MATH417 Matrix Algebra I

1 Introduction [1]

In this course we will study matrix algebra, or linear algebra.

The relation such as $2x - 1 = 0$ is said to be an equation. Let us consider the following multiple equations, or a system.

$$\begin{cases} 2x + 8y + 4z = 2, \\ 2x + 5y + z = 5, \\ 4x + 10y - z = 1. \end{cases} \tag{1}$$

We can write this system as

$$A\vec{x} = \vec{b}, \tag{2}$$

where

$$A = \begin{bmatrix} 2 & 8 & 4 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix}, \qquad \vec{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}. \tag{3}$$

Here A is a matrix, and $\vec{x}, \vec{b}$ are vectors. A vector can be regarded as a matrix with one column. The matrix A is a 3×3 matrix because it has 3 rows and 3 columns. The entry or element which belongs to the ith row and the jth column can be expressed as a_{ij}. For example, $a_{12} = 8$.

The $n \times n$ matrix I_n below is called the identity matrix.

$$I_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

where n diagonal entries are all 1 and other entries are zero.

Summer 2014, Math 417 Section 201 · Manabu Machida (University of Michigan)

[1] This section is related to Chapter 1 of the textbook.

If we find a matrix A^{-1} which satisfies

$$A^{-1}A = I_3,$$

then we can obtain $\vec{x}$ as

$$A^{-1}A\vec{x} = A^{-1}\vec{b} \quad \Rightarrow \quad \vec{x} = A^{-1}\vec{b}.$$

This A^{-1} is called the inverse of A.

The set of real numbers is denoted by $\mathbb{R}$. The set of complex numbers is denoted by $\mathbb{C}$. A number in $\mathbb{R}$ or $\mathbb{C}$ is said to be a scalar. The sets of n-dimensional real and complex numbers are denoted by $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively. For example, $\vec{x}, \vec{b} \in \mathbb{R}^3$. Similarly $\mathbb{R}^{n\times m}$ denotes the set of all real $n \times m$ matrices. For example, in (3), $A \in \mathbb{R}^{3\times 3}$.

2 Gauss-Jordan elimination [2]

Let us solve the linear system (2). First we write a 3×4 matrix $\left[A \middle| \vec{b} \right]$, which is called the augmented matrix.

$$\left[\begin{array}{ccc:c} 2 & 8 & 4 & 2 \\ 2 & 5 & 1 & 5 \\ 4 & 10 & -1 & 1 \end{array}\right]$$

We note that we usually solve (1) by dividing or multiplying an equation by a constant, and subtracting or adding a multiple of an equation from another equation. This means we can obtain $\vec{x}$ by simplifying the augmented matrix using the following elementary row operations:

- Divide or multiply a row by a nonzero scalar.
- Subtract or add a multiple of a row from another rows.
- Swap two rows.

Let us solve $A\vec{x} = \vec{b}$ using the augmented matrix. We want to change the left part of $\left[A \middle| \vec{b} \right]$ to I_3.

Step 1: (1st row)/2

[2] This section is related to Chapter 1 of the textbook.

$$\left[\begin{array}{rrr|r} 1 & 4 & 2 & 1 \\ 2 & 5 & 1 & 5 \\ 4 & 10 & -1 & 1 \end{array}\right].$$

Step 2: (2nd row) $-\, 2 \cdot$ (1st row), (3rd row) $-\, 4 \cdot$ (1st row)

$$\left[\begin{array}{rrr|r} 1 & 4 & 2 & 1 \\ 0 & -3 & -3 & 3 \\ 0 & -6 & -9 & -3 \end{array}\right].$$

Step 3: (2nd row)/(-3)

$$\left[\begin{array}{rrr|r} 1 & 4 & 2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & -6 & -9 & -3 \end{array}\right].$$

Step 4: (1st row) $-\, 4 \cdot$ (2nd row), (3rd row) $+\, 6 \cdot$ (2nd row)

$$\left[\begin{array}{rrr|r} 1 & 0 & -2 & 5 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & -3 & -9 \end{array}\right].$$

Step 5: (3rd row)/(-3)

$$\left[\begin{array}{rrr|r} 1 & 0 & -2 & 5 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

Step 6: (1st row) $+\, 2 \cdot$ (3rd row), (2nd row) $-$ (3rd row)

$$\left[\begin{array}{rrr|r} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

The last form is said to be the reduced row-echelon form (see below) and implies the matrix-vector equation

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 11 \\ -4 \\ 3 \end{bmatrix},$$

or

$$x = 11, \qquad y = -4, \qquad z = 3.$$

In this way we can obtain $\vec{x}$.

A matrix is said to be in reduced row-echelon form when the matrix satisfies the following conditions.

- If a row has nonzero entries, then the first nonzero entry or the pivot is 1. This 1 is called the leading 1 in the row.
- If a column contains a leading 1, then all the other entries in the column are 0.
- If a row contains a leading 1, then each row above it contains a leading 1 further to the left.

Let rref(A) denote the reduced row-echelon form of A.

Example 1.

$$\text{rref}(\left[\begin{array}{ccc:c} 2 & 8 & 4 & 2 \\ 2 & 5 & 1 & 5 \\ 4 & 10 & -1 & 1 \end{array}\right]) = \left[\begin{array}{ccc:c} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{array}\right], \quad \text{rref}(\begin{bmatrix} 2 & 8 & 4 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If the number of equations is less than the number of unknowns (underdetermined), the length of the solution vector $\vec{x}$ is shorter than the length of $\vec{b}$. If the number of equations is greater than the number of unknowns (overdetermined), the length of the solution vector $\vec{x}$ is longer than the length of $\vec{b}$. In either case, the coefficient matrix A becomes a rectangle.

Example 2 (Underdetermined). Let us consider

$$\begin{cases} 2x + 4y + z = 4, \\ x + 2y + z = 3, \\ x + 2y = 1. \end{cases}$$

We have

$$\text{rref}(\left[\begin{array}{ccc:c} 2 & 4 & 1 & 4 \\ 1 & 2 & 1 & 3 \\ 1 & 2 & 0 & 1 \end{array}\right]) = \left[\begin{array}{ccc:c} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The zero row at the bottom implies the third equation in the system is not independent. Thus we can drop this equation and consider

$$\begin{cases} 2x + 4y + z = 4, \\ x + 2y + z = 3, \end{cases}$$

and have

$$\mathrm{rref}(\left[\begin{array}{ccc|c} 2 & 4 & 1 & 4 \\ 1 & 2 & 1 & 3 \end{array}\right]) = \left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

That is, we have

$$\begin{cases} x + 2y = 1, \\ z = 2. \end{cases}$$

We note that y is a free variable and x, z are leading variables. There are infinitely many solutions depending on y. Using an arbitrary constant t, we can write

$$x = 1 - 2t, \quad y = t, \quad z = 2.$$

This example implies that in a linear system there are infinitely many solutions if there exist more than one solution.

———

Example 3 (Overdetermined). Let us consider

$$\begin{cases} x + 2y + z = 1, \\ 3x + 6y + 2z = 2, \\ x + 2y + z = 2, \\ 2x + 4y + 2z = 1. \end{cases}$$

The reduced row-echelon form is obtained as

$$\mathrm{rref}(\left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 3 & 6 & 2 & 2 \\ 1 & 2 & 1 & 2 \\ 2 & 4 & 2 & 1 \end{array}\right]) = \left[\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

On the third line, we have $0 = 1$. There is no solution.

———

If there is a unique solutions or there are infinitely many solutions, we say the system is consistent. The system is inconsistent if there is no solution.

3 Rank [3]

We define the rank of a matrix A as

$$\text{rank}(A) = \text{the number of leading 1's in}\,\text{rref}(A).$$

Example 4.

$$\text{rank}(\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}) = \text{rank}(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}) = 3,$$

$$\text{rank}(\begin{bmatrix} 1 & 2 & 1 \\ 3 & 6 & 2 \\ 1 & 2 & 1 \\ 2 & 4 & 2 \end{bmatrix}) = \text{rank}(\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}) = 2.$$

Suppose we have a system with n equations and m variables. Then the coefficient matrix A is an $n \times m$ matrix. We have

1. $\text{rank}(A) \leq n, \quad \text{rank}(A) \leq m$
2. $\text{rank}(A) = n \quad \Rightarrow$ consistent
3. $\text{rank}(A) = m \quad \Rightarrow$ one solution or inconsistent
4. $\text{rank}(A) < m \quad \Rightarrow$ infinitely many solutions or inconsistent

Furthermore,

- The number of free variables $= m - \text{rank}(A)$
- For an $n \times n$ matrix A, there is a unique solution if and only if $\text{rank}(A) = n$. In this case, $\text{rref}(A) = I$.

4 Linearity [4]

Let A be an $n \times m$ matrix, $\vec{x}, \vec{y} \in \mathbb{R}^m$, and α, β be scalars. We have

$$A\,(\alpha\vec{x} + \beta\vec{y}) = \alpha A\vec{x} + \beta A\vec{y}.$$

[3] This section is related to Chapter 1 of the textbook.

[4] This section is related to Chapter 1 of the textbook.

This is called linearity.

A vector $\vec{b} \in \mathbb{R}^n$ is called a linear combination or superposition of $\vec{v}_1, \dots, \vec{v}_m$ in $\mathbb{R}^n$ if $\vec{b}$ is given by

$$\vec{b} = x_1\vec{v}_1 + \cdots + x_m\vec{v}_m,$$

where $x_1, \dots, x_m$ are scalars.

5 Linear transformations [5]

Suppose output $\vec{y}$ is determined by some operations T from input $\vec{x}$:

$$\vec{x} \xrightarrow{T} \vec{y}.$$

If T satisfies

$$T(\alpha\vec{x} + \beta\vec{y}) = \alpha T(\vec{x}) + \beta T(\vec{y}),$$

then T is called a linear transformation.

If T is a linear transformation, we can express T as

$$T(\vec{x}) = A\vec{x}.$$

Conversely, a matrix A represents some linear transformation T.

Example 5 (Scaling). The transformation T:

$$\vec{x} \xrightarrow{T} k\vec{x} = \vec{y}, \qquad k \in \mathbb{R}$$

can be written as

$$T(\vec{x}) = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix} \vec{x}.$$

———

Example 6 (Rotation). The transformation T:

$$\vec{x} = \begin{bmatrix} \cos\theta_0 \\ \sin\theta_0 \end{bmatrix} \xrightarrow{T} \vec{y} = \begin{bmatrix} \cos(\theta + \theta_0) \\ \sin(\theta + \theta_0) \end{bmatrix},$$

can be written as

[5] This section is related to Chapter 2 of the textbook.

$$T(\vec{x}) = R_\theta \begin{bmatrix} \cos\theta_0 \\ \sin\theta_0 \end{bmatrix},$$

where

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Example 7. Let us consider matrix $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Since we can express A as

$$\begin{aligned} A &= r \begin{bmatrix} \frac{a}{r} & -\frac{b}{r} \\ \frac{b}{r} & \frac{a}{r} \end{bmatrix}, \qquad r = \sqrt{a^2+b^2}, \\ &= r \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \qquad \tan\theta = \frac{b}{a}, \\ &= \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}, \end{aligned}$$

this A represents a rotation combined with a scaling.

Example 8 (Composition). Let us consider the rotation through $\pi/2$:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \longrightarrow R_{\pi/2} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We have

$$R_{\pi/2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

But we should obtain the same vector by rotating the input vector through $\pi/3$ then rotating the resulting vector through $\pi/6$:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \longrightarrow R_{\pi/6} R_{\pi/3} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Let us check this. Indeed we obtain

$$R_{\pi/6} R_{\pi/3} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = R_{\pi/6} \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & -1 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Example 9 (Counterexamples). We consider the transformation T from $\mathbb{R}^2$ to $\mathbb{R}^2$ such that

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \xrightarrow{T} \begin{bmatrix} x_1 \\ x_2 + 1 \end{bmatrix}.$$

We have

$$T(\alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \beta \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}) = T(\begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \end{bmatrix}) = \begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 + 1 \end{bmatrix}.$$

On the other hand, we obtain

$$\alpha T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) + \beta T(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}) = \alpha \begin{bmatrix} x_1 \\ x_2 + 1 \end{bmatrix} + \beta \begin{bmatrix} y_1 \\ y_2 + 1 \end{bmatrix}.$$

The two results are different in general, and T is not a linear transformation. The transformation T cannot be represented by any matrix.

Let us also consider the transformation T from $\mathbb{R}^2$ to $\mathbb{R}$ that computes the length: The transformation T:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \xrightarrow{T} \sqrt{x_1^2 + x_2^2}.$$

It is enough to give one concrete example. We have

$$T(\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}) = T(\begin{bmatrix} 1 \\ 1 \end{bmatrix}) = \sqrt{1^2 + 1^2} = \sqrt{2}.$$

On the other hand, we obtain

$$T(\begin{bmatrix} 1 \\ 0 \end{bmatrix}) + T(\begin{bmatrix} 0 \\ 1 \end{bmatrix}) = \sqrt{1^2} + \sqrt{1^2} = 2.$$

The two results are different and T is not a linear transformation.

