San José State University

Math 250: Mathematical Data Visualization

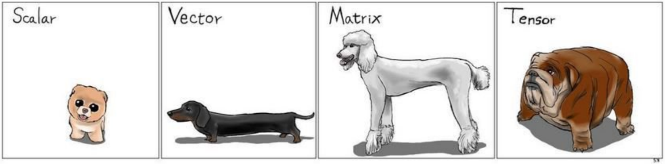
Matrix Algebra

Dr. Guangliang Chen

Introduction

This course is based on the following mathematical objects:

- **Vectors**: 1-D arrays;
- **Matrices**: 2-D arrays
- **Tensors**: 3-D arrays (or higher)



Notation: vectors

Vectors are denoted by **boldface** lowercase letters (such as $\mathbf{a}, \mathbf{b}, \mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}$):

$$\mathbf{a} = (1,2,3)^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$$

The ith element of a vector $\mathbf{a}$ is written as a_i or $\mathbf{a}(i)$.

For any $p \geq 1$, the ℓ_p **norm**, or simply p-**norm**, of a vector $\mathbf{a} \in \mathbb{R}^n$ is

$$\|\mathbf{a}\|_p = \left(\sum_{i=1}^n |a_i|^p \right)^{1/p}.$$

In the case of $p = 2$ (default), the norm is called the **Euclidean norm**.

Some special vectors:

- **The zero vector**: $\mathbf{0}_n = (0, 0, \ldots, 0)^T \in \mathbb{R}^n$
- **The vector of ones**: $\mathbf{1}_n = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$
- **The canonical basis vectors of $\mathbb{R}^n$**:

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

When the dimension of each of these vectors is not specified, it is implied by the context, e.g.,

$$\mathbf{a} \cdot \mathbf{1} \text{ (dot product)}, \qquad \text{where} \quad \mathbf{a} = (1, 2, 3)^T$$

Notation: matrices

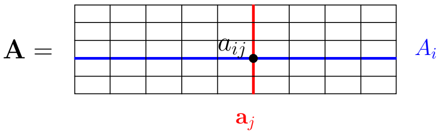
Matrices are denoted by **boldface** UPPERCASE letters (such as $\mathbf{A}, \mathbf{B}, \mathbf{U}, \mathbf{V}$).

We write $\mathbf{A} \in \mathbb{R}^{m\times n}$ to indicate that $\mathbf{A}$ has m rows and n columns.

The (i, j) entry of $\mathbf{A}$ is denoted by a_{ij}, or $\mathbf{A}(i, j)$, or $\mathbf{A}_{ij}$.

The ith row of $\mathbf{A}$ is denoted by A_i or $\mathbf{A}(i, :)$

The jth column is written as $\mathbf{a}_j$ or $\mathbf{A}(:, j)$.



Special matrices:

- **The zero matrix**: $\mathbf{O}_{m\times n} \in \mathbb{R}^{m\times n}$
- **The identity matrix**: $\mathbf{I}_n \in \mathbb{R}^{n\times n}$
- **The matrix of ones**: $\mathbf{J}_{m\times n} \in \mathbb{R}^{m\times n}$

Similarly, we may drop the subscripts when the size of the matrix is clear based on the context.

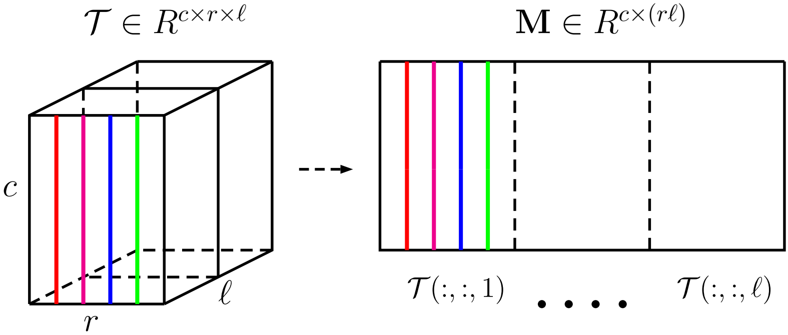
$$\mathbf{O} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{J} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Notation: tensors

Tensors are multidimensional arrays that generalize vectors and matrices.

