MATH240 Summer 2023

Haoran Li

Contents

1 Lecture 1 - System of linear equations

1.1 Linear systems

Throughout this course, we adopt the following notations

- Natural numbers: $\mathbb{N} = \{0, 1, 2, 3, \cdots\}$.
- Integers: $\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$.
- Rational numbers: $\mathbb{Q} = \left\{\frac{m}{n} \middle| m, n \in \mathbb{Z}, n \neq 0\right\}$ is the set of fractions. Here $\in$ means belong to.
- Real numbers: $\mathbb{R}$ is the set of numbers on the whole real number line. It includes
  - irrational numbers (like $\sqrt{2}$, $\sqrt[3]{3}$)
  - transcendental numbers (like π, e)
- Complex numbers: $\mathbb{C} = \{a + bi | a, b \in \mathbb{R}\}$, $i = \sqrt{-1}$ is the imaginary number such that $i^2 = -1$.
- $\mathbb{N} \subsetneqq \mathbb{Z} \subsetneqq \mathbb{Q} \subsetneqq \mathbb{R} \subsetneqq \mathbb{C}$.
- $\mathbb{R}^n = \{(r_1, r_2, r_3, \cdots, r_n) | r_1, r_2, r_3, \cdots, r_n \in \mathbb{R}\}$ is the set of all n-tuples of real numbers. Geometrically, it is the n-dimensional Euclidean space. For example
  - $\mathbb{R}^1 = \mathbb{R}$ is a line.
  - $\mathbb{R}^2$ is a plane.
  - $\mathbb{R}^3$ is our usual physical space.

Definition 1.1. A linear equation in the variables $x_1, x_2, x_3, \cdots, x_n$ is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = b \tag{1.1}$$

where the coefficients $a_1, a_2, a_3, \cdots, a_n$ and b are real or complex numbers, usually known in advance.

Example 1.2 (Examples and non-examples of linear equations)**.**

- $x_1 + \frac{1}{2}x_2 = 2$, ✓
- $\pi(x_1 + x_2) - 9.9x_3 = e$, ✓. Because if we expand it, we got $\pi x_1 + \pi x_2 - 9.9x_3 = e$ in which case $a_1 = \pi, a_2 = \pi, a_3 = -9.9, b = e$ as in the form of (1.1)
- $|x_2| - 1 = 0$, ✗
- $x_1 + x_2^2 = 9$, ✗
- $\sqrt{x_1} + \sqrt{x_2} = 1$, ✗

Definition 1.3. A system of linear equations (or a linear system) is a collection of one or more linear equations involving the same variables, say $x_1, x_2, x_3, \cdots, x_n$.

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = b_3 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{1.2}$$

Example 1.4. For $n = m = 2$, (1.2) is just

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases} \tag{1.3}$$

Example 1.5 (*The Nine Chapters on the Mathematical Art*)**.** In a cage full of chickens and rabbits. The total number of heads is 10 and the total number of legs is 26. Calculate the number of chickens and rabbits.

Solution. Let's assume the number of chickens and rabbits are x_1 and x_2, then we can write down a linear system

$$\begin{cases} x_1 + \ \ x_2 = 10 \\ 2x_1 + 4x_2 = 26 \end{cases} \tag{1.4}$$

Let's solve the linear system $\begin{cases} x_1 + \ \ x_2 = 10 \\ 2x_1 + 4x_2 = 26 \end{cases}$

Step 1. Replace `Row 2` by `Row 2 - 2(Row 1)`, we get $\begin{cases} x_1 + 2x_2 = 20 \\ \qquad 2x_2 = 6 \end{cases}$

Step 2. Divide `Row 2` by 2, we get $\begin{cases} x_1 + 2x_2 = 20 \\ \qquad\ x_2 = 3 \end{cases}$

Step 3. Replace `Row 1` by `Row 1 - 2(Row 2)`, we have the solution $\begin{cases} x_1 \qquad = 14 \\ \qquad x_2 = 3 \end{cases}$

□

Remark. This process is call the Gaussian elimination

Definition 1.6. A solution of the linear system (1.2) is

$$\begin{cases} x_1 = s_1 \\ x_2 = s_2 \\ x_3 = s_3 \\ \quad \vdots \\ x_n = s_n \end{cases} \qquad s_1, s_2, s_3, \cdots, s_n \text{ are numbers}$$

which makes (1.2) true. The set of all possible solutions is called the solution set of the linear system. To solve a linear system is to find all its solutions.