---

6 Orthogonal projections and reflections [6]

Let us consider a line L in the x-y plane running through the origin. Any vector $\vec{x} \in \mathbb{R}^2$ can be uniquely decomposed as

$$\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp},$$

[6] This section is related to Chapter 2 of the textbook.

where $\vec{x}^{\|}$ is parallel to L and $\vec{x}^{\perp}$ is perpendicular to L.

The transformation $\text{proj}_L(\vec{x}) = \vec{x}^{\|}$ from $\mathbb{R}^2$ to $\mathbb{R}^2$ is called the orthogonal projection of $\vec{x}$ onto L.

We note that if L is the x-axis, we have

$$\text{proj}_L(\vec{x}) = \vec{x}^{\|} = x_1\vec{e}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

where $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is the standard vector along the x-axis. Let θ be the angle between L and the x-axis. Then $\vec{u} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ is a unit vector parallel to L. We can compute $\text{proj}_L(\vec{x})$ as follows.

$$\begin{aligned} \text{proj}_L(\vec{x}) &= R_\theta \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} R_{-\theta}\vec{x} \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\vec{x} \\ &= \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \end{aligned}$$

We further proceed as

$$\begin{aligned} \text{proj}_L(\vec{x}) &= \begin{bmatrix} (x_1\cos\theta + x_2\sin\theta)\cos\theta \\ (x_1\cos\theta + x_2\sin\theta)\sin\theta \end{bmatrix} \\ &= \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}\right)\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \\ &= (\vec{x}\cdot\vec{u})\,\vec{u}. \end{aligned} \tag{4}$$

We note that since $\|\vec{u}\| = |\vec{u}| = \sqrt{\vec{u}\cdot\vec{u}} = 1$,

$$\vec{x}\cdot\vec{u} = \|\vec{x}\| \cos(\text{angle between } \vec{x} \text{ and } \vec{u}).$$

The reflection of $\vec{x}$ about L is a linear transformation $T(\vec{x}) = \text{ref}_L(\vec{x})$ which transforms $\vec{x}$ into its image on the opposite side of a mirror. We have

$$\text{ref}_L(\vec{x}) = \vec{x}^{\|} - \vec{x}^{\perp} = 2\,\text{proj}_L(\vec{x}) - \vec{x} = 2\,(\vec{x}\cdot\vec{u})\,\vec{u} - \vec{x}.$$

We note that

$$\begin{aligned}\mathrm{ref}_L(\vec{x}) &= 2\begin{bmatrix}\cos^2\theta & \cos\theta\sin\theta\\ \cos\theta\sin\theta & \sin^2\theta\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} - \begin{bmatrix}x_1\\ x_2\end{bmatrix}\\ &= \begin{bmatrix}\cos(2\theta) & \sin(2\theta)\\ \sin(2\theta) & -\cos(2\theta)\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}.\end{aligned}$$

Thus we obtain the matrix of ref_L. Conversely, any matrix of this form represents a reflection about a line.

Orthogonal projections and reflections in space can be considered in the same way. Let $\vec{u} \in \mathbb{R}^3$ is a unit vector parallel to L. For $\vec{x} \in \mathbb{R}^3$ we have

$$\mathrm{proj}_L(\vec{x}) = (\vec{x}\cdot\vec{u})\,\vec{u}, \qquad \mathrm{ref}_L(\vec{x}) = 2\,\mathrm{proj}_L(\vec{x}) - \vec{x}.$$

7 Inverse [7]

For a linear transformation

$$T(\vec{x}) = A\vec{x} = \vec{y},$$

let us consider its inverse T^{-1} such that

$$\vec{x} = T^{-1}(\vec{y}) = A^{-1}\vec{y}.$$

A linear transformation T is said to be invertible and T^{-1} exists if T is bijective or

$$\vec{y} = A\vec{x}$$

has a unique solution for all $\vec{y}$. Otherwise T is noninvertible. The matrix A is said to be invertible and A^{-1} exists if T is invertible.

Remark 1. The inverse T^{-1} of a linear transformation T is also linear.

Example 10. Consider

$$\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}\vec{x} = \vec{y}.$$

The inverse is calculated as

$$\vec{x} = \begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}^{-1}\vec{y} = \begin{bmatrix}-2 & 1\\ 3/2 & -1/2\end{bmatrix}\vec{y},$$

[7] This section is related to Chapter 2 of the textbook.

where we used the relation

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Here $\det(A) = ad - bc$ is called the determinant.

Example 11 (Noninvertible). Consider a linear transformation $T(\vec{x}) = A\vec{x}$, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}.$$

We have

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}, \qquad A \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}.$$

At least two vectors in the domain are transformed into one vector in the target space. We cannot obtain the inverse:

$$A^{-1} \begin{bmatrix} 3 \\ 9 \end{bmatrix} = ?$$

In the present case T is noninvertible. Indeed we have

$$\det(A) = ad - bc = 1 \cdot 6 - 2 \cdot 3 = 0,$$

and A^{-1} doesn't exist.

Example 12. The orthogonal projection $\mathrm{proj}_L(\vec{x})$ is not invertible but the reflection $\mathrm{ref}_L(\vec{x})$ is invertible. Although we can understand this geometrically, here let us calculate the determinants of the matrices for $\mathrm{proj}_L(\vec{x})$ and $\mathrm{ref}_L(\vec{x})$. For $\mathrm{proj}_L(\vec{x})$ we have

$$\begin{aligned} \det \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix} &= \det \begin{bmatrix} \frac{1+\cos(2\theta)}{2} & \frac{\sin(2\theta)}{2} \\ \frac{\sin(2\theta)}{2} & \frac{1-\cos(2\theta)}{2} \end{bmatrix} \\ &= \frac{1-\cos^2(2\theta)}{4} - \frac{\sin^2(2\theta)}{4} = \frac{1-1}{4} = 0. \end{aligned}$$

For $\mathrm{ref}_L(\vec{x})$ we have

$$\det \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} = -\cos^2(2\theta) - \sin^2(2\theta) = -1 \neq 0.$$

We know from Sec. 3 that in general a unique solution of the system of an $n \times m$ matrix A implies $\operatorname{rank}(A) = m$. In particular we studied that there is a unique solution to a system of an $n \times n$ matrix A if and only if $\operatorname{rank}(A) = n \Leftrightarrow \operatorname{rref}(A) = I_n$.

Theorem 1 (Invertibility). *We have*

$$\text{An } n \times n \text{ matrix } A \text{ is invertible} \quad \Leftrightarrow \quad \operatorname{rref}(A) = I_n \quad \Leftrightarrow \quad \operatorname{rank}(A) = n.$$

Remark 2. An $n \times n$ matrix A is invertible if and only if $A\vec{x} = \vec{b}$ has a unique solution $\vec{x} \in \mathbb{R}^n$ for a vector $\vec{b} \in \mathbb{R}^n$. Or we can say

$$A \text{ is noninvertible} \quad \Leftrightarrow \quad \text{infinitely many solutions or none}$$

Remark 3. Suppose $T(\vec{x}) = A\vec{x}$ is an invertible linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Let B be the $m \times n$ matrix of T^{-1}. If $BA = I_n$ and $AB = I_m$, then $n = m$.

We can compute A^{-1} as follows. Suppose that an $n \times n$ matrix A is invertible and $\operatorname{rref}(A) = I_n$. We write A and its inverse A^{-1} as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{n1} & & & a_{nn} \end{bmatrix}, \qquad A^{-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & & \\ \vdots & & \ddots & \\ \alpha_{n1} & & & \alpha_{nn} \end{bmatrix}$$

Let us begin with

$$A\vec{x} = \vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Note that

$$\vec{x} = A^{-1}\vec{e}_1 = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{n1} \end{bmatrix}.$$

By making use of Gauss-Jordan elimination we obtain

$$\operatorname{rref}([\,A\,|\,\vec{e}_1\,]) = \left[\begin{array}{c:c} & \alpha_{11} \\ I_n & \vdots \\ & \alpha_{n1} \end{array}\right]$$

We expand this calculation. Since $[\,\vec{e}_1\,\vec{e}_2\,\ldots\,\vec{e}_n\,] = I_n$, we have

$$\operatorname{rref}([\,A\,|\,I_n\,]) = \left[\,I_n\,|\,A^{-1}\,\right].$$

Thus A^{-1} is obtained. If the left part of the matrix on the right-hand side or $\operatorname{rref}(A)$ is not I_n, then A^{-1} doesn't exist.

For invertible $n \times n$ matrices A, B, we have the following properties.

- $A^{-1}A = AA^{-1} = I_n$
- $\left(A^{-1}\right)^{-1} = A$
- $(BA)^{-1} = A^{-1}B^{-1}$

The last formula above is obtained as follows. Consider $\vec{y} = BA\vec{x}$. From this we obtain $B^{-1}\vec{y} = A\vec{x}$. If we multiply A^{-1}, we obtain $A^{-1}B^{-1}\vec{y} = \vec{x}$. Thus we could construct the inverse of BA; the matrix $(BA)^{-1}$ exists and $\vec{x} = (BA)^{-1}\vec{y}$. We have $(BA)^{-1} = A^{-1}B^{-1}$.

Theorem 2. *We have the following criterion for invertibility. Let A, B be $n \times n$ matrices. If $BA = I_n$, then*

(a) A, B *are invertible*
(b) $A^{-1} = B$ *and* $B^{-1} = A$
(c) $AB = I_n$

Proof. Let us first prove (a) and (b). We consider $\vec{x}$ such that $A\vec{x} = \vec{0}$. Then

$$BA\vec{x} = B\vec{0} = \vec{0}.$$

Since $BA = I_n$, we see $\vec{x} = \vec{0}$. That is, $\vec{x}$ is the unique solution to $A\vec{x} = \vec{0}$. Hence A is invertible. We multiply A^{-1} by BA:

$$BAA^{-1} = I_nA^{-1} \qquad \therefore B = A^{-1}.$$

Thus B is also invertible and

$$B^{-1} = \left(A^{-1}\right)^{-1} = A.$$

We can prove (c) by

$$AB = AA^{-1} = I_n.$$

□

8 Image and kernel [8]

Let us define the image a linear transformation $T(\vec{x}) = A\vec{x}$ as

$$\text{im}(T) = \text{im}(A) = \{\vec{y} : T(\vec{x}) = \vec{y} \quad \text{for all } \vec{x}\}.$$

Example 13. Suppose a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ is expressed as $T(\vec{x}) = A\vec{x}$, where $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. We have

$$A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

Thus im(T) is $\mathbb{R}^2$, the set of all vectors in the x-y plane.

Example 14. Suppose a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ is expressed as $T(\vec{x}) = A\vec{x}$, where $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$. We have

$$A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 6 \end{bmatrix} = (x_1 + 2x_2) \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Thus im(T) is the line of all scalar multiples of $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

In general a vector can be expressed as a linear combination. For example,

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \vec{e}_1 + 2\vec{e}_2.$$

We define the span of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ as

$$\text{span}(\vec{v}_1, \ldots, \vec{v}_m) = \{c_1\vec{v}_1 + \cdots + c_m\vec{v}_m : c_1, \ldots, c_m \in \mathbb{R}\}.$$

Example 15. We have

[8] This section is related to Chapter 3 of the textbook.

$$\text{im}(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}) = \text{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}), \quad \text{im}(\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}) = \text{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix}).$$

Theorem 3. *For a linear transformation $T(\vec{x}) = A\vec{x}$ from $\mathbb{R}^m$ to $\mathbb{R}^n$, where $A = [\vec{v}_1 \cdots \vec{v}_m]$, we have*

$$\text{im}(T) = \text{span}(\vec{v}_1, \dots, \vec{v}_m).$$

The image of T is also called the column space of A because $\vec{v}_1, \dots, \vec{v}_m$ are the column vectors of A.

Proof. We note that

$$T(\vec{x}) = A\vec{x} = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ v_{21} & \cdots & v_{2m} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = x_1\vec{v}_1 + \cdots + x_m\vec{v}_m.$$

Thus the set of $T(\vec{x})$ for all $\vec{x}$ is the set of all linear combinations of $\vec{v}_1, \dots, \vec{v}_m$. □

The set of zeros of a linear transformation $T(\vec{x}) = A\vec{x}$ is called the kernel of T.

$$\ker(T) = \ker(A) = \left\{\vec{x} : T(\vec{x}) = \vec{0}\right\}.$$

The kernel of T is the solution set of $A\vec{x} = \vec{0}$. The kernel of T is also called the null space of A.

Remark 4. Suppose T is a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then $\text{im}(T)$ is a subset of the target space $\mathbb{R}^n$ of T and $\ker(T)$ is a subset of the domain $\mathbb{R}^m$ of T.

Example 16. Let us consider a linear transformation $T(\vec{x}) = A\vec{x}$ from $\mathbb{R}^3$ to $\mathbb{R}^2$, where $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$. We can calculate $\ker(T)$ as follows. By recalling that the kernel is the solution set, we consider

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \vec{0}.$$

Since

$$\text{rref}(\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 0 \end{array}\right]) = \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right],$$

we obtain

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix},$$

where t is an arbitrary constant. Hence, $\ker(T) = \text{span}(\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix})$.