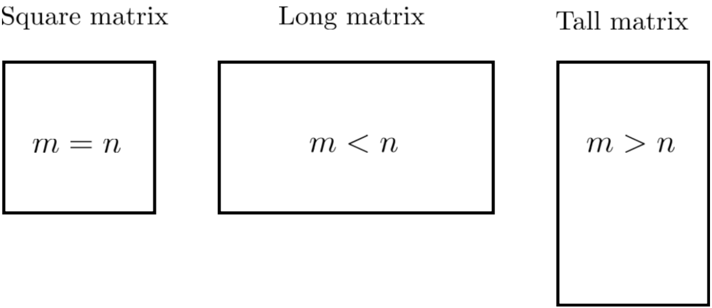
We use calligraphic UPPERCASE letters to denote them and write $\mathcal{T} \in \mathbb{R}^{c\times r\times l}$ to indicate the size of a 3D tensor $\mathcal{T}$.

Tensor algebra is a big, interesting filed on its own, but we will only use 3D tensors to store information for simple and efficient coding which requires knowing a little bit of how to unfold a 3D tensor to a matrix (and to assemble the tensor back).


$\mathcal{T} \in R^{c\times r\times \ell}$
c
r
ℓ
$\mathbf{M} \in R^{c\times (r\ell)}$
$\mathcal{T}(:,:,1)$
$\mathcal{T}(:,:,\ell)$

Descriptions of a matrix

One way to characterize a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is based on its shape:



A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to be positive/nonnegative if all of its entries are positive $(a_{ij} > 0,\ \forall i, j)$ / nonnegative $(a_{ij} \geq 0,\ \forall i, j)$.

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

A **diagonal** matrix is a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ whose off diagonal entries are all zero ($a_{ij} = 0$ for all $i \neq j$), e.g.,

$$\mathbf{A} = \begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix} = \mathrm{diag}(1, 2, 3)$$

Sometimes, a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is also said to be **diagonal** if $a_{ij} = 0$ for all $i \neq j$.

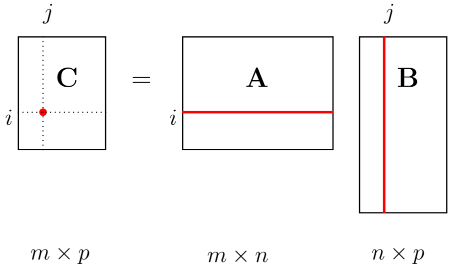
For example,

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

Matrix multiplication

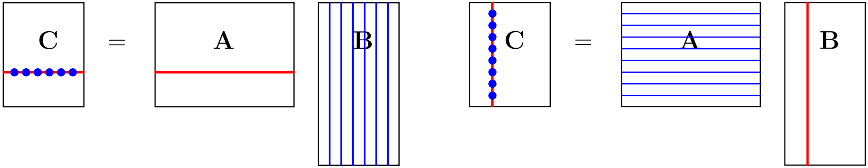
Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$. Their matrix product is an $m \times p$ matrix

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



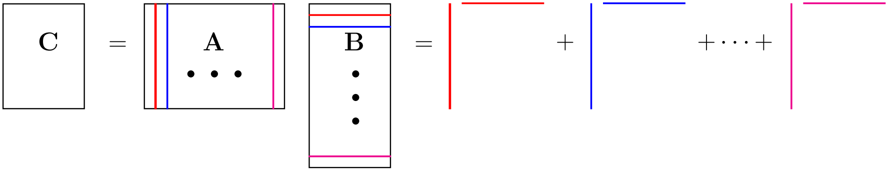
It is possible to obtain one full row (or column) of $\mathbf{C}$ at a time via matrix-vector multiplication:

$$C_i = A_i\mathbf{B}, \qquad \mathbf{c}_j = \mathbf{A}\mathbf{b}_j$$



The full matrix $\mathbf{C}$ can be written as a sum of rank-1 matrices:

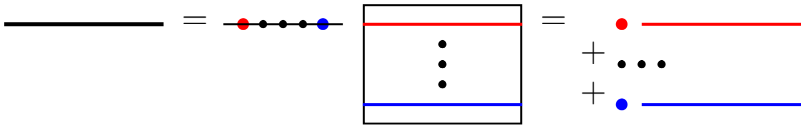
$$\mathbf{C} = \sum_{k=1}^{n} \mathbf{a}_k B_k.$$



