}$, and thus takes $m(2n-1)$ arithmetic operations.
4. Multiplying $\mathbf{AB}$ requires multiplying every row of $\mathbf{A}$ and every column of $\mathbf{B}$, which is mp dot product operations. Since each such dot product takes $2n-1$ operations, overall it takes $mp(2n-1)$ operations to multiply $\mathbf{A}$ and $\mathbf{B}$. □

In sum, the dot product operation between two vectors have a linear complexity, the matrix-vector product has a quadratic complexity, and the matrix-matrix product has a cubic complexity (which is the most expensive of the three and should be avoided, whenever possible, in the setting of large matrices).

1.2.1.2 An application of matrix associative law in computing

Matrix multiplications follow the associative and distributive laws (but there is no commutative law).

- **Associative law**: For any three matrices $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$ and $\mathbf{C} \in \mathbb{R}^{p\times q}$,

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \ . \tag{1.49}$$

- **Distributive law**: For any three matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ and $\mathbf{C} \in \mathbb{R}^{n\times p}$,

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \ . \tag{1.50}$$

Given two matrices $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$, and a vector $\mathbf{x} \in \mathbb{R}^{p}$, the associative law implies that

$$(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx}) \tag{1.51}$$

This equation has important applications in computing.

Important

Although mathematically the two sides of (1.51) are equivalent, the right-hand side (which is two matrix-vector multiplications in order) is computationally much more efficient:

- $(\mathbf{AB})\mathbf{x}$: This consists of a matrix-matrix multiplication and a matrix-vector operation in order. The total amount of arithmetic operations is $mnp + mp = O(mnp)$;
- $\mathbf{A}(\mathbf{Bx})$: This consists of two matrix-vector multiplications and requires $np + mn = O(n(m + p))$ operations in total.

We demonstrate in Fig. 1.12 their difference using a simulation in MATLAB (scripts can be found in Section).

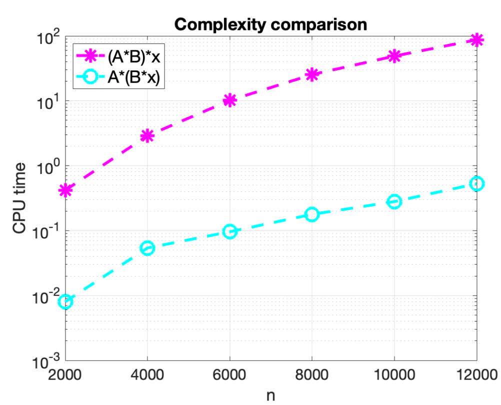

Fig. 1.12 We generate random matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ and a random vector $\mathbf{x} \in \mathbb{R}^{n}$ by uniformly sampling entries from the interval $(0, 1)$, for each value of $n = 2000, 4000, \ldots, 12000$, to compare the CPU times needed by the operations $(\mathbf{AB})\mathbf{x}$ and $\mathbf{A}(\mathbf{Bx})$. The plot shows that $\mathbf{A}(\mathbf{Bx})$ is much faster than $(\mathbf{AB})\mathbf{x}$ at all values of n. For example, when $n = 12,000$, the ratio of the two is about 165.1.

1.2.2 The Hadamard product

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ be two matrices of the same size. The *Hadamard product* of $\mathbf{A}$ and $\mathbf{B}$, also called the *entrywise product*, is another matrix $\mathbf{C}$ of the same size, with entries

$$c_{ij} = a_{ij} b_{ij}, \qquad 1 \le i \le m, \quad 1 \le j \le n\ . \tag{1.52}$$

The Hadamard product is denoted by $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$. The Hadamard product of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and itself is called the entrywise square of $\mathbf{A}$, and denoted as

$$\mathbf{A}^{\circ 2} = \mathbf{A} \circ \mathbf{A}. \tag{1.53}$$

Example 1.6

$$\begin{pmatrix} 0 & 2 & -3 \\ -1 & 0 & -4 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & -3 \\ 2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 9 \\ -2 & 0 & 4 \end{pmatrix}.$$

An important application of the Hadamard product is in efficiently computing the product of a diagonal matrix and a rectangular matrix when both are large in size.

$$\underbrace{\mathbf{A}}_{\text{diagonal}} \mathbf{B} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} = \begin{bmatrix} a_1 B_1 \\ \vdots \\ a_n B_n \end{bmatrix} \tag{1.54}$$

$$\mathbf{A} \underbrace{\mathbf{B}}_{\text{diagonal}} = [\mathbf{a}_1 \dots \mathbf{a}_n] \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} = [b_1 \mathbf{a}_1 \dots b_n \mathbf{a}_n] \tag{1.55}$$

We may implement the direct row/column operations using the Hadamard product. For example, in the former case, let $\mathbf{a} = \text{diag}(\mathbf{A}) = (a_1, \dots, a_n)^T \in \mathbb{R}^n$, which represents the diagonal of $\mathbf{A}$. Then

$$\underbrace{\mathbf{A}}_{n\times n} \cdot \underbrace{\mathbf{B}}_{n\times p} = \underbrace{[\mathbf{a} \dots \mathbf{a}]}_{p \text{ copies}} \circ \mathbf{B}. \tag{1.56}$$

See Fig. 1.13 for an illustration of the process. The ordinary way of multiplying out $\mathbf{AB}$ takes $O(n^2 p)$ operations, while the Hadamard product takes only $O(np)$ operations, which is one magnitude fewer.

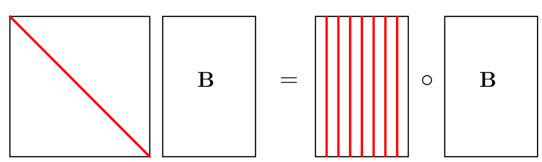

Fig. 1.13 Matrix multiplication (involving a diagonal matrix) through Hadamard product.

Example 1.7 Verify the following identity:

$$\begin{pmatrix} -1 & & \\ & 0 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 10 \\ 4 & 5 & 6 & 10 \\ 7 & 8 & 9 & 10 \end{pmatrix} = \begin{pmatrix} -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 10 \\ 4 & 5 & 6 & 10 \\ 7 & 8 & 9 & 10 \end{pmatrix}$$

1.2.3 Matrix transpose

The transpose of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is another matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ with $b_{ij} = a_{ji}$ for all i, j. We denote the transpose of $\mathbf{A}$ by $\mathbf{A}^T$. If a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ has a transpose that coincides with itself, i.e., $\mathbf{A}^T = \mathbf{A}$, then the matrix $\mathbf{A}$ is said to be symmetric.

The matrix transpose is a linear operator, that is, for any matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ and scalar $k \in \mathbb{R}$,

$$(k\mathbf{A})^T = k\mathbf{A}^T, \qquad \text{and} \quad (\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T. \tag{1.57}$$

Clearly, for any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$,

$$(\mathbf{A}^T)^T = \mathbf{A} \ . \tag{1.58}$$

Given any two matrices $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$, it can be shown that

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T \ . \tag{1.59}$$

1.2.4 Matrix trace

The trace of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the sum of all its diagonal entries:

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}. \tag{1.60}$$

Clearly,

$$\operatorname{tr}(\mathbf{O}) = 0, \quad \operatorname{tr}(\mathbf{I}_n) = \operatorname{tr}(\mathbf{J}_n) = n \ , \tag{1.61}$$

and for any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$,

$$\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{A}^T) \ . \tag{1.62}$$

Trace is also a linear operator: For any matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ and scalar $k \in \mathbb{R}$,

$$\operatorname{tr}(k\mathbf{A}) = k \cdot \operatorname{tr}(A), \qquad \text{and} \quad \operatorname{tr}(\mathbf{A} + \mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B}). \tag{1.63}$$

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times m$ matrix, then

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}) \; . \tag{1.64}$$

However, note that the matrix **AB** is not necessarily equal to **BA** (they don't even need to have the same size).

1.2.5 Matrix rank

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be an arbitrary matrix. The maximal number of linearly independent rows (or columns) contained in the matrix is called the **rank** of **A**, and denoted as rank(**A**):

- A long matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is said to have **full row rank** if $\mathrm{rank}(\mathbf{B}) = m$. In this case, all of its rows are linearly independent.
- A tall matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is said to have **full column rank** if $\mathrm{rank}(\mathbf{B}) = n$. In this case, all of its columns must be linearly independent.
- A square matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ is said to be **of full rank** (or nonsingular) if $\mathrm{rank}(\mathbf{P}) = n$; otherwise, it is said to be **rank deficient** (or singular).

Another way to define the rank of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is through a subspace associated to the matrix. Define

$$\mathrm{Col}(\mathbf{A}) = \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subseteq \mathbb{R}^m \tag{1.65}$$
$$\mathrm{Row}(\mathbf{A}) = \mathrm{span}\{A_1, \ldots, A_m\} \subseteq \mathbb{R}^n \tag{1.66}$$
$$\mathrm{Nul}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^n \tag{1.67}$$

which are respectively called the **column**, **row** and **null spaces** of **A**. See Figure 1.14 for an illustration.

The rank of **A** can be alternatively defined as the dimension of the column space of **A**:

$$\mathrm{rank}(\mathbf{A}) = \dim(\mathrm{Col}(\mathbf{A})). \tag{1.68}$$

Furthermore, if $\mathrm{rank}(\mathbf{A}) = r$, we must have

$$\dim(\mathrm{Row}(\mathbf{A})) = r, \qquad \text{and} \quad \dim(\mathrm{Nul}(\mathbf{A})) = n - r. \tag{1.69}$$

We mention some useful properties about the matrix rank. First of all, for any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$,

$$0 \le \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}^T) \le \min(m, n), \tag{1.70}$$

and $\mathrm{rank}(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{O}$. Second, for any two multiplicatively compatible matrices $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$,

$$\mathrm{rank}(\mathbf{AB}) \le \min(\mathrm{rank}(\mathbf{A}),\ \mathrm{rank}(\mathbf{B})). \tag{1.71}$$

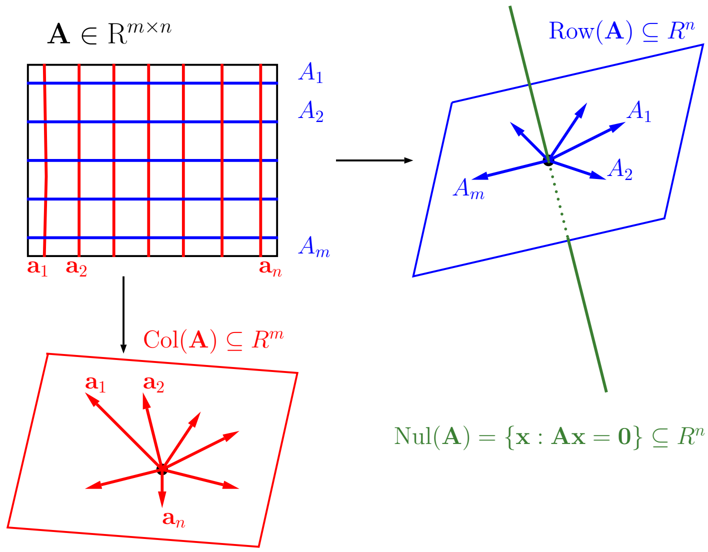

Fig. 1.14 Subspaces associated to a matrix. Note that Nul($\mathbf{A}$), the null space of $\mathbf{A}$, is the solution set of $\mathbf{A}\mathbf{x} = \mathbf{0}$. It is in the same space with, and also orthogonal to, Row($\mathbf{A}$), the row space of $\mathbf{A}$.

