Chapter 1

$\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$

The set $\mathbb{R}^n$

- $\mathbb{R}$ is the set of all real numbers.

- Let n be a positive integer.

- We can write $(x_1, x_2, \cdots, x_n)$ of real numbers, which is an ordered n-tuple of real numbers.

- $\mathbb{R}^n$ is the set of all ordered n-tuples of real numbers.

- An element $x \in \mathbb{R}^n$ can be written in

$$\text{row} \quad x = (x_1, x_2, x_3, \cdots, x_n) \quad \text{or in column} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{or any other form.}$$

As long as the order of listed n real numbers are seen, there would not be a problem.

Addition and Scalar multiplication

Addition

1. If $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, \cdots, y_n) \in \mathbb{R}^n$, we can add the two to obtain an element in $\mathbb{R}^n$:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ \vdots \\ x_n + y_n \end{pmatrix},$$

which is denoted by $x + y \in \mathbb{R}^n$.

Scalar multiplication

1. In this course, the scalar is a synonym of real number.

2. If λ is a real number and $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, then we can scale x to obtain λx:

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \\ \vdots \\ \lambda x_n \end{pmatrix} \in \mathbb{R}^n.$$

The set $\mathbb{R}^{m \times n}$

- Let $\mathbb{R}$ be again the set of all real numbers.

- Let m and n be two positive integers.

- The set of all $(A_{ij})_{i=1,j=1}^{i=m,j=n}$ of real numbers indexed by $(i,j) \in \{1,2,\cdots,m\} \times \{1,2,\cdots,n\}$ is denoted by $\mathbb{R}^{m \times n}$.

- The notation for an element $A \in \mathbb{R}^{m \times n}$ in this time is more specific. We distinguish two indices $i = 1,2,\cdots,m$ and $j = 1,2,\cdots,n$. We list mn real numbers in a box so that i is a row index, and j is a column index:

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\ A_{31} & A_{32} & A_{33} & \cdots & A_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mn} \end{pmatrix}.$$

- We call an element of $\mathbb{R}^{m \times n}$ an $(m \times n)$ matrix, which reads as "m by n matrix".

- Why ..?

Addition and Scalar multiplication

Addition

1. If $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times n}$, i.e., if the matrix A and B are in same shape, we can add the two to obtain an element in $\mathbb{R}^{m \times n}$:

$$
\begin{pmatrix}
A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\
A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\
A_{31} & A_{32} & A_{33} & \cdots & A_{3n} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mn}
\end{pmatrix}
+
\begin{pmatrix}
B_{11} & B_{12} & B_{13} & \cdots & B_{1n} \\
B_{21} & B_{22} & B_{23} & \cdots & B_{2n} \\
B_{31} & B_{32} & B_{33} & \cdots & B_{3n} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
B_{m1} & B_{m2} & B_{m3} & \cdots & B_{mn}
\end{pmatrix}
$$
$$
=
\begin{pmatrix}
A_{11} + B_{11} & A_{12} + B_{12} & A_{13} + B_{13} & \cdots & A_{1n} + B_{1n} \\
A_{21} + B_{21} & A_{22} + B_{22} & A_{23} + B_{23} & \cdots & A_{2n} + B_{2n} \\
A_{31} + B_{31} & A_{32} + B_{32} & A_{33} + B_{33} & \cdots & A_{3n} + B_{3n} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
A_{m1} + B_{m1} & A_{m2} + B_{m2} & A_{m3} + B_{m3} & \cdots & A_{mn} + B_{mn}
\end{pmatrix},
$$

   which is denoted by $A + B \in \mathbb{R}^{m \times n}$.

Scalar multiplication

1. If λ is a real number and $A \in \mathbb{R}^{m \times n}$, then we can scale A to obtain λA:

$$
\lambda
\begin{pmatrix}
A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\
A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\
A_{31} & A_{32} & A_{33} & \cdots & A_{3n} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mn}
\end{pmatrix}
=
\begin{pmatrix}
\lambda A_{11} & \lambda A_{12} & \lambda A_{13} & \cdots & \lambda A_{1n} \\
\lambda A_{21} & \lambda A_{22} & \lambda A_{23} & \cdots & \lambda A_{2n} \\
\lambda A_{31} & \lambda A_{32} & \lambda A_{33} & \cdots & \lambda A_{3n} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\lambda A_{m1} & \lambda A_{m2} & \lambda A_{m3} & \cdots & \lambda A_{mn}
\end{pmatrix}.
$$

In summary,

1. We have the set $\mathbb{R}^n$, equipped with the addition and the scalar multiplication, that we denote by $\left(\mathbb{R}^n, +, s\right)$.

$$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n,$$
$$s : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n.$$

2. We havethe set $\mathbb{R}^{m \times n}$, equipped with the addition and the scalar multiplication, that we denote by $\left(\mathbb{R}^{m \times n}, +, s\right)$.

$$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n,$$
$$s : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n.$$

1. That we work with $(\mathbb{R}^n, +, s)$, i.e., that $\mathbb{R}^n$ are equipped with the addition and the scalar multiplication may, more importantly, mean that <u>we do not perform other operations</u>.

   These are illegal expressions:

   (a) $\mathbb{R}^n + \mathbb{R}$, $n \geq 2$:

   (b) $\mathbb{R}^n + \mathbb{R}^m$, $n \neq m$.

   (c) product in general. $\times : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, though in some dimensions we can define meaningful product.

   (d) comparison in general.

2. Likewise, for the case of $\mathbb{R}^{m \times n}$, <u>we do not perform other operations</u>, unless otherwise defined later of this course.

- We have so far
$$(\mathbb{R}^n, +, s) \quad \text{and} \quad (\mathbb{R}^{m \times n}, +, s),$$
respectively a set equipped with the addition and scalar multiplication.

- Expression that makes use of the addition and scalar multiplication, and use only of them such as
$$\lambda_1 x + \lambda_2 y + \lambda_3 z + \lambda_4 w, \quad x, y, z, w \in \mathbb{R}^n$$
comes as an important expression to us.

- The expression is called a "Linear Combination" of x, y, z, w; Linear, meaning that the scalar multiplication $x \mapsto \lambda x$ is to make something proportional, and Combination, meaning here we are adding things.

The final comment on $(\mathbb{R}^n, +, s)$ and $(\mathbb{R}^{m \times n}, +, x)$:

As stated earlier, in some dimensions we may additionally define product, for example in $\mathbb{R}^2$ we can define

$$(x_1, x_2) \times (y_1, y_2) = (v_1, v_2), \quad \text{where} \quad v_1 + iv_2 = (x_1 + ix_2)(y_1 + iy_2).$$

However, all of such operations are treated in this course to be exceptional and come as addendum, emphasizing that there are only (i) <u>the addition</u> and (ii) <u>the scalar multiplication</u> we are allowed to operate so far.

Chapter 2

Matrix Matrix multiplication

- Emphasizing that there are no other general product operations, now we define the unique product operation between matrices.

- For given matrix $A \in \mathbb{R}^{\ell \times m}$ and matrix $B \in \mathbb{R}^{m \times n}$, we define the matrix $C \in \mathbb{R}^{\ell \times n}$ to be the product denoted by AB so that

$$C_{ij} = \sum_{\alpha=1}^{m} A_{i\alpha} B_{\alpha j} \quad \text{for } i = 1, 2, \cdots, \ell \text{ and } j = 1, 2, \cdots, n.$$

- Importantly, the product is defined only for the case where the second component of shape of A and the first component of shape of B are the same, and the product is not defined for all remaining cases.

- Why .. ?

The defined product can be regarded in a few different ways.

1. The product can be regarded as to build new n linear combinations of columns of A, as designated by numbers in B:

2. The product can be regarded as to build new ℓ linear combinations of rows of B, as designated by numbers in A:

3. The product can be regarded as to build new ℓn real numbers, out of rows of A and columns of B.

The product is useful in many places.

Perfect for the chain rule in multivariable calculus

Suppose that $X = \mathbb{R}^\ell$, $Y = \mathbb{R}^m$, and $Z = \mathbb{R}^n$, and

$$f : X \to Y, \quad g : Y \to Z.$$

Multivariable Calculus:

1. Also assume f and g are many times diffentiable functions.

2. For a given point $\bar{x}$, we collect numbers

$$B_{\alpha j} = \frac{\partial f^\alpha}{\partial x^j}(\bar{x}) = \lim_{h \to 0} \frac{f^\alpha(\bar{x} + h e_j) - f^\alpha(\bar{x})}{h}.$$

3. Similarly, for a given poit $\bar{y}$, we collect numbers

$$A_{i\alpha} = \frac{\partial g^i}{\partial y^\alpha}(\bar{y}) = \lim_{h \to 0} \frac{g^i(\bar{y} + h e_\alpha) - g^i(\bar{y})}{h}.$$

4. Now we consider the composition

$$g \circ f : X \to Z, \quad g \circ f(x) = g(f(x))$$

and collect numbers

$$C_{ij} = \frac{\partial (g \circ f)^i}{\partial x^j}(\bar{x}) = \lim_{h \to 0} \frac{(g \circ f)^i(\bar{x} + h e_j) - (g \circ f)^i(\bar{x})}{h}.$$

5. Then C turns out to equal to AB, for $\bar{y} = f(\bar{x})$.

Writing a system of linear equations

If we are given

$$\begin{cases} 3x - 7y + 4z = -2, \\ 9x - 2y - 6z = 0, \\ -5x + 3y - 11z = -8 \end{cases}$$

It is nice that we can write also

$$\begin{pmatrix} 3 & -7 & 4 \\ 9 & -2 & -6 \\ -5 & 3 & -11 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ -8 \end{pmatrix}$$

Using $\ell = 3$, $m = 3$, and $n = 1$ in the left-hand-side.

The question of why..?

1. Why does it have to be in that way?

2. Why not we define for $A, B \in \mathbb{R}^{m \times n}$ a product with the same shape in $\mathbb{R}^{m \times n}$?

$$
\begin{pmatrix}
A_{11}B_{11} & A_{12}B_{12} & A_{13}B_{13} & \cdots & A_{1n}B_{1n} \\
A_{21}B_{21} & A_{22}B_{22} & A_{23}B_{23} & \cdots & A_{2n}B_{2n} \\
A_{31}B_{31} & A_{32}B_{32} & A_{33}B_{33} & \cdots & A_{3n}B_{3n} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
A_{m1}B_{m1} & A_{m2}B_{m2} & A_{m3}B_{m3} & \cdots & A_{mn}B_{mn}
\end{pmatrix}
$$

   (a) Actually you can. In most of computer language, this entry-wise product in the same shape is provided by the broadcasting.

3. All why-questions come to the following one,

   Would other people be interested in the new definition ?

A few important observations

1. Important: You don't mess up with the order of multiplication here.

2. Unlike multiplication we know for two real numbers, for the Matrix-Matrix multiplication for $A \in \mathbb{R}^{\ell \times m}$ and $B \in \mathbb{R}^{m \times n}$, the product BA is not even defined in general. This is simply because in general $n \neq \ell$.

3. In the special case of that $\ell = m = n$, BA is defined. However even in such a case

$$AB \neq BA \quad \text{in general.}$$

Summary

1. We have $\left(\mathbb{R}^{m \times n}, +, s\right)$ for m and n positive integers.

2. Linear combinations of members of $\mathbb{R}^{m \times n}$ are expressions using $+$ and s such as

$$\lambda_1 A_1 + \lambda_2 A_2 + \cdots \lambda_j A_j \in \mathbb{R}^{m \times n}.$$

3. We have a definition of multiplication for $A \in \mathbb{R}^{\ell \times m}$ and matrix $B \in \mathbb{R}^{m \times n}$ resulting in $AB \in \mathbb{R}^{\ell \times n}$.

Knowledge so far will be applied to mathematics, science, and engineering, and will be extremely important and powerful.