The above example implies that for an $n \times m$ matrix A with $m > n$, there are nonzero vectors in $\ker(A)$. Note that the number of free variables is $m - \text{rank}(A) \geq m - n$ ($\text{rank}(A) \leq n$). Hence we have at least one free variable when $m - n > 0$. By contraposition we have

$$\ker(A) = \{\vec{0}\} \quad \Rightarrow \quad m \leq n.$$

For an $n \times m$ matrix A, $\ker(A) = \{\vec{0}\}$ means that $A\vec{x} = \vec{0}$ has a unique solution $\vec{x} = \vec{0}$, which implies $\text{rank}(A) = m$ (see Sec. 3). That is,

$$\ker(A) = \{\vec{0}\} \quad \Leftrightarrow \quad \text{rank}(A) = m.$$

Theorem 4. *For a square $n \times n$ matrix A,*

$$\ker(A) = \{\vec{0}\} \quad \Leftrightarrow \quad A \textit{ is invertible.}$$

We have the following equivalent statements related to invertibility in Theorem 1.

$$\begin{aligned} & \text{An } n \times n \text{ matrix } A \text{ is invertible} \\ \Leftrightarrow \quad & A\vec{x} = \vec{b} \text{ has a unique solution } \vec{x} \text{ for all } \vec{b} \in \mathbb{R}^n \\ \Leftrightarrow \quad & \text{rref}(A) = I_n \quad \Leftrightarrow \quad \text{rank}(A) = n \\ \Leftrightarrow \quad & \text{im}(A) = \mathbb{R}^n \quad \Leftrightarrow \quad \ker(A) = \{\vec{0}\}. \end{aligned} \tag{5}$$

9 Subspaces [9]

A set V of vectors is called a vector space[10] if for $\vec{x}, \vec{y} \in V$,

---

[9] This section is related to Chapter 3 of the textbook.

[10] More precisely, a set V which is closed under linear combinations is called a vector space or a linear space if the following rules are satisfied. Let $f, g, h \in V$ and $c, k \in \mathbb{R}$.

- $\vec{x} + \vec{y} \in V$,
- $\vec{0} \in V$,
- $-\vec{x} \in V$, and
- $k\vec{x} \in V$ for $k \in \mathbb{R}$.

For example, the set $\mathbb{R}^n$ of all (column) vectors with n components is a vector space. Also $\{\vec{0}\}$ is a vector space.

A subset W of a vector space V is called a subspace if

- W contains $\vec{0} \in V$,
- W is closed under addition ($\vec{w}_1 + \vec{w}_2 \in W$ for $\vec{w}_1, \vec{w}_2 \in W$), and
- W is closed under scalar multiplication ($k\vec{w} \in W$ for $\vec{w} \in W, k \in \mathbb{R}$).

Theorem 5. *For a linear transformation $T(\vec{x}) = A\vec{x}$ from $\mathbb{R}^m$ to $\mathbb{R}^n$,*

- $\ker(A)$ *is a subspace of* $\mathbb{R}^m$ *and*
- $\operatorname{im}(A)$ *is a subspace of* $\mathbb{R}^n$*.*

Proof. We can check that the solution set of $\vec{x} \in \mathbb{R}^m$ such that $A\vec{x} = \vec{0}$ satisfies the conditions for a subspace of $\mathbb{R}^m$. We can also confirm that the set of $\vec{y} \in \mathbb{R}^n$ such that $\vec{y} = A\vec{x}$ satisfies the conditions for a subspace of $\mathbb{R}^n$. □

Example 17. Consider $W = \{\vec{x} \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$. This set W is not a subspace of $\mathbb{R}^2$ because $k\vec{w} \notin W$ for $\vec{w} \in W$ and $k < 0$.

Example 18. The subspaces of $\mathbb{R}^2$ are $\mathbb{R}^2$, $\{\vec{0}\}$, and any of the lines through the origin.

We can show as follows that if W is a subspace of $\mathbb{R}^2$ which is neither $\{\vec{0}\}$ nor a line through the origin, then $W = \mathbb{R}^2$. Let $\vec{v}_1 \in W$ be a nonzero vector. Let L be the line spanned by $\vec{v}_1$. Since W is a subspace, $L \in W$. Since W is not a line, there exists a vector $\vec{v}_2 \in W$ which is not on L. We can express any vector $v \in \mathbb{R}^2$ as a linear combination of $\vec{v}_1$ and $\vec{v}_2$. However since W is a subspace, which is closed under linear combination, $\vec{v}$ is in W. That is, $W = \mathbb{R}^2$.

Remark 5. Similarly, the subspaces of $\mathbb{R}^3$ are $\mathbb{R}^3$, the planes through the origin, the lines through the origin, and $\{\vec{0}\}$.

(Addition) 1.1. $(f+g)+h = f+(g+h)$. 1.2. $f+g = g+f$. 1.3. There exists $0 \in V$ such that $f+0 = f$ for all f. 1.4. For each f there exists $-f$ such that $f+(-f) = 0$. (Scalar multiplication) 2.1. $k(f+g) = kf + kg$. 2.2. $(c+k)f = cf + kf$. 2.3. $c(kf) = (ck)f$. 2.4. $1f = f$. For example, a set of all $n \times m$ matrices is a vector space. Also, a set of all functions $f(x)$, $x \in \mathbb{R}$, is a vector space.

Example 19. Let us find $W_1 = \ker(\begin{bmatrix} 1 & 2 & 3 \end{bmatrix})$ and $W_2 = \operatorname{im}(\begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix})$.

Since $\vec{x} \in \ker(\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}) \subseteq \mathbb{R}^3$ satisfies

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0,$$

we see that W_1 is the plane $x_1 + 2x_2 + 3x_3 = 0$ in $\mathbb{R}^3$, and W_1 is a subspace of $\mathbb{R}^3$. Since

$$W_2 = \operatorname{span}(\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}),$$

and

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 0, \qquad \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 0,$$

we see that W_2 is the plane $x_1 + 2x_2 + 3x_3 = 0$. Thus $W_1 = W_2$. In general we can express a subspace as the kernel of the image of a linear transformation.

---

10 Bases [11]

Let us begin by recalling we have in Example 13

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix},$$

and in Example 14

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 6 \end{bmatrix} = (x_1 + 2x_2) \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

That is (Example 15),

$$\operatorname{im}(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}) = \operatorname{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}), \quad \operatorname{im}(\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}) = \operatorname{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix}).$$

[11] This section is related to Chapter 3 of the textbook.

Consider $\vec{v}_1, \dots, \vec{v}_m \in \mathbb{R}^n$.

- $\vec{v}_i$ is redundant if $\vec{v}_i$ is a linear combination of $\vec{v}_1, \dots, \vec{v}_{i-1}$.
- $\vec{v}_1, \dots, \vec{v}_m$ are linearly independent if none of them is redundant.
- $\vec{v}_1, \dots, \vec{v}_m$ form a basis of a subspace V of $\mathbb{R}^n$ if V is spanned by $\vec{v}_1, \dots, \vec{v}_m$ which are linearly independent.

Example 20. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3$ be $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 6 \end{bmatrix}$, and $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$, respectively. Then, the vector $\vec{v}_2$ is redundant, and the vectors $\vec{v}_1, \vec{v}_3$ form a basis of $\mathbb{R}^2$.

The above example shows that linearly independent column vectors of A form a basis of $\mathrm{im}(A)$.

Example 21. Consider the following four vectors.

$$\vec{v}_1 = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 7 \\ 0 \\ 4 \\ 0 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 6 \\ 0 \\ 3 \\ 7 \end{bmatrix}, \quad \vec{v}_4 = \begin{bmatrix} 3 \\ 5 \\ 2 \\ 4 \end{bmatrix}.$$

By looking at the second components and the fourth components, we find that $\vec{v}_1, \dots, \vec{v}_4$ are linearly independent.

Example 22. Consider the following three vectors.

$$\vec{v}_1 = \begin{bmatrix} 4 \\ 1 \\ 7 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 7 \\ 4 \\ 6 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix}.$$

To find their linear dependence, let us check if there exist solutions c_1, c_2 to the equation

$$c_1\vec{v}_1 + c_2\vec{v}_2 = \vec{v}_3 \quad \Leftrightarrow \quad \begin{bmatrix} 4 & 7 \\ 1 & 4 \\ 7 & 6 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix}$$

Since

$$\mathrm{rref}\left(\left[\begin{array}{cc:c} 4 & 7 & 6 \\ 1 & 4 & 3 \\ 7 & 6 & 5 \end{array}\right]\right) = \left[\begin{array}{cc:c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right],$$

we have $0 = 1$ in the third row. That is, there is no solution, and $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent.

Theorem 6. *The vectors $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{R}^n$ are linearly independent if and only if there exists only the trivial relation among them, i.e., the following (linear) relation holds only when $c_1 = \cdots = c_m = 0$.*

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}.$$

Proof. The theorem is proved if we prove that some of $\vec{v}_1, \ldots, \vec{v}_m$ are redundant if and only if there are nontrivial relations among them.

Suppose $\vec{v}_i$ is redundant. Then, using some constants $c_1, \ldots, c_{i-1}$ we have $\vec{v}_i = c_1\vec{v}_1 + \cdots + c_{i-1}\vec{v}_{i-1}$. Thus we have a nontrivial relation

$$c_1\vec{v}_1 + \cdots + c_{i-1}\vec{v}_{i-1} + (-1)\vec{v}_i + 0\vec{v}_{i+1} + \cdots + 0\vec{v}_m = 0.$$

Conversely, suppose that there is a nontrivial relation $c_1\vec{v}_1 + \cdots + c_{i-1}\vec{v}_{i-1} + \cdots + c_m\vec{v}_m = \vec{0}$, where i is the highest index such that $c_i \neq 0$, i.e., $c_{i+1} = \cdots = c_m = 0$. Then we have

$$\vec{v}_i = -\frac{c_1}{c_i}\vec{v}_1 - \cdots - \frac{c_{i-1}}{c_i}\vec{v}_{i-1},$$

and $\vec{v}_i$ is redundant. Thus the theorem is proved. □

Theorem 7. *We consider $\vec{v}_1, \ldots, \vec{v}_m$ in a subspace V of $\mathbb{R}^n$. Then, $\vec{v}_1, \ldots, \vec{v}_m$ form a basis of V if and only if any $\vec{v} \in V$ can be expressed uniquely as a linear combination*

$$\vec{v} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m,$$

where $c_1, \ldots, c_m$ are constants.

Proof. ($\Rightarrow$) Let us save this for homework.

($\Leftarrow$) Since $\vec{0} \in V$, there exist $c_1, \ldots, c_m$ such that

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}.$$

Since this representation is unique and the trivial relation with $c_1 = \cdots = c_m = 0$ satisfies the above relation, we have

$$c_1 = \cdots = c_m = 0.$$

Therefore $\vec{v}_1, \ldots, \vec{v}_m$ are linearly independent, and they form a basis of V. □

Let us consider an $n \times m$ matrix $A = [\vec{v}_1 \cdots \vec{v}_m]$. For $\vec{x} \in \ker(A)$ we have

$$\begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \vec{0} \quad \Leftrightarrow \quad x_1\vec{v}_1 + \cdots + x_m\vec{v}_m = \vec{0}.$$

That is, the column vectors of A are linearly independent if and only if $\ker(A) = \{\vec{0}\}$. In this case, there is no free variable and $\text{rank}(A) = m$. This condition implies $m \leq n$. Indeed, $\vec{v}_1, \ldots, \vec{v}_m$ cannot be linearly independent when $m > n$.

Example 23. Consider $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^2$ ($m = 3$, $n = 2$, and $m > n$). Let us suppose $\vec{v}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. Then any vector $\vec{v}_3$ is redundant.

Let us summarize equivalent statements for linear independence.

$$\begin{aligned} & \vec{v}_1, \ldots, \vec{v}_m \in \mathbb{R}^n \text{ are linearly independent} \\ \Leftrightarrow\ & \text{None of } \vec{v}_1, \ldots, \vec{v}_m \text{ is redundant} \\ \Leftrightarrow\ & \text{There doesn't exist any } \vec{v}_i \text{ such that } \vec{v}_i \text{ is a linear combination of} \\ & \vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_m \\ \Leftrightarrow\ & c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0} \text{ holds only when } c_1 = \cdots = c_m = 0 \\ \Leftrightarrow\ & \ker([\vec{v}_1 \cdots \vec{v}_m]) = \{\vec{0}\} \\ \Leftrightarrow\ & \text{rank}([\vec{v}_1 \cdots \vec{v}_m]) = m \end{aligned}$$

11 Dimension [12]

The word "dimension" was already used on page 2. Now we define the word as follows.

For a subspace V of $\mathbb{R}^n$, the number of vectors in a basis of V is called the dimension of V, denoted by $\dim(V)$.

We note that all bases of a subspace V of $\mathbb{R}^n$ consist of the same number of vectors. If $\dim(V)$ vectors in V are linearly independent, then they form a basis of V.

We have

$$\text{rank}(A) = \#\text{ of leading 1's} = \dim(\text{span(columns of } A)).$$

[12] This section is related to Chapter 3 of the textbook.