This shows that matrix multiplication never increases the rank. Third, multiplying a *square, nonsingular* matrix on either side of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ always preserves its rank:

$$\mathrm{rank}(\mathbf{PA}) = \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{AQ}). \tag{1.72}$$

where $\mathbf{P} \in \mathbb{R}^{m\times m}, \mathbf{Q} \in \mathbb{R}^{n\times n}$ are any two invertible matrices. Lastly, for any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, it is always true that

$$\mathrm{rank}(\mathbf{A}\mathbf{A}^T) = \mathrm{rank}(\mathbf{A}^T\mathbf{A}) = \mathrm{rank}(\mathbf{A}). \tag{1.73}$$

It is also worth mentioning a few different ways of characterizing rank-1 matrices, which are simple but often encountered in practical applications.

Theorem 1.2 (Characterizations of rank-1 matrices)

1. *Any nonzero row or column vector has rank 1 (as a matrix).*
2. *A nonzero matrix is of rank 1 if and only if its nonzero rows (or columns) are multiples of each other.*
3. *A nonzero matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *has rank 1 if and only if there exist nonzero vectors* $\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n$*, such that* $\mathbf{A} = \mathbf{u}\mathbf{v}^T$.

The proof of the theorem is rather straightforward, so we omit it but use the following example to illustrate it.

Example 1.8 The following shows a rank-1 matrix, as well as its decomposition into two vectors:

$$\begin{pmatrix} 1 & 0 & 3 \\ 2 & 0 & 6 \\ 3 & 0 & 9 \\ 4 & 0 & 12 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 3 \end{pmatrix} \tag{1.74}$$

1.2.6 Matrix inverse

A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be invertible if there exists another square matrix of the same size $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. In this case, $\mathbf{B}$ is called the matrix inverse of $\mathbf{A}$ and denoted as $\mathbf{B} = \mathbf{A}^{-1}$.

For any 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad - bc \neq 0$, it is invertible and its inverse is

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{1.75}$$

In general, a sufficient and necessary condition for a square matrix $\mathbf{A}$ to be invertible is that it is nonsingular (i.e., of full rank).

Let $\mathbf{A}$ be an invertible matrix. The matrix $\mathbf{A}^T$ is also invertible, with

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T, \tag{1.76}$$

If $\mathbf{A}$ is also symmetric, then this equation reduces to

$$\mathbf{A}^{-1} = (\mathbf{A}^{-1})^T \ .$$

which shows that $\mathbf{A}^{-1}$ is also symmetric.

If $k \neq 0$, then $k\mathbf{A}$ is also invertible and the inverse is

$$(k\mathbf{A})^{-1} = \frac{1}{k}\mathbf{A}^{-1} \ . \tag{1.77}$$

For any two invertible matrices $\mathbf{A}, \mathbf{B}$ of the same size, their product matrix is also invertible:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \ . \tag{1.78}$$

Note, however, that $\mathbf{A} + \mathbf{B}$ is not necessarily invertible (even though individually they are invertible).

Lastly, we mention a special case where the inverse of a square matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ coincides with its transpose, i.e.,

$$\mathbf{Q}^{-1} = \mathbf{Q}^T, \qquad \text{or equivalently,} \quad \mathbf{QQ}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}. \tag{1.79}$$

Such matrices are called **orthogonal** matrices. For example, the following are orthogonal matrices:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \tag{1.80}$$

Geometrically, an orthogonal matrix multiplying a vector (i.e., $\mathbf{Qx} \in \mathbb{R}^n$) represents a rotation of the vector in the space.

The following theorem shows that orthogonal matrices are square matrices with orthonormal vectors in columns.

Theorem 1.3 *A square matrix $\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_n]$ is orthogonal if and only if its columns form an orthonormal basis for $\mathbb{R}^n$.*

Proof. Since

$$\mathbf{Q}^T\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \dots & \mathbf{q}_n \end{bmatrix} = (\mathbf{q}_i^T \mathbf{q}_j), \tag{1.81}$$

we conclude that $\mathbf{Q}$ is an orthogonal matrix if and only if

$$\mathbf{q}_i^T \mathbf{q}_j = \delta_{ij}, \text{ for all } i, j.$$

That is, $\mathbf{q}_1, \dots, \mathbf{q}_n \in \mathbb{R}^n$ are unit vectors and orthogonal to each other, thus forming an orthonormal basis for $\mathbb{R}^n$.

1.2.7 Matrix determinant

The matrix determinant is a rule to evaluate square matrices to numbers (in order to determine if they are nonsingular):

$$\det : \mathbf{A} \in \mathbb{R}^{n\times n} \;\mapsto\; \det(\mathbf{A}) \in \mathbb{R}.$$

Its general definition is quite complicated and we thus omit it but refer the reader to [cite]. Here, we focus on reviewing the many cases where matrix determinants can be easily obtained.

For example, if $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonal ($a_{ij} = 0$ for all $i \neq j$), upper triangular ($a_{ij} = 0$ for all $i > j$), or lower triangular ($a_{ij} = 0$ for all $i < j$), then the determinant of $\mathbf{A}$ is the product of its diagonal entries:

$$\det(\mathbf{A}) = \prod_{i=1}^{n} a_{ii} \; . \tag{1.82}$$

Similarly, if $\mathbf{A} \in \mathbb{R}^{n\times n}$ is block diagonal, block upper-triangular or block lower-triangular,

$$\mathbf{A} = \begin{pmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{kk} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1k} \\ & A_{22} & \dots & A_{2k} \\ & & \ddots & \vdots \\ & & & A_{kk} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \vdots & \vdots & \ddots & \\ A_{k1} & A_{k2} & \dots & A_{kk} \end{pmatrix} \quad (1.83)$$

then the determinant of $\mathbf{A}$ is the product of the determinants of the main blocks:

$$\det(\mathbf{A}) = \prod_{i=1}^{k} \det(A_{ii}) \ . \quad (1.84)$$

A remarkable property about the use of the matrix determinant is the following.

Theorem 1.4 *A square matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *is invertible (or nonsingular) if and only if it has a nonzero determinant.*

Example 1.9 Find the determinant of the following matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix},$$

and use it to determine of the matrix is invertible.

Solution 1.2 This matrix is block lower-triangular, with two main blocks:

$$A_{11} = (3), \quad A_{22} = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \ .$$

Thus,

$$\det(\mathbf{A}) = \det(3) \cdot \det \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} = 3 \cdot (1 \cdot 4 - (-1) \cdot 2) = 3 \cdot 6 = 18.$$

Since the determinant is nonzero, we conclude that the matrix is nonsingular and invertible with $\text{rank}(\mathbf{A}) = 3$.

We mention a few useful properties about the matrix determinant. For a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and scalar $k \in \mathbb{R}$,

$$\det(\mathbf{A}^T) = \det(\mathbf{A}), \qquad \text{and} \quad \det(k\mathbf{A}) = k^n \det(\mathbf{A}) \ . \quad (1.85)$$

Given another square matrix of the same size, $\mathbf{B} \in \mathbb{R}^{n\times n}$, we have

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}) \ . \quad (1.86)$$

This implies that the product of two square matrices of the same size is invertible (nonsingular) if and only if each of them is invertible (nonsingular). Additionally, for any invertible matrix $\mathbf{A}$,

$$\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1} \ . \tag{1.87}$$

1.2.8 Eigenvalues and eigenvectors

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ by any square matrix. A scalar $\lambda_0 \in \mathbb{R}$ is called an *eigenvalue* of $\mathbf{A}$ if there exists a nonzero vector $\mathbf{v}_0 \in \mathbb{R}^n$ such that

$$\mathbf{A}\mathbf{v}_0 = \lambda_0\mathbf{v}_0, \qquad \text{or equivalently,} \quad (\mathbf{A} - \lambda_0\mathbf{I})\mathbf{v}_0 = \mathbf{0} \tag{1.88}$$

The vector $\mathbf{v}_0$ is called an *eigenvector* of $\mathbf{A}$ (associated to the eigenvalue λ_0). In some cases, we say for short that $(\lambda_0, \mathbf{v}_0)$ is an *eigenpair* of $\mathbf{A}$.

For the given matrix $\mathbf{A}$, let

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}), \quad \text{for all } \lambda \in \mathbb{R}. \tag{1.89}$$

It is a polynomial in λ of order n, called the *characteristic polynomial*[2] of $\mathbf{A}$. Eigenvalues of $\mathbf{A}$ must be roots of the characteristic equation $p_{\mathbf{A}}(\lambda) = 0$. Note that the equation could also have several complex roots, which are eigenvalues of the matrix in the complex number domain. In this section and throughout the book, for simplicity and to present only what is needed later, we only work with real eigenvalues (note that eigenvectors corresponding to real eigenvalues must be real as well) and will completely avoid complex eigenvalues. Thus, whenever we mention eigenvalues again, they should be understood as real eigenvalues.

Let $\lambda_0 \in \mathbb{R}$ be an eigenvalue of $\mathbf{A} \in \mathbb{R}^{n\times n}$. All eigenvectors associated to λ_0 span a linear subspace of $\mathbb{R}^n$, called the *eigenspace* of $\mathbf{A}$ corresponding to the eigenvalue:

$$\mathrm{E}(\lambda_0) = \{\mathbf{v} \in \mathbb{R}^n \mid (\mathbf{A} - \lambda_0\mathbf{I})\mathbf{v} = \mathbf{0}\}. \tag{1.90}$$

The dimension g_0 of $\mathrm{E}(\lambda_0)$ is called the *geometric multiplicity* of λ_0, while the degree a_0 of the factor $(\lambda - \lambda_0)^{a_0}$ in the characteristic polynomial $p(\lambda)$ is called the *algebraic multiplicity* of λ_0.[3] Note that we must have

$$1 \le g_0 \le a_0. \tag{1.91}$$

Remark 1.2 For any matrix $\mathbf{A}$ with a zero eigenvalue, the corresponding eigenspace is identical to the null space of the matrix, i.e.,

$$\mathrm{E}(0) = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{v} = \mathbf{0}\} = \mathrm{Nul}(\mathbf{A}). \tag{1.92}$$

---

[2] Other books may define the characteristic polynomial as $\tilde{p}_{\mathbf{A}}(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$. The two definitions are identical when n is even and will differ by a negative sign when n is odd. This is because $\tilde{p}_{\mathbf{A}}(\lambda) = \det(-(\mathbf{A} - \lambda\mathbf{I})) = (-1)^n \det(\mathbf{A} - \lambda\mathbf{I}) = (-1)^n p_{\mathbf{A}}(\lambda)$.

[3] Note that a_0 must be as large as possible in the sense that $(\lambda - \lambda_0)^{a_0} \mid p(\lambda)$, but $(\lambda - \lambda_0)^{a_0+1} \nmid p(\lambda)$).

Example 1.10 Find all the eigenvalues and associated eigenvectors of the following matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix},$$

What are their algebraic and geometric multiplicities?

We start by finding the characteristic polynomial of $\mathbf{A}$

$$p(\lambda) = \det \begin{pmatrix} 3-\lambda & 0 & 0 \\ 5 & 1-\lambda & -1 \\ -2 & 2 & 4-\lambda \end{pmatrix} = (3-\lambda)((1-\lambda)(4-\lambda)+2) = -(\lambda-3)^2(\lambda-2).$$

This shows that the matrix $\mathbf{A}$ has two eigenvalues $\lambda_1 = 3, \lambda_2 = 2$ with algebraic multiplicities $a_1 = 2, a_2 = 1$. To find the eigenvectors of $\mathbf{A}$ corresponding to $\lambda_1 = 3$, we need to solve the following linear system:

$$(\mathbf{A} - 3\mathbf{I})\mathbf{x} = \mathbf{0} : \begin{pmatrix} 0 & 0 & 0 \\ 5 & -2 & -1 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \longrightarrow \quad \begin{pmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Clearly, $x_1 = 0$ and $2x_2 + x_3 = 0$. It follows that

$$\mathbf{x} = \begin{pmatrix} 0 \\ x_2 \\ -2x_2 \end{pmatrix} = x_2 \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}$$

where x_2 is a free variable. This shows that the eigenspace corresponding to λ_1 is a one-dimensional subspace consisting of vectors that are all multiples of $\mathbf{v}_1 = (0, 1, -2)^T$ and thus the geometric multiplicity of λ_1 is $g_1 = 1$.