1.2 Geometric interpretation

Example 1.7. $\begin{cases} x_1 + \ \ x_2 = 10 \\ 2x_1 + 4x_2 = 26 \end{cases}$ describe two lines in $\mathbb{R}^2$, and the solution set is the intersection.

Question. How many solutions does $\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases}$ have?

Answer. It may have

- a *unique* solution if these two lines *intersect*.
- (uncountably) *infinitely many* solutions if these two lines *overlap*.
- *no* solutions if these two lines are *parallel* but not overlapping.

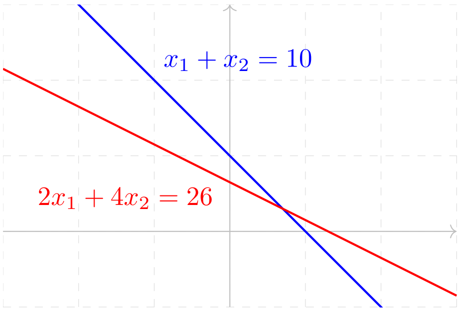
Example 1.8. Compare the following three linear systems

$$\begin{cases} x_1 + x_2 = 10 \\ 2x_1 + 4x_2 = 26 \end{cases} \tag{1.5a}$$

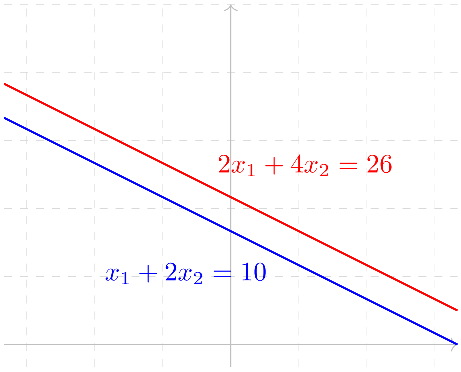
$$\begin{cases} x_1 + 2x_2 = 10 \\ 2x_1 + 4x_2 = 26 \end{cases} \tag{1.5b}$$

$$\begin{cases} x_1 + 2x_2 = 13 \\ 2x_1 + 4x_2 = 26 \end{cases} \tag{1.5c}$$

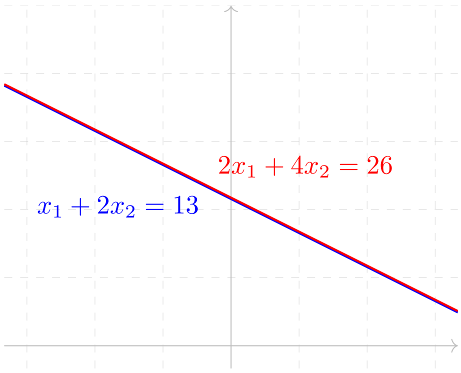
- (1.5a) has a unique solution $\begin{cases} x_1 = 7 \\ x_2 = 3 \end{cases}$
- (1.5b) has no solutions since the 1st equation contradicts the 2nd.
- (1.5c) has infinitely many solutions since the 2nd equation is twice of the 1st.