Example 24. Consider $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, which has $\text{rref}(A) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\text{rank}(A) = 2$. For $V = \text{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix})$, we have $\dim(V) = 2$.

Example 25. Consider $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$, which has $\text{rref}(A) = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ and $\text{rank}(A) = 1$. Since $V = \text{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix}) = \text{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix})$, we have $\dim(V) = 1$.

We note that a basis of $\text{im}(A)$, where $A = [\vec{v}_1 \cdots \vec{v}_m]$ is an $n \times m$ matrix, is formed by some vectors among $\vec{v}_1, \dots, \vec{v}_m$ (see Theorem 3, $\text{im}(A) = \text{span}(\vec{v}_1, \dots, \vec{v}_m)$) that correspond to the columns of $\text{rref}(A)$ containing the leading 1's. This fact implies th following theorem.

Theorem 8. *For an $n \times m$ matrix A,*

$$\text{rank}(A) = \dim(\text{im}(A)).$$

This gives another definition of the rank.

Theorem 9 (Rank-nullity theorem). *For an $n \times m$ matrix A,*

$$\dim(\ker(A)) = m - \dim(\text{im}(A)).$$

We call $\dim(\ker(A))$ *nullity. That is,*

$$\textit{(nullity of A)} + \textit{(rank of A)} = m.$$

Proof. We will show that $\dim(\ker(A))$ is the number of free variables. We suppose $\text{rank}(A) = k$. Let us consider a linear system $A\vec{x} = \vec{0}$, $\vec{x} \in \ker(A)$. Noting that $\ker(A) = \ker(\text{rref}(A))$ (by the way, $\text{im}(A) \neq \text{im}(\text{rref}(A))$ in general), we see that there are $m - k$ free variables. The solution vector $\vec{x}$ is written as

$$\vec{x} = t_1 \vec{w}_1 + \cdots + t_{m-k} \vec{w}_{m-k},$$

where $t_1, \dots, t_{m-k}$ are constants and $\vec{w}_1, \dots, \vec{w}_{m-k}$ form a basis of $\ker(A)$. Thus,

$$\dim(\ker(A)) = m - k = m - \text{rank}(A).$$

Together with the previous theorem, the proof is completed. $\square$

Example 26. Let us find bases of im(A) and ker(A), where

$$A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 & \vec{v}_5 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 1 & 2 \\ 1 & 2 & 0 & 2 & 3 \\ 1 & 2 & 0 & 3 & 4 \\ 1 & 2 & 0 & 4 & 5 \end{bmatrix}.$$

We see that $\vec{v}_2, \vec{v}_3, \vec{v}_5$ are redundant because

$$\vec{v}_2 = 2\vec{v}_1, \quad \vec{v}_3 = \vec{0}, \quad \vec{v}_5 = \vec{v}_1 + \vec{v}_4.$$

Since $\vec{v}_1, \vec{v}_4$ are linearly independent and

$$\operatorname{im}(A) = \operatorname{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5) = \operatorname{span}(\vec{v}_1, \vec{v}_4),$$

the vectors $\vec{v}_1, \vec{v}_4$ form a basis of im(A).

We can generate vectors in ker(A) using the redundant vectors $\vec{v}_2, \vec{v}_3, \vec{v}_5$. Since

$$-2\vec{v}_1 + \vec{v}_2 = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_5 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \vec{0},$$

$$\vec{v}_3 = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \vec{0},$$

and

$$-\vec{v}_1 - \vec{v}_4 + \vec{v}_5 = \begin{bmatrix} \vec{v}_1 & \vec{v}_4 & \vec{v}_5 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_5 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \vec{0},$$

we have $A\vec{w}_2 = \vec{0}$, $A\vec{w}_3 = \vec{0}$, $A\vec{w}_5 = \vec{0}$, where

$$\vec{w}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{w}_5 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

Note that $\vec{w}_2, \vec{w}_3, \vec{w}_5$ belong to $\ker(A)$. Moreover they are linearly independent (the components of $\vec{w}_i$ are zero below the ith component). Since there are three free variables, they span $\ker(A)$. Indeed, $\dim(\ker(A)) = m - \dim(\text{im}(A)) = 5 - 2 = 3$. Thus the three vectors $\vec{w}_2, \vec{w}_3, \vec{w}_5$ form a basis of $\ker(A)$.

———

A basis of the image of a matrix A is formed by linearly independent column vectors of A, and a basis of the kernel of A is formed by the vectors generated from redundant column vectors in A.

Theorem 10. *The vectors $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^n$ form a basis of $\mathbb{R}^n$ if and only if the matrix $[\vec{v}_1 \cdots \vec{v}_n]$ is invertible.*

Proof. According to Theorem 7, the vectors $\vec{v}_1, \ldots, \vec{v}_n$ form a basis of $\mathbb{R}^n$ if and only if each $\vec{b} \in \mathbb{R}^n$ can be uniquely written as

$$\vec{b} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

In order for this system to have a unique solution, the matrix $[\vec{v}_1 \cdots \vec{v}_n]$ must be invertible. □

At this point we can add three more equivalent statements to invertibility (5):

$$\begin{aligned} & \text{An } n \times n \text{ matrix } A \text{ is invertible} \\ \Leftrightarrow\ & \text{The column vectors of } A \text{ form a basis of } \mathbb{R}^n \\ \Leftrightarrow\ & \text{span(the column vectors of } A) = \mathbb{R}^n \\ \Leftrightarrow\ & \text{The column vectors of } A \text{ are linearly independent} \end{aligned} \tag{6}$$

12 Orthonormal bases [13]

Two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ are perpendicular or orthogonal if

$$\vec{v} \cdot \vec{w} = v_1 w_1 + \cdots + v_n w_n = 0.$$

A vector $\vec{x} \in \mathbb{R}^n$ is orthogonal to a subspace V of $\mathbb{R}^n$ if $\vec{x} \cdot \vec{v} = 0$ for all $\vec{v} \in V$.

The length of $\vec{v} \in \mathbb{R}^n$ is

$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{v_1^2 + \cdots + v_n^2}.$$

The length is also called the magnitude or norm. For a vector $\vec{v}$, we consider

$$\vec{u} = \frac{\vec{v}}{\|\vec{v}\|}.$$

We have $\|\vec{u}\| = 1$ and the vector $\vec{u}$ is a unit vector parallel to $\vec{v}$.

The vectors $\vec{u}_1, \ldots, \vec{u}_m \in \mathbb{R}^n$ are orthonormal if

$$\vec{u}_i \cdot \vec{u}_j = \delta_{ij}.$$

Here δ_{ij} is called the Kronecker delta and

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Example 27. Let $\vec{e}_1, \vec{e}_2, \vec{e}_3$ be the standard vectors:

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

We have $\vec{e}_i \cdot \vec{e}_j = \delta_{ij}$ $(i, j = 1, 2, 3)$.

Example 28. Consider

$$\vec{u}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \qquad \vec{u}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

[13] This section is related to Chapter 5 of the textbook.

We have $\|\vec{u}_1\| = \|\vec{u}_2\| = 1$ and $\vec{u}_1 \cdot \vec{u}_2 = 0$, and the vectors $\vec{u}_1, \vec{u}_2$ are orthonormal.

Theorem 11. *Orthonormal vectors $\vec{u}_1, \ldots, \vec{u}_m \in \mathbb{R}^n$ are linearly independent.*

Proof. Consider the linear combination

$$c_1\vec{u}_1 + \cdots + c_m\vec{u}_m = \vec{0},$$

where $c_1, \ldots, c_m$ are scalars. We form the dot product with $\vec{u}_i$:

$$c_1\vec{u}_1 \cdot \vec{u}_i + \cdots + c_m\vec{u}_m \cdot \vec{u}_i = \vec{0} \cdot \vec{u}_i.$$

Since $\vec{u}_1, \ldots, \vec{u}_m$ are orthonormal, we obtain

$$c_i = 0.$$

Since this holds for all $i = 1, \ldots, m$, we obtain $c_1 = \cdots = c_m = 0$. That is, $\vec{u}_1, \ldots, \vec{u}_m$ are linearly independent. □

Theorem 12. *Orthonormal vectors $\vec{u}_1, \ldots, \vec{u}_n \in \mathbb{R}^n$ form a basis of $\mathbb{R}^n$.*

Proof. Since n vectors $\vec{u}_1, \ldots, \vec{u}_n$ are linearly independent, they form a basis of $\mathbb{R}^n$. □

Theorem 13 (Orthogonal projection). *For $\vec{x} \in \mathbb{R}^n$ and a subspace V of $\mathbb{R}^n$, we can uniquely write*

$$\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp},$$

where $\vec{x}^{\parallel} \in V$, and $\vec{x}^{\perp}$ is perpendicular to V. The vector $\vec{x}^{\parallel} = \operatorname{proj}_V(\vec{x})$ is called the orthogonal projection of $\vec{x}$ onto V.

Theorem 14. *If V is a subspace of $\mathbb{R}^n$ with an orthonormal basis $\vec{u}_1, \ldots, \vec{u}_m$, then for all $\vec{x} \in \mathbb{R}^n$*

$$\operatorname{proj}_V(\vec{x}) = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_m \cdot \vec{x})\vec{u}_m.$$

Proof. With some scalars $c_1, \ldots, c_m$ we can write

$$\vec{x}^{\parallel} = c_1\vec{u}_1 + \cdots + c_m\vec{u}_m.$$

Since $\vec{x}^{\perp}$ is perpendicular to V, we have $\vec{u}_i \cdot \vec{x}^{\perp} = \vec{u}_i \cdot (\vec{x} - \vec{x}^{\parallel}) = 0$ and

$$\vec{u}_i \cdot \vec{x} - c_i = 0.$$

□

Remark 6. Recall $\operatorname{proj}_L(\vec{x}) = (\vec{u} \cdot \vec{x})\vec{u}$ in the two-dimensional case (4).

The above theorem implies that for an orthonormal basis $\vec{u}_1, \dots, \vec{u}_n$ of $\mathbb{R}^n$, we have

$$\vec{x} = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_n \cdot \vec{x})\vec{u}_n, \qquad \vec{x} \in \mathbb{R}^n.$$

Consider a subspace V of $\mathbb{R}^n$. The set of $\vec{x}^{\perp}$, i.e.,

$$\{\vec{x} \in \mathbb{R}^n : \vec{v} \cdot \vec{x} = 0 \quad \text{for all } \vec{v} \in V\}$$

is called the orthogonal complement $V^{\perp}$ of V.

Theorem 15. *Let us write* $\text{proj}_V(\vec{x}) = A\vec{x}$*, where* V *is a subspace of* $\mathbb{R}^n$*. Then,*

$$V^{\perp} = \ker(A).$$

Moreover, $V^{\perp}$ *is a subspace of* $\mathbb{R}^n$*, and* $V \cap V^{\perp} = \{\vec{0}\}$.

Proof. We note that $V = \text{im}(A)$. We have $\text{proj}_V(\vec{x}) = \vec{0}$ if $\vec{x} \in \ker(A)$. Therefore, for any such $\vec{x}$

$$\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp} = \vec{x}^{\perp}.$$

Thus $V^{\perp} = \ker(A)$. Because the kernel is a subspace, $V^{\perp}$ is a subspace of $\mathbb{R}^n$. Let us suppose there exists $\vec{x} \in \mathbb{R}^n$ such that $\vec{x} \in V$ and $\vec{x} \in V^{\perp}$. Since $\vec{x} \in V^{\perp}$, we have $\vec{v} \cdot \vec{x} = 0$ for all $\vec{v} \in V$, and in particular $\vec{x} \cdot \vec{x} = \|\vec{x}\|^2 = 0$. That is, $\vec{x} = \vec{0}$. □

Example 29. For example, V is a line through the origin in $\mathbb{R}^2$ or $\mathbb{R}^3$, or a plane through the origin in $\mathbb{R}^3$. Let A be the matrix of the orthogonal projection onto V. Then for any $\vec{x}$ we have $A\vec{x} = \vec{x}^{\parallel}$, which implies $V = \text{im}(A)$. We note that $A\vec{x}^{\parallel} = \vec{x}^{\parallel}$. Since $A\vec{x}^{\perp} = A(\vec{x} - \vec{x}^{\parallel}) = \vec{0}$, the vector $\vec{x}^{\perp}$ belongs to $\ker(A)$. The set of $\vec{x}^{\perp}$ is $V^{\perp}$. Hence $V^{\perp} = \ker(A)$.

Theorem 16 (Rank-nullity theorem). *We have*

$$\dim(V) + \dim(V^{\perp}) = \dim(\mathbb{R}^n) = n.$$

We also have

$$\left(V^{\perp}\right)^{\perp} = V.$$

Proof. We use the rank-nullity theorem for $A\vec{x} = \text{proj}_V(\vec{x})$ from $\mathbb{R}^n$ to $\mathbb{R}^n$:

$$\dim(\text{im}(A)) + \dim(\ker(A)) = n.$$

Note that $\text{im}(A) = V$ and $\ker(A) = V^{\perp}$. We note that $V \subseteq \left(V^{\perp}\right)^{\perp}$. But we have $\dim(V^{\perp}) + \dim(\left(V^{\perp}\right)^{\perp}) = n$, which implies $\dim(V) = \dim(\left(V^{\perp}\right)^{\perp})$. Therefore $\left(V^{\perp}\right)^{\perp} = V$. □

13 Gram-Schmidt process [14]

Consider a subspace V of $\mathbb{R}^n$ with $\dim(V) = m$. Let us construct an orthonormal basis of V, $\vec{u}_1, \ldots, \vec{u}_m$ from a given basis $\vec{v}_1, \ldots, \vec{v}_m$ of V.