Chapter 3

Linear Independence and Basis

Let us consider elements in $\mathbb{R}^2$,

$$x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad z = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

It is easy to notice that if we are with x and y in our hands, z can be reproduced by

$$z = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

a linear combination of x and y.

Linear Combinations and Span

For a given finite number of elements $a_1, a_2, \cdots, a_k \in \mathbb{R}^n$, the set of all linear combinations of them is denoted by

$$\mathsf{span} \left\langle a_1, a_2, \cdots, a_k \right\rangle = \left\{ x \in \mathbb{R}^n \mid x \text{ is a linear combination of } a_1, a_2, \cdots, a_k \right\}.$$

Example

Let

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad \cdots \quad e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Then, $e_1, e_2, \cdots, e_n$ span all elements in $\mathbb{R}^n$, i.e., $\mathsf{span} \left\langle e_1, e_2, \cdots, e_n \right\rangle = \mathbb{R}^n$.

Linear Independence

As far as reproducibility is concerned, we notice that we keep only two elements

$$x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

and every element in $\mathbb{R}^2$ can be reproduced by linear combination of x and y.

This necessiates the notion of linear independece of a few elements in $\mathbb{R}^n$.

Example 1 z above is linearly dependent with respect to x and y.

Example 2 On the other hand, in $\mathbb{R}^3$,

$$x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \quad z = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$$

Then, x, y, z are linearly independent. By no chance, z is a linear combination of x and y.

Definition 1. *We say $a_1, a_2, \cdots, a_k \in \mathbb{R}^n$ are linearly independent if*

$$\lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_k a_k = 0 \quad \implies \quad \lambda_1 = \lambda_2 = \cdots = \lambda_k = 0.$$

$a_1, a_2, \cdots, a_k \in \mathbb{R}^n$ are said to be linearly dependent if they are not linearly independent.

Basis

Definition 2. *We say $a_1, a_2 \cdots, a_k$ of $\mathbb{R}^n$ form a basis of $\mathbb{R}^n$ if*

1. *$a_1, a_2 \cdots, a_k$ span $\mathbb{R}^n$,*

2. *$a_1, a_2 \cdots, a_k$ are linearly independent.*

Examples

$e_1, e_2, \cdots, e_n$ form a basis.

We will prove in later section the following seemingly natural facts:

1. Elements in $\mathbb{R}^n$ fewer than n cannot span <u>all</u> of $\mathbb{R}^n$.

2. Elements in $\mathbb{R}^n$ exceeding n must linearly <u>dependent</u>.

3. n elements in $\mathbb{R}^n$ are linearly independent if and only if they <u>do span all</u> of $\mathbb{R}^n$.

Chapter 4

Power and Importance of $\mathbb{R}^{m \times n}$

We saw that $\left(\mathbb{R}^{m \times n}, +, s\right)$ is so simply defined. The power of $\mathbb{R}^{m \times n}$ comes from that, (i) there are so many instances in mathematics, science, engineering, and etc., where entities of interests are such simple that we can add and scale them and that (ii) we will learn nice tools for $\mathbb{R}^{m \times n}$.

Example 1 In fact, this is an example of $\mathbb{C}^{m \times n}$, with $\mathbb{C}$ of complex numbers. Wolfgang Pauli, a physicist, introduced *Pauli matrices* in developing theory of quantum mechanics in early 20c.

We consider a multiplication table for a 4 elements set $\{1, \sigma_x, \sigma_y, \sigma_z\}$:

Table 4.1: A multiplicative table

	1	σ_x	σ_y	σ_z
1	1	σ_x	σ_y	σ_z
σ_x	σ_x	-1	σ_z	$-\sigma_y$
σ_y	σ_y	$-\sigma_z$	-1	σ_x
σ_z	σ_z	σ_y	$-\sigma_x$	-1

Note that $\sigma_x \sigma_y \neq \sigma_y \sigma_x$, and this set cannot be represented by elements in $\mathbb{R}$.

One of the way we achieve the table:

We consider
$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \right\}$$
for $1, \sigma_x, \sigma_y, \sigma_z$ from the left.

The set $\mathsf{span}\left\langle \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \right\rangle$ contains all the elements that can be generated.

note: The role played by $\mathbb{R}$, the scalar factor, can be taken by the mathematical object of *field*. But we will stick to $\mathbb{R}$ in this course.

Example 2 Colors (Light)

Although there are so many colors such as yellow, green, blue, white, black, $\cdots$, it turns out that the only three of them, red, green, and blue are independent. All others are linear combination of the three.

We choose $[0,1]^3 \subset \mathbb{R}^3$ represents the colors, with

1. $(1,0,0) \simeq$ red,

2. $(0,1,0) \simeq$ green,

3. $(0,0,1) \simeq$ blue,

4. $(0,0,0) \simeq$ black,

5. $(1,1,1) \simeq$ white.

The value ranged in $[0,1]$ represents the intensity.

Example 3 Bacteria with manipulatable genes.

In Bio-Engineering, ceartain bacateria is called a *vector*. This reflects that a certain kind of bacteria may or may not contain a few genes of our interests, and can be manipulated in the laboratory in the way of addition.

By a reproduction, a bacteria with gene A and a bacteria with gene B can result in a bacteria with both genes A and B. The number of such genes that can be maniplulated in lab. and bacteria can be represented by $\mathbb{R}^n$.

Example 4 Sauce

Sauce can be perfectly represented by $\mathbb{R}^n$. They can be added and scaled:

We linear combinate the sugar, salt, vinegar, soy sauce, water, what you want, in grams.

So a recipe you have to remember for a certain sauce is an element in $\mathbb{R}^n$, and if you want to make sauce a lot, you scale the element.

1. We saw many examples of *entities* that can be added and scaled, and sometimes we employ multiplication.

2. They are well-represented by $\mathbb{R}^{m \times n}$ or $\mathbb{R}^n$.

3. In all of examples, we see that there are two different roles:

$$\text{Respresentation} \quad \text{vs.} \quad \text{Essential Entities}$$

4. The theory has been developed in Linear Algebra so that the two roles are explicitly detached from each other. The latter, of essential entities, is defined in mathematics as *vector space* and *algebra*.

5. In our course, we use the word *vector* to denote the essential entity, while we simply use the word element of $\mathbb{R}^n$, or the matrix in $\mathbb{R}^{m \times n}$ for the representational use.

6. The connection from the set of essential entities to for instance $\mathbb{R}^n$, is done by fixing the basis. Let us be clear on this below.

- The definition of vector space X over the scalar $\mathbb{R}$ can be found in math textbooks and webpages. For completeness, we also contained the mathematical definition in the last page of this chapter.

- If you are not familiar with mathematical definitions, consider a vector space X as a set of symbols that can be added and scaled.

- k-dimensional vector space for k a positive integer.

  Definition 3. *A vector space X that admits a basis consists of k elements is called a k-dimensional vector space.*

  Example The space of colors admits a basis of three elements. Hence the space is 3-dimensional vector space.

- Connection from the *essential entities* to *representing elements* is done once we fix what are the basis elements in X.

  Example The space of sauces consists of $a = sugar$, $b = salt$, $c = vinegar$, and $d = water$.

  Once we fix a, b, c, and d to be the basis, we will understand an element

  $$\begin{pmatrix} 5 \\ 5 \\ 3 \\ 100 \end{pmatrix} \in \mathbb{R}^4$$

  to represent a sauce consisting of 5-sugar, 5-salt, 3-vinegar, and 100 water, in grams.

  We may want to take the diluted ingredients $a + 100d$, $b + 100d$, $c + 100d$, and d to be the basis. Then

  $$\begin{pmatrix} 5 \\ 5 \\ 3 \\ 100 \end{pmatrix} \in \mathbb{R}^4$$

  to represent a sauce consisting of 5-sugar, 5-salt, 3-vinegar, and 1400 water, in grams.

- You can see the detached roles of *representation* and *essential entities*.

- And how the basis choice connects the two.

A vector space X over the scalar $\mathbb{R}$ is mathematically defined as follows.

1. X is a nonempty set.

2. $+ : X \times X \to X$ is well-defined with following properties.

   (a) For any $x, y, z \in X$, $x + (y + z) = (x + y) + z$.

   (b) For any $x, y \in X$, $x + y = y + x$.

   (c) $\exists 0 \in X$ such that for any $x \in X$ $x + 0 = x$.

   (d) For any $x \in X$, $\exists (-x) \in X$ such that $x + (-x) = 0$.

3. $s : \mathbb{R} \times X \to X$ is well-defined with follwoing properties.

   (a) For any $\lambda_1, \lambda_2 \in \mathbb{R}$ and $x \in X$, $\lambda_1(\lambda_2 x) = (\lambda_1 \lambda_2)x$

   (b) For any $x \in X$, $1x = x$.

4. $+$ and s work in compatible ways satisfying the following.

   (a) For any $\lambda \in \mathbb{R}$ and $x, y \in X$, $\lambda(x + y) = \lambda x + \lambda y$.

   (b) For any $\lambda_1, \lambda_2 \in \mathbb{R}$ and $x \in X$, $(\lambda_1 + \lambda_2)x = \lambda_1 x + \lambda_2 x$.

Summary

1. For an m-dimensional vector space $(X, +, s, \{b_1, b_2, \cdots, b_m\})$ with m basis vectors $b_1, b_2, \cdots, b_m$, we consider the representation of its element for example

$$\begin{pmatrix} 3 \\ -5 \\ 0 \\ \vdots \\ 0.8 \end{pmatrix} \in \mathbb{R}^{m \times 1}$$

   to represent $3b_1 - 5b_2 + 0b_3 + \cdots + 0.8b_m$.

2. Be careful when X itself is $\mathbb{R}^m$, which is the very possible example. One way we will follow in this course is that, for the representational use, we make use of

$$\mathbb{R}^{m \times 1} \quad \text{of one column matrices.}$$

   and try best to distinguish the essential entity X and its representation.

Chapter 5

Transformation of Solid shapes

We will spend long times studying transformations of geometrical shapes.

Examples

Cube, Airplane, Right Hand.

1. Rotation

   - How do we make an airplane heading to a certain direction ?

2. Dilation

- Dilation in all direction.
- Dilation differently for semi-axes.
- Dilation differently for a few non-orthogonal directions.

3. Reflection with repect to a hyperplane.

- Right hand vs. Left hand.
- Composition of many reflections.

4. Projection

- Same projection twice ?
- Composition of many projections.

5. Shearing

For those solid shapes and transformations, we do the following.

For solid shapes:

- We let the set

$$X = \{\text{Locations}\} \quad \text{with respect to a fixed origin.}$$

- Location with respect to an origin can be scaled by a real number.

  We let locations be added using the parallelogram law.

  You can check X of locations to a fixed origin is a vector space.

  More precisely, X is equipped with the basis locations e_1, e_2, e_3 for the case of 3-dimensions. In general, up to e_n for n-dimensional locations.

- A solid shape is then a subset occupying a region of locations.

$$G \subset X$$

A transformation of a shape can be then a mapping from Location space to itself.

- A transformation is then a mapping

$$T : X \to X.$$

- We let the set of all <u>linear</u> transformations from Location space X to itself:

$$L_{X,X} = \big\{ T : X \to X \ \big| \ T \text{ is linear} \big\}.$$

- Of course, in general we can think of

$$
\begin{aligned}
X, &\quad \text{the space of } n\text{-dimensional locations to some fixed origin,} \\
Y, &\quad \text{the space of } m\text{-dimensional locations to some fixed origin,} \\
L_{X,Y}, &\quad \text{the space of linear mappings } T : X \to Y.
\end{aligned}
$$

Representation:

- We will use
$$\mathbb{R}^{n\times 1}, \quad \mathbb{R}^{n\times n} \quad \text{for } X \text{ and } L_{X,X}.$$

- Indeed, $n \times n$ matrix A and $n \times 1$ column x can be multiplied to produce another $n \times 1$ column Ax, representing a new location.