(a) Two lines intersect



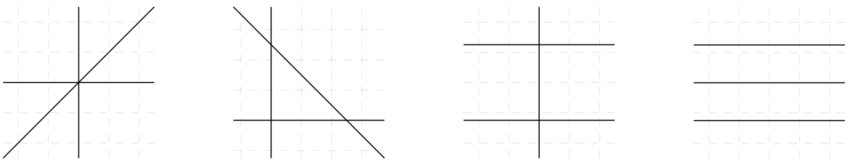
(b) Two lines parallel but not overlapping



(c) Two lines overlap

Figure 1.1: Figure for Example 1.8

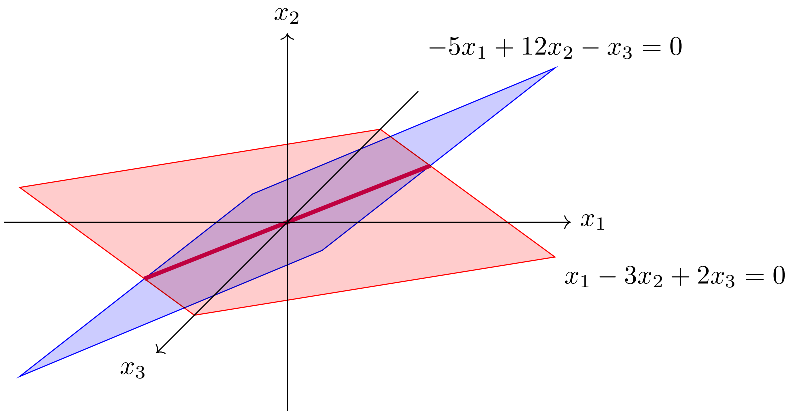
If we increase the number of equations, we get more lines, it might looks like



If we increase the number of variables, we get

- $a_1x_1 + a_2x_2 + a_3x_3 = b$ describes a plane in $\mathbb{R}^3$.
- $a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = b$ describes a *hyperplane* in $\mathbb{R}^n$.
- Therefore the solution set of (1.2) is the intersection of m hyperplanes.

Example 1.9. Geometric interpretation of $\begin{cases} x_1 - 3x_2 + 2x_3 = 0 \\ -5x_1 + 12x_2 - x_3 = 0 \end{cases}$



Remark. It is geometrically clear that for a system of 2 equations in 3 variables, there are either no solutions or infinitely many, since two planes either intersects at a line, or overlap, or simply parallel.

Definition 1.10. We say a linear system is consistent if it has solution(s), and inconsistent if it has none.

Example 1.11. In the previous Example 1.8, (1.5a) and (1.5c) are consistent, while (1.5b) is inconsistent

Exercise 1.12.

1. Try Gaussian elimination on the following linear systems

   (a) $\begin{cases} x_1 + 5x_2 = 7 \\ -2x_1 - 7x_2 = -5 \end{cases}$

   (b) $\begin{cases} 2x_1 + 4x_2 = -4 \\ 5x_1 + 7x_2 = 11 \end{cases}$

   (c) $\begin{cases} x_1 - x_2 + x_3 = 1 \\ 2x_1 - x_3 = 1 \\ x_1 + x_2 + x_3 = 3 \end{cases}$

2. Find the point of intersection of the lines $x_1 - 5x_2 = 1$ and $3x_1 - 7x_2 = 5$.

3. For what values of h and k is the following system consistent?

$$\begin{cases} 2x_1 - x_2 = h \\ -6x_1 + 3x_2 = k \end{cases}$$

2 Lecture 2 - Matrices and row echelon form

2.1 Matrices

Definition 2.1. A m by n (or $m \times n$) matrix is a rectangular array of numbers with m rows and n columns

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{2.1}$$

We use the (i, j)-th entry to mean the entry on the i-th row and j-column (i.e. a_{ij}).

Definition 2.2. A matrix is

- a zero matrix is a matrix with all entries zeros.

$$\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

- a square matrix is a matrix with the same number of rows and columns, i.e. $m = n$.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

- a vector if it only has one column, i.e. $n = 1$.

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

- a row vector if it only has one row, i.e. $m = 1$.

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

- the identity matrix if it is a square matrix with diagonal elements 1's, and 0's otherwise. Here the diagonal are the (i, i)-th entries.