Step 1

$$\vec{u}_1 = \frac{1}{\|\vec{v}_1\|}\vec{v}_1.$$

Step 2

We write $\vec{v}_2 = \vec{v}_2^{\|} + \vec{v}_2^{\perp}$, where

$$\vec{v}_2^{\|} = \text{proj}_{\vec{u}_1}(\vec{v}_2) = (\vec{u}_1 \cdot \vec{v}_2)\vec{u}_1,$$

and

$$\vec{v}_2^{\perp} = \vec{v}_2 - \vec{v}_2^{\|} = \vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\vec{u}_1.$$

Step 3

$$\vec{u}_2 = \frac{1}{\|\vec{v}_2^{\perp}\|}\vec{v}_2^{\perp}.$$

Step 4

Repeat Step 2 and Step 3. Note that when we write $\vec{v}_3 = \vec{v}_3^{\|} + \vec{v}_3^{\perp}$,

$$\vec{v}_3^{\|} = \text{proj}_{(\vec{u}_1, \vec{u}_2)}(\vec{v}_3) = (\vec{u}_1 \cdot \vec{v}_3)\vec{u}_1 + (\vec{u}_2 \cdot \vec{v}_3)\vec{u}_2.$$

In general we have ($j \geq 2$)

$$\vec{v}_j^{\|} = \text{proj}_{(\vec{u}_1, \ldots, \vec{u}_{j-1})}(\vec{v}_j) = (\vec{u}_1 \cdot \vec{v}_j)\vec{u}_j + \cdots + (\vec{u}_{j-1} \cdot \vec{v}_j)\vec{u}_{j-1}.$$

Example 30. Consider a plane V spanned by

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

[14] This section is related to Chapter 5 of the textbook.

That is, any $\vec{v} \in V$ is expressed as $\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2$ with some scalars c_1, c_2. Let us obtain an orthonormal basis $\vec{u}_1, \vec{u}_2$ of V.

We first obtain

$$\vec{u}_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Next we compute $\vec{v}_2^{\parallel}$ as

$$\vec{v}_2^{\parallel} = \left(\frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right) \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \frac{7}{5} \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Thus we have

$$\vec{v}_2^{\perp} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \frac{7}{5} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

Finally we obtain

$$\vec{u}_2 = \frac{1}{\sqrt{10}/5} \frac{1}{5} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

We can readily check that

$$\|\vec{u}_1\| = \|\vec{u}_2\| = 1, \qquad \vec{u}_1 \cdot \vec{u}_2 = 0.$$

---

14 Orthogonal matrices [15]

A linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$ is said to be orthogonal if

$$\|T(\vec{x})\| = \|\vec{x}\| \qquad \text{for all } \vec{x} \in \mathbb{R}^n.$$

Then the matrix A of T is called an orthogonal matrix.

Example 31. Let us consider the matrix A of the counterclockwise rotation T in $\mathbb{R}^2$ through angle θ. We have

$$T(\vec{x}) = A\vec{x} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \vec{x}.$$

[15] This section is related to Chapter 5 of the textbook.

The linear transformation T is orthogonal and A is an orthogonal matrix. Indeed,

$$\begin{aligned}\|T(\vec{x})\| &= \left\|\begin{bmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \end{bmatrix}\right\| = \sqrt{(x_1\cos\theta - x_2\sin\theta)^2 + (x_1\sin\theta + x_2\cos\theta)^2} \\ &= \sqrt{x_1^2 + x_2^2} = \|\vec{x}\|.\end{aligned}$$

Theorem 17. *An $n \times n$ matrix A is orthogonal if and only if the column vectors of A form an orthonormal basis of $\mathbb{R}^n$.*

Proof. ($\Rightarrow$) Let us write the column vectors as $\vec{v}_1, \ldots, \vec{v}_n$, where $A = [\vec{v}_1 \cdots \vec{v}_n]$. We note that $\vec{v}_i = A\vec{e}_i$ $(i = 1, \ldots, n)$. First we observe that

$$\|A\vec{e}_i\| = \|\vec{e}_i\| = 1.$$

We also have

$$\|A\vec{e}_i + A\vec{e}_j\|^2 = \|A(\vec{e}_i + \vec{e}_j)\|^2 = \|\vec{e}_i + \vec{e}_j\|^2 = \|\vec{e}_i\|^2 + \|\vec{e}_j\|^2 = \|A\vec{e}_i\|^2 + \|A\vec{e}_j\|^2.$$

By the Pythagorean theorem, $A\vec{e}_1, A\vec{e}_2, \ldots, A\vec{e}_n$ are orthonormal.

($\Leftarrow$) We have

$$\begin{aligned}\|A\vec{x}\| &= \left\|\begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right\| \\ &= \|x_1\vec{v}_1 + \cdots + x_n\vec{v}_n\| \\ &= \left[(x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) \cdot (x_1\vec{v}_1 + \cdots + x_n\vec{v}_n)\right]^{1/2} \\ &= \left(x_1^2 + \cdots + x_n^2\right)^{1/2} \\ &= \|\vec{x}\|.\end{aligned}$$

□

15 Transpose [16]

The transpose A^T of A is a matrix such that

$$\{A^T\}_{ij} = A_{ji}.$$

Example 32.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

- A is said to be symmetric if $A^T = A$.
- A is said to be skew-symmetric if $A^T = -A$.

There are the following properties.

- $(A+B)^T = A^T + B^T$,
- $(kA)^T = kA^T$,
- $(AB)^T = B^T A^T$,
- $\mathrm{rank}(A^T) = \mathrm{rank}(A)$,
- $(A^T)^{-1} = (A^{-1})^T$.

The last property can be understood from the fact that $AA^{-1} = I_n$ and $(AA^{-1})^T = (A^{-1})^T A^T = I_n$.

If the columns $\vec{v}_1, \ldots, \vec{v}_n$ of A are orthonormal, then we have

$$A^T A = \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = I_n.$$

For an $n \times n$ matrix, we have the following equivalent statements.

$$\begin{aligned} & A \text{ is an } n \times n \text{ orthogonal matrix} \\ \Leftrightarrow\ & \|A\vec{x}\| = \|\vec{x}\| \text{ for all } \vec{x} \in \mathbb{R}^n \\ \Leftrightarrow\ & \text{The columns of } A \text{ form an orthonormal basis of } \mathbb{R}^n \\ \Leftrightarrow\ & A^T A = I_n \\ \Leftrightarrow\ & A^{-1} = A^T \end{aligned}$$

[16] This section is related to Chapter 5 of the textbook.

Example 33. Recall the rotation in the x-y plane through θ is represented by

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

We have

$$R_\theta^{-1} = R_{-\theta} = R_\theta^T.$$

---

16 Least squares [17]

Let $A = [\vec{v}_1 \cdots \vec{v}_m]$ be an $n \times m$ matrix.

Theorem 18.

$$\ker(A^T) = (\mathrm{im}(A))^\perp.$$

Proof. The image of A is a subspace V of $\mathbb{R}^n$.

$$\begin{aligned} V^\perp &= \{\vec{x} \in \mathbb{R}^n : \vec{v}_i \cdot \vec{x} = 0,\ i = 1, \dots, m\} \\ &= \left\{ \vec{x} \in \mathbb{R}^n : \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_m^T \end{bmatrix} \vec{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \right\} \\ &= \left\{ \vec{x} \in \mathbb{R}^n : A^T\vec{x} = \vec{0} \right\} \\ &= \ker(A^T). \end{aligned}$$

□

Theorem 19.

$$\ker(A) = \ker(A^TA).$$

Proof. (i) If $A\vec{x} = \vec{0}$, then $A^TA\vec{x} = \vec{0}$. Therefore $\ker(A) \subseteq \ker(A^TA)$. (ii) If $A^TA\vec{x} = \vec{0}$, then $A\vec{x} \in \ker(A^T)$. However $\ker(A^T) = (\mathrm{im}(A))^\perp$. Of course, $A\vec{x} \in \mathrm{im}(A)$. Hence $A\vec{x} = \vec{0}$. This means $\vec{x} \in \ker(A)$ and $\ker(A) \supseteq \ker(A^TA)$. By (i) and (ii), we obtain $\ker(A) = \ker(A^TA)$. □

Theorem 20. *If* $\ker(A) = \{\vec{0}\}$*, then* A^TA *is invertible.*

Proof. The matrix A^TA is an $m \times m$ square matrix. Using Theorem 19, $\ker(A^TA) = \{\vec{0}\}$. Thus A^TA is invertible. □

[17] This section is related to Chapter 5 of the textbook.

Theorem 21 (Orthogonal projection). *For $\vec{x} \in \mathbb{R}^n$ and a subspace V of $\mathbb{R}^n$,*

$$\|\vec{x} - \text{proj}_V(\vec{x})\| \le \|\vec{x} - \vec{v}\|$$

for all $\vec{v} \in V$.

Consider an inconsistent system (there is no solution) $A\vec{x} = \vec{b}$. That is, $\vec{b} \notin \text{im}(A)$. We look for an approximate solution $\vec{x}^*$ by minimizing the error $\|\vec{b} - A\vec{x}\|$.

Consider $A\vec{x} = \vec{b}$ with an $n \times m$ matrix A. Then $\vec{x}^* \in \mathbb{R}^m$ is called a least-squares solution if

$$\|\vec{b} - A\vec{x}^*\| \le \|\vec{b} - A\vec{x}\|$$

for all $\vec{x} \in \mathbb{R}^m$.

Remark 7. If $A\vec{x} = \vec{b}$ is consistent, then $\vec{x}^*$ is a solution.

We can find $\vec{x}^*$ as follows.

$$\begin{aligned}
& \|\vec{b} - A\vec{x}^*\| \le \|\vec{b} - A\vec{x}\| \quad \text{for all } \vec{x} \in \mathbb{R}^m \\
\Leftrightarrow \quad & A\vec{x}^* = \text{proj}_V(\vec{b}), \quad V = \text{im}(A) \\
\Leftrightarrow \quad & \vec{b} = \vec{b}^{\parallel} + \vec{b}^{\perp} = \text{proj}_V(\vec{b}) + \vec{b}^{\perp}, \quad \vec{b} - \text{proj}_V(\vec{b}) \in V^{\perp}, \quad \text{and} \quad \vec{b} - A\vec{x}^* \in \ker(A^T) \\
\Leftrightarrow \quad & A^T\left(\vec{b} - A\vec{x}^*\right) = \vec{0} \\
\Leftrightarrow \quad & A^T A\vec{x}^* = A^T\vec{b}
\end{aligned}$$

The equation $A^T A\vec{x}^* = A^T\vec{b}$ is said to be the normal equation of $A\vec{x} = \vec{b}$. If $\ker(A) = \{\vec{0}\}$, then

$$\vec{x}^* = \left(A^T A\right)^{-1} A^T\vec{b}.$$

Remark 8. The matrix $A^+ = \left(A^T A\right)^{-1} A^T$ is called the pseudoinverse.

Example 34. Prof. M eats w_i (lb) ice cream and w_s (lb) steak every month and his weight increases by Δw (lb).

	w_i	w_s	Δw
May	1	4	2
June	1	8	4
July	1	12	5
August	3	8	?

Predict Δw for August by finding a formula $c_i w_i + c_s w_s = \Delta w$.