- By techniques in $\mathbb{R}^{n\times n}$ (or in general $\mathbb{R}^{m\times n}$), you will learn the surprising fact

  that every linear transformation is a few compositions of rotation, dialation, reflection, shearing, and projections. No others.

The composition of linear transformations will be stacking matrix multiplications

$$T = A_k A_{k-1} A_{k-2} \cdots A_2 A_1$$

of $n \times n$ matrices.

Home Work

Suppose $A \in \mathbb{R}^{k \times \ell}$, $B \in \mathbb{R}^{\ell \times m}$, and $C \in \mathbb{R}^{m \times n}$ so that we can multiply A and B, and multiply B and C. Try to check that

$$(AB)C = A(BC).$$

A map $T : X \to Y$ is linear if

for a linear combination $\lambda_1 x_1 + \lambda_2 x_2 + \cdots \lambda_k x_k$

$T\big(\lambda_1 x_1 + \lambda_2 x_2 + \cdots \lambda_k x_k\big)$ must equal to $\lambda_1 T(x_1) + \lambda_2 T(x_2) + \cdots \lambda_k T(x_k).$

- We will write and analyze matrices for rotations, reflections, dilations, shearings, and projections one by one from next lecture.

- This is to study orthogonal matrices, diagonal matrices, symmetric matrices, triangular matrices, and so on, which we will introduce.

We have $x \in \mathbb{R}^{n \times 1}$ and $A \in \mathbb{R}^{n \times n}$, by which, the new location

$$y = Ax \in \mathbb{R}^{n \times 1}$$

We proceed with $n = 3$.

Today, we introduce a few kinds of matrices. We will need only one more technique to deal with axes that are not coordinate axes. The point is that other than ones we introduce today + the technique, you do not need to know any further kinds of matrices.

Identity Matrix

1. We consider the matrix
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \in \quad \mathbb{R}^{3 \times 3}$$

2. It is a special element in $\mathbb{R}^{3 \times 3}$, and denoted by I.

3. Note that for any $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$,

$$y = Ax = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

4. Be careful that I is not the one below !

$$I \neq \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

- Diagonal entries are numbers in diagonal of the square.
- Off diagonal entries are entries not in diagonal.
- All of diagonal entries of I are 1
- All of off diagonal entries of I are 0.

Diagonal Matrices

1. Let $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}^3$. We consider a matrix

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad \in \quad \mathbb{R}^{3 \times 3}.$$

2. First, let us consider the special cases $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$

   (a) Example 1 : $x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad Ax =$

   (b) Example 2 : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad Ax =$

   (c) Example 3 : $x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad Ax =$

   (d) Example 4 : $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad Ax =$

3. $\lambda_i = 0$: One component is shrinken to 0.

4. $\lambda_i < 0$: Dilate with sign flip.

5. More matrices for the dilations below will be introduced later.

   (a) Principal axes are not coordinate axes.

   (b) Dilation with directions not orthogonal to each other.

Elementary Rotation Matrices

1. Let $\theta \in \mathbb{R}$. We consider a matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \in \quad \mathbb{R}^{3\times 3}.$$

(a) Example 1 : $x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad Ax =$

(b) Example 2 : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad Ax =$

(c) Example 3 : $x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad Ax =$

(d) Example 4 : $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \qquad Ax =$

2. We consider a matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \quad \in \quad \mathbb{R}^{3\times3}.$$

(a) Example 1 : $x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $Ax =$

(b) Example 2 : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $Ax =$

(c) Example 3 : $x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ $Ax =$

(d) Example 4 : $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ $Ax =$

3. We consider a matrix

$$\begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \quad \in \quad \mathbb{R}^{3\times3}.$$

(a) Example 1 : $x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $Ax =$

(b) Example 2 : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $Ax =$

(c) Example 3 : $x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ $Ax =$

(d) Example 4 : $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ $Ax =$

We will introduce more rotation matrices later.

Elementary Orthogonal Projection Matrices

1. We consider a matrix
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \in \quad \mathbb{R}^{3\times 3}.$$

    (a) Example 1 : $x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $Ax =$

    (b) Example 2 : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $Ax =$

    (c) Example 3 : $x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ $Ax =$

    (d) Example 4 : $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ $Ax =$

2. We will introduce more orthogonal projection matrices onto the projecting plane that is not aligned with coordinate.

3. There are non-orthogonal projections, but we will not use them much. We will introduce them later.

Elementary Reflection Matrices

1. We consider a matrix
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \in \quad \mathbb{R}^{3\times 3}.$$

   (a) Example 1 : $x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $\quad Ax =$

   (b) Example 2 : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $\quad Ax =$

   (c) Example 3 : $x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ $\quad Ax =$

   (d) Example 4 : $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ $\quad Ax =$

2. We will introduce more reflection matrices whose symmetry hyperplane is not aligned with coordinate.

Elementary Shearing Matrices

1. We consider a matrix
$$\begin{pmatrix} \lambda_1 & \alpha & \beta \\ 0 & \lambda_2 & \gamma \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad \in \quad \mathbb{R}^{3 \times 3}.$$

2. This kind of matrices are called upper triangular matrices.

(a) Example 1 : $x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $Ax =$

(b) Example 2 : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $Ax =$

(c) Example 3 : $x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ $Ax =$

(d) Example 4 : $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ $Ax =$

1. As soon we are able to <u>freely go back and forth</u> from one representation to another representation w.r.t choice of basis vectors, we will know in full generality about matrices of above kinds.

2. Again, you recall that <u>you do not need to know more kinds</u> other than those.

1. Now, we learn the technique to deal with axes oblique to coordinate axes.

2. To be concrete, let us be given an objective to come up with dilation but with repect to three axes, specified by three location points

$$\begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

with dilation factors $\lambda_1 = \frac{1}{2}$, $\lambda_2 = 2$, and $\lambda_3 = 1$.

3. We will use the three as our basis and seek a *new representations* of locations.

4. Typically, as seen in the example, we would know old representations of your preferred temporal basis.

I know B. The Old representation of your new basis

Let us list those old representations of your temporal basis in columns of a $n \times n$ matrix B:

$$B = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{pmatrix}.$$

All you need to remember is the following equality:

If your new representation of a certaion location x is $\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix}$,

$$\begin{pmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

is nothing but its old representation. To remember:

$$B_{ij'} a_{j'} = a_i.$$

Because

$$\begin{pmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

(recall that B is the Old representations of you new basis)

if we want to know what are $\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix}$, the temporal representation, we get it by having

$$B^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

We will re-visit the inverse matrix later to study it properly, but here we just admit that B^{-1} of our matrix B exists with the property we want.

Definition 4. *For a given $A \in \mathbb{R}^{3\times3}$, if there exists a matrix $L \in \mathbb{R}^{3\times3}$ such that*

$$AL = LA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which must be then unique, L is said to be the inverse matrix of A, and denoted by A^{-1}.

If we multiply B^{-1} both sides,

- (RHS) $= B^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$,

- (LHS) $= B^{-1}B \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = $ (RHS).

Now, we are able to achieve our objective of non-elementary dialation:

- Get the new representation, do the <u>elementary dilation</u> in new representation, and don't forget to return the result, <u>not in your representation</u>, but in old representation.

This is to conduct:

$$x = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mapsto \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$$\mapsto \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$$\mapsto \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = y$$

Because I know that the inverse matrix

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

I can compute one by one.

Example 2: non-elementary reflection

- We want to compose a matrix that does the reflection with respect to a hyperplane we want.

- That I know what the hyplane is: I know representations of normal vector and tangent vectors of the hyperplane.

- What we have to do:

  1. Get the new reprentation of x in temporal basis of normal and tangents.
  2. Flip the sign of normal. (This is an elementary reflection).
  3. Return back to the old representation.

- For example, let the hyperplane is normal to $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

  1. I know one choice of basis:

  $$b_1 = \begin{pmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{pmatrix}, \quad b_2 = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}, \quad b_3 = \begin{pmatrix} -\frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{6} \\ \frac{2\sqrt{6}}{6} \end{pmatrix}.$$

  2. The matrix B of old reprentation of new temporal basis is

  $$B = \begin{pmatrix} \frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & 0 & \frac{2\sqrt{6}}{6} \end{pmatrix}$$

  3. We then conduct

  $$x = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mapsto \begin{pmatrix} \frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & 0 & \frac{2\sqrt{6}}{6} \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

  $$\mapsto \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & 0 & \frac{2\sqrt{6}}{6} \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

  $$\mapsto \begin{pmatrix} \frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & 0 & \frac{2\sqrt{6}}{6} \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{3}}{3} & 0 & \frac{2\sqrt{6}}{6} \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = y$$

Example 3: non-elementary, non-orthogonal projection

- Goal: We want to compose a matrix that does the projection of a point onto the two-dimensional, or one-dimensional plane, in the manner described below.

- We let

$$b_1 \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

1. Get the new reprentation of x in temporal basis.
2. Project on to the axes. (This is an elementary projection).
3. Return back to the old representation.

- In Matrix Multiplications:

  1. The matrix B of old reprentation of new temporal basis is

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

  2. Let us admit that

  $$B^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

  3. We then conduct for example for $x = \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix}$

  $$x = \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

  $$\mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

  $$\mapsto \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$$

  4. We also compute this non-elementary, non-orthogonal projection matrix

  $$A = \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

You do the examples on Rotation, and Shearing that takes places with respect to axes you choose.

Home Work 2.

Consider

$$b_1 = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}, \quad b_2 = \begin{pmatrix} -\frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{6} \\ \frac{2\sqrt{6}}{6} \end{pmatrix}, \quad b_3 = \begin{pmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{pmatrix}.$$

Write the 3×3 rotation matrix, that makes a rotation θ counter-clock-wisely in the plane specified by b_1 and b_2, and that keeps the remaining component unchanged.

Optional Facts.

Question. How do I recognize which $n \times n$ matrix does which ?

- There are a few recognizable patterns. We introduce <u>FACTS</u> without proofs:

  1. Dilations with respect to orthogonal axes.

  $$\mathsf{S} = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is symmetric}\} \subset \mathbb{R}^{n \times n}.$$

  This includes, of course, diagonal and identity matrix.

  We saw in the lecture the example of $A = \begin{pmatrix} \frac{5}{4} & -\frac{3}{4} & 0 \\ -\frac{3}{4} & \frac{5}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$

  2. Projections are those: transform twice does not count. $PP = P$.
     (a) Projecting twice does not count:

     $$\mathsf{P} = \{A \in \mathbb{R}^{n \times n} \mid AA = A\} \subset \mathbb{R}^{n \times n}.$$

     (b) Indeed, If $P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, you can see $PP = P$.

     (c) If P is projection, then complementary matrix $I - P$ automatically has the same property, namely,

     $$(I - P)(I - P) = I\,I - P - P + PP = I - 2P + P = I - P.$$

     (d) Indeed, we check $A = BPB^{-1}$, where P is an elementary projection.

     $$AA = BPB^{-1}BPB^{-1} = BPIPB^{-1} = BPPB^{-1} = BPB^{-1}.$$

3. Rotations are those:

   $\mathsf{SO} = \{A \in \mathbb{R}^{n \times n} \mid (i) \text{ columns of } A \text{ are orthonormal. } (ii) \text{ columns are right-handed.}\}.$

   We saw the example of an elementary rotation $A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$

4. Reflections are those:

   $\mathsf{H} = \{A \in \mathbb{R}^{n \times n} \mid (i) \text{ columns of } A \text{ are orthonormal. } (ii) \text{ columns are left-handed.}\}.$

   We saw the example of an elementary reflection $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$

5. We know that upper triangular matrix does the shearing.

   $$\mathsf{T} = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is upper triangular.}\}$$

   As a matter of fact, every $n \times n$ matrix is a rotation of an upper triangular matrix.