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Definition 2.3. Soon we will be getting tired of writing all these equations in the linear system (1.2), instead we write down its augmented matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{1n} & b_3 \\ \vdots & \vdots & \vdots & & \vdots & \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m \end{bmatrix}$$

Which is obtained by omitting x_i's, pluses, and equal signs. If we delete the last column, we will get the coefficient matrix

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

Example 2.4.

- For (1.4), its augmented matrix and coefficient matrix are

$$\begin{bmatrix} 1 & 1 & 10 \\ 2 & 4 & 26 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}$$

- For $\begin{cases} x_1 - x_2 + x_3 = 1 \\ 2x_1 \qquad - x_3 = 1 \\ x_1 + x_2 + x_3 = 3 \end{cases}$, its augmented matrix and coefficient matrix are

$$\begin{bmatrix} 1 & -1 & 1 & 1 \\ 2 & 0 & -1 & 1 \\ 1 & 1 & 1 & 3 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

- In general, a linear system of m equations in n variables has a m by $(n+1)$ augmented matrix and a m by n coefficient matrix.

Definition 2.5. Inspired by Gaussian elimination, we define the following three elementary row operations

- Replacement: Replace one row by the sum of itself and a multiple of another row.
- Interchangement: Interchange two rows.
- Scaling: Multiply all entries in a row by a *nonzero* constant.

We say matrices A, B are row equivalent ($A \sim B$) if B can obtained by applying a sequence of elementary row operations to A (or vise versa).

Example 2.6. Let's rewrite the process in Example 1.5

$$\begin{bmatrix} 1 & 1 & 10 \\ 2 & 4 & 26 \end{bmatrix} \xrightarrow{R2 \to R2 - 2R1} \begin{bmatrix} 1 & 1 & 10 \\ 0 & 2 & 6 \end{bmatrix} \xrightarrow{R2 \to \frac{R2}{2}} \begin{bmatrix} 1 & 1 & 10 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow{R1 \to R1 - R2} \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 3 \end{bmatrix}$$

Example 2.7. Solve $\begin{cases} x_1 - x_2 + x_3 = 1 \\ 2x_1 \qquad - x_3 = 1 \\ x_1 + x_2 + x_3 = 3 \end{cases}$ with augmented matrix.

$$\begin{bmatrix} 1 & -1 & 1 & 1 \\ 2 & 0 & -1 & 1 \\ 1 & 1 & 1 & 3 \end{bmatrix} \xrightarrow{R2 \to R2 - 2R1} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 1 & 1 & 1 & 3 \end{bmatrix} \xrightarrow{R3 \to R3 - R1} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 2 & 0 & 2 \end{bmatrix}$$

$$\xrightarrow{R3 \to R3 - R2} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 3 & 3 \end{bmatrix} \xrightarrow{R3 \to \frac{R3}{3}} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow[R2 \to R2 + 3R3]{R1 \to R1 - R3} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow{R2 \to \frac{R2}{2}} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow{R1 \to R1 + R2} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

This gives the unique solution $\begin{cases} x_1 = 1 \\ x_2 = 1 \\ x_3 = 1 \end{cases}$

2.2 Row echelon form

Definition 2.8.

- A leading entry of a row refers to the leftmost nonzero entry (in a nonzero row).
- A matrix is of row echelon form (REF) if it is of a "staircase shape".

$$\overset{\text{REF}}{\begin{bmatrix} \blacksquare & * & * & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \end{bmatrix}}$$

$\blacksquare$ are the leading entries, $*$ are some unknown numbers.

- The leading entries of an REF matrix are called pivots.
- The position of pivots are called pivot positions.
- The column pivots are in are called pivot columns.
- An REF of reduced row echelon form (RREF) if all its pivots are 1's and in each pivot column, every entry except the pivot are 0's.

$$\overset{\text{RREF}}{\begin{bmatrix} 1 & * & 0 & * & 0 & 0 & * & * \\ 0 & 0 & 1 & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * \end{bmatrix}}$$

Example 2.9. In Example 2.7, $\begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 2 & 0 & 2 \end{bmatrix}$ is not an REF. $\begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 3 & 3 \end{bmatrix}$ is an REF, but not an RREF. $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ is an RREF

Theorem 2.10. Every matrix is row equivalent to some REF matrix (which is not in general unique), but it is row equivalent to some unique RREF matrix.