Further interpretation of $\mathbf{AB}$ when one of the matrices is actually a vector:

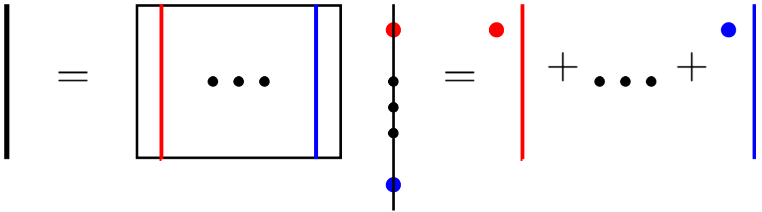
- If $\mathbf{A} = (a_1, \ldots, a_n)$ is a row vector ($m = 1$):

$$\mathbf{AB} = \sum_{i=1}^{n} a_i B_i \quad \longleftarrow \text{linear combination of rows of } \mathbf{B}$$



- If $\mathbf{B} = (b_1, \ldots, b_n)^T$ is a column vector ($p = 1$):

$$\mathbf{AB} = \sum_{j=1}^{n} b_j \mathbf{a}_j \quad \longleftarrow \text{linear combination of columns of } \mathbf{A}$$



Finally, below are some identities involving the vector $\mathbf{1} \in \mathbb{R}^n$:

$$\mathbf{1}^T\mathbf{1} = n, \qquad \mathbf{1}\mathbf{1}^T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} = \mathbf{J}_n$$

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

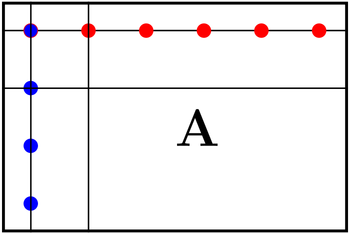
$$\mathbf{A}\mathbf{1} = \sum_j \mathbf{a}_j, \qquad \text{(vector of row sums)}$$

$$\mathbf{1}^T\mathbf{A} = \sum_i A_i, \qquad \text{(horizontal vector of column sums)}$$

$$\mathbf{1}^T\mathbf{A}\mathbf{1} = \sum_i \sum_j a_{ij} \qquad \text{(total sum of all entries)}$$

Graphical illustration

$(1, 1, 1, 1)$



$$\begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix}$$

The Hadamard product

A different way to multiply two matrices of the same size, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$, is through the **Hadamard product**, also called the **entrywise product**:

$$\mathbf{C} = \mathbf{A} \circ \mathbf{B} \in \mathbb{R}^{m \times n}, \qquad \text{with} \quad c_{ij} = a_{ij} b_{ij}.$$

For example,

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

Hadamard products are very useful in computing, as we shall see.

Matrix algebra

Use Math 39 course webpage[1] to review the following matrix operations:

- Transpose
- Rank
- Trace*
- Determinant*
- Inverse*

*Defined only for square matrices

[1] https://www.sjsu.edu/faculty/guangliang.chen/Math39.html

Characterization of rank-1 matrices

Rank-1 matrices are the simplest matrices (besides the zero matrices), and can be used as building blocks for getting more complicated matrices.

- Any *nonzero* row or column vector (as a matrix) has rank 1.
- A *nonzero* matrix is of rank 1 if and only if all of its nonzero rows (or columns) are multiples of each other.
- 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 $\mathbf{A} = \mathbf{u}\mathbf{v}^T$.

For example, the following is a rank-1 matrix:

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

Its rows are multiples of each other, and so are its 2 nonzero columns.

Another example of rank-1 matrices is $\mathbf{J}_n = \mathbf{1}\mathbf{1}^T$.

Eigenvalues and eigenvectors

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. The **characteristic polynomial** of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}).$$

We define the **eigenvalues** of $\mathbf{A}$ as the real roots of the characteristic equation $p(\lambda) = 0$ (and will never work with complex eigenvalues).