For the other eigenvalue $\lambda_2 = 2$, by similar steps, we can obtain that $g_2 = 1$ and the associated eigenvectors are all multiples of $\mathbf{v}_2 = (0, 1, -1)^T$. This is left to the reader to verify.

Interestingly, for any given square matrix, eigenvectors corresponding to distinct eigenvalues must be linearly independent.

Theorem 1.5 *For a given square matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$*, let* $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ *be its eigenvectors corresponding to* k *distinct eigenvalues* $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$*, where* k *is a positive integer. Then* $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ *is a linearly independent set.*

Proof Suppose there exist constants $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$\mathbf{0} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \; .$$

We would like to show that $c_1 = \cdots = c_k = 0$.

For this goal, we first multiply both sides of the above equation by $\mathbf{A}$ to get

$$\mathbf{0} = c_1(\mathbf{A}\mathbf{v}_1) + \cdots + c_k(\mathbf{A}\mathbf{v}_k) = c_1\lambda_1\mathbf{v}_1 + \cdots + c_k\lambda_k\mathbf{v}_k \; .$$

and then repeat this operation $k-2$ times to get

$$\begin{aligned}\mathbf{0} &= c_1\lambda_1^2\mathbf{v}_1 + \cdots + c_k\lambda_k^2\mathbf{v}_k\\ &\vdots\\ \mathbf{0} &= c_1\lambda_1^{k-1}\mathbf{v}_1 + \cdots + c_k\lambda_k^{k-1}\mathbf{v}_k\end{aligned}$$

Collectively, these k equations may be written in matrix form:

$$\mathbf{O} = [c_1\mathbf{v}_1 \ldots c_k\mathbf{v}_k]\cdot\begin{pmatrix}1 & \lambda_1 & \cdots & \lambda_1^{k-1}\\ \vdots & \vdots & \ldots & \vdots\\ 1 & \lambda_k & \cdots & \lambda_k^{k-1}\end{pmatrix}$$

The matrix containing the λ_i's, i.e.,

$$\mathbf{M} = \begin{pmatrix}1 & \lambda_1 & \cdots & \lambda_1^{k-1}\\ \vdots & \vdots & \ldots & \vdots\\ 1 & \lambda_k & \cdots & \lambda_k^{k-1}\end{pmatrix} \tag{1.93}$$

is called a Vandermonde matrix and is known to have the following determinant (In Problem you are asked to verify the identity in the case of $k=3$):

$$\det(\mathbf{M}) = \prod_{1\le i<j\le k}(\lambda_j - \lambda_i) \tag{1.94}$$

Since all the eigenvalues are distinct ($\lambda_i \neq \lambda_j$ for all $1 \le i \neq j \le k$), we conclude that $\mathbf{M}$ is nonsingular and consequently,

$$[c_1\mathbf{v}_1 \ldots c_k\mathbf{v}_k] = \mathbf{O}\ .$$

This implies that $c_1 = \cdots = c_k = 0$ (because $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are nonzero vectors). □

1.3 Diagonalization of square matrices

In this part we discuss diagonalization of real, square matrices by real, invertible matrices, which is an important tool in studying matrices.

1.3.1 Similar matrices

Two square matrices of the same size, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, are said to be *similar* if there exists an invertible matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1} \ .$$

Similar matrices have a lot of things in common, as stated in the following theorem.

Theorem 1.6 *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ *be two similar matrices. Then they must have the rank, trace, determinant and characteristic polynomials:*

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}), \quad \text{tr}(\mathbf{A}) = \text{tr}(\mathbf{B}), \quad \det(\mathbf{A}) = \det(\mathbf{B}), \quad p_{\mathbf{A}}(\lambda) = p_{\mathbf{B}}(\lambda).$$

Furthermore, they must have the same eigenvalues, with identical algebraic and geometric multiplicities, but their eigenvectors are not necessarily the same.

Proof Let $\mathbf{P} \in \mathbb{R}^{n\times n}$ be an invertible matrix such that $\mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$. The first three identities can be verified directly by using corresponding properties about matrix rank, trace and determinant. This is left to the reader as an exercise.

We prove the identity concerning the characteristic equation:

$$\begin{aligned} p_{\mathbf{B}}(\lambda) &= \det(\mathbf{B} - \lambda\mathbf{I}) = \det(\mathbf{P}\mathbf{A}\mathbf{P}^{-1} - \lambda\mathbf{I}) \\ &= \det\left(\mathbf{P}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{P}\right) = \det(\mathbf{P}^{-1})\det(\mathbf{A} - \lambda\mathbf{I})\det(\mathbf{P}) \\ &= \det(\mathbf{A} - \lambda\mathbf{I}) = p_{\mathbf{A}}(\lambda). \end{aligned}$$

As a result, $\mathbf{A}$ and $\mathbf{B}$ have identical eigenvalues with exactly the same algebraic multiplicities. To see that the geometric multiplicities are also the same, let λ be an arbitrary eigenvalue shared by $\mathbf{A}$ and $\mathbf{B}$ and consider the equation

$$\mathbf{0} = (\mathbf{B} - \lambda\mathbf{I}) \cdot \mathbf{v} = (\mathbf{P}^{-1}\mathbf{A}\mathbf{P} - \lambda\mathbf{I}) \cdot \mathbf{v} = \mathbf{P}^{-1}(\mathbf{A} - \lambda\mathbf{I}) \cdot \mathbf{P}\mathbf{v},$$

Multiplying both sides by $\mathbf{P}$ gives that

$$\mathbf{0} = (\mathbf{A} - \lambda\mathbf{I}) \cdot \mathbf{P}\mathbf{v} \ .$$

This shows that $\mathbf{v}$ is an eigenvector of $\mathbf{B}$ corresponding to eigenvalue λ if and only if $\mathbf{P}\mathbf{v}$ is an eigenvector of $\mathbf{A}$ corresponding to λ. Therefore, $\mathbf{A}$ and $\mathbf{B}$ have the same number of linearly independent eigenvectors corresponding to the same eigenvalue λ (note that $\mathbf{P}$ is invertible), and thus the geometric multiplicity of λ is the same for both $\mathbf{A}$ and $\mathbf{B}$. □

1.3.2 Diagonalizable matrices

A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be *diagonalizable* (in the real number domain) if it is similar to a diagonal matrix, i.e., there exist an invertible matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \qquad \text{or equivalently,} \quad \mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda} \ . \tag{1.95}$$

To better understand this definition, write

$$\mathbf{P} = \left[\mathbf{p}_1 \ \dots \ \mathbf{p}_n\right], \qquad \text{and} \quad \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n) \ .$$

The above equation can be rewritten as

$$\mathbf{A} \ \left[\mathbf{p}_1 \ \dots \ \mathbf{p}_n\right] = \left[\mathbf{p}_1 \ \dots \ \mathbf{p}_n\right] \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} .$$

By comparing columns of the two sides, we get that

$$\mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i, \ 1 \le i \le n. \tag{1.96}$$

This shows that each λ_i is an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\mathbf{p}_i$. Therefore, any diagonalizable matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ must have n real eigenvalues (and thus no complex eigenvalue), and the full list of those eigenvalues, which are not necessarily distinct, is called the *spectrum* of the matrix $\mathbf{A}$. Accordingly, the factorization of $\mathbf{A}$ in (1.95) is called the *eigendecomposition* of $\mathbf{A}$, or the *spectral decomposition* of $\mathbf{A}$.

Example 1.11 The matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix}$$

is diagonalizable because

$$\begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 3 & \\ & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}^{-1}$$

but the matrix

$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$$

is not diagonalizable (we will justify this conclusion later).

1.3.3 Why diagonalization is important

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a diagonalizable matrix, that is, it is similar to a diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ of the same size. Since similar matrices share a lot of things in common, we can use $\mathbf{\Lambda}$ instead to compute the rank, determinant and eigenvalues (and their algebraic multiplicities) of $\mathbf{A}$, which is a lot simpler as shown in the theorem below.

Theorem 1.7 *For any diagonalizable matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ (some of which could repeat each other), we have*

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i, \quad \mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i, \text{ and } \mathrm{rank}(\mathbf{A}) = \#\textit{nonzero eigenvalues}. \tag{1.97}$$

Proof Since $\mathbf{A}$ is diagonalizable, there exists an invertible matrix $\mathbf{P}$ of the same size such that

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \qquad \text{where} \quad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

It follows that

$$\det(\mathbf{A}) = \det(\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}) = \det(\mathbf{P})\det(\mathbf{\Lambda})\det(\mathbf{P}^{-1}) = \det(\mathbf{\Lambda}) = \prod_{i=1}^{n} \lambda_i$$

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}) = \mathrm{tr}(\mathbf{P}^{-1}\mathbf{P}\mathbf{\Lambda}) = \mathrm{tr}(\mathbf{\Lambda}) = \sum_{i=1}^{n} \lambda_i$$

$$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}) = \mathrm{rank}(\mathbf{\Lambda}) = \sum_{i=1}^{n} 1_{\lambda_i \neq 0}$$

where $1_{\lambda_i \neq 0}$ is 1 if the statement $\lambda_i \neq 0$ is true, or 0 otherwise. □

Remark 1.3 In fact, the three identities in the above theorem hold true for any square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ that has n real eigenvalues. The diagonalizability assumption in the theorem is simply there to ensure that the matrix has n real eigenvalues.

Another application of diagonalization is to help compute matrix powers ($\mathbf{A}^k$) and exponential ($e^{\mathbf{A}}$), when $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a diagonalizable matrix. To see how to get different powers of $\mathbf{A}$, suppose $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ for some invertible matrix $\mathbf{P}$ and diagonal matrix $\mathbf{\Lambda}$. Then

$$\mathbf{A}^2 = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \cdot \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} = \mathbf{P}\mathbf{\Lambda}^2\mathbf{P}^{-1}$$
$$\mathbf{A}^3 = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \cdot \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \cdot \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} = \mathbf{P}\mathbf{\Lambda}^3\mathbf{P}^{-1}\,,$$

and more generally, for any positive integer k,

$$\mathbf{A}^k = \mathbf{P}\mathbf{\Lambda}^k\mathbf{P}^{-1}, \qquad \text{where} \quad \mathbf{\Lambda}^k = \mathrm{diag}\left(\lambda_1^k, \ldots, \lambda_n^k\right)\,. \tag{1.98}$$

So the matrix power just falls upon the diagonal matrix $\mathbf{\Lambda}$!