Example 2.11. In Example 2.7, $\begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 3 & 3 \end{bmatrix}$ is an REF that is row equivalent to the original matrix $\begin{bmatrix} 1 & -1 & 1 & 1 \\ 2 & 0 & -1 & 1 \\ 1 & 1 & 1 & 3 \end{bmatrix}$, and $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ is its unique row equivalent RREF.

Example 2.12. Solve $\begin{cases} x_1 - x_2 + x_3 = 1 \\ 2x_1 \phantom{{}+x_2} - x_3 = 1 \\ x_1 + x_2 - 2x_3 = 1 \end{cases}$ with augmented matrix.

$$\begin{bmatrix} 1 & -1 & 1 & 1 \\ 2 & 0 & -1 & 1 \\ 1 & 1 & -2 & 1 \end{bmatrix} \xrightarrow[R3\to R3-R1]{R2\to R2-2R1} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 2 & -3 & 0 \end{bmatrix} \xrightarrow{R3\to R3-R2} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

You might notice that the last row represents $0x_1 + 0x_2 + 0x_3 = 1$, this is a contradiction, therefore the linear system is inconsistent.

Example 2.13. $\begin{cases} x_1 - x_2 + x_3 = 1 \\ 2x_1 \phantom{{}-x_2} - x_3 = 1 \end{cases}$, we write down its augmented matrix

$$\begin{bmatrix} 1 & -1 & 1 & 1 \\ 2 & 0 & -1 & 1 \end{bmatrix} \xrightarrow{R2\to R2-2R1} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \end{bmatrix}$$
$$\overset{\frac{R2}{2}}{\sim} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 1 & -\frac{3}{2} & -\frac{1}{2} \end{bmatrix} \xrightarrow{R1\to R1+R2} \begin{bmatrix} 1 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{3}{2} & -\frac{1}{2} \end{bmatrix}$$

This gives the solution set

$$\begin{cases} x_1 \phantom{{}+x_2} - \frac{1}{2}x_3 = \frac{1}{2} \\ \phantom{x_1 +{}} x_2 - \frac{3}{2}x_3 = \frac{1}{2} \end{cases} \Rightarrow \begin{cases} x_1 = \frac{1}{2}x_3 + \frac{1}{2} \\ x_2 = \frac{3}{2}x_3 - \frac{1}{2} \end{cases} \tag{2.2}$$

Let's formalize these as row reduction algorithm

Step 1. Begin with the leftmost nonzero column. This is a pivot column. The pivot position should be at the top.

Step 2. Select a nonzero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

Step 3. Use row replacement operations to create zeros in all positions below the pivot.

Step 4. Cover (or ignore) the rows containing the pivot positions. Apply Steps 1-3 to the rows that remains. Repeat the process until you are left with an REF.

Step 5. Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. If a pivot is not 1, make it 1 by a scaling operation.

Steps 1-4 are call forward phase, after which you get an REF. Step 5 is called backward phase, after which you get the RREF.

Example 2.14. Consider the augmented matrix $\begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}$

- Forward phase

$$\begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix} \xrightarrow[\text{Step 1,2}]{R1\leftrightarrow R4} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \xrightarrow[\text{Step 3}]{\substack{R2\to R2+R1 \\ R3\to R3+2R1}} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \xrightarrow[\text{Step 4,1,2,3}]{\substack{R3\to R3-\frac{5}{2}R2 \\ R4\to R4+\frac{3}{2}R2}} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix} \xrightarrow[\text{Step 4,1}]{R3\leftrightarrow R4} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- Backward phase

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{Step 5}]{R3\to R3/(-5)} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{Step 5}]{\substack{R1\to R1+9R3 \\ R2\to R2+6R3}} \begin{bmatrix} 1 & 4 & 5 & 0 & -7 \\ 0 & 2 & 4 & 0 & -6 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow[\text{Step 5}]{R2\to R2/2} \begin{bmatrix} 1 & 4 & 5 & 0 & -7 \\ 0 & 1 & 2 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{Step 5}]{R1\to R1-4R2} \begin{bmatrix} 1 & 0 & -3 & 0 & 5 \\ 0 & 1 & 2 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Definition 2.15. The variables corresponding to pivot columns in a matrix are called basic variables, the other variables are called free variables. In a solution set, basic variables are expressed in terms of free variables, and a free variable can take any value.