- More importantly and as a matter of fact, every $n \times n$ matrix A can be written in the form
  $$A = B \ E_k E_{k-1} E_{k-2} \cdots E_2 E_1 \ B^{-1},$$
  where $E_k E_{k-1} E_{k-2} \cdots E_2 E_1$ is the finite consecutive applications of elementary transforms.

- General answer to the question is thus: We run the decomposition algorithm to obtain $A = B \ E_k E_{k-1} E_{k-2} \cdots E_2 E_1 \ B^{-1}$. Then you know what it does.

Final Remark

You should be able to tell clearly the differences of the following two:

If you can, then you understood very well what is going on by manipulating matrices.

1. Transformation 1

   - Let the 3×3 matrix $A_1 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$,

   - Let the 3×3 matrix $A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$,

   - Let the 3×3 matrix $A_2 = \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

   - We suppose A_1, A_2 and A_3 are the three representations of some linear transformations.

   - Then for a location x represented by $\begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix}$, we consider the three consecutive

     linear transformations:

     $$x = \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

     $$\mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

     $$\mapsto \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$$

   - But, what is being essentially done here is to obtain:

     $$\begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = y_1 \quad \text{the 1st new location in the same representation}$$

     $$\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = y_2 \quad \text{the 2nd new location in the same representation}$$

     $$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} = y_3 \quad \text{the 3rd new location in the same representation.}$$

2. Transformation 2 : the projection we did today.

   - Here, what is being essentially done is different things:

     Get the new representation of $x \mapsto$ Get the new location $y \mapsto$ Get y's old representation

$\implies$ What are being done essentially are different. Their representational matrix multiplications are the same. But you understand what's going on here now.

Chapter 6

Linear mappings from X to Y

- We will re-visit $\mathbb{R}^{n \times n}$ of square matrices, to discuss the determinat of a square matrix, the inverse if exists, how we can calculate them, and so on.

- Continuing discussions on the question: What does the matrix A do ?

- On the matrix $A \in \mathbb{R}^{m \times n}$ that is not square.

- A matrix in $\mathbb{R}^{n \times n}$ is called a square matrix. A matrix in $\mathbb{R}^{m \times n}$ is called a rectangular matrix. Of course, a square matrix is a rectangular matrix.

- Because we have developed the same on : "what does a square matrix A do?",

- We can also easily understand "what does a rectangular matrix do?" too.

- This is done through the notion of the rank of matrix.

First of all, in this chapter,

- $\mathbb{R}^{n \times 1}$ represent a vector space X of dimension n,

- $\mathbb{R}^{m \times 1}$ represent another vector space Y of dimension m,

- $\mathbb{R}^{m \times n}$ represent the space of linear mappings from X to Y.

Rank of a matrix A

Let $A \in \mathbb{R}^{m \times n}$, i.e., it has m rows and n columns, in the way

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\ A_{31} & A_{32} & A_{33} & \cdots & A_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mn} \end{pmatrix}.$$

Motivation and observation: when the number "one" is relevant

Consider a special matrix $A \in \mathbb{R}^{3 \times 5}$,

$$A = \begin{pmatrix} 1 & 2 & 10 & 200 & 3 \\ 2 & 4 & 20 & 400 & 6 \\ 3 & 6 & 30 & 600 & 9 \end{pmatrix}$$

so that every columns are constant multiple of the first column.

But, it is also such that every rows are constant multiple of the first row.

Try differently. Is this always so ?

Motivation and observation: when the number "two" is relevant

Consider a special matrix $A \in \mathbb{R}^{2 \times 3}$,

$$A = \begin{pmatrix} 3 & 0 & -9 \\ 1 & 2 & 1 \end{pmatrix}$$

Certainly, $(3, 0, -9)$ and $(1, 2, 1)$ are the two independent rows.

Are the three columns independent ?

No. Certainly, $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ are linearly independent.

However, $\begin{pmatrix} -9 \\ 1 \end{pmatrix} = -3 \begin{pmatrix} 3 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 2 \end{pmatrix}$, i.e., $\begin{pmatrix} -9 \\ 1 \end{pmatrix} \in \text{span} \langle \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix} \rangle$.

Thus, the number of linearly independent columns coincide the number "two".

It is not obvious to see why, but the number of linearly independent rows and the number of linearly independent columns seem to be the same.

To be able to check the statement in mathematical way, we do the followings.

1. We collect
$$\mathsf{span}\,\langle n \text{ columns of } A\rangle \subset \mathbb{R}^{m\times 1}.$$

2. And we also collect
$$\mathsf{span}\,\langle m \text{ rows of } A\rangle \subset \mathbb{R}^{1\times n}.$$

The question is how many number of linearly independent elements in them ?

Before we proceed, we introduce two notations.

- For $A \in \mathbb{R}^{m\times n}$, A^T, the transpose of A, is the matrix in $\mathbb{R}^{n\times m}$ where the rows and columns are interchanged, i.e.,

$$(A^T)_{ij} = A_{ji}, \quad j = 1, 2, \cdots, m, \quad i = 1, 2, \cdots, n.$$

- For two columns $x, y \in \mathbb{R}^{\ell\times 1}$, we say x and y are orthogonal to each other if

$$x^T y = 0.$$

We now give a name for the above two we collected.

1. We call the first one, as

$$\mathrm{Ran}\,(A), \quad \text{the range of } A$$

2. We consider the *transpose* the second one,

$$\mathrm{Trow}(A) = \mathsf{span}\,\langle m \text{ transpose of rows of } A\rangle \subset \mathbb{R}^{n\times 1}.$$

The range $\text{Ran}(A)$ is literally the range of A.

- Check: if $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, then according to the <u>Matrix Multiplication rule</u>,

  $Ax = x_1 c_1 + x_2 c_2 + \cdots x_n c_n$, a linear combination of columns of A.

- Here, $c_1, c_2, \cdots, c_n$ are columns of A.

A subset $E \subset X$ is called a vector subspace of X if E is also a vector space.

$\text{Ran}(A) \subset \mathbb{R}^{m \times 1}$ is a *vector subspace* of $\mathbb{R}^{m \times 1}$, and $\text{Trow}(A) \in \mathbb{R}^{n \times 1}$ is a *vector subspace* of $\mathbb{R}^{n \times 1}$.

- Indeed, any linear combination of elements in the set, is again in the set, i.e.,

  if $\mathbb{R}^{m \times 1}$ columns $a_1, a_2, \cdots, a_k \in \text{Ran}(A) \implies \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_k a_k \in \text{Ran}(A)$.

  if $\mathbb{R}^{n \times 1}$ columns $b_1, b_2, \cdots, b_\ell \in \text{Trow}(A) \implies \lambda_1 b_1 + \lambda_2 b_2 + \cdots + \lambda_\ell b_\ell \in \text{Trow}(A)$.

- We also check other conditions to be a vector space.

The objective of the next lecture is to reveal the following:

1. $\text{Trow}(A)$ and $\text{Ran}(A)$ are both r-dimensional, and

2. "The question we posed" can be understood by looking at the <u>square</u> matrix of $\mathbb{R}^{r \times r}$

   $$\text{Trow}(A) \to \text{Ran}(A).$$

   between the two.

3. What does the square matrix do? We completely know.

This is known as the Fundamental Theorem of Linear Algebra. We will go through this in the next lecture in detail.

Chapter 7

Fundamental Theorem of Linear Algebra

Here, we verify the picture:

through the fundamental theorem of linear algebra. We divide it into two parts.

We recall the orthogonality.

Theorem 1. *Let $x_0 \in \mathbb{R}^{n \times 1}$.*

$$x_0 \perp Trow(A) \quad \Longleftrightarrow \quad Ax_0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Proof.

$$\begin{aligned} x_0 \perp \text{Trow}(A) \quad &\Longleftrightarrow \quad x_0^T \left(\lambda_1 b_1 + \lambda_2 b_2 + \cdots + \lambda_m b_m \right) = 0, \quad b_1, b_2, \cdots, b_m \text{ are transpose of rows of } A \\ &\Longleftrightarrow \quad x_0^T b_i = 0 \quad \text{for every } i = 1, 2, \cdots, m \\ &\Longleftrightarrow \quad i\text{-th component of } Ax_0 = 0 \text{ for every } i = 1, 2, \cdots, m, \end{aligned}$$

according to the Matrix Multiplication Rule. $\square$

The theorem tells the following.

1. In $\mathbb{R}^{n \times 1}$, we choose orthonormal basis elements of $\mathbb{R}^{n \times 1}$ v_1, v_2, $\cdots$, v_r, v_{r+1}, $\cdots$, v_n such that

2. The first r basis elements v_1, v_2, $\cdots$, v_r span $\text{Trow}(A)$.

3. We let

$$V = \begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & \cdots & | \end{pmatrix}.$$

4. Then we can think of the $n \times n$ projection matrix P:

$$P = V \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} V^{-1}.$$

   This gives, for any $x \in \mathbb{R}^{m \times 1}$, the projections

$$x_0 = (I - P)x, \quad x' = Px, \quad x = x_0 + x'.$$

   x' is an element of $\text{Trow}(A)$.

5. x_0 part of x does not do anything under the multiplication of A:

$$Ax = A(x_0 + x') = 0 + Ax' = Ax'.$$

Theorem 2. *For any matrix $A \in \mathbb{R}^{m \times n}$,*

the number of lin. indep. elements in $Trow(A)$ = the number of lin. indep. elements in $Ran(A)$.

- In fact we only need to check that for every matrix $A \in \mathbb{R}^{m \times n}$,

  the number of lin. indep. elements in $\text{Trow}(A)$ $\leq$ the number of lin. indep. elements in $\text{Ran}(A)$.

- This is because, if that is true, then do the same on another matrix $B = A^T$,

  the number of lin. indep. elements in $\text{Trow}(A^T)$ $\leq$ the number of lin. indep. elements in $\text{Ran}(A^T)$.

Proof. We only check the inequality:

1. To do that: Suppose $b_1, b_2, \cdots, b_r \in \text{Trow}(A)$ are linearly independent. We show that $a_1 = Ab_1$, $a_2 = Ab_2$, $\cdots$, $a_r = Ab_r$ are linearly independent.

2. Method by proving contrapositive statement.

3. Suppose that $a_1, a_2, \cdots a_r$ are linearly dependent. i.e., we can find coefficients λ_1, λ_2, $\cdots$, λ_r, not all of them are zero, such that

$$\lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_r a_r = 0.$$

4. In other words, by linearity,

$$A(\lambda_1 b_1 + \lambda_2 b_2 + \cdots \lambda_r b_r) = 0.$$

5. But Theorem 1 says that then

$$x_0 = \lambda_1 b_1 + \lambda_2 b_2 + \cdots \lambda_r b_r \in \text{Trow}(A) \quad \text{is orthogonal to all of Trow}(A).$$

6. The only element in $\text{Trow}(A)$ that is orthogonal to all of $\text{Trow}(A)$ is only

$$x_0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0. \end{pmatrix}.$$

7. This concludes that $b_1, b_2, \cdots, b_r \in \text{Trow}(A)$ are linearly dependent.

$\square$

The theorem tells the following.