Let us write $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & 4 \\ 1 & 8 \\ 1 & 12 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} c_i \\ c_s \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}.$$

We have

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 8 & 12 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 8 \\ 1 & 12 \end{bmatrix} = \begin{bmatrix} 3 & 24 \\ 24 & 224 \end{bmatrix}.$$

Thus,

$$(A^T A)^{-1} = \frac{1}{96} \begin{bmatrix} 224 & -24 \\ -24 & 3 \end{bmatrix}.$$

We obtain

$$\begin{aligned} \vec{x}^* &= \frac{1}{96} \begin{bmatrix} 224 & -24 \\ -24 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 4 & 8 & 12 \end{bmatrix} \vec{b} \\ &= \frac{1}{96} \begin{bmatrix} 128 & 32 & -64 \\ -12 & 0 & 12 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} \\ &= \frac{1}{96} \begin{bmatrix} 64 \\ 36 \end{bmatrix} \\ &= \begin{bmatrix} 2/3 \\ 3/8 \end{bmatrix}. \end{aligned}$$

This means

$$\frac{2}{3} w_i + \frac{3}{8} w_s = \Delta w.$$

Finally we obtain

$$\Delta w \text{ (Aug)} = \frac{2}{3} \cdot 3 + \frac{3}{8} \cdot 8 = 5 \, \text{lb}.$$

Remark 9. We can see the relation to linear regression as follows. Let us use the above example. For simplicity we assume $w_i = 1$ for every month. We will find

$$c_i^* + c_s^* w_s = \Delta w.$$

We write

$$A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ 1 & a_3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

We then obtain

$$\begin{aligned}
\begin{bmatrix} c_i^* \\ c_s^* \end{bmatrix} &= (A^T A)^{-1} A^T \vec{b} = \left(\begin{bmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ 1 & a_3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \\
&= \begin{bmatrix} 3 & \sum_{i=1}^3 a_i \\ \sum_{i=1}^3 a_i & \sum_{i=1}^3 a_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^3 b_i \\ \sum_{i=1}^3 a_i b_i \end{bmatrix} \\
&= \frac{1}{3\left(\sum_{i=1}^3 a_i^2\right) - \left(\sum_{i=1}^3 a_i\right)^2} \begin{bmatrix} \sum_{i=1}^3 a_i^2 & -\sum_{i=1}^3 a_i \\ -\sum_{i=1}^3 a_i & 3 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^3 b_i \\ \sum_{i=1}^3 a_i b_i \end{bmatrix}.
\end{aligned}$$

Therefore we obtain

$$\begin{aligned}
c_s^* &= \frac{3\left(\sum_{i=1}^3 a_i b_i\right) - \left(\sum_{i=1}^3 a_i\right)\left(\sum_{i=1}^3 b_i\right)}{3\left(\sum_{i=1}^3 a_i^2\right) - \left(\sum_{i=1}^3 a_i\right)^2} = \frac{\sum_{i=1}^3 (a_i - \bar{a})(b_i - \bar{b})}{\sum_{i=1}^3 (a_i - \bar{a})^2}, \\
c_i^* &= \frac{\left(\sum_{i=1}^3 a_i^2\right)\left(\sum_{i=1}^3 b_i\right) - \left(\sum_{i=1}^3 a_i\right)\left(\sum_{i=1}^3 a_i b_i\right)}{3\left(\sum_{i=1}^3 a_i^2\right) - \left(\sum_{i=1}^3 a_i\right)^2} = \bar{b} - c_s^* \bar{a},
\end{aligned}$$

where $\bar{a} = (1/3)\sum_{i=1}^3 a_i$ and $\bar{b} = (1/3)\sum_{i=1}^3 b_i$. These formulae appear in linear regression. Recall $a_1 = 4$, $a_2 = 8$, $a_3 = 12$, $b_1 = 2$, $b_2 = 4$, $b_3 = 5$. The above formulae give $c_i^* = 2/3$ and $c_s^* = 3/8$.

17 Determinants [18]

Let us consider a $n \times n$ matrix A. We can compute the reduced row-echelon form $\mathrm{rref}(A)$ with elementary row operations:

$$A \quad \to \quad B_1 \quad \to \quad B_2 \quad \to \quad \cdots \quad \to \quad \mathrm{rref}(A).$$

If $\mathrm{rref}(A) \neq I_n$, then

$$\det(A) = 0. \tag{7}$$

This implies that a square matrix A is invertible if and only if $\det(A) \neq 0$. We can also say that if column vectors are not linearly independent, $\det(A) = 0$.

[18] This section is related to Chapter 6 of the textbook.

Hereafter we assume $\text{rref}(A) = I_n$.

$$\det(I_n) = 1.$$

The determinant of A is obtained as follows. Suppose we divide some row by k when moving from B_{i-1} to B_i. Then,

$$\det(B_{i-1}) = k \det(B_i).$$

If we swap two rows, then

$$\det(B_{i-1}) = -\det(B_i).$$

The determinant doesn't change by addition or subtraction of a scalar multiple of one row. We obtain

$$\det(A) = (-1)^{\#\text{ of swaps}} \prod k.$$

Example 35. For $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, we know that $\det(A) = 1 \cdot 4 - 2 \cdot 3 = -2$. Let us obtain $\det(A)$ using the above method.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{\text{2nd}-3\cdot\text{1st}} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} \xrightarrow{\text{2nd}/(-2)} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \xrightarrow{\text{1st}-2\cdot\text{2nd}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We have

$$\det(A) = (-1)^{\#\text{ of swaps}} \prod k = (-1)^0 \,(-2) = -2.$$

Example 36. Let us calculate $\det(A)$, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 4 & 5 \end{bmatrix}.$$

We have

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 4 & 5 \end{bmatrix} \xrightarrow{\text{2nd}-2\cdot\text{1st and 3rd}-3\cdot\text{1st}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & -1 \\ 0 & -2 & -4 \end{bmatrix}$$

$$\xrightarrow{\text{2nd}/(-1)\text{ and 3rd}/(-2)} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\xrightarrow{\text{interchange 2nd and 3rd}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{\text{1st}-2\cdot\text{2nd}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{\text{1st}+\text{3rd and 2nd}-2\cdot\text{3rd}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence we obtain

$$\det(A) = (-1)^{\#\text{ of swaps}} \prod k = (-1)^1\ (-1)(-2) = -2.$$

From the above examples we see the following theorem.

Theorem 22. *The determinant of an (upper or lower) triangular matrix is the product of the diagonal entries of the matrix. In particular, the determinant of a diagonal matrix is the product of its diagonal entries.*

For 3×3 matrices, there is a well-known formula called Sarrus's rule.

The determinant of the matrix A, where

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

is obtained as

$$\det(A) = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}.$$

Example 37. Let us calculate $\det(A)$, where

$$\begin{aligned}\det\begin{bmatrix}1&2&3\\2&4&5\\3&4&5\end{bmatrix} &= 1\cdot4\cdot5+2\cdot5\cdot3+3\cdot2\cdot4-3\cdot4\cdot3-1\cdot5\cdot4-2\cdot2\cdot5\\ &= 20+30+24-36-20-20=-2.\end{aligned}$$

Let us consider an $n\times n$ matrix A, where

$$A=\begin{bmatrix}a_{11}&\cdots&a_{1n}\\\vdots&&\vdots\\a_{n1}&\cdots&a_{nn}\end{bmatrix}.$$

There is no useful formula as Sarrus's rule if $n\geq 4$. But we have the following theorem.

Theorem 23 (Laplace expansion (cofactor expansion)). *We can compute* $\det(A)$ *by Laplace expansion down the (any)* j*th column*

$$\det(A)=\sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det(A_{ij}),$$

or by Laplace expansion along the (any) i*th row*

$$\det(A)=\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det(A_{ij}).$$

The matrix A_{ij} is the $(n-1)\times(n-1)$ matrix obtained by omitting the ith row and the jth column of an $n\times n$ matrix A. A minor of A is $\det(A_{ij})$, and $(-1)^{i+j}\det(A_{ij})$ is called a cofactor of A.

Example 38. Let us use the Laplace expansion down the 1st column.

$$\begin{aligned}\det\begin{bmatrix}1&2\\3&4\end{bmatrix} &= (-1)^{1+1}a_{11}\det(A_{11})+(-1)^{2+1}a_{21}\det(A_{21})\\ &= 1\cdot\det(4)-3\cdot\det(2)=1\cdot4-3\cdot2=-2.\end{aligned}$$

Example 39. Sarrus's rule can be written as

$$\det(A) = a_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{21}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix} + a_{31}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}.$$

There are the following properties for determinants.

- $\det(A^T) = \det(A)$,
- $\det(AB) = \det(A)\det(B)$,
- $\det(A^{-1}) = 1/\det(A)$.

The last property is shown as follow. We take determinants of $I_n = AA^{-1}$. The left-hand side is $\det(I_n) = 1$. The right-hand side is $\det(AA^{-1}) = \det(A)\det(A^{-1})$.

If A is orthogonal ($A^{-1} = A^T$), then $\det(A) = \pm 1$ because

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

Let us consider the following calculations. The first one is

$$\begin{aligned} \det\begin{bmatrix} a & b \\ c & d \end{bmatrix} &\overset{\text{transpose}}{=} \det\begin{bmatrix} a & c \\ b & d \end{bmatrix} \overset{\text{swap}}{=} -\det\begin{bmatrix} b & d \\ a & c \end{bmatrix} \\ &\overset{\text{transpose}}{=} -\det\begin{bmatrix} b & a \\ d & c \end{bmatrix}, \end{aligned}$$

and the second one is

$$\begin{aligned} \det\begin{bmatrix} a & b+b' \\ c & d+d' \end{bmatrix} &= a\det\begin{bmatrix} 1 & \frac{b+b'}{a} \\ c & d+d' \end{bmatrix} = a\det\begin{bmatrix} 1 & \frac{b+b'}{a} \\ 0 & d+d'-c\frac{b+b'}{a} \end{bmatrix} \\ &= a\left[d+d'-(b+b')\frac{c}{a}\right]\det\begin{bmatrix} 1 & \frac{b+b'}{a} \\ 0 & 1 \end{bmatrix} \\ &= a(d+d')-(b+b')c \\ &= ad-bc+ad'-b'c \\ &= \det\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \det\begin{bmatrix} a & b' \\ c & d' \end{bmatrix}. \end{aligned}$$

In general we have the following properties about column vectors.

- Alternating:

$$\det([\cdots \vec{v}_i \cdots \vec{v}_j \cdots]) = -\det([\cdots \vec{v}_j \cdots \vec{v}_i \cdots]),$$

- Multilinear:

$$\det([\cdots (\alpha \vec{v}_i + \beta \vec{v}_{i'}) \cdots]) = \alpha \det([\cdots \vec{v}_i \cdots]) + \beta \det([\cdots \vec{v}_{i'} \cdots]).$$

The above two relations imply that if there is a redundant vector, then the determinant is zero.

Using the multilinearity we can derive the Laplace expansion as follows. Let us use the Laplace expansion down the 1st column for a 3×3 matrix.

$$\begin{aligned}
&\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\
&= a_{11} \det \begin{bmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} + a_{21} \det \begin{bmatrix} 0 & a_{12} & a_{13} \\ 1 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} + a_{31} \det \begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 1 & a_{32} & a_{33} \end{bmatrix} \\
&= a_{11} \det \begin{bmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} - a_{21} \det \begin{bmatrix} 1 & a_{22} & a_{23} \\ 0 & a_{12} & a_{13} \\ 0 & a_{32} & a_{33} \end{bmatrix} + a_{31} \det \begin{bmatrix} 1 & a_{32} & a_{33} \\ 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \end{bmatrix} \\
&= a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{21} \det \begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix} + a_{31} \det \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix} \\
&= (-1)^{1+1} a_{11} \det(A_{11}) + (-1)^{2+1} a_{21} \det(A_{21}) + (-1)^{3+1} a_{31} \det(A_{31}).
\end{aligned}$$

Using determinants and cofactors we can write the inverse of an $n \times n$ matrix A.

Theorem 24. *For an invertible $n \times n$ matrix A,*

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A),$$

where $\operatorname{adj}(A)$ *is the classical adjoint of A defined by*

$$\operatorname{adj}(A) = \begin{bmatrix} (-1)^{1+1} \det(A_{11}) & (-1)^{1+2} \det(A_{21}) & \cdots \\ (-1)^{2+1} \det(A_{12}) & (-1)^{2+2} \det(A_{22}) & \cdots \\ \vdots & & \end{bmatrix}.$$

Note that $\{\mathrm{adj}(A)\}_{ij} = (-1)^{i+j}\det(A_{ji})$.

Example 40. The inverse of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

$$\begin{aligned} A^{-1} &= \frac{1}{\det(A)}\begin{bmatrix} (-1)^{1+1}\det(A_{11}) & (-1)^{1+2}\det(A_{21}) \\ (-1)^{2+1}\det(A_{12}) & (-1)^{2+2}\det(A_{22}) \end{bmatrix} \\ &= \frac{1}{ad-bc}\begin{bmatrix} a & -b \\ -c & d \end{bmatrix}. \end{aligned}$$

---

18 Geometrical interpretations of determinants [19]

Let us consider geometrical interpretations of determinants. We begin with a 2×2 matrix with linearly independent column vectors:

$$A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Let us recall the Gram-Schmidt process and construct orthonormal vectors $\vec{u}_1, \vec{u}_2$. We have

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}, \qquad \vec{u}_2 = \frac{\vec{v}_2^{\perp}}{\|\vec{v}_2^{\perp}\|},$$

where

$$\vec{v}_2^{\perp} = \vec{v}_2 - \mathrm{proj}_{\vec{u}_1}(\vec{v}_2) = \vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\vec{u}_1.$$

We can write the matrix A as

$$A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} = QR = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}\begin{bmatrix} \|\vec{v}_1\| & \vec{u}_1 \cdot \vec{v}_2 \\ 0 & \|\vec{v}_2^{\perp}\| \end{bmatrix}.$$

Indeed this is called the QR factorization, where $Q = [\vec{u}_1\vec{u}_2]$ is an orthogonal matrix and R is an upper triangular matrix with positive diagonal entries.

We have

[19] This section is related to Chapter 6 of the textbook.

$$\begin{aligned}\sqrt{\det(A^TA)} &= \sqrt{(\det(A))^2} = |\det(A)| \\ &= |\det(Q)|\ |\det(R)| \\ &= |\det(R)| \\ &= \|\vec{v}_1\|\ \left\|\vec{v}_2^{\perp}\right\|.\end{aligned}$$

This is the area of the parallelogram defined by $\vec{v}_1, \vec{v}_2$.