Example 2.16. In Example 2.13, x_1, x_2 are basic variables and x_3 is a free variable. And we formally write our solution set as $\begin{cases} x_1 = \frac{1}{2}x_3 + \frac{1}{2} \\ x_2 = \frac{3}{2}x_3 - \frac{1}{2} \\ x_3 \text{ is free} \end{cases}$

Exercise 2.17. Find the general solution of the system $\begin{cases} x_1 - 2x_2 - x_3 + 3x_4 = 0 \\ -2x_1 + 4x_2 + 5x_3 - 5x_4 = 3 \\ 3x_1 - 6x_2 - 4x_3 + 8x_4 = 2 \end{cases}$

Solution.

$$\begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ -2 & 4 & 5 & -5 & 3 \\ 3 & -6 & -4 & 8 & 2 \end{bmatrix} \xrightarrow[R3 \to R3-3R1]{R2\to R2+2R1} \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix} \xrightarrow{R3\to(-1)\cdot R3} \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & 1 & 1 & -2 \end{bmatrix}$$

$$\xrightarrow{R2\to R2-3R3} \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 0 & -2 & 9 \\ 0 & 0 & 1 & 1 & -2 \end{bmatrix} \xrightarrow{R2\leftrightarrow R3} \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & -2 & 9 \end{bmatrix}$$

$$\xrightarrow{R3\to \frac{R3}{-2}} \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 1 & -\frac{9}{2} \end{bmatrix} \xrightarrow[R1\to R1-3R3]{R2\to R2-R3} \begin{bmatrix} 1 & -2 & -1 & 0 & \frac{27}{2} \\ 0 & 0 & 1 & 0 & \frac{5}{2} \\ 0 & 0 & 0 & 1 & -\frac{9}{2} \end{bmatrix} \xrightarrow{R1\to R1+R2} \begin{bmatrix} 1 & -2 & 0 & 0 & 16 \\ 0 & 0 & 1 & 0 & \frac{5}{2} \\ 0 & 0 & 0 & 1 & -\frac{9}{2} \end{bmatrix}$$

Write this as solution set, we get

$$\begin{cases} x_1 - 2x_2 = 16 \\ \qquad\quad x_3 = \frac{5}{2} \\ \qquad\quad x_4 = -\frac{9}{2} \end{cases} \Rightarrow \begin{cases} x_1 = 2x_2 + 16 \\ x_2 \text{ is free} \\ x_3 = \frac{5}{2} \\ x_4 = -\frac{9}{2} \end{cases}$$

□

Theorem 2.18. Suppose the augmented matrix of a linear system is $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$, and its RREF is $\begin{bmatrix} U & \mathbf{d} \end{bmatrix}$, then the linear system has

- no solutions $\iff$ $\mathbf{d}$ is a pivot column, i.e. contains a pivot.
- has solutions $\iff$ $\mathbf{d}$ is not a pivot column
  - a unique solution $\iff$ every column of U is a pivot column.
  - infinitely many solutions $\iff$ some columns of U is not a pivot column.

Example 2.19.