For a specific eigenvalue λ_0, any nonzero vector $\mathbf{v}_0 \in \mathbb{R}^n$ satisfying

$$(\mathbf{A} - \lambda_0 \mathbf{I})\mathbf{v}_0 = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{A}\mathbf{v}_0 = \lambda_0 \mathbf{v}_0$$

is called an **eigenvector** of $\mathbf{A}$ (associated to the eigenvalue λ_0).

All eigenvectors of $\mathbf{A}$ associated to an eigenvalue λ_0 span a linear subspace, called the **eigenspace** of $\mathbf{A}$ corresponding to λ_0:

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

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 $p(\lambda)$ is called the **algebraic multiplicity** of λ_0.

Note that for any matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with k distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, we have

$$1 \leq g_i \leq a_i, \quad 1 \leq i \leq k$$

Example 0.1. For the matrix $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix}$, find its eigenvalues and associated eigenvectors.

Answer. The eigenvalues are $\lambda_1 = 3, \lambda_2 = 2$ with algebraic multiplicities $a_1 = 2, a_2 = 1$, and geometric multiplicities $g_1 = g_2 = 1$. In fact, $\mathrm{E}(\lambda_1) = \mathrm{span}\{(0,1,-2)^T\}$ and $\mathrm{E}(\lambda_2) = \mathrm{span}\{(0,1,-1)^T\}$.

The following theorem indicates that for any $n \times n$ matrix with n eigenvalues, all of its rank, trace, and determinant can be computed from the eigenvalues of the matrix.

Theorem 0.1. For any $\mathbf{A} \in \mathbb{R}^{n\times n}$ with n eigenvalues, $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ (not necessarily distinct),

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

$$\mathrm{trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$$

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

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 the same

- rank, trace, determinant
- characteristic polynomial
- eigenvalues and their multiplicities (but not eigenvectors)

Diagonalizability of square matrices

Def 0.1. A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is **diagonalizable** 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

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

Remark. If we write $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_n]$ and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, then the above equation can be rewritten as

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda},$$

or in columns

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

From this we get that

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

This shows that $\mathbf{A}$ has n eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ (not necessarily distinct) with corresponding eigenvectors $\mathbf{p}_1, \dots, \mathbf{p}_n \in \mathbb{R}^n$.

Thus, the above factorization of a diagonalizable matrix $\mathbf{A}$ is called the **eigendecomposition**, or **spectral decomposition**, of $\mathbf{A}$.

Example 0.2. 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 (we will see why later).

Why are diagonalizable matrices important?

Every diagonalizable matrix is similar to a diagonal matrix (that consists of its eigenvalues), and is easy to deal with in a lot of ways.

For example, it can help compute **matrix powers** ($\mathbf{A}^k$). To see this, suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonalizable, that is, $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ for some invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{\Lambda}$. Then

$$\begin{aligned}
\mathbf{A}^2 &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \cdot \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} = \mathbf{P}\mathbf{\Lambda}^2\mathbf{P}^{-1} \\
\mathbf{A}^3 &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \cdot \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \cdot \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} = \mathbf{P}\mathbf{\Lambda}^3\mathbf{P}^{-1} \\
\mathbf{A}^k &= \mathbf{P}\mathbf{\Lambda}^k\mathbf{P}^{-1} \quad \text{(for any positive integer } k\text{)}
\end{aligned}$$

where $\mathbf{\Lambda}^k = \mathrm{diag}(\lambda_1^k, \ldots, \lambda_n^k)$.

If a diagonalizable matrix is also invertible, then we must have

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

where

$$\mathbf{\Lambda}^{-1} = \operatorname{diag}\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_n}\right).$$

Checking diagonalizability of a square matrix

Theorem 0.2. A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if it has n linearly independent eigenvectors (i.e., $\sum g_i = n$).

Proof.

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \iff \mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda} \iff \mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i, 1 \le i \le n$$

The $\mathbf{p}_i$'s are linearly independent if and only if $\mathbf{P}$ is is nonsingular.

Remark. A diagonalizable matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ must have n eigenvalues. Additionally, for each distinct eigenvalue, we must have $a_i = g_i$, because

$$n = \sum g_i \le \sum a_i \le n$$

Example 0.3. The matrix $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$ is not diagonalizable because it has only one distinct eigenvalue $\lambda_1 = 1$ with $a_1 = 2$ and $g_1 = 1$.