The matrix exponential of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as

$$e^{\mathbf{A}} = \mathbf{I} + \frac{1}{1!}\mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}\mathbf{A}^k. \tag{1.99}$$

Since $\mathbf{A}$ is diagonalizable, we can apply the above formula for matrix powers to obtain that

$$e^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\mathbf{P}\mathbf{\Lambda}^k\mathbf{P}^{-1}\right) = \mathbf{P}\left(\sum_{k=0}^{\infty} \frac{1}{k!}\mathbf{\Lambda}^k\right)\mathbf{P}^{-1} = \mathbf{P}e^{\mathbf{\Lambda}}\mathbf{P}^{-1}, \tag{1.100}$$

where

$$e^{\mathbf{\Lambda}} = \sum_{k=0}^{\infty} \frac{1}{k!}\mathbf{\Lambda}^k = \operatorname{diag}\left(\sum_{k=0}^{\infty} \frac{1}{k!}\lambda_1^k, \ldots, \sum_{k=0}^{\infty} \frac{1}{k!}\lambda_n^k\right) = \operatorname{diag}\left(e^{\lambda_1}, \ldots, e^{\lambda_n}\right). \quad (1.101)$$

Example 1.12 For the diagonalizable matrix in the preceding example,

$$\underbrace{\begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}}_{\mathbf{P}} \underbrace{\begin{pmatrix} 3 & \\ & -1 \end{pmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}^{-1}}_{\mathbf{P}^{-1}}$$

the 10th power of $\mathbf{A}$ is

$$\mathbf{A}^{10} = \mathbf{P}\mathbf{\Lambda}^{10}\mathbf{P}^{-1} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 3^{10} & \\ & (-1)^{10} \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} \end{pmatrix} = \begin{pmatrix} 14763 & 14762 \\ 44286 & 44287 \end{pmatrix}.$$

and the matrix exponential of $\mathbf{A}$ is

$$e^{\mathbf{A}} = \mathbf{P}e^{\mathbf{\Lambda}}\mathbf{P}^{-1} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} e^3 & \\ & e^{-1} \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{4}e^3 + \frac{3}{4}e^{-1} & \frac{1}{4}e^3 - \frac{1}{4}e^{-1} \\ \frac{3}{4}e^3 - \frac{3}{4}e^{-1} & \frac{3}{4}e^3 + \frac{1}{4}e^{-1} \end{pmatrix}.$$

1.3.4 Checking diagonalizability

The following theorem lays out a way for checking the diagonalizability of square matrices.

Theorem 1.8 *A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonalizable in the real number domain if and only if it has n linearly independent eigenvectors in $\mathbb{R}^n$.*

Proof We prove both directions:

1. Suppose $\mathbf{A}$ is diagonalizable, that is, there exist an invertible matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ such that $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$. We have already shown that the columns of $\mathbf{P}$ are all eigenvectors of $\mathbf{A}$. Since $\mathbf{P}$ is invertible, $\mathbf{P}$ has linearly independent columns. Thus, $\mathbf{A}$ has n linearly independent eigenvectors.
2. Suppose $\mathbf{A}$ has n linearly independent eigenvectors $\mathbf{p}_1, \ldots, \mathbf{p}_n \in \mathbb{R}^n$ corresponding respectively to eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ (some of them could be equal to each other):

   $$\mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i,\ 1 \le i \le n.$$

   Let

   $$\mathbf{P} = \left[\mathbf{p}_1 \ \ldots \ \mathbf{p}_n\right], \qquad \text{and} \quad \mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

   Then $\mathbf{P}$ is invertible, and satisfies

   $$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}, \qquad \text{or equivalently,} \quad \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}.$$

This shows that $\mathbf{A}$ is diagonalizable. □

Example 1.13 In Example we mentioned that the matrix

$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$$

is not diagonalizable. The reason is that this matrix has a repeated eigenvalue $\lambda_1 = 1$ ($a_1 = 2$) but there is only one linearly independent eigenvector $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1, 1)^T$:

$$\mathbf{0} = (\mathbf{B} - \lambda_1 \mathbf{I})\mathbf{v} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \mathbf{v} \quad \longrightarrow \quad \mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Example 1.14 Consider the matrix $\mathbf{A}$ defined in Example 1.10

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix},$$

It has been shown that $\mathbf{A}$ has two distinct eigenvalues $\lambda_1 = 3, \lambda_2 = 2$ with algebraic multiplicities $a_1 = 2, a_2 = 1$ and geometric multiplicities $g_1 = g_2 = 1$. Since $g_1 + g_2 < 3$, we conclude that the matrix is not diagonalizable. However, because the 3×3 matrix has 3 eigenvalues ($a_1 + a_2 = 3$), we still have $\det(\mathbf{A}) = 3 \cdot 3 \cdot 2 = 18$, $\text{tr}(\mathbf{A}) = 3 + 3 + 2 = 8$ and $\text{rank}(\mathbf{A}) = 3$. Clearly, they coincide with the results we have previously obtained.

An immediate consequence of the above theorem, when coupled with Theorem 1.5, is the following result.

Theorem 1.9 *Any square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with n distinct real eigenvalues must be diagonalizable.*

Two special classes of real, square matrices that are always diagonalizable in the real number domain are *idempotent* matrices and *symmetric* matrices. We study each of them in some detail next.

1.4 Idempotent matrices

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square matrix. It is said to be *idempotent* if $\mathbf{A}^2 = \mathbf{A}$. We denote by $I^n(\mathbb{R})$ the collection of all real, idempotent matrices of size $n \times n$, i.e.,

$$I^n(\mathbb{R}) = \{\mathbf{A} \in \mathbb{R}^{n\times n} \mid \mathbf{A}^2 = \mathbf{A}\}. \tag{1.102}$$

Example 1.15 The following are some examples of idempotent matrices:

$$\mathbf{O}, \quad \mathbf{I}, \quad \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

In general, any 2×2 matrix of the form

$$\begin{pmatrix} a & b \\ c & 1-a \end{pmatrix}, \qquad \text{with} \quad a^2 + bc = a \tag{1.103}$$

is idempotent:

$$\begin{pmatrix} a & b \\ c & 1-a \end{pmatrix} \begin{pmatrix} a & b \\ c & 1-a \end{pmatrix} = \begin{pmatrix} a^2 + bc & ab + b(1-a) \\ ca + (1-a)c & bc + (1-a)^2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & 1-a \end{pmatrix}.$$

It should also be noted that the constant matrix of 1's, $\mathbf{J}_n$, is not idempotent:

$$\mathbf{J}_n^2 = (\mathbf{1}\mathbf{1}^T)(\mathbf{1}\mathbf{1}^T) = \mathbf{1}\underbrace{(\mathbf{1}^T\mathbf{1})}_{=n}\mathbf{1}^T = n\mathbf{1}\mathbf{1}^T = n\mathbf{J}_n, \tag{1.104}$$

but a normalized version of it, $\frac{1}{n}\mathbf{J}_n$, is idempotent:

$$\left(\frac{1}{n}\mathbf{J}_n\right)^2 = \frac{1}{n^2}\mathbf{J}_n^2 = \frac{1}{n^2} \cdot n\mathbf{J}_n = \frac{1}{n}\mathbf{J}_n. \tag{1.105}$$

Idempotent matrices have many nice properties. For example, all the matrix powers of an idempotent matrix $\mathbf{A}$ are equal to itself: For any integer $k \geq 3$,

$$\mathbf{A}^k = \mathbf{A}^2 \cdot \mathbf{A}^{k-2} = \mathbf{A} \cdot \mathbf{A}^{k-2} = \mathbf{A}^{k-1} = \cdots = \mathbf{A}.$$

As a result, the matrix exponential of $\mathbf{A}$ is

$$e^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{1}{k!}\mathbf{A}^k = \mathbf{I} + \sum_{k=1}^{\infty} \frac{1}{k!}\mathbf{A} = \mathbf{I} + (e-1)\mathbf{A}.$$

Idempotent matrices must have a determinant of 0 or 1, because

$$\det(\mathbf{A}) = \det(\mathbf{A}^2) = [\det(\mathbf{A})]^2.$$

If an idempotent matrix $\mathbf{A}$ satisfies $\det(\mathbf{A}) = 1$, then it must be nonsingular. We can thus multiply both sides of $\mathbf{A}^2 = \mathbf{A}$ by the inverse of $\mathbf{A}$ to obtain that

$$\mathbf{A}^{-1}\mathbf{A}^2 = \mathbf{A}^{-1}\mathbf{A} \qquad \longrightarrow \qquad \mathbf{A} = \mathbf{I}$$

This shows that the only nonsingular idempotent matrices are the identity matrices. In other words, all non-identity idempotent matrices have a determinant of 0 and are thus singular.

Idempotent matrices can only have eigenvalues 0 or 1 or both. Specifically, the zero matrix $\mathbf{O}$ only has the eigenvalue of 0, the identity matrix $\mathbf{I}$ only has the eigenvalue of 1, and all other idempotent matrices must have exactly two distinct eigenvalues, 0 and 1. To see this, let $\mathbf{A}$ be an idempotent matrix with an eigenvalue λ and corresponding eigenvector $\mathbf{v} \neq \mathbf{0}$. That is, $\mathbf{Av} = \lambda\mathbf{v}$. Using $\mathbf{A} = \mathbf{A}^2$, we then get

$$\lambda\mathbf{v} = \mathbf{Av} = \left(\mathbf{A}^2\right)\mathbf{v} = \mathbf{A}(\mathbf{Av}) = \mathbf{A}(\lambda\mathbf{v}) = \lambda(\mathbf{Av}) = \lambda(\lambda\mathbf{v}) = \lambda^2\mathbf{v}.$$

Since $\mathbf{v} \neq 0$, we must have $\lambda = \lambda^2$ and thus $\lambda = 0$ or 1.

Using the theory of minimal polynomials [Horn and Johnson, matrix analysis, 2nd ed, section 3.3], it can be proved that every idempotent matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonalizable (in the real number domain). That is, there exists an invertible matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ such that $\mathbf{A} = \mathbf{P\Lambda P}^{-1}$. Since this also represents an eigenvalue decomposition, the diagonal of $\mathbf{\Lambda}$ must be a collection of several 0's and several 1's. Let a_0 and a_1 be the frequencies of 0's and 1's on the diagonal of $\mathbf{\Lambda}$, which are also the algebraic multiplicities of the eigenvalues 0 and 1, with $0 \le a_0, a_1 \le n$ and $a_0 + a_1 = n$:

- $a_0 = n, a_1 = 0$: $\mathbf{A} = \mathbf{O}$;
- $a_0 = 0, a_1 = n$: $\mathbf{A} = \mathbf{I}$;
- $1 \le a_0, a_1 \le n - 1$: All other idempotent matrices

It follows that $\mathrm{tr}(\mathbf{A}) = a_1$ and $\mathrm{rank}(\mathbf{A}) = a_1$.

Example 1.16 Consider the following two matrices:

$$\mathbf{A} = \begin{pmatrix} 3 & -6 \\ 1 & -2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{pmatrix}$$

which are both idempotent. Since $\mathrm{tr}(\mathbf{A}) = 3 - 2 = 1$, we conclude that $\mathrm{rank}(\mathbf{A}) = 1$, $\mathbf{A}$ has two eigenvalues 0, 1 (each with algebraic multiplicity 1), and $\det(\mathbf{A}) = 0$. Similarly, $\mathrm{tr}(\mathbf{B}) = 2 + 3 - 3 = 2$, and thus we further obtain that $\mathrm{rank}(\mathbf{B}) = 2$, $\mathbf{B}$ has two eigenvalues 0, 1 (with algebraic multiplicities $a_0 = 1, a_1 = 2$), and $\det(\mathbf{A}) = 0$.

Example 1.17 Since $\frac{1}{n}\mathbf{J}_n \in \mathbb{R}^{n\times n}$ is idempotent, and obviously has a rank of 1, we have that it has an eigenvalue of 1 with algebraic multiplicity $a_1 = 1$, and the other eigenvalue is 0 with $a_0 = n - 1$. This implies that $\mathbf{J}_n$ has eigenvalues n and 0 with algebraic multiplicities $a_1 = 1, a_0 = n - 1$:

$$\frac{1}{n}\mathbf{J}_n \cdot \mathbf{v} = \lambda \cdot \mathbf{v} \quad \Longleftrightarrow \quad \mathbf{J}_n \cdot \mathbf{v} = n\lambda \cdot \mathbf{v}.$$

Example 1.18 Here we introduce an important and useful matrix,

$$\mathbf{T}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{n\times n} \tag{1.106}$$

It is called the *centering* matrix because, when applied to a vector $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, it would center the components of the vector:

$$\mathbf{T}_n\mathbf{x} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\mathbf{x} = \mathbf{x} - \frac{1}{n}\underbrace{\mathbf{1}\mathbf{1}^T}_{\mathbf{J}_n}\mathbf{x}$$

$$= \mathbf{x} - \mathbf{1}\cdot\frac{1}{n}\mathbf{1}^T\mathbf{x} = \mathbf{x} - \mathbf{1}\bar{x} = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$

where

$$\bar{x} = \frac{1}{n}\mathbf{1}^T\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the mean value of the components of $\mathbf{x}$. We will present more of its geometric interpretations in Chapter 6, and its applications in the dimension reduction chapters such as Chapter 9.