- In Example 2.12, the linear system has no solutions since

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & 1 \\ 2 & 0 & -1 & 1 \\ 1 & 1 & -2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} U & \mathbf{d} \end{bmatrix}$$

- In Example 2.7, the linear system has a unique solution since

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} U & \mathbf{d} \end{bmatrix}$$

- In Example 2.17, the linear system has infinitely many solutions since

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ -2 & 4 & 5 & -5 & 3 \\ 3 & -6 & -4 & 8 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 & 0 & 16 \\ 0 & 0 & 1 & 0 & \frac{5}{2} \\ 0 & 0 & 0 & 1 & -\frac{9}{2} \end{bmatrix} = \begin{bmatrix} U & \mathbf{d} \end{bmatrix}$$

Exercise 2.20. Find the general solutions of the system with given augmented matrix, name the pivot columns, pivot positions, basic and free variables.

1. $\begin{bmatrix} 0 & 1 & -6 & 5 \\ 1 & -2 & 7 & -4 \end{bmatrix}$

2. $\begin{bmatrix} 1 & -7 & 0 & 6 & 5 \\ 0 & 0 & 1 & -2 & -3 \\ -1 & 7 & -4 & 2 & 7 \end{bmatrix}$

Question. How does the size of the augmented matrix affect the solution set?

3 Lecture 3 - Matrix algebra

3.1 Matrix addition and scalar multiplication

Definition 3.1. Let's use $M_{m\times n}(\mathbb{R})$ to denote the set of all (real-valued) m by n matrices.

Definition 3.2. Suppose A, B are $m\times n$ matrices, c is a scalar (i.e. a number), then we can define

- Addition

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

- Scalar multiplication

$$c\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix}$$

Example 3.3.

- $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$
- $2\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$

3.2 Matrix multiplication

Definition 3.4. Suppose A is a $m\times n$ matrix, and B is a $n\times p$ matrix, we can define matrix multiplication AB to be the $m\times p$ matrix, computed via the row-column rule: The (i,j)-entry is to multiply the i-row and j-th column

$$\begin{bmatrix} \\ \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \\ \\ \end{bmatrix}\begin{bmatrix} & b_{1j} & \\ & b_{2j} & \\ & \vdots & \\ & b_{nj} & \end{bmatrix} = \begin{bmatrix} \\ \\ & \blacksquare & \\ \\ \\ \end{bmatrix}$$

Where the (i,j)-entry $\blacksquare = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$.

If A is a square matrix, then we could define matrix power A^k to be simply $\overbrace{AA\cdots A}^{k \text{ times}}$

Example 3.5.

- $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11}+a_{12}b_{21}+a_{13}b_{31} & a_{11}b_{12}+a_{12}b_{22}+a_{13}b_{32} \\ a_{21}b_{11}+a_{22}b_{21}+a_{23}b_{31} & a_{21}b_{12}+a_{22}b_{22}+a_{23}b_{32} \end{bmatrix}$

- $\begin{bmatrix}1 & 2 & 2\\ 2 & 1 & 1\end{bmatrix}\begin{bmatrix}3 & 1\\ 1 & 2\\ 2 & 3\end{bmatrix} = \begin{bmatrix}1\cdot 3+2\cdot 1+2\cdot 2 & 1\cdot 1+2\cdot 2+2\cdot 3\\ 2\cdot 3+1\cdot 1+1\cdot 2 & 2\cdot 1+1\cdot 2+1\cdot 3\end{bmatrix} = \begin{bmatrix}9 & 11\\ 9 & 7\end{bmatrix}$

- $$\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}a_{11}x_1+a_{12}x_2+a_{13}x_3\\ a_{21}x_1+a_{22}x_2+a_{23}x_3\end{bmatrix} = x_1\begin{bmatrix}a_{11}\\ a_{21}\end{bmatrix}+x_1\begin{bmatrix}a_{12}\\ a_{22}\end{bmatrix}+x_1\begin{bmatrix}a_{13}\\ a_{23}\end{bmatrix} \tag{3.1}$$

- $$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}0\cdot 1+1\cdot 0 & 0\cdot 0+1\cdot 0\\ 0\cdot 1+0\cdot 0 & 0\cdot 0+0\cdot 0\end{bmatrix} = \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} \tag{3.2}$$

- $$\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} = \begin{bmatrix}1\cdot 0+0\cdot 0 & 1\cdot 1+0\cdot 0\\ 0\cdot 0+0\cdot 0 & 0\cdot 1+0\cdot 0\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} \tag{3.3}$$

- $$\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

- $$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

Example 3.6.