1. In $\mathbb{R}^{n \times 1}$, we choose orthonormal basis elements of $\mathbb{R}^{n \times 1}$ $v_1, v_2, \cdots, v_r, v_{r+1}, \cdots,$ v_n such that

2. The first r basis elements $v_1, v_2, \cdots, v_r$ span $\mathrm{Trow}\,(A)$.

3. In $\mathbb{R}^{m \times 1}$, we do the similar, choose orthonormal basis elements of $\mathbb{R}^{m \times 1}$ $w_1, w_2,$ $\cdots, w_r, w_{r+1}, \cdots, w_m$ such that

4. The first r basis elements $w_1, w_2, \cdots, w_r$ span $\mathrm{Ran}\,(A)$.

5. Importantly the number r is shared !

6. The mapping goes in the three steps: Let $x \in \mathbb{R}^{n \times 1}$.

step 1 x first undergoes the <u>reduced projection</u>:

$$x \mapsto \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} V^{-1} x = x''.$$

x'' is in temporal representation with new basis, but for convenience, we let the components from $(r+1)$-th is removed. Thus, $x'' \in \mathbb{R}^{r \times 1}$.

step 2 The next step is the mapping from $\mathrm{Trow}\,(A)$ to $\mathrm{Ran}\,(A)$. We divide this into two steps, one is to take *internal transformation*

$$x'' \mapsto Tx'', \quad T : \mathbb{R}^{r \times 1} \to \mathbb{R}^{r \times 1}.$$

step 3 The last step is the <u>image copy</u>. In the last stage, the $\mathbb{R}^{r \times 1}$ column $Tx'' = x'''$ is in temporal representation with $v_1, v_2 \cdots, v_r$.

$$\begin{pmatrix} x_1''' \\ x_2''' \\ \vdots \\ x_r''' \end{pmatrix}.$$

This is mapped to

$$x_1''' w_1 + x_2''' w_2 + \cdots + x_r''' w_r = \begin{pmatrix} | & | & \cdots & | \\ w_1 & w_2 & \cdots & w_r \\ | & | & \cdots & | \end{pmatrix} \begin{pmatrix} x_1''' \\ x_2''' \\ \vdots \\ x_r''' \end{pmatrix}.$$

This ends the whole picture. A rectangular matrix $A \in \mathbb{R}^{m \times n}$ can be understood by the $r \times r$ matrix T in the following decomposition

$$A = W\,T\,P\,V^{-1}.$$

V^{-1}	: $n \times n$ matrix	basis selection w.r.t $\mathrm{Trow}\,(A)$.
P	: $r \times n$ matrix	projection onto $\mathrm{Trow}\,(A)$, reduced
T	: $r \times r$ matrix	some $\mathbb{R}^{r \times r}$ transformation of $\mathrm{Trow}\,(A)$
W	: $m \times r$ matrix	the final image mapping to $\mathrm{Ran}\,(A)$

Chapter 8

Computing by Hands:
The Picture of Fundamental
Theorem of Linear Algebra

In this chapter, we discuss about following questions. Suppose a matrix $A \in \mathbb{R}^{m \times n}$ is given.

1. How to pick up basis elements of $\operatorname{Ran}(A)$?

2. How to pick up basis elements of $\operatorname{Trow}(A)$?

3. How to pick up basis elements of $\operatorname{Trow}(A)^{\perp}$?

4. How to pick up basis elements of $\operatorname{Ran}(A)^{\perp}$?

5. How to calculate the rank r?

In case the shape of the matrix A is of very big size, in general, those questions are subjected to running some algorithm using computer.

However, if the size of the matrix A is not too large, you would want to do it by your hands with pen and paper. Today, we introduce one technique starting with two Good News.

1. In fact, you already know this.

2. The way it achieves is so nice that you would extract good enough information from a certain amount of work.

Row operation: Replacement of A by A'

- Here, we consider A' of same shape, instead of A, anyone satisfying $\text{Trow}(A') = \text{Trow}(A)$, i.e., the **span** of rows of A' and the **span** of rows of A are the same.

- You are free to change the order of rows.

- One typical example is to pick up one of the row of A, and replace it by the row itself + a linear combination of remaining rows of A.

- We could equivalently do the Column operation. Traditionally, this technique has been done in the way of Row Operation.

Here comes the first good news, that you already know this:

Example
$$\begin{cases} 2x - 3y = 5 \\ x - 2y = 2 \end{cases} \qquad \text{Let} \quad A = \begin{pmatrix} 2 & -3 & 5 \\ 1 & -2 & 2 \end{pmatrix}$$

1. According to the Matrix Multiplication Rule, Row operation is implemented by multiplying $m \times m$ matrix from the left.

2. For our purposes, it is matter of choice whether you involves those $m \times m$ matrices or you omit them. Do what are conveninent for you.

The purpose of the replacement, and why it is easy

1. We are here, replacing the matrix A sequentially by

$$A \to A' \to A'' \to A''' \to \cdots$$

while keeping $\text{Trow}(A) = \text{Trow}(A') = \text{Trow}(A'') = \text{Trow}(A''') = \cdots$.

2. Of course, those matrices appear later have to be simpler than ones appeared before, in what so ever senses. For example, there are more 0 in the entries. Or we can easily extract some basis elements of $\text{Trow}(A)$ out of it.

3. By contrast, the information on $\text{Ran}(A)$ are completely forgotten. In general,

$$\text{Ran}(A) \neq \text{Ran}(A'), \quad \text{Ran}(A') \neq \text{Ran}(A''), \quad \cdots$$

and in this stage we simply do not care whatever happens in range.

$\implies$ The consequence then will be:

- We will see much simpler picture of $\text{Trow}(A)$,

- We will still keep the same $\text{rank}(A)$, thanks to the Fundamental Theorem.

- We will completely lose the picture of $\text{Ran}(A)$.

- But once the Row operation is finished, column picture will be so nicely recovered. This is the second good news.

- We do this one by one.

- The goal of this is known as the Column-Row (CR)-decomposition, that can be seen in the Lecture of G. Strang.

What do we do ?:

1. Introduce 0 as many as possible.

2. Turn it to the row echelon form.

3. Turn it to the reduced row echelon form.

4. Check if calculations went correct: Complete the Column-Row decomposition.

Example 2

$$A = \begin{pmatrix} 1 & 3 & 10 & 200 \\ 2 & 6 & 20 & 400 \\ 3 & 9 & 30 & 600 \end{pmatrix}$$

Example 3

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 3 & 7 \\ 1 & 4 & 6 \end{pmatrix}$$

Example 4

$$A = \begin{pmatrix} 0 & -7 & -4 & 2 \\ 2 & 4 & 6 & 12 \\ 3 & 1 & -1 & -2 \end{pmatrix}$$

Example 5

$$A = \begin{pmatrix} 0 & 1 & 2 & 1 & 2 \\ 1 & 0 & 0 & 0 & 1 \\ 2 & 1 & 2 & 1 & 5 \\ 1 & 2 & 4 & 3 & 6 \end{pmatrix}$$

More on computation by hands

Example: Complete the CR decomposition of

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 3 & 7 \\ 1 & 4 & 6 \end{pmatrix}$$

Summarize: For example, we arrived at

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 3 & 7 \\ 1 & 4 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}.$$

1. $\operatorname{rank}(A)$: (RHS) is a product of (3×2) and (2×3) matrices. $\operatorname{rank}(A) = 2$.

2. $\operatorname{Trow}(A)$: $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$ is in much simpler form. Big improvement.

3. $\operatorname{Ran}(A)$: $\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 4 \end{pmatrix}$ at least reveals which columns are independent. Small improve-
ment

1. We already know the $\operatorname{rank}(A)$.

2. We select basis elements of the domain $\mathbb{R}^{n \times 1}$, adapted to $\operatorname{Trow}(A)$.

3. We select basis elements of the codomain $\mathbb{R}^{m \times 1}$, adapted to $\operatorname{Ran}(A)$.

Selecting basis elements of the domain $\mathbb{R}^{n \times 1}$, adapted to Trow (A)

1. We select basis of Trow (A).

   Out of the last form
   $$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$$

   We select two elements
   $$\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

   that span Trow (A).

2. Now, we select basis of Trow $(A)^{\perp}$.

   (a) This is to find element $x_0 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ such that
   $$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

   (b) This equation is super easy to solve.

   (c) Recalling that the $1st$ and $2nd$ columns are marked independent columns, we put $c = -1$, and let $\begin{pmatrix} a \\ b \end{pmatrix}$ be the precisely $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$. This gives rise to
   $$\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}.$$

We have selected all the basis for the domain:
$$\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}.$$

The first two span Trow (A), and the last span Trow $(A)^{\perp}$. They are not orthogonal to each other though.

Selecting basis elements of the codomain $\mathbb{R}^{m \times 1}$, adapted to $\mathbf{Ran}\,(A)$

One possible way:

1. We select basis of $\mathrm{Ran}\,(A)$.
   Out of the last form
   $$\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 4 \end{pmatrix}$$
   We may select two elements
   $$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}.$$
   that span $\mathrm{Ran}\,(A)$.

2. To select basis of $\mathrm{Ran}\,(A)^{\perp}$, we solve
   $$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

One another way I would do:

- Swapping the role of row and column, I consider A^T, and turn A^T into the Reduce Row Echelon Form.
- In fact, it suffices to turn C^T into the Reduced Row Echelon Form.
- Let us do that. Out of C^T,

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

- We turn it into
$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} 1 & 0 & -5 \\ 0 & 1 & 3 \end{pmatrix}$$

1. Now, the we select basis of $\mathrm{Ran}\,(A)$.
   $$\begin{pmatrix} 1 \\ 0 \\ -5 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}.$$

2. Select basis of $\mathrm{Ran}\,(A)^{\perp}$:
   $$\begin{pmatrix} -5 \\ 3 \\ -1 \end{pmatrix}.$$

We have selected all the basis for the domain:
$$\begin{pmatrix} 1 \\ 0 \\ -5 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}, \quad \begin{pmatrix} -5 \\ 3 \\ -1 \end{pmatrix}.$$

The first two span $\mathrm{Ran}\,(A)$, and the last span $\mathrm{Ran}\,(A)^{\perp}$. They are not orthogonal to each other though.

(Optional Material)

If you also have completed the decomposition of C^T,

$$C^T = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & -5 \\ 0 & 1 & 3 \end{pmatrix},$$

You may write

$$A = CR = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -5 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}.$$

Home Work

Do each of following steps for the matrices written below.

1. Turn the matrix A into the Reduced Row Echelon Form.

2. Complete the $A = CR$ decomposition, checking if the calculation was correct.

3. Turn the matrix C^T into the Reduced Row Echelon Form.

4. Compute the rank of A.

5. Select basis elements of domain $\mathbb{R}^{n \times 1}$:

   (a) basis elements of $\mathrm{Trow}\,(A)$,

   (b) basis elements of $\mathrm{Trow}\,(A)^{\perp}$.

6. Select basis elements of codomain $\mathbb{R}^{m \times 1}$:

   (a) basis elements of $\mathrm{Ran}\,(A)$,

   (b) basis elements of $\mathrm{Ran}\,(A)^{\perp}$.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 5 \\ 2 & 4 & 6 & 9 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Chapter 9

Invertible Square Matrices

We get back to the special cases of $n \times n$ square matrices. In this chapter, unless otherwise specified, all matrice are square $n \times n$ matrices.

We begin with observing the following nice feature on $n \times n$ matrices.

1. If A and B are $n \times n$ matrices, we can perform multiplications both AB and BA, although in general $AB \neq BA$.

2. AB and BA are again $n \times n$ matrices. Hence, we can for instance perform $ABCDEF$ of many $n \times n$ matrices.

3. There exists unique $n \times n$ matrix $I \in \mathbb{R}^{n \times n}$ such that

$$\text{for every } A \in \mathbb{R}^{n \times n}, \quad AI = IA = A.$$

Hence, one naturally asks a question that for a given A

1. Is there a G such that $GA = I$?

2. Is there a H such that $AH = I$?

3. If so, how can I compute G or H ?

We do the following in this chapter:

- We give the definition of invertibility of a matrix $A \in \mathbb{R}^{n \times n}$.

- Calculation of A^{-1} in case A is invertible, by the Row Reduction.

Let us state two important theorems about the invertibility of A.