Algebraically, $\mathbf{T}_n$ has many nice properties. First, it is both symmetric (because it is the difference of two symmetric matrices) and idempotent:

$$\begin{aligned}\mathbf{T}_n^2 &= \left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right) \\ &= \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n - \frac{1}{n}\mathbf{J}_n + \frac{1}{n^2}\underbrace{\mathbf{J}_n^2}_{=n\mathbf{J}_n} \\ &= \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n \\ &= \mathbf{T}_n.\end{aligned}$$

Also,

$$\operatorname{tr}(\mathbf{T}_n) = \operatorname{tr}(\mathbf{I}_n) - \frac{1}{n}\operatorname{tr}(\mathbf{J}_n) = n - \frac{1}{n}\cdot n = n - 1. \tag{1.107}$$

According to the above theory, we can immediately conclude that

- $\operatorname{rank}(\mathbf{T}_n) = n - 1$ and $\det(\mathbf{T}_n) = 0$;
- $\mathbf{T}_n$ has two distinct eigenvalues, 0 and 1, with algebraic multiplicities $a_0 = 1$ and $a_1 = n - 1$ respectively.

Furthermore, the unique eigenvalue 0 has a corresponding eigenvector $\mathbf{1}$, because

$$\mathbf{T}_n\mathbf{1} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{1} = \mathbf{I}_n\mathbf{1} - \frac{1}{n}\mathbf{1}\underbrace{\mathbf{1}^T\mathbf{1}}_{n} = \mathbf{1} - \mathbf{1} = \mathbf{0} = 0\cdot\mathbf{1}, \tag{1.108}$$

Another interpretation is that all the rows of $\mathbf{T}_n$ sum to zero (and because of the symmetry of $\mathbf{T}_n$, all its columns sum to zero as well).

Lastly, we display a few special instances of the matrix $\mathbf{T}_n$ when $n = 1, 2, 3$:

$$\mathbf{T}_1 = (0), \quad \mathbf{T}_2 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \quad \mathbf{T}_3 = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \tag{1.109}$$

1.5 Symmetric matrices

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square matrix. It is called a *symmetric* matrix if its transpose coincides with itself, i.e., $\mathbf{A}^T = \mathbf{A}$. We denote by $S^n(\mathbb{R})$ the collection of all real, symmetric matrices of size $n \times n$, i.e.,

$$S^n(\mathbb{R}) = \{\mathbf{A} \in \mathbb{R}^{n\times n} \mid \mathbf{A}^T = \mathbf{A}\}. \tag{1.110}$$

Symmetric matrices have many nice properties, for example, all their eigenvalues are real (there is no complex eigenvalue) and eigenvectors corresponding to different eigenvalues are not only linearly independent, but also orthogonal to each other. The following theorem indicates that symmetric matrices are always diagonalizable and furthermore, they can be diagonalized through orthogonal matrices.

Theorem 1.10 *For any symmetric matrix* $\mathbf{A} \in S^n(\mathbb{R})$*, there exist an orthogonal matrix* $\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_n] \in \mathbb{R}^{n\times n}$ *and a diagonal matrix* $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n\times n}$*, such that*

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T. \tag{1.111}$$

This theorem is often called the *Spectral Theorem*, for example in Textbook [cite]. We omit the proof of this theorem (the interested reader is referred to [cite]) but make a few comments:

- The factorization of $\mathbf{A}$ in (1.111) is the spectral decomposition of $\mathbf{A}$: The λ_i's represent the eigenvalues of $\mathbf{A}$ while the $\mathbf{q}_i$'s are the associated eigenvectors (with unit norm and orthogonal to each other).
- Because $\mathbf{Q}$ is an orthogonal matrix, $\mathbf{A}$ is said to be *orthogonally diagonalizable*.
- The converse of the theorem is also true, i.e., orthogonally diagonalizable matrices must be symmetric: If $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ for some orthogonal matrix $\mathbf{Q}$ and diagonal matrix $\mathbf{\Lambda}$, then

$$\mathbf{A}^T = (\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T)^T = (\mathbf{Q}^T)^T\mathbf{\Lambda}^T\mathbf{Q}^T = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \mathbf{A}\ . \tag{1.112}$$

- One can rewrite the matrix decomposition $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ into a sum of rank-1 matrices:

$$\mathbf{A} = \begin{bmatrix}\mathbf{q}_1 & \dots & \mathbf{q}_n\end{bmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n \end{bmatrix} = \sum_{i=1}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^T \tag{1.113}$$

For convenience, we often sort the diagonal elements of $\mathbf{\Lambda}$ in decreasing order:

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n. \tag{1.114}$$

Note that the columns of $\mathbf{Q}$ must be rearranged accordingly to match with the positions of the λ_i's.

Example 1.19 Find the spectral decomposition of the matrix $\mathbf{A} = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}$.

The matrix $\mathbf{A}$ is symmetric, and thus orthogonally diagonalizable. By direct calculation we obtain the following eigenvalues and eigenvectors:

$$\lambda_1 = 4, \ \mathbf{q}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$
$$\lambda_2 = -1, \ \mathbf{q}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

Putting them together, we have

$$\begin{aligned} \mathbf{A} &= \underbrace{\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}}_{\mathbf{Q}} \cdot \underbrace{\begin{pmatrix} 4 & \\ & -1 \end{pmatrix}}_{\boldsymbol{\Lambda}} \cdot \underbrace{\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}^T}_{\mathbf{Q}^T} \\ &= 4 \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} + (-1) \begin{pmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix} \end{aligned}$$

The spectral theorem is a fundamental result in linear algebra and has many applications. We will use it to develop the singular value decomposition of general matrices later in this book. Here we briefly mention its use in defining quadratic forms.

1.6 Quadratic forms

An important way of using symmetric matrices is to define the so-called quadratic forms. Given a matrix $\mathbf{A} \in S^n(\mathbb{R})$, the *quadratic form* based on $\mathbf{A}$ is a function $Q : \mathbb{R}^n \mapsto \mathbb{R}$ with

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}, \qquad \text{for all} \quad \mathbf{x} \in \mathbb{R}^n \ . \tag{1.115}$$

Remark 1.4 When given an expression $\mathbf{x}^T \mathbf{A} \mathbf{x}$ where $\mathbf{A}$ is not symmetric, we can rewrite it into the standard form in the definition as follows:

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \frac{1}{2} \left(\mathbf{x}^T \mathbf{A} \mathbf{x} + (\mathbf{x}^T \mathbf{A} \mathbf{x})^T \right) = \frac{1}{2} \left(\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{A}^T \mathbf{x} \right) = \mathbf{x}^T \cdot \underbrace{\frac{1}{2} (\mathbf{A} + \mathbf{A}^T)}_{\text{symmetric}} \cdot \mathbf{x}. \tag{1.116}$$

Therefore, we only need to focus on symmetric matrices when studying quadratic forms.

A quadratic form is a second order polynomial (with no linear or constant term):

$$\mathbf{x}^T \mathbf{A}\mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \ . \tag{1.117}$$

Because of the symmetry assumption on the matrix $\mathbf{A}$, we may also write it as

$$\mathbf{x}^T \mathbf{A}\mathbf{x} = \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{1 \le i < j \le n} 2a_{ij} x_i x_j \ . \tag{1.118}$$

This equation is very useful for getting back and forth between the vector form and the expanded polynomial. For example, if

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}, \ \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

then

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x} = x_1^2 + 2x_2^2 + 6x_1 x_2 \ .$$

Conversely, if

$$Q(\mathbf{x}) = x_1^2 + 2x_2^2 + 3x_3^2 + 6x_1 x_2 - 4x_1 x_3 + 10x_2 x_3$$

then

$$a_{11} = 1, \quad a_{22} = 2, \quad a_{33} = 3,$$
$$a_{12} = a_{21} = 6/2 = 3, \quad a_{13} = a_{31} = -4/2 = -2, \quad a_{23} = a_{32} = 10/2 = 5.$$

Putting them together, we get that

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & -2 \\ 3 & 2 & 5 \\ -2 & 5 & 3 \end{pmatrix} .$$

One way of using quadratic forms is to define positive definiteness of square matrices. We present the concept in the next section.

1.7 Positive semidefinite matrices

A symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ is said to be *positive semidefinite* if the corresponding quadratic form is nonnegative everywhere in $\mathbb{R}^n$, that is,

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x} \ge 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{1.119}$$

We denote by $S^n_{0+}(\mathbb{R})$ the set of all $n \times n$ positive semidefinite matrices.

For any $\mathbf{A} \in S^n_{0+}(\mathbb{R})$, if the equality holds true only for $\mathbf{x} = \mathbf{0}$ (i.e., $\mathbf{x}^T \mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$), then $\mathbf{A}$ is said to be *positive definite*. The set of all $n \times n$ positive definite matrices is denoted as $S^n_+(\mathbb{R})$. Clearly, a positive definite matrix is always positive semidefinite, but no the other way. See Figure 1.15 for an illustration.

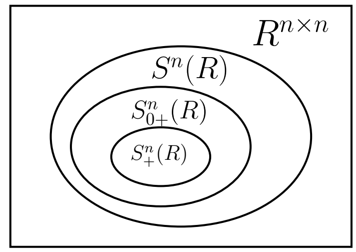

Fig. 1.15 Increasingly larger classes of matrices $S^n_+(\mathbb{R}) \subset S^n_{0+}(\mathbb{R}) \subset S^n(\mathbb{R}) \subset \mathbb{R}^{n\times n}$.

A symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ is said to be *negative definite* (resp. semidefinite) if $-\mathbf{A}$ is positive definite (resp. semidefinite). Similarly, we denote by $S^n_{0-}(\mathbb{R})$ and $S^n_-(\mathbb{R})$ the sets of negative semidefinite and of negative definite matrices, respectively. If a symmetric matrix is neither positive semidefinite nor negative semidefinite, then we say that it is *indefinite*. The set of indefinite matrices $\mathbf{A} \in S^n(\mathbb{R})$ can be expressed as $\left(S^n_{0+}(\mathbb{R}) \cup S^n_{0-}(\mathbb{R})\right)^c$.

The following theorem provides a way to check positive definiteness or semidefiniteness of a symmetric matrix via its eigenvalues.

Theorem 1.11 *A symmetric matrix is positive definite (resp. positive semidefinite) if and only if all of its eigenvalues are strictly positive (resp. nonnegative).*

Proof Consider a symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ and its spectral decomposition: $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where $\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_n] \in \mathbb{R}^{n\times n}$ is orthogonal and $\mathbf{Q} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. For any $\mathbf{x} \in \mathbb{R}^n$, let $\mathbf{y} = \mathbf{Q}^T\mathbf{x} \in \mathbb{R}^n$. Then

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{x} = \mathbf{y}^T\mathbf{\Lambda}\mathbf{y} = \sum_{i=1}^n \lambda_i y_i^2. \tag{1.120}$$

Clearly, if all $\lambda_i \geq 0$, then $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x}$ and thus $\mathbf{A}$ is positive semidefinite. In the special case that all $\lambda_i > 0$, $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ if and only if $\mathbf{y} = \mathbf{0}$ and correspondingly, $\mathbf{x} = \mathbf{0}$. The completes the proof of the sufficiency part.