Two special classes of square matrices are always diagonalizable:

- **Idempotent** matrices:

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

  For example,

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 3 & -6 \\ 1 & -2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{pmatrix}$$

- **Symmetric** matrices:

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

Idempotent matrices

The following are some exemplar idempotent matrices:

$$\mathbf{O}, \quad \mathbf{I}, \quad \frac{1}{n}\mathbf{J}_n, \quad \text{and} \quad \mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T.$$

Note that $\mathbf{J}_n$ alone is not idempotent, because $\mathbf{J}_n^2 = n\mathbf{J}_n$.

To see why $\mathbf{C}_n$ is idempotent:

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

Important fact: $\mathbf{C}_n$ **is a centering matrix**.

For any point $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$,

$$\begin{aligned}\mathbf{C}_n\mathbf{x} &= \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{x} \\ &= \mathbf{x} - \frac{1}{n}\mathbf{1}(\mathbf{1}^T\mathbf{x}) \\ &= \mathbf{x} - \mathbf{1}\bar{x} \\ &= [x_1 - \bar{x}, \ldots, x_n - \bar{x}]^T\end{aligned}$$

where

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

An important result is the following. A proof based on minimal polynomials can be found in [Horn and Johnson, matrix analysis, 2nd ed].

Theorem 0.3. Every idempotent matrix $\mathbf{A} \in I^n(\mathbb{R})$ is diagonalizable, 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 $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$.

Additionally, idempotent matrices can only have eigenvalues 0 or 1 or both. To see this, suppose $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Using $\mathbf{A} = \mathbf{A}^2$, we then get

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

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

For any $\mathbf{A} \in I^n(\mathbb{R})$, let a_0 and a_1 be the algebraic multiplicities of the eigenvalues 0 and 1.

Because $\mathbf{A}$ is diagonalizable, we must have $a_0 + a_1 = n$, and

$$\mathrm{trace}(\mathbf{A}) = a_1 = \mathrm{rank}(\mathbf{A}).$$

Consider the following cases:

- $a_0 = n, a_1 = 0$: $\mathbf{A} = \mathbf{O}$;
- $a_0 = 0, a_1 = n$: $\mathbf{A} = \mathbf{I}$ (the only nonsingular matrix in $I_n(\mathbb{R})$);
- $1 \le a_0, a_1 \le n - 1$: All other idempotent matrices

Example 0.4. Since $\frac{1}{n}\mathbf{J}_n \in \mathbb{R}^{n\times n}$ is idempotent and

$$\operatorname{rank}\left(\frac{1}{n}\mathbf{J}_n\right) = 1 = \operatorname{trace}\left(\frac{1}{n}\mathbf{J}_n\right),$$

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 $1, n-1$ respectively:

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

Example 0.5. Consider the centering matrix $\mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$. Since

$$\mathrm{trace}(\mathbf{C}_n) = \mathrm{trace}(\mathbf{I}_n) - \frac{1}{n}\,\mathrm{trace}(\mathbf{J}_n) = n - \frac{1}{n} \cdot n = n - 1.$$

we conclude that

- $a_0 = 1$ and $a_1 = n - 1$.
- $\mathrm{rank}(\mathbf{C}_n) = n - 1$ and $\det(\mathbf{C}_n) = 0$.

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

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

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

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

Symmetric matrices

Symmetric matrices have many nice properties. For example, all their eigenvalues are real and they can be diagonalized via orthogonal matrices.

Theorem 0.4 (The Spectral Theorem). Let $\mathbf{A} \in S^n(\mathbb{R})$. Then there exist an orthogonal matrix $\mathbf{Q} = [\mathbf{q}_1 \ldots \mathbf{q}_n] \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, such that

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \qquad \text{(we say that } \mathbf{A} \text{ is orthogonally diagonalizable)}$$

Remark. The above factorization also represents a *spectral decomposition* of $\mathbf{A}$: The λ_i's represent eigenvalues of $\mathbf{A}$ while the $\mathbf{q}_i$'s are the associated eigenvectors (with unit norm and orthogonal to each other).