Theorem 1. *Let $A \in \mathbb{R}^{n \times n}$. Then followings are all equivalent.*

1. *There exists G such that $AG = I$.*

2. *There exists H such that $HA = I$.*

3. *There exists E such that $AE = EA = I$.*

4. *A has full rank, that is n.*

Theorem 2 (uniqueness of inverse)**.** *Suppose that $EA = AE = I$ and $FA = AF = I$. Then $E = F$.*

Proof.

$$EA = I \quad \implies \quad EAF = F \quad \implies \quad E = F.$$

$\square$

Based on the two theorems, we give the following definition.

Definition 3. *For a given $A \in \mathbb{R}^{n \times n}$, we say A is invertible if there exists a matrix $E \in \mathbb{R}^{n \times n}$ such that*

$$AE = EA = I,$$

which must then be unique. E is said to be the inverse matrix of A, and denoted by A^{-1}.

Because this course is more about practical aspects of Linear Algebra, we give proof of the Theorem 1 after we learn

- how to know if A is invertible or not,

- if invertible, how to compute A^{-1}.

Second look : Turning A to RREF

- Although we perform each of replacement turning A to RREF in the high speed and easy-going manner,

- All we did consists of the following two:

    1. Swapping two rows of A.

    2. Replace a row of A by

        The row itself $\times a$ $(a \neq 0)$ + Linear Combination of the remaining rows.

- The each of above two is implemented by multiplying a square matrix E from left: Indeed, for example if $A \in \mathbb{R}^{4,n}$,

    1. Swapping 1st and 3rd row in $\mathbb{R}^{4 \times 4}$:

    $$E = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{to } A$$

    2. Replace the 2nd row in $\mathbb{R}^{4 \times 4}$:

    $$E' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \lambda_1 & a & \lambda_3 & \lambda_4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{to } A \quad a \neq 0$$

    implements the operations.

- Hence, RREF after a finite number of such implementations to A are described by

$$E_k E_{k-1} E_{k-2} \cdots E_2 E_1 A = R, \quad R \text{ is the RREF of } A.$$

    Associativity law gives that we perform all of the implementation first,

$$R = \left(E_k E_{k-1} E_{k-2} \cdots E_2 E_1 \right) A = EA, \quad E = E_k E_{k-1} E_{k-2} \cdots E_2 E_1.$$

- Now, let us think about what will be the RREF of A.

    1. Suppose A is invertible. What is its RREF?
       It must be the identity matrix, no other choice.

    $$EA = I.$$

       If we record E of the accumulation of all implementations, then it is the inverse.

    2. On the other hands, if A is not invertible, its RREF must not be the identity matrix.

How to record RREF implementations?

- We do the exactly same operations parallely to both A and the identity matrix I.

- Then all the implementations will turn A into RREF (identity matrix if A is invertible),

- and will turn I into all its accumulation E, that must be A^{-1}.

Example 1

$$A = \begin{pmatrix} 1 & 3 & 200 \\ 2 & 6 & 400 \\ 3 & 9 & 600 \end{pmatrix}$$

Example 2

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 3 & 7 \\ 1 & 4 & 6 \end{pmatrix}$$

Example 3

$$A = \begin{pmatrix} 0 & -7 & -4 \\ 2 & 4 & 6 \\ 3 & 1 & -1 \end{pmatrix}$$

Example 4

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

Example 5

$$A = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 0 & 1 \\ 2 & 1 & 1 & 5 \\ 1 & 2 & 3 & 6 \end{pmatrix}$$

Proof of Theorem 1

Theorem 1. *Let $A \in \mathbb{R}^{n \times n}$. Then followings are all equivalent.*

1. *There exists G such that $AG = I$.*

2. *There exists H such that $HA = I$.*

3. *There exists E such that $AE = EA = I$.*

4. *A has full rank, that is n.*

Proof. We will prove

1. $(1) \implies (4) \implies (3) \implies (1)$,

2. $(2) \implies (4) \implies (3) \implies (2)$.

Then $(1),(2),(3),(4)$ are all equivalent.

- $(1) \implies (4)$.

  We show that $\operatorname{Ran}(A) = \mathbb{R}^{n \times 1}$.

  For every $y \in \mathbb{R}^{n \times 1}$, let $x = Gy$. Then $Ax = AGy = Iy = y$.

- $(4) \implies (3)$.

  - Because A is $n \times n$ square matrix with full rank, its Reduced Row Echelon Form must be precisely the identity matrix. In other words, there exists $HA = I$.

  - Because A^T is $n \times n$ square matrix with full rank, its Reduced Row Echelon Form must be precisely the identity matrix. In other words, there exists $G^T A^T = I$.

  - If so,
  $$HA = I \quad \implies \quad HAG = G \quad \implies \quad H = G.$$

  Hence, by letting $E = H = G$, (3) is proven.

- $(3) \implies (1)$

  This is obvious.

- $(2) \implies (4)$

  We show that $\operatorname{Ran}(A^T) = \operatorname{Trow}(A) = \mathbb{R}^{n \times 1}$. We know that $A^T H^T = I$.

  For every $y \in \mathbb{R}^{n \times 1}$, let $x = H^T y$. Then $A^T x = A^T H^T y = Iy = y$.

- $(4) \implies (3)$.

  This has been done.

- $(3) \implies (2)$

  This is obvious.

$\square$

Chapter 10

Determinant of a square Matrix

Let $A \in \mathbb{R}^{m \times n}$.

- $\operatorname{rank}(A) \in \mathbb{Z}$,
- $\operatorname{Trow}(A) \subset \mathbb{R}^{n \times 1}$,
- $\operatorname{Ran}(A) \subset \mathbb{R}^{m \times 1}$

then provide clearer picture on the roles of A played between $\mathbb{R}^{n \times 1}$ and $\mathbb{R}^{m \times 1}$.

We now introduce a real number, out of a matrix A when A is a square $n \times n$ matrix.

$$\det(A) \in \mathbb{R}, \quad \text{for } A \text{ square matrix.}$$

It is called the determinant of A.

We will do:

1. What is the determinant ?

2. Calculation method

3. What it is used for

4. Properties.

From now on in this chapter, every matrix is square.

What is the determinant of A?

- We use notation $\det(A)$, and in addition, sometimes it is useful to use notation

  $\det(a_1, a_2, \cdots, a_n)$, $a_1, a_2, \cdots a_n$ are rows of A from top to bottom.

- This is to interprete det as a function of n rows of A.

- One can also view det as a function of n columns of A

  $\det(b_1, b_2, \cdots, b_n)$ $b_1, b_2, \cdots b_n$ are columns of A from left to right.

  which gives the same real number.

- The use of rows is just for traditional reason.

The real-valued function $\det : (a_1, a_2, \cdots, a_n) \mapsto \mathbb{R}$ we are interested in is, for some reason, a function such that

1. Linear in a certain sense:

$$\det\left(\left(\sum_{k=1}^{m} \lambda_k x_k\right), a_2, a_3, \cdots, a_n\right) = \sum_{k=1}^{m} \lambda_k \det(x_k, a_2, a_2, \cdots, a_n)$$

$$\det\left(a_1, \left(\sum_{k=1}^{m} \lambda_k x_k\right), a_2, a_3, \cdots, a_n\right) = \sum_{k=1}^{m} \lambda_k \det(a_1, x_k, a_3, a_4, \cdots, a_n)$$

$$\vdots$$

$$\det\left(a_1, a_2, a_3, \cdots, a_{n-1}, \left(\sum_{k=1}^{m} \lambda_k x_k\right)\right) = \sum_{k=1}^{m} \lambda_k \det(a_1, a_2, a_3, \cdots, a_{n-1}, x_k)$$

$\Big($This is not to say $\det(A + B) = \det(A) + \det(B)$!, in general, $\det(A + B) \neq \det(A) + \det(B)$!$\Big)$

2. Anti-symmetric about consecutive row swap, i.e.,

   for any i, $\det(\cdots, a_i, a_{i+1}, \cdots) = -\det(\cdots, a_{i+1}, a_i, \cdots)$.

For the moment, we will not discuss why.

Let us take an example calculation what this real number should be for a given matrix.

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Using the item 1,

$$\det(A) = a \det \begin{pmatrix} 1 & 0 \\ c & d \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 \\ c & d \end{pmatrix}$$

$$= ac \det \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + ad \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + bc \det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + bd \det \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Note that the item 2 implies that if two consecutive rows are same, its det is zero:

$$\det(\cdots, a, a, \cdots) = -\det(\cdots, a, a, \cdots) \quad \text{can happen only if the number is zero.}$$

Therefore, using item 2,

$$\det(A) = ad \det \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + bc \det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$= (ad - bc) \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

What number for the det of the identity matrix would you want to assign ?

Thus, in addition to the item 1 and 2,

3. We want the det function to be such that

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

Let us also compute the determinant of 3×3 matrix

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

From the calculation, we see that we only need to know the determinant of the very special form of matrices such as

$$
\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad
\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad
\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \cdots
$$

These are <u>permutations</u> that swaps the rows if multiplied from the left.

How many permutations exists from row index $i \in \{1, 2, \cdots, n\}$ to the new row index $i' \in \{1, 2, \cdots, n\}$? $n!$

In fact, we only need to know, how many consecutive swaps are needed to turn a permutation matrix to the identity matrix.

Proposition 1. *For a permutation matrix, the number of consecutive swaps to turn to the identity matrix is either even or odd, and the evenness or the oddness is independent of the way we turn.*

Definition 2. *Let $A \in \mathbb{R}^{n \times n}$. We define a real number*

$$
\det(A) = \sum_{\substack{among\ every\ permutation\ \sigma}} \left(A_{1,\sigma(1)} A_{2,\sigma(2)} A_{3,\sigma(3)} \cdots A_{n,\sigma(n)} \right) \det(P_\sigma)
$$

$$
= \sum_{\substack{among\ every\ permutation\ \sigma}} \left(A_{1,\sigma(1)} A_{2,\sigma(2)} A_{3,\sigma(3)} \cdots A_{n,\sigma(n)} \right) sgn(P_\sigma).
$$

We will show that the real number $\det(A)$ is the number:

1. Its sign tells that rows of A from the top to bottom is right-handed or left-handed, together with fact that

2. $\det(A) = 0$ if and only if rows of A is not independent.

3. Its maginitude is precisely the volume of the parallelpiped the n rows of A with origin makes.

Chapter 11

Determinant and RREF

Row operations and RREF again plays important role in proving many properties about the det.

Let us consider the two elementary row operations, the matrix E multiplied from the left on A, say in turnig A into RREF:

1. Swapping rows, i.e., E is a permutation. For example,

$$E = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

2. Replacing i-th row with $a \times row_i$ + linear combination of the remaining row, with $a \neq 0$. For example,

$$E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \lambda_3 \neq 0.$$

The effect of muliplying E on A from left, in terms of det change from $\det(A)$ to $\det(EA)$, is in fact very simple.

Recall that If any of two rows of A are same, then the det is 0. It is easy to see then

Proposition 1. .