To prove the necessary direction, we suppose that $\mathbf{A}$ is positive semidefinite but with a negative eigenvalue λ_j. Let $\mathbf{y} = \mathbf{e}_j$ and correspondingly, $\mathbf{x} = \mathbf{Q}\mathbf{y} = \mathbf{q}_j \neq \mathbf{0}$. It follows that

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{\Lambda}\mathbf{y} = \lambda_j < 0.$$

This shows that $\mathbf{A}$ is not positive semidefinite, thus a contraction. The proof is similar if $\mathbf{A}$ is positive definite but with an eigenvalue $\lambda_i \leq 0$. □

Remark 1.5 For any matrix $\mathbf{A} \in S^n(\mathbb{R})$, $(\lambda, \mathbf{v})$ is an eigenpair of $\mathbf{A}$ if and only if $(-\lambda, \mathbf{v})$ is an eigenpair of $-\mathbf{A}$:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \quad \Longleftrightarrow \quad (-\mathbf{A})\mathbf{v} = (-\lambda)\mathbf{v} \tag{1.121}$$

This shows that $\mathbf{A}$ and $-\mathbf{A}$ have completely opposite (in sign) eigenvalues (but the same eigenvectors). Thus, by the theorem, $\mathbf{A}$ is negative definite (resp. negative semidefinite) if and only if all of its eigenvalues are strictly negative (resp. nonpositive).

Remark 1.6 A matrix is indefinite if and only if it has both positive and negative eigenvalues. Another equivalent characterization of an indefinite matrix $\mathbf{A} \in S^n(\mathbb{R})$ is that it can be written as a sum of two nonzero symmetric matrices, one being positive semidefinite and the other being negative semidefinite (the proof is left to Problem):

$$\mathbf{A} = \mathbf{A}^+ + \mathbf{A}^-, \quad \mathbf{A}^+ \in S^n_{0+}(\mathbb{R}) - \{\mathbf{O}\},\ \mathbf{A}^- \in S^n_{0-}(\mathbb{R}) - \{\mathbf{O}\} \tag{1.122}$$

We call $\mathbf{A}^+, \mathbf{A}^-$ respectively the positive and negative semidefinite components of $\mathbf{A}$.

This theorem implies the following pattern for eigenvalues of a positive semidefinite matrix $\mathbf{A}$:

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n, \tag{1.123}$$

where $r = \text{rank}(\mathbf{A})$. Correspondingly, we may obtain the following reduced form of the spectral decomposition in (1.113):

$$\mathbf{A} = \sum_{i=1}^{r} \lambda_i \mathbf{q}_i \mathbf{q}_i^T = \begin{bmatrix} \mathbf{q}_1 & \dots & \mathbf{q}_r \end{bmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_r \end{bmatrix} = \mathbf{Q}_r \mathbf{\Lambda}_r \mathbf{Q}_r^T \tag{1.124}$$

where $\mathbf{Q}_r = \begin{bmatrix} \mathbf{q}_1 & \dots & \mathbf{q}_r \end{bmatrix} \in \mathbb{R}^{n\times r}$ is a tall matrix with orthonormal columns, and $\mathbf{\Lambda}_r = \text{diag}(\lambda_1, \dots, \lambda_r) \in \mathbb{R}^{r\times r}$.

Using Theorem 1.11 together with the following two identities:

$$\det(\mathbf{A}) = \lambda_1\lambda_2 \qquad \text{and} \qquad \text{tr}(\mathbf{A}) = \lambda_1 + \lambda_2$$

where λ_1, λ_2 are the two eigenvalues of $\mathbf{A}$, we present in Table 1.1 a simple, fast test based on the determinant and trace to determine the positive definiteness of 2×2 matrices. Note, however, that the test is not valid for larger matrices.

det($\mathbf{A}$)	tr($\mathbf{A}$)	positive definiteness	sufficient/necessary
+	+	$\mathbf{A} \in S^2_+(\mathbb{R})$	both
0	0 or +	$\mathbf{A} \in S^2_{0+}(\mathbb{R})$	only sufficient
+	-	$\mathbf{A} \in S^2_-(\mathbb{R})$	both
0	0 or -	$\mathbf{A} \in S^2_{0-}(\mathbb{R})$	only sufficient
-	any	$\mathbf{A} \notin S^2_{0+}(\mathbb{R}) \cup S^2_{0-}(\mathbb{R})$	only sufficient

Table 1.1 Positive definiteness test for a 2×2 symmetric matrix $\mathbf{A}$ based on its rank and determinant

Example 1.20 The following are all symmetric matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}.$$

According to the test in Table 1.1,

- $\mathbf{A}$ is positive definite, because $\det(\mathbf{A}) = 1 > 0$ and $\text{tr}(\mathbf{A}) = 6$. The matrix has two positive eigenvalues $\lambda_{1,2} = 3 \pm 2\sqrt{2}$;
- $\mathbf{B}$ is positive semidefinite (but not positive definite), because $\det(\mathbf{A}) = 0$ and $\text{tr}(\mathbf{A}) = 5 > 0$. The two eigenvalues are $\lambda_1 = 5, \lambda_2 = 0$;
- $\mathbf{C}$ is indefinite, because $\det(\mathbf{A}) = -7 < 0$. The matrix has both positive and negative eigenvalues $\lambda_{1,2} = \frac{3\pm\sqrt{37}}{2}$.

The following result is also an simple one but will be needed later when deriving the matrix singular value decomposition.

Theorem 1.12 *For any rectangular matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$*, both* $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ *and* $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ *are square, symmetric, and positive semidefinite.*

Proof It is obvious that $\mathbf{A}^T\mathbf{A}$ is square ($n \times n$) and symmetric:

$$(\mathbf{A}^T\mathbf{A})^T = \mathbf{A}^T(\mathbf{A}^T)^T = \mathbf{A}^T\mathbf{A}. \tag{1.125}$$

To show that it is positive semidefinite, consider the associated quadratic form for any $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x}^T(\mathbf{A}^T\mathbf{A})\mathbf{x} = (\mathbf{x}^T\mathbf{A}^T)(\mathbf{A}\mathbf{x}) = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2 \geq 0. \tag{1.126}$$

The proof for the other product $\mathbf{A}\mathbf{A}^T$ is similar. □

Remark 1.7 Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. We discuss when the two positive semidefinite matrices, $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ and $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$, are positive definite, depending on the shape of the matrix $\mathbf{A}$:

- $m > n$ (tall matrix): $\mathbf{A}^T\mathbf{A}$ is positive definite if and only if $\text{rank}(\mathbf{A}^T\mathbf{A}) = n$, which is further equivalent to $\text{rank}(\mathbf{A}) = n$ (i.e., $\mathbf{A}$ is of full column rank). In contrast, $\mathbf{A}\mathbf{A}^T$ is never positive definite because $\text{rank}(\mathbf{A}\mathbf{A}^T) = \text{rank}(\mathbf{A}) \leq n < m$.
- $m < n$ (long matrix): Similarly, $\mathbf{A}\mathbf{A}^T$ is positive definite if and only if $\mathbf{A}$ is of full row rank (i.e., $\text{rank}(\mathbf{A}) = m$). In contrast, $\mathbf{A}^T\mathbf{A}$ is never positive definite because $\text{rank}(\mathbf{A}^T\mathbf{A}) = \text{rank}(\mathbf{A}) \leq m < n$.
- $m = n$ (square matrix): $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are (simultaneously) positive definite if and only if $\mathbf{A}$ is nonsingular.

Remark 1.8 A notable result about the two product matrices, i.e., $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ for any given matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, is that they must have the same positive eigenvalues (with the same algebraic multiplicities), due to the following result:

$$\lambda^n \det(\lambda\mathbf{I} - \mathbf{A}\mathbf{A}^T) = \lambda^m \det(\lambda\mathbf{I} - \mathbf{A}^T\mathbf{A}). \tag{1.127}$$

Regarding the number zero as an eigenvalue, all of the following cases can occur:

1. Exactly one of the two matrices $\mathbf{A}\mathbf{A}^T, \mathbf{A}^T\mathbf{A}$ has a zero eigenvalue.
2. Both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have a zero eigenvalue.
3. None of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ has a zero eigenvalue.

In Problem [cite] we ask the reader to give concrete matrix examples in each case.

1.8 Matrix square roots

An interesting aspect of positive semidefinite matrices is that they have square roots (which are also matrices), just like nonnegative numbers have square roots (which are still numbers).

Definition 1.1 Let $\mathbf{A}, \mathbf{R} \in S^n_{0+}(\mathbb{R})$ be two positive semidefinite matrices of the same size. The matrix $\mathbf{R}$ is called the *square root* of $\mathbf{A}$ if $\mathbf{A} = \mathbf{R}^2$. We denote it by $\mathbf{R} = \mathbf{A}^{1/2}$.

Remark 1.9 If $\mathbf{A}$ is actually strictly positive definite, then its square root $\mathbf{R}$ must be strictly positive definite because

$$0 \neq \det(\mathbf{A}) = \det(\mathbf{R}^2) = \det(\mathbf{R})^2 \quad \longrightarrow \quad \det(\mathbf{R}) \neq 0.$$

If a positive semidefinite matrix happens to be diagonal, then there is an easy way to find its square root. Assume a diagonal, positive semidefinite matrix

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \qquad \text{where} \quad \lambda_1, \ldots, \lambda_n \geq 0.$$

Define

$$\mathbf{\Lambda}^{1/2} = \mathrm{diag}\left(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2}\right), \tag{1.128}$$

which is still diagonal and positive semidefinite. Clearly, $\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2} = \mathbf{\Lambda}$. This shows that $\mathbf{\Lambda}^{1/2}$ is indeed a square root of $\mathbf{\Lambda}$. Note that without the positive semidefiniteness requirement in the definition of matrix square roots, we can arbitrarily modify the signs of the diagonals of $\mathbf{\Lambda}^{1/2}$ without violating the equality condition.

The following theorem presents a general formula for finding the square root of any positive semidefinite matrix.

Theorem 1.13 *Let* $\mathbf{A} \in S^n_+(\mathbb{R})$ *be an arbitrary positive semidefinite matrix with spectral decomposition* $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$*, where* $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *is an orthogonal matrix and* $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ *a diagonal matrix with* $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$*. Then* $\mathbf{A}$ *has a unique matrix square root*

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T. \tag{1.129}$$

Proof First, such defined matrix $\mathbf{R}$ is also positive semidefinite. By direct calculation,

$$\mathbf{R}^2 = (\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T)(\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T) = \mathbf{Q}\underbrace{\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}}_{\mathbf{\Lambda}}\mathbf{Q}^T = \mathbf{A}.$$

This shows that $\mathbf{R}$ is a square root of $\mathbf{A}$. It remains to show that it is the only square root of $\mathbf{A}$.