Theorem 25. *Consider an $n \times m$ matrix $A = [\vec{v}_1 \cdots \vec{v}_m]$, where $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{R}^n$. $\sqrt{\det(A^TA)}$ is the m-volume of the m-parallelepiped defined by $\vec{v}_1, \ldots, \vec{v}_m$. We note that $\sqrt{\det(A^TA)} = |\det(A)|$ if $m = n$.*

Proof. By using the Gram-Schmidt process we obtain the QR factorization

$$A = QR = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{bmatrix} \begin{bmatrix} \|\vec{v}_1\| & \vec{u}_1 \cdot \vec{v}_2 & \cdots & \vec{u}_1 \cdot \vec{v}_m \\ & \left\|\vec{v}_2^{\perp}\right\| & \cdots & \vec{u}_2 \cdot \vec{v}_m \\ & & \ddots & \vdots \\ & & & \left\|\vec{v}_m^{\perp}\right\| \end{bmatrix}.$$

Thus we have

$$\begin{aligned}\sqrt{\det(A^TA)} &= \sqrt{\det(R^TQ^TQR)} = \sqrt{\det(R^T)\det(Q^T)\det(Q)\det(R)} \\ &= \sqrt{(\det(R))^2} \\ &= |\det(R)| \\ &= \|\vec{v}_1\|\ \left\|\vec{v}_2^{\perp}\right\|\ \cdots\ \left\|\vec{v}_m^{\perp}\right\|.\end{aligned}$$

□

Example 41. The determinant $|\det(A)| = |\det[\vec{v}_1\vec{v}_2\vec{v}_3]| = \|\vec{v}_1\|\ \left\|\vec{v}_2^{\perp}\right\|\ \left\|\vec{v}_3^{\perp}\right\|$ is the volume of the parallelepiped.

Consider a linear transformation $T(\vec{x}) = A\vec{x}$ from $\mathbb{R}^2$ to $\mathbb{R}^2$. Let Ω be the parallelogram defined by $\vec{v}_1$ and $\vec{v}_2$. The parallelogram defined by $A\vec{v}_1$ and $A\vec{v}_2$ is denoted by $T(\Omega)$. If we define $B = [\vec{v}_1\vec{v}_2]$, then

$$\text{area of } \Omega = |\det(B)|,$$

and

$$\text{area of } T(\Omega) = \left|\det \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 \end{bmatrix}\right| = |\det(AB)| = |\det(A)|\,|\det(B)|.$$

We obtain

$$\frac{\text{area of } T(\Omega)}{\text{area of } \Omega} = |\det(A)|.$$

The ratio on the left-hand side is called the expansion factor.

Theorem 26. *For a linear transformation* $T(\vec{x}) = A\vec{x}$ *from* $\mathbb{R}^n$ *to* $\mathbb{R}^n$, $|\det(A)|$ *is the expansion factor of* T *on* n*-parallelepipeds. That is,*

$$V(A\vec{v}_1, \ldots, A\vec{v}_n) = |\det A| V(\vec{v}_1, \ldots, \vec{v}_n),$$

for all $\vec{v}_1, \ldots, \vec{v}_n$ *in* $\mathbb{R}^n$.

Example 42. The rotation matrix $R_y(\theta)$ about the y-axis in space is given by

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}.$$

Since $\det(R_y(\theta)) = 1$, this matrix doesn't change the volume.

19 Eigenvalues and eigenvectors [20]

Let us suppose the relation

$$A\vec{v} = \lambda\vec{v},$$

holds for an $n \times n$ matrix A, a scalar λ, and a nonzero vector $\vec{v}$. λ is said to be an eigenvalue and $\vec{v}$ is said to be the eigenvector associated with λ. A basis $\vec{v}_1, \ldots, \vec{v}_n$ of $\mathbb{R}^n$ is called an eigenbasis for A if these vectors are eigenvectors of A.

Theorem 27. *For an* $n \times n$ *matrix* A*, a scalar* λ *is an eigenvalue of* A *if and only if*

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

Here $f_A(\lambda)$ *is the characteristic polynomial, and* $f_A(\lambda) = 0$ *is called the characteristic equation or the secular equation.*

[20] This section is related to Chapter 7 of the textbook.

Proof.

$$\begin{aligned} & A\vec{v} = \lambda\vec{v} \quad (\vec{v} \neq \vec{0}) \\ \Leftrightarrow \quad & A\vec{v} - \lambda\vec{v} = \vec{0} \\ \Leftrightarrow \quad & (A - \lambda I_n)\vec{v} = \vec{0} \\ \Leftrightarrow \quad & \ker(A - \lambda I_n) \neq \{\vec{0}\} \\ \Leftrightarrow \quad & A - \lambda I_n \text{ is not invertible} \\ \Leftrightarrow \quad & \det(A - \lambda I_n) = 0. \end{aligned}$$

□

Remark 10. If $\lambda = 0$, then A is noninvertible.

$$A \text{ is invertible} \quad \Leftrightarrow \quad \text{eigenvalue } \lambda \neq 0.$$

Together with (7), we can add two more equivalent statements to (5) and (6):

$$\begin{aligned} & \text{An } n \times n \text{ matrix } A \text{ is invertible} \\ \Leftrightarrow \quad & \det(A) \neq 0 \\ \Leftrightarrow \quad & 0 \text{ is not an eigenvalue of } A \end{aligned} \tag{8}$$

Example 43. Consider $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ $(n = 2)$. We have

$$\begin{aligned} f_A(\lambda) = \det(A - \lambda I_2) &= \det\left(\begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}\right) \\ &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} \\ &= \lambda^2 - \mathrm{tr}(A)\lambda + \det(A). \end{aligned}$$

Here, $\mathrm{tr}(A)$ is the sum of the diagonal entries of a square matrix A and called the trace of A.

Remark 11. In general, $f_A(\lambda)$ of an $n \times n$ matrix A is a polynomial of degree n:

$$f_A(\lambda) = (-\lambda)^n + \mathrm{tr}(A)(-\lambda)^{n-1} + \cdots + \det(A).$$

That is, there are at most n eigenvalues.

Example 44. Let us find the eigenvalues of $A = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix}$. We have

$$\det(A-\lambda I_3)=\det\begin{bmatrix}2-\lambda & 3 & 4\\ 0 & 3-\lambda & 4\\ 0 & 0 & 4-\lambda\end{bmatrix}=(2-\lambda)(3-\lambda)(4-\lambda)=0.$$

Thus we obtain

$$\lambda=2,3,4.$$

The eigenvalues of a triangular matrix are its diagonal entries.

———

An eigenvalue λ_0 of A has algebraic multiplicity k if

$$f_A(\lambda)=(\lambda_0-\lambda)^k g(\lambda),$$

for some polynomial $g(\lambda)$ with $g(\lambda_0)\neq 0$.

Example 45. Let us find the eigenvalues of A with their algebraic multiplicities, where $A=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$. We have

$$f_A(\lambda)=\det(A-\lambda I_3)=\det\begin{bmatrix}1-\lambda & 1 & 1\\ 1 & 1-\lambda & 1\\ 1 & 1 & 1-\lambda\end{bmatrix}=\lambda^2(3-\lambda).$$

Therefore we obtain

$$\lambda=\begin{cases}0 & \text{with algebraic multiplicity 2,}\\ 3 & \text{with algebraic multiplicity 1.}\end{cases}$$

———

Theorem 28. *If an $n\times n$ matrix A has eigenvalues $\lambda_1,\dots,\lambda_n$, then*

$$\det(A)=\lambda_1\lambda_2\cdots\lambda_n,\qquad \operatorname{tr}(A)=\lambda_1+\lambda_2+\cdots+\lambda_n.$$

Proof.

$$f_A(\lambda)=\det(A-\lambda I_n)=(\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda).$$

On the other hand,

$$f_A(\lambda)=(-\lambda)^n+\operatorname{tr}(A)(-\lambda)^{n-1}+\cdots+\det(A).$$

□

Example 46. For the matrix $A = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix}$, we have

$\det(A-\lambda I_3) = (2-\lambda)(3-\lambda)(4-\lambda) = (-\lambda)^3+(2+3+4)(-\lambda)^2+(2\cdot3+3\cdot4+4\cdot2)(-\lambda)+2\cdot3\cdot4.$

Example 47. For the matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, we have $\operatorname{tr}(A) = 3$, $\det(A) = 0$, and $\lambda = 0, 0, 3$. The characteristic polynomial is obtained as

$$f_A(\lambda) = \lambda^2(3-\lambda) = (0-\lambda)(0-\lambda)(3-\lambda).$$

Hence we obtain

$$\operatorname{tr}(A) = 0 + 0 + 3 = 3, \qquad \det(A) = 0 \cdot 0 \cdot 3 = 0.$$

---

20 Eigenspaces [21]

The eigenspace E_λ associated with the eigenvalue λ of an $n \times n$ matrix A is

$$E_\lambda = \ker(A - \lambda I_n) = \{\vec{v} \in \mathbb{R}^n \; : \; A\vec{v} = \lambda\vec{v}\}.$$

The eigenvectors associated with λ are the nonzero vectors in E_λ.

The geometric multiplicity is the dimension of the eigenspace, i.e.,

$$\text{geometric multiplicity} = \dim(E_\lambda) = \text{nullity}(A-\lambda I_n) = n-\text{rank}(A-\lambda I_n).$$

Example 48. Let find geometric multiplicities for $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix}$. By finding

[21] This section is related to Chapter 7 of the textbook.

$$\det(A - \lambda I_2) = (\lambda - 2)(\lambda - 3) = 0,$$

we obtain $\lambda = 2, 3$. Therefore,

$$\begin{aligned} E_2 &= \ker(A - 2I_2) = \ker \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix}, \\ E_3 &= \ker(A - 3I_2) = \ker \begin{bmatrix} -2 & 2 \\ -1 & 1 \end{bmatrix}. \end{aligned}$$

We obtain

$$E_2 = \text{span}(\begin{bmatrix} 2 \\ 1 \end{bmatrix}), \qquad E_3 = \text{span}(\begin{bmatrix} 1 \\ 1 \end{bmatrix}).$$

Since $\dim(E_2) = \dim(E_3) = 1$, the geometric multiplicities are 1 and 1. We note that

$$\sum_\lambda \dim(E_\lambda) = 2.$$

The bases in each eigenspace $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ form an eigenbasis.

———

Example 49. Next let find geometric multiplicities for $A = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}$. By finding

$$\det(A - \lambda I_2) = (\lambda - 2)^2 = 0,$$

we obtain $\lambda = 2$ with algebraic multiplicity 2. Therefore,

$$E_2 = \ker(A - 2I_2) = \ker \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} = \text{span}(\begin{bmatrix} 1 \\ 1 \end{bmatrix}).$$

Since $\dim(E_2) = 1$, the geometric multiplicity is 1. We note that

$$\sum_\lambda \dim(E_\lambda) = 1 < 2.$$

In this case there is no eigenbasis.

———

From the above two examples we can understand the following theorems.

Theorem 29. *Suppose A is an $n \times n$ matrix. If $\sum_\lambda \dim(E_\lambda) < n$, then there is no eigenbasis for A.*

Proof. See Theorem 31.

Theorem 30. *If an $n \times n$ matrix A has n distinct eigenvalues, then there exists an eigenbasis for A.*

Proof. Let us write $A\vec{v}_i = \lambda_i \vec{v}_i$ $(i = 1, 2, \dots, n)$. If these eigenvectors are linearly independent, then they form an eigenbasis. Let us assume there is at least one redundant eigenvector. Let $\vec{v}_m$ be the first redundant vector:

$$\vec{v}_m = c_1\vec{v}_1 + \cdots + c_{m-1}\vec{v}_{m-1},$$

with some scalars $c_1, \dots, c_{m-1}$. We have

$$(A - \lambda_m I_n)\vec{v}_m = (\lambda_1 - \lambda_m)c_1\vec{v}_1 + \cdots + (\lambda_{m-1} - \lambda_m)c_{m-1}\vec{v}_{m-1} = \vec{0}.$$

Suppose $c_{m-1} \neq 0$. By defining

$$d_i = -\frac{(\lambda_i - \lambda_m)c_i}{(\lambda_{m-1} - \lambda_m)c_{m-1}},$$

we obtain

$$\vec{v}_{m-1} = d_1\vec{v}_1 + \cdots + d_{m-2}\vec{v}_{m-2}.$$

The above relation shows $\vec{v}_{m-1}$ is redundant. However, this contradicts the assumption. If $c_{m-1} = 0$, we can define d_i using another nonzero constant c_k. That is, there is no redundant vector, and $\vec{v}_1, \dots, \vec{v}_n$ are linearly independent. Moreover these vectors span $\mathbb{R}^n$. Hence they form a basis. □

Even if we don't have n distinct eigenvalues for an $n \times n$ matrix A, we can have an eigenbasis (Theorem 31 explains when there exists an eigenbasis).

Example 50. Let us find an eigenbasis for $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. The eigenvalue $\lambda = 2$ and

$$E_2 = \ker(A - 2I_2) = \ker \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \operatorname{span}(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}).$$

Therefore $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ form a basis.