1. If rows of A is dependent, then $\det(A) = 0$.

2. If there is 0 row in A, then $\det(A) = 0$.

Proof. If i-th row can be written by linear combination of remaning rows,

$$a_i = \sum_{j \neq i} \lambda_j a_j \quad \text{then} \quad \det(\cdots, a_{i-1}, \sum_{j \neq i} \lambda_j a_j, \cdots) = \sum_{j \neq i} \lambda_j \det(\cdots, a_{i-1}, a_j, \cdots)$$

but all the terms in (RHS) has same rows (i-th and j-th), and thus sum must be 0. □

Having observe those, we can now see:

1. If E is a permutation, then

$$\det(EA) = \text{sign}\,(E) \det(A).$$

2. If E is to replace i-row by $\sum_{j=1}^{n} \lambda_j a_j$, $\lambda_i \neq 0$, then

$$\begin{aligned}
\det(EA) &= \det(\cdots, a_{i-1}, \sum_{j=1}^{n} \lambda_j a_j, \cdots) \\
&= \det(\cdots, \lambda_i a_i, \cdots) \\
&= \lambda_i \det(\cdots, a_i, \cdots) = \lambda_i \det(A).
\end{aligned}$$

Moreover,

1. If E is a permutation, by definition

$$\text{sign}\,(E) = \det(E).$$

2. If E is to replace $\sum_{j=1}^{n} \lambda_j a_j$, $\lambda_i \neq 0$, then

$$\begin{aligned}
\det(E) &= \det(e_1, e_2, \cdots, e_{i-1}, \sum_{j=1}^{n} \lambda_j e_j, e_{i+1}, \cdots, e_n) \\
&= \det(e_1, e_2, \cdots, e_{i-1}, \lambda_i e_i, e_{i+1}, \cdots, e_n) \\
&= \lambda_i \det(e_1, e_2, \cdots, e_{i-1}, e_i, e_{i+1}, \cdots, e_n) = \lambda_i \det(I) = \lambda_i.
\end{aligned}$$

We proved that (i): $\det(E)$ of above two types never is 0. Also we proved that (ii):

Proposition 2.
$$\det(EA) = \det(E) \det(A).$$

Determinant of A and determinant of its RREF

Now, let

$$E_k E_{k-1} E_{k-2} \cdots E_1 A = R,$$

where R is the RREF of A.

Using the Proposition 2, we have

$$
\begin{aligned}
\det(R) &= \det\left(E_k E_{k-1} E_{k-2} \cdots E_1 A\right) \\
&= \det(E_k) \det\left(E_{k-1} E_{k-2} \cdots E_1 A\right) \\
&= \cdots \\
&= \det(E_k) \det(E_{k-1}) \cdots \det(E_1) \det(A).
\end{aligned}
$$

Hence we proved that:

Proposition 3.

$$\det(A) = \frac{\det(R)}{\det(E_k) \det(E_{k-1}) \cdots \det(E_1)},$$

where we used that $\det(E_k) \det(E_{k-1}) \cdots \det(E_1) \neq 0$.

Proposition 4. $\det(A) = 0$ *if and only if rows of A is dependent.*

Proof. .

- We already checked $\det(A) = 0$ if rows of A is dependent, begining this chapter.

- We need to show that rows of A is dependent if $\det(A) = 0$. This is obvious from the formula in Proposition 3.

$\square$

Continuing investigation of det,

Proposition 5.
$$\det(AB) = \det(A)\det(B).$$

Proof. .

1. Suppose $\det(A) = 0$.

   Then $\text{rank}(A) < n$. In other words, $\text{Ran}(A) \neq \mathbb{R}^{n \times 1}$. Because any linear combination of columns of AB is a linear combination of columns A, $\text{Ran}(AB) \neq \mathbb{R}^{n \times 1}$ too. Hence $\det(AB) = 0$.

2. Suppose $\det(B) = 0$.

   Then $\text{rank}(B) < n$. In other words, $\text{Trow}(B) \neq \mathbb{R}^{n \times 1}$, and there is nontrivial x_0 such that $Bx_0 = 0$. Because $ABx_0 = 0$. $\text{Trow}(AB) \neq \mathbb{R}^{n \times 1}$ too. Hence $\det(AB) = 0$.

3. Suppose $\det(A) \neq 0$ and $\det(B) \neq 0$.

   A and B are both invertible. Let $C = AB$ and write $B = A^{-1}C$.

   Recall that there exist row operations so that

   $$A^{-1} = E_k E_{k-1} E_{k-2} \cdots E_1.$$

   Therefore,

   $$\det(B) = \det(E_k)\det(E_{k-1}) \cdots \det(E_1)\det(C)$$
   $$\iff \det(B) = \frac{\det(RREF(A))}{\det(A)}\det(C).$$

   But, since A is invertible, its RREF is identity matrix. Hence we derived

   $$\det(B) = \frac{\det(C)}{\det(A)}.$$

   $\square$

Using Proposition 5, if A is invertible, $\det(A^{-1}) = \frac{1}{\det(A)}$.

Next, we verify that $\det(A)$ can be computed column perspective too:

Proposition 6. $\det(A^T) = \det(A)$.

Proof. .

1. From the definition of $\det(A)$,

$$\det(A) = \sum_{\substack{\text{among every permutation} \\ \sigma:\{1,2,\cdots,n\}\to\{1,2,\cdots,n\}}} \left(A_{1,\sigma(1)} A_{2,\sigma(2)} A_{3,\sigma(3)} \cdots A_{n,\sigma(n)} \right) \mathsf{sgn}(P_\sigma).$$

2. We note one observation: For each iteration, the product of n real numbers $A_{1,\sigma(1)} A_{2,\sigma(2)} A_{3,\sigma(3)} \cdots A_{n,\sigma(n)}$ is same as the product of the following n real numbers

$$A_{\sigma^{-1}(1),1} A_{\sigma^{-1}(2),2} A_{\sigma^{-1}(3),3} \cdots A_{\sigma^{-1}(n),n}.$$

3. Hence,

$$\det(A) = \sum_{\substack{\text{among every permutation} \\ \sigma:\{1,2,\cdots,n\}\to\{1,2,\cdots,n\}}} \left(A_{\sigma^{-1}(1),1} A_{\sigma^{-1}(2),2} A_{\sigma^{-1}(3),3} \cdots A_{\sigma^{-1}(n),n} \right) \mathsf{sgn}(P_\sigma)$$

(writing $s = \sigma^{-1}$)

$$= \sum_{\substack{\text{among every permutation} \\ s:\{1,2,\cdots,n\}\to\{1,2,\cdots,n\}}} \left(A_{s(1),1} A_{s(2),2} A_{s(3),3} \cdots A_{s(n),n} \right) \mathsf{sgn}(P_{s^{-1}})$$

$$= \sum_{\substack{\text{among every permutation} \\ s:\{1,2,\cdots,n\}\to\{1,2,\cdots,n\}}} \left(A_{s(1),1} A_{s(2),2} A_{s(3),3} \cdots A_{s(n),n} \right) \mathsf{sgn}(P_s)$$

$$= \sum_{\substack{\text{among every permutation} \\ s:\{1,2,\cdots,n\}\to\{1,2,\cdots,n\}}} \left(A^T_{1,s(1)} A^T_{2,s(2)} A^T_{3,s(3)} \cdots A^T_{n,s(n)} \right) \mathsf{sgn}(P_s)$$

$$= \det(A^T),$$

4. where we admited the fact that $\text{sign}(P_s) = \text{sign}(P_{s^{-1}})$.

$\square$

$|\det(A)|$ is the n-dimensional volume.

We do this for the simple 2×2 case, and generalize to $n \times n$ case.

Notations

We do this from column perspective.

- Let A be a 2×2 matrix with columns x_1 and x_2.

$$A = \begin{pmatrix} | & | \\ x_1 & x_2 \\ | & | \end{pmatrix}.$$

- The locations in $\mathbb{R}^2$ from x_1 and x_2, decides a parallelogram, together with the origin 0. We denote the area of this parallelogram by

$$\mathsf{vol}(0, x_1, x_2).$$

  We define $\mathsf{vol}(0, x_1, x_2) = 0$ if A has dependent columns.

Upper triangular 2×2 matrix

- Consider an invertible 2×2 upper triangular matrix,

$$U = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}, \quad a \neq 0, d \neq 0.$$

- Let x_1 and x_2 to be its two columns from left. Then, the parallelogram we see is such that

$$\text{base length} = |a| \quad \text{height length} = |d|,$$
$$\mathsf{vol}(0, x_1, x_2) = |ad|.$$

- On the other hand, the determinant of the triangular matrix U is simply

$$\det(U) = ad.$$

- Hence, we conclude that

  If U is 2×2 upper triangular matrix, then $\quad \mathsf{vol}(0, x_1, x_2) = |\det(U)|.$

  This is true even when U is not invertible.

Rotation

- Now, let A be a general 2×2 matrix written as

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

- In this 2×2 case, it is obvious that there is a rotation matrix Q so that the rotated location

$$x_1' = Qx_1 \quad \text{is aligned in the 1st-axis.}$$

  In other words,

$$QA = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix} = U.$$

- By a rotation, the area of parallelogram must not change:

$$\mathsf{vol}(0, x_1, x_2) = \mathsf{vol}(0, x_1', x_2').$$

- Therefore,

$$\mathsf{vol}(0, x_1, x_2) = \mathsf{vol}(0, x_1', x_2') = |\det(U)| = |\det Q \det A| = |\det A|$$

  and we are done.

A **is** $n \times n$ **matrix**

- Let

$$A = \begin{pmatrix} | & | & \cdots & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & \cdots & | \end{pmatrix}$$

- The locations in $\mathbb{R}^n$ from x_1, x_2, $\cdots$, x_n decides a n-dimensional parallelpiped, together with the origin 0. We consider similarly

$$\mathsf{vol}(0, x_1, x_2, \cdots x_n).$$

- Now, we consider $n \times n$ upper triangular matrix U:

$$U = \begin{pmatrix} U_{11} & U_{12} & U_{13} & \cdots & U_{1n} \\ 0 & U_{22} & U_{23} & \cdots & U_{2n} \\ 0 & 0 & U_{33} & \cdots & U_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & U_{nn} \end{pmatrix}$$

Proposition 7. *The determinant of upper triangular or lower triangular matrix is the product of diagonal components.*

Proof. This is obvious from the definition of

$$\det(A) = \sum_{\substack{\text{among every permutation} \\ \sigma:\{1,2,\cdots,n\} \to \{1,2,\cdots,n\}}} \left(A_{1,\sigma(1)} A_{2,\sigma(2)} A_{3,\sigma(3)} \cdots A_{n,\sigma(n)} \right) \mathsf{sgn}(P_\sigma).$$

$\square$

We have that for the upper triangular matrix case

$$\mathsf{vol}(0, x_1, x_2, \cdots x_n) = |U_{11} U_{22} U_{33} \cdots U_{nn}|.$$

- Hence, we conclude that

    If U is $n \times n$ upper triangular matrix, then $\quad \mathsf{vol}(0, x_1, x_2, \cdots, x_n) = |\det(U)|.$

- We admit the following fact without proof:

    Theorem. *For any $n \times n$ matrix A, there exists an $n \times n$ rotation matrix Q with $\det(Q) = 1$ such that QA is an upper triangular matrix.*

- Finally, since the volume must not change under the rotation,

$$\mathsf{vol}(0, x_1, x_2, \cdots, x_n) = \mathsf{vol}(0, x_1', x_2', \cdots, x_n') = |\det(U)| = |\det(Q)\det(A)| = |\det(A)|.$$

Chapter 12

Collecting $n!$ numbers

Here, we re-visit the calculation of det, which is to collect $n!$ numbers.

We recall the computation of $\det(A)$ of 3×3 matrix

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

1. We first expand $\det(A)$ into n terms

$$\det(A) = a \det \begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + c \det \begin{pmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{pmatrix}$$

2. Each of n terms, one more depth expansion only involves $n-1$ terms, since two same rows implies zero det: From the first term above,

$$a \det \begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} = ae \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ g & h & i \end{pmatrix} + af \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ g & h & i \end{pmatrix}.$$

3. Recursively, each of $n-1$ terms, one more depth expansion only involves $n-2$ terms: From the first term above,

$$ae \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ g & h & i \end{pmatrix} = aei \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

Collecting exactly $n!$ numbers for the det.