Suppose that a positive semidefinite matrix $\mathbf{T} = \mathbf{PSP}^T$, for some orthogonal matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ and diagonal matrix $\mathbf{S} = \text{diag}(s_1, \ldots, s_n)$ with $s_1 \geq \cdots \geq s_n \geq 0$, is another square root of $\mathbf{A}$. That is,

$$\mathbf{A} = \mathbf{T}^2 = \mathbf{PS}^2\mathbf{P}^T.$$

Observe that $\mathbf{S}^2$ must contain the eigenvalues of $\mathbf{A}$ sorted in decreasing order. Therefore, we conclude that $\mathbf{S}^2 = \mathbf{\Lambda}$, or equivalently, $\mathbf{S} = \mathbf{\Lambda}^{1/2}$. It follows that

$$\mathbf{Q\Lambda Q}^T = \mathbf{R}^2 = \mathbf{A} = \mathbf{T}^2 = \mathbf{P\Lambda P}^T,$$

and further that

$$\mathbf{P}^T\mathbf{Q\Lambda} = \mathbf{\Lambda P}^T\mathbf{Q}.$$

Let $\mathbf{U} = \mathbf{P}^T\mathbf{Q}$, which is also an orthogonal matrix. Then the last equation can be rewritten as

$$\mathbf{U\Lambda} = \mathbf{\Lambda U}, \qquad \text{or} \quad \text{in entrywise form, } u_{ij}\lambda_j = \lambda_i u_{ij}, \text{ for all } i, j.$$

This shows that $u_{ij} = 0$ whenever $\lambda_i \neq \lambda_j$. It follows that

$$u_{ij}\lambda_j^{1/2} = \lambda_i^{1/2}u_{ij}, \text{ for all } i, j, \qquad \text{or} \quad \text{in matrix form, } \mathbf{U\Lambda}^{1/2} = \mathbf{\Lambda}^{1/2}\mathbf{U}.$$

Consequently, using $\mathbf{Q} = \mathbf{PU}$, we get that

$$\begin{aligned} \mathbf{R} &= (\mathbf{PU})\mathbf{\Lambda}^{1/2}(\mathbf{PU})^T \\ &= \mathbf{P}(\mathbf{U\Lambda}^{1/2})\mathbf{U}^T\mathbf{P}^T \\ &= \mathbf{P}(\mathbf{\Lambda}^{1/2}\mathbf{U})\mathbf{U}^T\mathbf{P}^T \\ &= \mathbf{P\Lambda}^{1/2}\mathbf{P}^T \\ &= \mathbf{T}. \end{aligned}$$

This thus completes the proof of uniqueness of matrix square roots. □

Remark 1.10 For a positive definite matrix $\mathbf{A} \in S_+^n(\mathbb{R})$, we can also define the reciprocal square root $\mathbf{A}^{-1/2}$ (just like $x^{-1/2}$ for a positive real number x):

$$\mathbf{A}^{-1/2} = \left(\mathbf{A}^{1/2}\right)^{-1}. \tag{1.130}$$

Let $\mathbf{A} = \mathbf{Q\Lambda Q}^T$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \cdots \geq \lambda_n > 0$. Then we have

$$\mathbf{A}^{-1/2} = \mathbf{Q\Lambda}^{-1/2}\mathbf{Q}^T, \qquad \text{where} \quad \mathbf{\Lambda}^{-1/2} = \text{diag}\left(\lambda_1^{-1/2}, \ldots, \lambda_n^{-1/2}\right). \tag{1.131}$$

Clearly, like $\mathbf{A}$ and $\mathbf{A}^{1/2}$, $\mathbf{A}^{-1/2}$ is also a positive definite matrix.

Remark 1.11 Using the reduced form of the eigendecomposition of **A** in (1.124), we obtain the following reduced form for the square root of a positive semidefinite matrix **A**:

$$\mathbf{A} = \mathbf{Q}_r \mathbf{\Lambda}_r \mathbf{Q}_r^T \quad \longrightarrow \quad \mathbf{A}^{1/2} = \mathbf{Q}_r \mathbf{\Lambda}_r^{1/2} \mathbf{Q}_r^T. \tag{1.132}$$

This formula is more efficient for computing the matrix square roots, as it only requires computing the eigenvectors corresponding to the positive eigenvalues.

Example 1.21 Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, which is positive semidefinite because $\lambda_1 = 5, \lambda_2 = 0$. To find the matrix square root of **A**, we need to first find its orthogonal diagonalization (in reduced form):

$$\mathbf{A} = \mathbf{Q}_1 \mathbf{\Lambda}_1 \mathbf{Q}_1^T = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} (5) \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}.$$

It follows that

$$\mathbf{A} = \mathbf{Q}_1 \mathbf{\Lambda}_1^{1/2} \mathbf{Q}_1^T = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \left(\sqrt{5}\right) \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \end{pmatrix}.$$

The reader is asked to verify that the square of this matrix is **A** that is given at the beginning of the question.

Example 1.22 Let $\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. It can be directly verified that $\mathbf{B}^2 = \mathbf{A}$. Thus, **B** is the unique matrix square root of **A**. Moreover, since **A** is positive definite, its reciprocal square root exists and is

$$\mathbf{A}^{-1/2} = \mathbf{B}^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

1.9 The generalized eigenvalue problem

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ be two square matrices of the same size. We say that $\lambda \in \mathbb{R}$ is a *generalized eigenvalue* of $(\mathbf{A}, \mathbf{B})$ if there exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ such that

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v}. \tag{1.133}$$

In this case, **v** is called a *generalized eigenvector* of $(\mathbf{A}, \mathbf{B})$ corresponding to λ. In some cases, we say for short that $(\lambda, \mathbf{v})$ is *generalized eigenpair* of $(\mathbf{A}, \mathbf{B})$.

In the above definition, if we let $\mathbf{B} = \mathbf{I}$, then the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$ would reduce to the ordinary eigenvalues of **A**:

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

This shows that the generalized eigenvalue problem is indeed a generalization of the ordinary eigenvalue problem.

We mention two other special cases of the generalized eigenvalue problem (1.133), each of which also reduces to an ordinary eigenvalue problem:

- If the matrix $\mathbf{B}$ happens to be nonsingular, then (1.133) can be rewritten as

$$(\mathbf{B}^{-1}\mathbf{A})\mathbf{v} = \lambda \mathbf{v}. \tag{1.134}$$

  This shows that the generalized eigenvalue λ of $(\mathbf{A}, \mathbf{B})$ is an ordinary eigenvalue of the matrix $\mathbf{B}^{-1}\mathbf{A}$, with corresponding eigenvector.
- If the matrix $\mathbf{A}$ is nonsingular, since $\mathbf{v} \neq \mathbf{0}$, we must have $\lambda \neq 0$ (otherwise we would have $\mathbf{A}\mathbf{v} = \lambda\mathbf{B}\mathbf{v} = \mathbf{0}$, which is a contradiction). In this case, (1.133) can be rewritten as

$$(\mathbf{A}^{-1}\mathbf{B})\mathbf{v} = (1/\lambda)\mathbf{v}. \tag{1.135}$$

  This shows that $1/\lambda$ is an ordinary eigenvalue for the matrix $\mathbf{A}^{-1}\mathbf{B}$, with corresponding eigenvector $\mathbf{v}$.

Now, let us rewrite (1.133) as

$$(\mathbf{A} - \lambda\mathbf{B})\mathbf{v} = 0. \tag{1.136}$$

Note that there exists a nonzero solution $\mathbf{v}$ if and only if $\mathbf{A} - \lambda\mathbf{B}$ is singular. We have thus obtained that λ is a generalized eigenvalue of $(\mathbf{A}, \mathbf{B})$ if and only if

$$\det(\mathbf{A} - \lambda\mathbf{B}) = 0. \tag{1.137}$$

Let $p_{\mathbf{A},\mathbf{B}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{B})$, which is also a polynomial in λ. We call $p_{\mathbf{A},\mathbf{B}}(\lambda)$ the characteristic polynomial of $(\mathbf{A}, \mathbf{B})$.

We discuss the number of (real) generalized eigenvalues that two $n \times n$ matrices $(\mathbf{A}, \mathbf{B})$ may have in general. Recall that for the square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ alone, the characteristic polynomial $p_{\mathbf{A}}(\lambda)$ must be of order-n and thus may have no more than n real roots. Accordingly, an $n \times n$ matrix may have zero to n eigenvalues. However, for a pair of $n \times n$ matrices $(\mathbf{A}, \mathbf{B})$, the characteristic polynomial $p_{\mathbf{A},\mathbf{B}}(\lambda)$ may have an order less than n. In the special case when the order is zero (i.e., $p_{\mathbf{A},\mathbf{B}}(\lambda) = 0$ for all $\lambda \in \mathbb{R}$), the matrix pair have infinitely many generalized eigenvalues!

Example 1.23 Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

To find the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$, compute

$$\det(\mathbf{A} - \lambda\mathbf{B}) = \begin{vmatrix} 1-\lambda & 2-\lambda \\ 2-\lambda & 4-\lambda \end{vmatrix} = (1-\lambda)(4-\lambda) - (2-\lambda)^2 = -\lambda.$$

Thus, $(\mathbf{A}, \mathbf{B})$ has a generalized eigenvalue of $\lambda = 0$, with corresponding generalized eigenvectors

$$\mathbf{0} = (\mathbf{A} - 0 \cdot \mathbf{B})\mathbf{v} = \mathbf{A}\mathbf{v} \qquad \longrightarrow \qquad \mathbf{v} = k \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \ k \in \mathbb{R}.$$

Example 1.24 Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}.$$

To find the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$, compute

$$\det(\mathbf{A} - \lambda \mathbf{B}) = \begin{vmatrix} 1-\lambda & 2-2\lambda \\ 2-3\lambda & 4-6\lambda \end{vmatrix} = (1-\lambda)(4-6\lambda) - (2-2\lambda)(2-3\lambda) = 0.$$

Thus, any scalar λ is a generalized eigenvalue of $(\mathbf{A}, \mathbf{B})$. This pair of matrices has infinitely many generalized eigenvalues!

For an arbitrary generalized eigenvalue $\lambda \in \mathbb{R}$, we find its corresponding generalized eigenvector as follows:

$$\mathbf{0} = (\mathbf{A} - \lambda \cdot \mathbf{B})\mathbf{v} = \begin{pmatrix} 1-\lambda & 2-2\lambda \\ 2-3\lambda & 4-6\lambda \end{pmatrix} \mathbf{v} \qquad \longrightarrow \qquad \mathbf{v} = k \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \ k \in \mathbb{R}.$$

This indicates that all the generalized eigenvalues share the same generalized eigenvector!

Lastly, we mention an important class of generalized eigenvalue problems where $\mathbf{A} \in S^n(\mathbb{R})$ is symmetric and $\mathbf{B} \in S^n_+(\mathbb{R})$ is positive definite. Such problems, called *generalized symmetric-definite eigenvalue* problems, have very nice properties and also occur a lot in applications.

Theorem 1.14 *For any two matrices* $\mathbf{A} \in S^n(\mathbb{R})$ *and* $\mathbf{B} \in S^n_+(\mathbb{R})$*, the generalized eigenvalue problem* $\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v}$ *has n real eigenvalues* $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ *with linearly independent generalized eigenvectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ *which can be normalized such that*

$$\mathbf{v}_i^T \mathbf{B} \mathbf{v}_j = \delta_{ij}, \quad \textit{for all } 1 \le i, j \le n.$$

Proof For any such given matrices $\mathbf{A}, \mathbf{B}$, consider the equation $\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v}$ where $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$ are the unknowns. Since $\mathbf{B}$ is positive definite, there exists a unique square root $\mathbf{B}^{1/2}$ which is a positive definite matrix such that $\mathbf{B}^{1/2} \cdot \mathbf{B}^{1/2} = \mathbf{B}$. Let $\mathbf{u} = \mathbf{B}^{1/2}\mathbf{v}$. Then $\mathbf{v} = \mathbf{B}^{-1/2}\mathbf{u}$, where $\mathbf{B}^{-1/2} = (\mathbf{B}^{1/2})^{-1}$ is also positive definite. Rewrite the equation in $\mathbf{u}$ to get that

$$\mathbf{A}(\mathbf{B}^{-1/2}\mathbf{u}) = \lambda \mathbf{B}(\mathbf{B}^{-1/2}\mathbf{u}) = \lambda \mathbf{B}^{1/2}\mathbf{u}$$

and further that

$$(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2})\mathbf{u} = \lambda \mathbf{u}.$$

This shows that $(\lambda, \mathbf{u})$ must be an eigenpair for $\widetilde{\mathbf{A}} = \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} \in \mathbb{R}^{n \times n}$, which is a symmetric matrix. By the Spectral Theorem, there exist n orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^n$, corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, such that

$$\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad i = 1, \ldots, n$$

or equivalently,

$$\mathbf{A}(\mathbf{B}^{-1/2}\mathbf{u}_i) = \lambda_i \mathbf{B}(\mathbf{B}^{-1/2}\mathbf{u}_i), \quad i = 1, \ldots, n$$

This shows that $\lambda_1, \ldots, \lambda_n$ are the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$ with corresponding generalized eigenvectors

$$\mathbf{v}_i = \mathbf{B}^{-1/2}\mathbf{u}_i, \quad \text{for } i = 1, \ldots, n \tag{1.138}$$

Lastly, since $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are orthonormal vectors, the generalized eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ must be linearly independent (because they only differ from the $\mathbf{u}_i$'s by a nonsingular matrix factor $\mathbf{B}^{-1/2}$). Additionally, we have

$$\delta_{ij} = \mathbf{u}_i^T \mathbf{u}_j = (\mathbf{B}^{1/2}\mathbf{v}_i)^T (\mathbf{B}^{1/2}\mathbf{v}_j) = \mathbf{v}_i^T \mathbf{B} \mathbf{v}_j, \quad \text{for all } i, j.$$

This thus completes the proof. □

The proof of the theorem implies that the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$, where $\mathbf{A} \in S^n(\mathbb{R})$ and $\mathbf{B} \in S^n_+(\mathbb{R})$, are the same as the eigenvalues of the symmetric matrix $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} \in S^n(\mathbb{R})$.