Remark. One can rewrite the matrix decomposition

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$$

into a sum of rank-1 matrices:

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

For convenience, the diagonal elements of $\mathbf{\Lambda}$ are often sorted in decreasing order (and the columns of $\mathbf{Q}$ should be arranged in matching order):

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Example 0.6. Find the spectral decomposition of the following matrix

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

Answer.

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

$$= 4\begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} + (-1)\begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

Quadratic forms

Symmetric matrices can be used to define the so-called quadratic forms.

Def 0.2. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. A **quadratic form** based on $\mathbf{A}$ is a function $Q : \mathbb{R}^n \mapsto \mathbb{R}$ with

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

Remark. A quadratic form is a second-order polynomial in the components of $\mathbf{x}$ without linear or constant terms:

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

For example, if $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, then

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

Question: Which symmetric matrix corresponds to

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

Positive (semi)definite matrices

A *symmetric* matrix $\mathbf{A} \in S^n(\mathbb{R})$ is said to be **positive semidefinite (PSD)** if $Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

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 further said to be **positive definite (PD)**.

We denote by $S^n_{0+}(\mathbb{R})$ and $S^n_{+}(\mathbb{R})$ the sets of positive semidefinite and of positive definite matrices of size $n \times n$, respectively.

Note that we must have

$$S^n_{+}(\mathbb{R}) \subset S^n_{0+}(\mathbb{R}) \subset S^n(\mathbb{R}) \subset \mathbb{R}^{n\times n}.$$

Theorem 0.5. A symmetric matrix is positive definite (positive semidefinite) if and only if all of its eigenvalues are strictly positive (nonnegative).

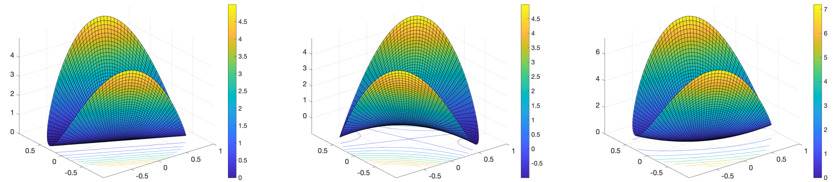
Remark. There is a quick way of determining the positive (semi)definiteness of a 2×2 nonzero matrix $\mathbf{A}$:

- $\mathbf{A} \in S^2_+(\mathbb{R})$ if and only if $\det(\mathbf{A}) > 0$ and $\mathrm{trace}(\mathbf{A}) > 0$;
- $\mathbf{A} \in S^2_{0+}(\mathbb{R})$ if and only if $\det(\mathbf{A}) = 0$ and $\mathrm{trace}(\mathbf{A}) > 0$.

This is due to $\det(\mathbf{A}) = \lambda_1\lambda_2$ and $\mathrm{trace}(\mathbf{A}) = \lambda_1 + \lambda_2$.

Example 0.7. Determine the positive definiteness of each of the following matrices:

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$



Spectral decomposition of PSD matrices in reduced form

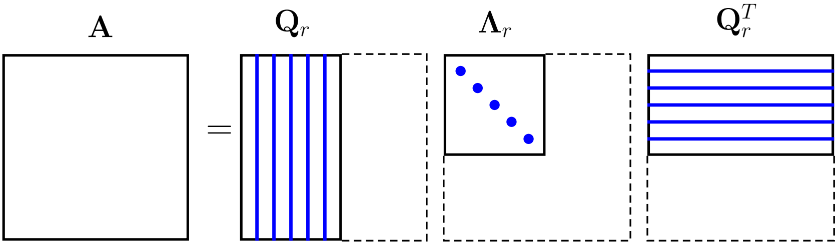
The preceding theorem implies that for a PSD matrix $\mathbf{A} \in S^n_{0+}(\mathbb{R})$,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n, \quad r = \operatorname{rank}(\mathbf{A}).$$

Correspondingly, we may obtain

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

where $\mathbf{Q}_r = [\mathbf{q}_1 \ldots \mathbf{q}_r] \in \mathbb{R}^{n \times r}$ is a tall matrix with orthonormal columns, and $\mathbf{\Lambda}_r = \operatorname{diag}(\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^{r \times r}$.