Example 51. Let us find an eigenbasis for $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. The eigenvalues are $0, 0, 3$. We obtain

$$E_0 = \ker\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \ker\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \text{span}(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}),$$

$$E_3 = \ker\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} = \ker\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} = \text{span}(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}).$$

The eigenvectors $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ form a basis.

Theorem 31. *Consider an $n \times n$ matrix A with eigenvalues $\lambda_1, \lambda_2, \ldots$. Suppose we find eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_s$ which form a basis of each eigenspace $E_{\lambda_1}, E_{\lambda_2}, \ldots$ (i.e., s is the sum of the geometric multiplicities of $\lambda_1, \lambda_2, \ldots$). Then these vectors are linearly independent even though they belong to different eigenspaces. This implies $s \leq n$. There exists an eigenbasis if and only if $s = n$.*

Proof. The proof is similar to the proof of Theorem 30. Let $\vec{v}_m$ be the first redundant vector:

$$\vec{v}_m = c_1\vec{v}_1 + \cdots + c_{m-1}\vec{v}_{m-1},$$

with some scalars $c_1, \ldots, c_{m-1}$. If all $\vec{v}_1, \ldots, \vec{v}_s$ belong to the same eigenspace E_{λ_m}, they are linearly independent. Hence there is a vector $\vec{v}_k$ associated with λ_k, which belongs to E_{λ_k} ($\neq E_{\lambda_m}$). We have

$$(A-\lambda_m I_m)\vec{v}_m = (\lambda_1-\lambda_m)c_1\vec{v}_1+\cdots+(\lambda_k-\lambda_m)c_k\vec{v}_k+\cdots+(\lambda_{m-1}-\lambda_m)c_{m-1}\vec{v}_{m-1} = \vec{0}.$$

Since $\lambda_k \neq \lambda_m$, the above equation is a nontrivial relation among $\vec{v}_1, \ldots, \vec{v}_{m-1}$. This contradicts the assumption. □

21 Diagonalization [22]

Consider two $n \times n$ matrices A and B. We say that A is similar to B if there exists an invertible matrix S such that

$$AS = SB, \qquad B = S^{-1}AS.$$

[22] This section is related to Chapter 7 of the textbook.

An $n \times n$ matrix A is said to be diagonalizable if A is similar to a diagonal matrix D:
$$S^{-1}AS = D.$$

Example 52. Consider $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$. With $S = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$, we have
$$S^{-1}AS = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} = D.$$
We note that
$$A\vec{v}_1 = 0\vec{v}_1, \qquad A\vec{v}_2 = 3\vec{v}_2.$$

Example 53. Consider $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. We note that
$$A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$
There is no eigenbasis. Indeed in this case, we cannot find S. That is, A is not diagonalizable.

Theorem 32. *A square matrix A is diagonalizable if and only if there exists an eigenbasis. If the eigenvalues are all distinct, then A is diagonalizable.*

We can diagonalize an $n \times n$ matrix A as follows.

Step 1

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

Step 2

Find the eigenspace $E_\lambda = \ker(A - \lambda I_n)$ for each λ.

Step 3

Determine if A is diagonalizable or not $\left(\sum_\lambda \dim(E_\lambda) \stackrel{?}{=} n\right)$. If eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ form an eigenbasis, S is obtained as

$$S = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}.$$

Then we have the relation

$$S^{-1}AS = D, \qquad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

We can determine as follows whether a given $n \times n$ matrix A is diagonalizable. We first find the eigenvalues of A by solving $f_A(\lambda) = \det(A - \lambda I_n) = 0$. Then for each eigenvalue λ, we find a basis of the eigenspace $E_\lambda = \ker(A - \lambda I_n)$. The matrix A is diagonalizable if and only if the dimensions of the eigenspaces add up to n.

Let us consider powers of a square matrix A. If A is diagonalizable, we have $A = SDS^{-1}$, where D is a diagonal matrix. We obtain

$$A^t = \left(SDS^{-1}\right)^t = SDS^{-1}SDS^{-1} \cdots SDS^{-1} = SD^tS^{-1}, \qquad t = 1, 2, \ldots.$$

Example 54. Let us find

$$\begin{bmatrix} -0.5 & 0.5 \\ -3 & 2 \end{bmatrix}^\infty.$$

By solving $f_A(\lambda) = 0$, where $A = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ -3 & 2 \end{bmatrix}$, we obtain $\lambda = 1, \frac{1}{2}$. Since $E_1 = \mathrm{span}(\begin{bmatrix} 1 \\ 3 \end{bmatrix})$ and $E_{1/2} = \mathrm{span}(\begin{bmatrix} 1 \\ 2 \end{bmatrix})$, we can diagonalize A as $A = SDS^{-1}$, where $S = \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$. We obtain

$$A^\infty = SD^\infty S^{-1} = \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1^\infty & 0 \\ 0 & \left(\frac{1}{2}\right)^\infty \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} -2 & 1 \\ -6 & 3 \end{bmatrix}.$$

———

22 Symmetric matrices [23]

In this section, we will consider a (real) symmetric matrix A, which satisfies

$$A^T = A.$$

Theorem 33. *Consider a symmetric matrix A, and two eigenvalues and eigenvectors,*

$$A\vec{v}_1 = \lambda_1\vec{v}_1, \qquad A\vec{v}_2 = \lambda_2\vec{v}_2, \qquad \lambda_1 \neq \lambda_2.$$

Then, $\vec{v}_1 \cdot \vec{v}_2 = 0$.

Proof. We note that

$$\vec{v}_1^T A\vec{v}_2 = \vec{v}_1 \cdot (A\vec{v}_2) = \vec{v}_1 \cdot (\lambda_2\vec{v}_2) = \lambda_2\vec{v}_1 \cdot \vec{v}_2.$$

On the other hand we have

$$\vec{v}_1^T A\vec{v}_2 = \vec{v}_1^T A^T\vec{v}_2 = (A\vec{v}_1)^T\vec{v}_2 = (A\vec{v}_1) \cdot \vec{v}_2 = (\lambda_1\vec{v}_1) \cdot \vec{v}_2 = \lambda_1\vec{v}_1 \cdot \vec{v}_2.$$

By subtraction we obtain

$$(\lambda_1 - \lambda_2)\vec{v}_1 \cdot \vec{v}_2 = 0.$$

Therefore, $\vec{v}_1 \cdot \vec{v}_2 = 0$. □

Theorem 34. *A symmetric $n \times n$ matrix A has n real eigenvalues if they are counted with their algebraic multiplicities.*

Proof. See below.

Example 55. We have seen different examples of eigenvalues of symmetric matrices. For example, in Sec. 20,

$$\lambda \text{ of } \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2, 2, \qquad \lambda \text{ of } \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = 0, 0, 3.$$

A matrix A is said to be orthogonally diagonalizable if there exist an orthogonal matrix S and a diagonal matrix D such that

$$S^{-1}AS = S^TAS = D.$$

[23] This section is related to Chapter 8 of the textbook.

Theorem 35 (Spectral theorem). *A matrix A is orthogonally diagonalizable if and only if A is symmetric.*

Proof. ($\Rightarrow$) There exist an orthogonal S and a diagonal D such that

$$S^{-1}AS = D \quad \text{or} \quad A = SDS^{-1} = SDS^T.$$

Hence,

$$A^T = (SDS^T)^T = SD^TS^T = SDS^T = A.$$

Therefore A is symmetric:

$$A^T = A.$$

($\Leftarrow$) We consider a symmetric $n \times n$ matrix A. We give a proof by induction on n. When $n = 1$, we can set $S = \begin{bmatrix} 1 \end{bmatrix}$.

Let us assume that the claim is true for n. We will show that it holds for $n+1$. With Theorem 34, we pick a real eigenvalue λ and choose an eigenvector $\vec{v}_1$ of length 1 for λ. We can find an orthonormal basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n+1} \in \mathbb{R}^{n+1}$. Form the orthogonal matrix $P = [\vec{v}_1 \cdots \vec{v}_{n+1}]$, and compute $P^{-1}AP$. We note that

- The first column of $P^{-1}AP$ is $\lambda\vec{e}_1$. We note that

$$\begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2 \\ \vdots \end{bmatrix} A \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots \end{bmatrix} = \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2 \\ \vdots \end{bmatrix} \begin{bmatrix} \lambda\vec{v}_1 & A\vec{v}_2 & \cdots \end{bmatrix} = \begin{bmatrix} \lambda\vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1^T A\vec{v}_2 & \cdots \\ \lambda\vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2^T A\vec{v}_2 & \cdots \\ \vdots & & \end{bmatrix}.$$

- $P^{-1}AP = P^TAP$ is symmetric because

$$(P^TAP)^T = P^TA^TP = P^TAP.$$

Combining these two statements, we conclude that $P^{-1}AP$ has the block form

$$P^{-1}AP = \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix},$$

where B is a symmetric $n \times n$ matrix. By the induction hypothesis, B is orthogonally diagonalizable, i.e.,

$$Q^{-1}BQ = D,$$

where Q is an orthogonal $n \times n$ matrix and D is a diagonal $n \times n$ matrix. Let us introduce an orthogonal $(n+1) \times (n+1)$ matrix

$$R = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}.$$

We have

$$R^{-1} \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix} R = \begin{bmatrix} 1 & 0 \\ 0 & Q^{-1} \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & D \end{bmatrix}.$$

That is,

$$R^{-1}P^{-1}APR = \begin{bmatrix} \lambda & 0 \\ 0 & D \end{bmatrix}.$$

Let us define $S = PR$. Since P, Q are both orthogonal, for any vector $\vec{x}$, we have

$$\|S\vec{x}\| = \|P(R\vec{x})\| = \|R\vec{x}\| = \|\vec{x}\|.$$

That is, S is also orthogonal. Finally we obtain

$$S^{-1}AS = \begin{bmatrix} \lambda & 0 \\ 0 & D \end{bmatrix}.$$

Thus A is diagonalized and the claim is proved. □

Example 56. Let us diagonalize a symmetric matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$. We have $\lambda = 1 \pm \sqrt{2}$ and

$$E_{1\pm\sqrt{2}} = \operatorname{span}(\begin{bmatrix} 1 \pm \sqrt{2} \\ 1 \end{bmatrix}).$$

By defining

$$S = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{2}}{\sqrt{4+2\sqrt{2}}} & \frac{1-\sqrt{2}}{\sqrt{4-2\sqrt{2}}} \\ \frac{1}{\sqrt{4+2\sqrt{2}}} & \frac{1}{\sqrt{4-2\sqrt{2}}} \end{bmatrix}, \qquad D = \begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix},$$

we obtain $S^{-1}AS = D$. Indeed, $\vec{u}_1$ and $\vec{u}_2$ are orthonormal [24].

Finally we prove Theorem 34.

[24] The matrix A must be real symmetric. The matrix $A = \begin{bmatrix} 2i & 1 \\ 1 & 0 \end{bmatrix}$ is symmetric. But since $\lambda = i$ and $E_i = \operatorname{span}(\begin{bmatrix} 1 \\ -i \end{bmatrix})$, A is not diagonalizable.

Proof (Theorem 34). Suppose that there is an eigenvector $\vec{v}$ with the associated eigenvalue $\lambda = p + iq$ is complex[25]:

$$A\vec{v} = (p + iq)\vec{v}, \qquad p, q \in \mathbb{R}.$$

We can rewrite the above equation as

$$(A - pI_n)\vec{v} = iq\vec{v}.$$

We note that $A - pI_n$ is symmetric. It suffices to show that any symmetric matrix doesn't have a purely imaginary eigenvalue iq. Let us assume that there exist an eigenvalue iq and an eigenvector $\vec{v}$ such that

$$A\vec{v} = iq\vec{v}.$$

We have[26]

$$(A\vec{v})^T\overline{\vec{v}} = A\vec{v} \cdot \overline{\vec{v}} = iq\vec{v} \cdot \overline{\vec{v}}.$$

On the other hand we obtain

$$(A\vec{v})^T\overline{\vec{v}} = \vec{v}^T A^T\overline{\vec{v}} = \vec{v}^T A\overline{\vec{v}} = \vec{v}^T\overline{A\vec{v}} = \vec{v}^T\overline{iq\vec{v}} = -iq\vec{v}^T\overline{\vec{v}} = -iq\vec{v} \cdot \overline{\vec{v}}.$$

Therefore, $iq\vec{v} \cdot \overline{\vec{v}} = -iq\vec{v} \cdot \overline{\vec{v}}$. However, $\vec{v} \cdot \overline{\vec{v}} > 0$ because $\vec{v} \neq \vec{0}$. We obtain $q = 0$.

The characteristic polynomial $f_A(\lambda)$, which is a polynomial of degree n, has n complex roots if they are properly counted with their multiplicities (the fundamental theorem of algebra). Since any eigenvalue of a symmetric matrix A is real, this means that A has n real eigenvalues. □

---

[25] i is the imaginary unit and $i \cdot i = -1$.

[26] For a complex number $z = a + ib$ $(a, b \in \mathbb{R})$, we define its complex conjugate by $\overline{z} = a - ib$.