We do the $\det(A)$ of 4×4 matrix

$$A = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix}$$

1. We first expand $\det(A)$ into 4 terms

$$\det(A) = a \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ e & f & g & h \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 & 0 & 0 \\ e & f & g & h \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix} + c \det \begin{pmatrix} 0 & 0 & 1 & 0 \\ e & f & g & h \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix} + d \det \begin{pmatrix} 0 & 0 & 0 & 1 \\ e & f & g & h \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix}$$

2. Each of 4 terms, one more depth expansion only involves 3 terms, since two same rows implies zero det: From the first term above,

$$a \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ e & f & g & h \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix} = af \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix} + ag \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix} + ah \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix}.$$

3. Recursively, for each of 3 terms, one more depth expansion only involves 2 terms: From the first term above,

$$af \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ i & j & k & \ell \\ m & n & o & p \end{pmatrix} = afk \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ m & n & o & p \end{pmatrix} + af\ell \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ m & n & o & p \end{pmatrix}.$$

4. Finally, for each of 2 terms, one more depth expansion only involves 1 term: From the first term above,

$$afk \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ m & n & o & p \end{pmatrix} = afkp \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Collecting exactly $n!$ numbers for the det.

We recall,

Swapping two consecutive rows, or swapping two consecutive columns makes sign flipped.

We do the collection, taking care of the sign factor at the same time.

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

1. We first expand $\det(A)$ into n terms

$$\det(A) = a \det \begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + c \det \begin{pmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{pmatrix}$$

$$= a \det \begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} - b \det \begin{pmatrix} 1 & 0 & 0 \\ e & d & f \\ h & g & i \end{pmatrix} + c \det \begin{pmatrix} 1 & 0 & 0 \\ f & e & d \\ i & h & g \end{pmatrix}$$

2. Each of n terms, one more depth expansion only involves $n - 1$ terms: From the second term above,

$$-b \det \begin{pmatrix} 1 & 0 & 0 \\ e & d & f \\ h & g & i \end{pmatrix} = -bd \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h & g & i \end{pmatrix} - bf \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ h & g & i \end{pmatrix}$$

$$= -bd \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h & g & i \end{pmatrix} + bf \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h & i & g \end{pmatrix}$$

3. Recursively, each of $n - 1$ terms, one more depth expansion only involves $n - 2$ terms: From the second term above,

$$bf \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h & i & g \end{pmatrix} = bfg \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = bfg.$$

We keep the parts we have already taken care of, as the smaller size identity block, and pass to the next depth.

In practice, we can write

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

1. We first expand $\det(A)$ into n terms

$$\det(A) = a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} e & d \\ h & g \end{pmatrix}$$

2. Each of n terms, one more depth expansion only involves $n-1$ terms: From the second term above,

$$-b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} = -bd \det \begin{pmatrix} i \end{pmatrix} + bf \det \begin{pmatrix} g \end{pmatrix}$$

3. Recursively, each of $n-1$ terms, one more depth expansion only involves $n-2$ terms: From the second term above,

$$bf \det \begin{pmatrix} g \end{pmatrix} = bfg.$$

Examples of calculation

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

- As much as you can recognize, and as much as you can keep sign factors correct,

  Do the row swap and row replacement towards a triangular matrix, introducing as many as zeroes in the resulting matrix.

- Then do the $n!$ expansion.

$\operatorname{cof}(A)$, A^{-1} formula, and Cramer's formula.

Let us go back to the first depth expansion

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

1. We first expand $\det(A)$ into n terms

$$\det(A) = a \det \begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{pmatrix} + c \det \begin{pmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{pmatrix}$$

2. Assuming the $\det(A)$ calculation has already met end, in other words, assuming we have calculated the three numbers

$$\det \begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{pmatrix}$$

Let us take this three numbers to be yet another row,

$$c_1 = \left(\det \begin{pmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{pmatrix} \right)$$

We can write then

$$a_1 \cdot c_1 = \det(A)$$

In fact, we may do the following too.

1. We first expand $\det(A)$ into n terms

$$\det(A) = d \det \begin{pmatrix} a & b & c \\ 1 & 0 & 0 \\ g & h & i \end{pmatrix} + e \det \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ g & h & i \end{pmatrix} + f \det \begin{pmatrix} a & b & c \\ 0 & 0 & 1 \\ g & h & i \end{pmatrix}$$

2. Assuming we have calculated the three numbers

$$\det \begin{pmatrix} a & b & c \\ 1 & 0 & 0 \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} a & b & c \\ 0 & 0 & 1 \\ g & h & i \end{pmatrix}$$

Let us take this three numbers to be yet another row,

$$c_2 = \left(\det \begin{pmatrix} a & b & c \\ 1 & 0 & 0 \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ g & h & i \end{pmatrix}, \quad \det \begin{pmatrix} a & b & c \\ 0 & 0 & 1 \\ g & h & i \end{pmatrix} \right)$$

We can write then

$$a_2 \cdot c_2 = \det(A)$$

3. Importantly, we note

$$a_1 \cdot c_2 = 0$$

This is simply because

$$a_1 \cdot c_2 = a \det \begin{pmatrix} a & b & c \\ 1 & 0 & 0 \\ g & h & i \end{pmatrix} + b \det \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ g & h & i \end{pmatrix} + c \det \begin{pmatrix} a & b & c \\ 0 & 0 & 1 \\ g & h & i \end{pmatrix} = \det \begin{pmatrix} a & b & c \\ a & b & c \\ g & h & i \end{pmatrix} = 0.$$

We can collect n new rows c_i that has the property:

$$a_i \cdot c_j = \begin{cases} \det(A) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Provided we have calculated all those c_i, $i = 1, 2, \cdots, n$, and also provided $\det(A) \neq 0$, we then have equality

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

where C is the matrix with n rows c_i from the top to bottom. It is called the cofactor matrix of A and denoted by $\mathrm{cof}(A)$.

If we want to compute A^{-1} we have two means:

1. Do the RREF row operation on $\left(A \mid \mathrm{I} \right)$.

2. Do compute $\det(A)$ and collect rows c_i.

Actually, computing $\det(A)$ and $\mathrm{cof}(A)$ in general does not reduce computations.

Home Work

1. Compute the det for the above four examples.

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

2. Let

$$
A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} - & a_1 & - \\ - & a_2 & - \\ - & a_3 & - \end{pmatrix}
$$

consider the columnwise expansion, for example

$$
\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \det \begin{pmatrix} 1 & b & c \\ 0 & e & f \\ 0 & h & i \end{pmatrix} + d \begin{pmatrix} 0 & b & c \\ 1 & e & f \\ 0 & h & i \end{pmatrix} + g \begin{pmatrix} 0 & b & c \\ 0 & e & f \\ 1 & h & i \end{pmatrix},
$$

in other words, we collect this times d_1

$$
d_1 = \left(\det \begin{pmatrix} 1 & b & c \\ 0 & e & f \\ 0 & h & i \end{pmatrix}, \begin{pmatrix} 0 & b & c \\ 1 & e & f \\ 0 & h & i \end{pmatrix}, \begin{pmatrix} 0 & b & c \\ 0 & e & f \\ 1 & h & i \end{pmatrix} \right).
$$

and also similarly d_j, $j = 1, 2, \cdots, n$.

Show that d_{ji}, the i-th component of d_j, and c_{ij}, the j-th component of c_i are the same number.

3. Using the Home Work 2, prove the theorem on Cramer's formula:

Theorem 1. *Let A be an $n \times n$ invertible matrix whose columns are denoted by $b_1, b_2, \cdots, b_n$ from the left. Suppose $y \in \mathbb{R}^{n \times 1}$ is given and $x \in \mathbb{R}^{n \times 1}$ solves the equation $Ax = y$.*

Then the i-th component of the solution x

$$
x_i = \frac{\det(\cdots, b_{i-1}, y, b_{i+1}, \cdots)}{\det A}.
$$

Chapter 13

System of m equations and n unknowns

We consider a system of linear equations:

$$A_{1,1}x_1 + A_{1,2}x_2 + A_{1,3}x_3 + \cdots + A_{1,n}x_n = b_1$$
$$A_{2,1}x_1 + A_{2,2}x_2 + A_{2,3}x_3 + \cdots + A_{2,n}x_n = b_2$$
$$\vdots$$
$$A_{m,1}x_1 + A_{m,2}x_2 + A_{m,3}x_3 + \cdots + A_{m,n}x_n = b_m$$

We write the system

$$Ax = b,$$

where

$$A \in \mathbb{R}^{m \times n}, \quad x \in \mathbb{R}^{n \times 1} = X, \quad b \in \mathbb{R}^{m \times 1} = Y.$$

One might think that a meaningful solution map

$$R : Y \to X, \quad b \mapsto x = Rb$$

is only possible when $m = n$ and actually $m = n = r$, with $r = \text{rank}(A)$.

However, this is not true. You know everything to say that in fact, you can always single out one element $\tilde{x} \in X$, defining a meaningful map

$$\tilde{R} : Y \to X, \quad b \mapsto \tilde{x} = \tilde{R}b,$$

using the fact that the map $x \mapsto Ax$ is a bijection between $\text{Trow}(A)$ and $\text{Ran}(A)$.

Today, we outline the program how we study the problem of solving linear system:

1. Overview

2. Solving skill by hand: REF and RREF

3. Other methods such as Cholesky methods.

4. If possible, iterative methods.

Two Obvious Facts

1. If $b \notin \mathrm{Ran}\,(A)$, then there can be no way to have any solution.

2. Suppose $b \in \mathrm{Ran}\,(A)$.

   If $\mathrm{Trow}\,(A) \neq \mathbb{R}^{n \times 1}$, i.e., $\mathrm{Ker}\,(A) = \mathrm{Trow}\,(A)^{\perp}$ is not trivial, then there are infinitely many solutions.

   This is because, that $b \in \mathrm{Ran}\,(A)$ implies there exists $x \in \mathbb{R}^{n \times 1}$ with $Ax = b$. Now, for infinitely many choices of $x_0 \in \mathrm{Ker}\,(A)$, $x + x_0$ is again a solution.

What do we want?

1. We make a choice, in Y, a projection onto $\mathrm{Ran}\,(A)$ for any $b \in Y$.

2. Not surprisingly, $b \mapsto b' \in \mathrm{Ran}\,(A)$ that is by the orthogonal projection is the one used mostly.

3. Now, for given the projection b', we completely know that:

   (a) There exists unique element $x' \in \mathrm{Trow}\,(A)$ such that $Ax' = b'$.

   (b) The complete set of *solutions* are $x' + x_0$ for any $x_0 \in \mathrm{Ker}\,(A)$.

4. In most cases, the map

$$\tilde{R} : b \mapsto b' \mapsto x' \in \mathrm{Trow}\,(A)$$

   will be the meaningful map, singling out unique element.

5. The $\tilde{x}$ it returns is the one where

$$A\tilde{x} = b', \quad b' \text{ is the element in } \mathrm{Ran}\,(A) \text{ such that } |b - b'|_2 \text{ is minimized.}$$

Solutions of $Ax = 0$, by REF or RREF

Let

$$A = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix}$$

- We already know how to obtain basis of $\mathrm{Trow}\,(A)$ and $\mathrm{Ker}\,(A) = \mathrm{Trow}\,(A)^{\perp}$.

- Let us repeat the procedure in detail again though.

We remark one thing in row operations:

Suppose we make a row operation into

$$A' = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 3 & 6 & 8 & 10 \end{pmatrix}$$

This is to do

$$EA = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 3 & 6 & 8 & 10 \end{pmatrix} = A'$$

and we recall that

1. We multiply E such that

   $$\mathrm{Trow}\,(A) = \mathrm{Trow}\,(EA), \quad \text{and consequently} \quad \mathrm{Trow}\,(A)^{\perp} = \mathrm{Trow}\,(EA)^{\perp}.$$

2. However, because all the linear combinations of columns of A is passed to E to produce linear combinations of columns of E

   $$\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

   we have that

   $$\mathrm{Ran}\,(EA) = \big\{ b \in Y \mid b = Ew \text{ for some } w \in \mathrm{Ran}\,(A) \big\}$$

   and this is keep changing by multiplying E.

Having this in minds, we completely solve the equation $Ax = 0$.