$\mathbf{A}$
$=$
$\mathbf{Q}_r$
$\mathbf{\Lambda}_r$
$\mathbf{Q}_r^T$

Example 0.8. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \in S^2_{0+}(\mathbb{R})$, which has a rank of $r = 1$.

The spectral decomposition is

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 & \\ & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \quad \longleftarrow \text{reduced from}$$

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).

Def 0.3. Let $\mathbf{A} \in S^n_{0+}(\mathbb{R})$. The **square root** of $\mathbf{A}$ is defined as the matrix $\mathbf{R} \in S^n_{0+}(\mathbb{R})$ such that $\mathbf{R}^2 = \mathbf{A}$. We denote it by $\mathbf{R} = \mathbf{A}^{1/2}$.

Note that if $\mathbf{A} \in S^n_{+}(\mathbb{R})$, then $\mathbf{R} = \mathbf{A}^{1/2} \in S^n_{+}(\mathbb{R})$ because

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

In such a case, we can further define the **reciprocal square root** of $\mathbf{A}$ as $\mathbf{A}^{-1/2} = (\mathbf{A}^{1/2})^{-1} \in S^n_{+}(\mathbb{R})$.

Special case: If $\mathbf{A} \in S^n_{0+}(\mathbb{R})$ happens to be diagonal, i.e.,

$$\mathbf{A} = \mathrm{diag}(a_1, \ldots, a_n), \qquad \text{where} \quad a_1, \ldots, a_n \geq 0,$$

then there is an easy way to find its square root. Define

$$\mathbf{R} = \mathrm{diag}\left(a_1^{1/2}, \ldots, a_n^{1/2}\right) \in S^n_{0+}(\mathbb{R}).$$

Clearly, $\mathbf{R}^2 = \mathbf{A}$. This shows that $\mathbf{R}$ is indeed a square root of $\mathbf{\Lambda}$.

Note that without the positive semidefiniteness requirement in the definition of matrix square roots, it won't be unique as we can arbitrarily modify the signs of the diagonals $a_i^{1/2}$ without violating the equality condition.

Theorem 0.6. Let $\mathbf{A} \in S^n_{0+}(\mathbb{R})$ 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 $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Then $\mathbf{A}$ has a unique matrix square root

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

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

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

We omit the proof of the uniqueness part in class.

Remark. For any $\mathbf{A} \in S^n_{+}(\mathbb{R})$ with eigendecomposition $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$,

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

Example 0.9. Consider

$$\underbrace{\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{Q}} \underbrace{\begin{pmatrix} 9 & \\ & 1 \end{pmatrix}}_{\mathbf{\Lambda}} \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{Q}^T}$$

The square root of $\mathbf{A}$ is

$$\mathbf{A}^{1/2} = \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{Q}} \underbrace{\begin{pmatrix} 3 & \\ & 1 \end{pmatrix}}_{\mathbf{\Lambda}^{1/2}} \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{Q}^T} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

and the reciprocal square root of $\mathbf{A}$ is

$$\mathbf{A}^{-1/2} = \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{Q}} \underbrace{\begin{pmatrix} \frac{1}{3} & \\ & 1 \end{pmatrix}}_{\mathbf{\Lambda}^{-1/2}} \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{\mathbf{Q}^T} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

Remark. Using the reduced form of the eigendecomposition of $\mathbf{A} \in S^n_{0+}(\mathbb{R})$, we obtain the following reduced form for the square root of $\mathbf{A}$:

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

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

Example 0.10. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \in S^2_{0+}(\mathbb{R})$, which has two nonnegative eigenvalues $\lambda_1 = 5, \lambda_2 = 0$. To find the matrix square root of $\mathbf{A}$, we only need to find its orthogonal diagonalization in reduced form:

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$$

It follows that

$$\mathbf{A}^{1/2} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{5} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \end{pmatrix}$$

The 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

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

The vector $\mathbf{v}$ is called a **generalized eigenvector** of $(\mathbf{A}, \mathbf{B})$ corresponding to λ.

Remark. 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 $\mathbf{A}$:

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

Now, let us rewrite the definition as

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

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

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

Let $p_{\mathbf{A},\mathbf{B}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{B})$, the **characteristic polynomial** of $(\mathbf{A}, \mathbf{B})$.