Since $\mathbf{B}$ is invertible, we may rewrite the generalized symmetric-definite eigenvalue problem as an ordinary eigenvalue problem

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v} \quad \Longleftrightarrow \quad \mathbf{B}^{-1}\mathbf{A}\mathbf{v} = \lambda \mathbf{v}. \tag{1.139}$$

This shows that the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$ are also the eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$.

There are thus two important matrices involved in the generalized symmetric-definite eigenvalue problem: $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ and $\mathbf{B}^{-1}\mathbf{A}$. The former is symmetric but the latter is not in general. It turns out that the two matrices are similar to each other (this is why their eigenvalues are both the same as the generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$):

$$\mathbf{B}^{-1}\mathbf{A} = \mathbf{B}^{-1/2} \cdot (\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}) \cdot \mathbf{B}^{1/2} \tag{1.140}$$

In sum, in a generalized symmetric-definite eigenvalue problem $(\mathbf{A}, \mathbf{B})$, the following three are always the same:

- Generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$;
- Eigenvalues of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ (which is a symmetric matrix);
- Eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$;

Note that regarding eigenvectors, we only have that the generalized eigenvectors of $(\mathbf{A}, \mathbf{B})$ are the same as the eigenvectors $\mathbf{B}^{-1}\mathbf{A}$, as indicated by (1.139). However, as shown in (1.138), the generalized eigenvectors of $(\mathbf{A}, \mathbf{B})$ differ from the eigenvectors of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ by a fixed matrix factor.

There is also a direct way to relate the eigenvectors of $\mathbf{B}^{-1}\mathbf{A}$ and the eigenvectors of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$:

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \quad \Longleftrightarrow \quad (\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2})(\mathbf{B}^{1/2}\mathbf{v}) = \lambda(\mathbf{B}^{1/2}\mathbf{v}). \tag{1.141}$$

This equation shows that $\mathbf{v}$ is an eigenvector of $\mathbf{B}^{-1}\mathbf{A}$ (and also a generalized eigenvector of $(\mathbf{A},\mathbf{B})$) if and only if $\mathbf{B}^{1/2}\mathbf{v}$ is an eigenvector of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$.

We derive next the matrix representation of the result in the preceding theorem, which can provide more insights. Let $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ and $\mathbf{V} = [\mathbf{v}_1,\ldots,\mathbf{v}_n]$, which contain the generalized eigenvalues and eigenvectors of $(\mathbf{A},\mathbf{B})$, respectively. Then

$$\mathbf{v}_i^T\mathbf{B}\mathbf{v}_j = \delta_{ij} \qquad \longrightarrow \qquad \mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{I} \tag{1.142}$$

and

$$\mathbf{A}[\mathbf{v}_1,\ldots,\mathbf{v}_n] = [\mathbf{B}\mathbf{v}_1,\ldots,\mathbf{B}\mathbf{v}_n]\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix} \qquad \longrightarrow \qquad \mathbf{A}\mathbf{V} = \mathbf{B}\mathbf{V}\mathbf{\Lambda}. \tag{1.143}$$

Furthermore, we may obtain that

$$\mathbf{V}^T\mathbf{A}\mathbf{V} = \mathbf{V}^T(\mathbf{A}\mathbf{V}) = \mathbf{V}^T(\mathbf{B}\mathbf{V}\mathbf{\Lambda}) = (\mathbf{V}^T\mathbf{B}\mathbf{V})\mathbf{\Lambda} = \mathbf{I}\mathbf{\Lambda} = \mathbf{\Lambda}. \tag{1.144}$$

This indicates that the generalized eigenvectors matrix $\mathbf{V}$ simultaneously diagonalizes both $\mathbf{A}$ and $\mathbf{B}$, but for different purposes:

- Diagonalization of $\mathbf{A}$ is for the derivation of the generalized eigenvalues in $\mathbf{\Lambda}$;
- Diagonalization of $\mathbf{B}$ is for the normalization of the generalized eigenvectors in $\mathbf{V}$.

We summarize these matrix representations in the following theorem.

Theorem 1.15 *For any symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ and positive definite matrix $\mathbf{B} \in S_+^n(\mathbb{R})$, there exist a diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ and an invertible matrix $\mathbf{V} \in \mathbb{R}^{n\times n}$, such that*

$$\mathbf{V}^T\mathbf{A}\mathbf{V} = \mathbf{\Lambda}, \quad \mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{I}, \qquad \text{and} \quad \mathbf{A}\mathbf{V} = \mathbf{B}\mathbf{V}\mathbf{\Lambda}. \tag{1.145}$$

In particular, $\mathbf{\Lambda}$ and $\mathbf{V}$ contain the generalized eigenvalues and eigenvectors of $(\mathbf{A},\mathbf{B})$, respectively.

Remark 1.12 The generalized eigenvalue problems can be easily solved in MATLAB:

- $E = eig(A,B)$ produces a column vector E containing the generalized eigenvalues of square matrices A and B.
- $[V,D] = eig(A,B)$ produces a diagonal matrix D of generalized eigenvalues and a full matrix V whose columns are the corresponding eigenvectors.

Example 1.25 Let $\mathbf{A} = \begin{pmatrix}2 & 3\\ 3 & 2\end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix}2 & 3\\ 3 & 5\end{pmatrix}$, the former being symmetric and the latter positive definite. Since $\mathbf{A},\mathbf{B}$ are only of size 2×2, we find their generalized eigenvalues and eigenvectors are ordinary eigenvalues and eigenvectors of

$$\mathbf{B}^{-1}\mathbf{A} = \begin{pmatrix}5 & -3\\ -3 & 2\end{pmatrix}\begin{pmatrix}2 & 3\\ 3 & 2\end{pmatrix} = \begin{pmatrix}1 & 9\\ 0 & -5\end{pmatrix}.$$

By direct calculation, this matrix has two distinct eigenvalues and corresponding eigenvectors:

$$\lambda_1 = 1,\ \mathbf{v}_1 = \begin{pmatrix}1\\0\end{pmatrix}, \quad \lambda_2 = -5,\ \mathbf{v}_2 = \begin{pmatrix}3\\-2\end{pmatrix}.$$

Letting

$$\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2] = \begin{pmatrix}1 & 3\\0 & -2\end{pmatrix}, \qquad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix}1 & \\ & -5\end{pmatrix}$$

we must have $\mathbf{AV} = \mathbf{BV\Lambda}$. However, the arbitrary scaling used in $\mathbf{v}_1, \mathbf{v}_2$ do not guarantee $\mathbf{V}^T\mathbf{BV} = \mathbf{I}$ (even if we had normalized $\mathbf{v}_2$ to have unit length). We examine

$$\mathbf{V}^T\mathbf{BV} = \begin{pmatrix}1 & 3\\0 & -2\end{pmatrix}^T \begin{pmatrix}2 & 3\\3 & 5\end{pmatrix} \begin{pmatrix}1 & 3\\0 & -2\end{pmatrix} = \begin{pmatrix}2 & \\ & 2\end{pmatrix} = 2\mathbf{I}.$$

An easy fix is to normalize $\mathbf{V}$ by $\frac{1}{\sqrt{2}}$ to have

$$\widetilde{\mathbf{V}} = \frac{1}{\sqrt{2}}\mathbf{V} = \begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{3}{\sqrt{2}}\\ 0 & -\sqrt{2}\end{pmatrix}.$$

The reader can verify that with such a choice of generalized eigenvectors, we now have $\widetilde{\mathbf{V}}^T\mathbf{B}\widetilde{\mathbf{V}} = \mathbf{I}$, besides $\mathbf{A}\widetilde{\mathbf{V}} = \mathbf{B}\widetilde{\mathbf{V}}\mathbf{\Lambda}$, and thus also $\widetilde{\mathbf{V}}^T\mathbf{A}\widetilde{\mathbf{V}} = \mathbf{\Lambda}$.

Problems

1. Let $\mathbf{A} = \mathbf{u}\mathbf{v}^T$, where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Show that

$$\mathrm{tr}(\mathbf{A}) = \mathbf{u} \cdot \mathbf{v}. \tag{1.146}$$

2. Show that for any two matrices of the same size, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$,

$$\mathrm{rank}(\mathbf{A} + \mathbf{B}) \le \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B}). \tag{1.147}$$

3. Give an example of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ in each of the following cases:
   a. Exactly one of the two matrices $\mathbf{A}\mathbf{A}^T, \mathbf{A}^T\mathbf{A}$ has a zero eigenvalue.
   b. Both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have a zero eigenvalue.
   c. None of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ has a zero eigenvalue.
4. Show that $\mathbf{A} \in \mathbb{R}^{n\times n}$ is idempotent if and only if $\mathbf{I} - \mathbf{A}$ is idempotent.
5. Show that $\mathbf{J}_n$ is positive semidefinite and then find its matrix square root.
6. Let a symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ be indefinite. Show that there exist two nonzero matrices, $\mathbf{A}^+ \in S^n_{0+}(\mathbb{R}) - \{\mathbf{O}\}, \mathbf{A}^- \in S^n_-(\mathbb{R}) - \{\mathbf{O}\}$ such that

$$\mathbf{A} = \mathbf{A}^+ + \mathbf{A}^-. \tag{1.148}$$

Hint: consider the spectral decomposition of $\mathbf{A}$ *in rank-1 form.*

7. In Example 1.25 we computed the generalized eigenvalues and eigenvectors of $(\mathbf{A}, \mathbf{B})$ as the ordinary eigenvalues and eigenvectors of $\mathbf{B}^{-1}\mathbf{A}$. We had to rescale the eigenvectors in some way to satisfy $\mathbf{VBV} = \mathbf{I}$. Here, you are asked to use the other matrix $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ and follow the procedure outlined in the proof of Theorem 1.14 to redo the problem. That is, first find the eigenvectors of $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ and normalize each of them to have unit length, and then obtain the generalized eigenvectors of $(\mathbf{A}, \mathbf{B})$ using (1.138). Such obtained generalized eigenvectors must automatically satisfy $\mathbf{VBV} = \mathbf{I}$ and thus require no further normalization.
8. Theorem 1.12 states that the product of any matrix with its transpose (in either order) is positive semidefinite. Here you are asked to prove something opposite. Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be a positive semidefinite matrix of rank r. Show that there exists a matrix $\mathbf{A} \in \mathbb{R}^{n \times r}$ such that $\mathbf{B} = \mathbf{A}\mathbf{A}^T$. *Hint: Consider the reduced form of the spectral decomposition of* $\mathbf{B}$.
9. Let
$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$
Show that $(\mathbf{A}, \mathbf{B})$ has no real generalized eigenvalues. Find also the corresponding generalized eigenvectors.
10. Let
$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}.$$
Show that $(\mathbf{A}, \mathbf{B})$ has two real generalized eigenvalues. Find also the corresponding generalized eigenvectors.