Interestingly, $p_{\mathbf{A},\mathbf{B}}(\lambda)$ is also a polynomial in λ, but it can have an arbitrary order between 0 and n, as we show next.

Example 0.11. 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 0.12. 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!

Generalized symmetric-definite eigenvalue problems

Let $\mathbf{A} \in S^n(\mathbb{R})$ and $\mathbf{B} \in S^n_+(\mathbb{R})$. The generalized eigenvalue problem

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

is called a **generalized symmetric-definite eigenvalue problem**. Such problems have very nice properties and have a lot of applications.

Theorem 0.7. The above generalized symmetric-definite eigenvalue problem has n generalized 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

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

Remark. We derive a few more results from the theorem. Let $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbb{R}^{n\times n}$. Then

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

Next, using $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{B}\mathbf{v}_i,\ 1 \le i \le n$, we have

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

Lastly, $\mathbf{V}$ also diagonalizes $\mathbf{A}$:

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

Proof of the theorem. Since $\mathbf{B} \in S_+^n(\mathbb{R})$, we can rewrite

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v} \quad \Longrightarrow \quad \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} \cdot \mathbf{B}^{1/2}\mathbf{v} = \lambda \cdot \mathbf{B}^{1/2}\mathbf{v}$$

Letting

$$\widetilde{\mathbf{A}} = \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}, \qquad \text{and} \quad \widetilde{\mathbf{v}} = \mathbf{B}^{1/2}\mathbf{v}$$

we further obtain that

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

Since $\widetilde{\mathbf{A}} \in S^n(\mathbb{R})$, there are n eigenpairs $(\lambda_i, \widetilde{\mathbf{v}}_i), 1 \le i \le n$, with

$$\delta_{ij} = \widetilde{\mathbf{v}}_i^T \widetilde{\mathbf{v}}_j = \left(\mathbf{B}^{1/2}\mathbf{v}_i\right)^T \mathbf{B}^{1/2}\mathbf{v}_j = \mathbf{v}_i^T \mathbf{B}\mathbf{v}_j.$$

Consequently, $(\mathbf{A}, \mathbf{B})$ has n generalized eigenvalues λ_i with associated generalized eigenvectors $\mathbf{v}_i = \mathbf{B}^{-1/2}\widetilde{\mathbf{v}}_i$.

Some observations:

- The generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$, for $\mathbf{A} \in S^n(\mathbb{R}), \mathbf{B} \in S^n_+(\mathbb{R})$, are identical to the eigenvalues of $\widetilde{\mathbf{A}} = \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} \in S^n(\mathbb{R})$.
- The generalized eigenvectors of $(\mathbf{A}, \mathbf{B})$ are $\mathbf{v}_i = \mathbf{B}^{-1/2}\widetilde{\mathbf{v}}_i$, where $\widetilde{\mathbf{v}}_i$ are the unit-norm eigenvectors of $\widetilde{\mathbf{A}}$.

Furthermore, the generalized eigenvalues/eigenvectors of $(\mathbf{A}, \mathbf{B})$ coincide with the eigenvalues/eigenvectors of $\mathbf{B}^{-1}\mathbf{A}$:

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

In fact, $\widetilde{\mathbf{A}} = \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ and $\mathbf{B}^{-1}\mathbf{A}$ are two similar matrices:

$$\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} = \mathbf{B}^{1/2} \cdot \mathbf{B}^{-1}\mathbf{A} \cdot \mathbf{B}^{-1/2}$$

Example 0.13. Let $\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix} \in S^2(\mathbb{R})$ and $\mathbf{B} = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} \in S^2_+(\mathbb{R})$. Find the generalized eigenvalues and eigenvectors of $(\mathbf{A}, \mathbf{B})$.

Next time: Matrix Computing in MATLAB

Be sure to complete the following activities before next class:

- Install MATLAB on your computer with the Statistics and Machine Learning Toolbox[2]
- MATLAB fundamentals[3]
- Introduction to Linear Algebra with MATLAB[4]

---

[2] https://www.mathworks.com/products/statistics.html
[3] https://matlabacademy.mathworks.com/details/matlab-fundamentals/mlbe
[4] https://matlabacademy.mathworks.com/details/introduction-to-linear-algebra-with-matlab/linalg