1. $$\begin{bmatrix}1 & 0 & -3\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} = \begin{bmatrix}a_{11}-3a_{31} & a_{12}-3a_{32} & a_{13}-3a_{33}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

2. $$\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 2 & 0 & 1\end{bmatrix} = \begin{bmatrix}a_{11}+2a_{13} & a_{12} & a_{13}\\ a_{21}+2a_{23} & a_{22} & a_{23}\\ a_{31}+2a_{33} & a_{32} & a_{33}\end{bmatrix}$$

3. $$\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12} & a_{13}\\ 2a_{21} & 2a_{22} & 2a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

4. $$\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}a_{11} & 3a_{12} & a_{13}\\ a_{21} & 3a_{22} & a_{23}\\ a_{31} & 3a_{32} & a_{33}\end{bmatrix}$$

5. $$\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{bmatrix}\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} = \begin{bmatrix}a_{31} & a_{32} & a_{33}\\ a_{21} & a_{22} & a_{23}\\ a_{11} & a_{12} & a_{13}\end{bmatrix}$$

6. $$\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}a_{12} & a_{11} & a_{13}\\ a_{22} & a_{21} & a_{23}\\ a_{32} & a_{31} & a_{33}\end{bmatrix}$$

Exercise 3.7. Suppose $A = \begin{bmatrix}1 & 1 & 1\\ 2 & 2 & 1\\ 2 & 1 & 1\end{bmatrix}, B = \begin{bmatrix}1 & 1 & 1\\ 2 & 2 & 1\\ 2 & 1 & 1\end{bmatrix}$, computes matrix multiplication AB

Fact 3.8. Suppose A, B, C, D are matrices, c is a scalar, 0 is the zero matrix, I is the identity matrix. we have the following facts

a) Matrix multiplication is generally *NOT commutative*, i.e. $AB \neq BA$

b) Matrix multiplication is *associative*, i.e. the order of multiplication doesn't matter, in other words $(AB)C = A(BC)$, so it makes sense to write successive multiplication $A_1A_2A_3 \cdots A_n$

c) Scalar multiplication and matrix multiplication commutes, $A(cB) = c(AB) = (cA)B$. so it makes sense to write $cA_1A_2A_3 \cdots A_n$

d) Matrix multiplication is *distributive* over addition, i.e. $A(B+C) = AB+AC$, $(A+B)C = AC + BC$

e) Zero matrix and identity matrix acts as 0 and 1, i.e. $A + 0 = 0 + A = A$, $A0 = 0A = 0$, $IA = AI = A$

f) Even if $A \neq 0$, $B \neq 0$, AB could still be 0, take (3.2) for an example

g) $AB = AC$ does NOT imply $B = C$

Remark. Some of the properties of matrices are really similar to that of numbers, so we dub this the name of *matrix algebra*

3.3 Partitioned matrix

Definition 3.9. A is a partitioned (or block) matrix if is divided into smaller submatrix by some horizontal and vertical lines. And the submatrices are the blocks

$$\left[\begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array}\right] = \left[\begin{array}{cc|cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} \end{array}\right]$$

Here the blocks are $A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $A_{12} = \begin{bmatrix} a_{13} & a_{14} & a_{15} & a_{16} \\ a_{23} & a_{24} & a_{25} & a_{26} \end{bmatrix}$, $A_{13} = \begin{bmatrix} a_{17} \\ a_{27} \end{bmatrix}$, $A_{21} = \begin{bmatrix} a_{31} & a_{32} \end{bmatrix}$, $A_{22} = \begin{bmatrix} a_{33} & a_{34} & a_{35} & a_{36} \end{bmatrix}$, $A_{23} = \begin{bmatrix} a_{37} \end{bmatrix}$, $A_{31} = \begin{bmatrix} a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{bmatrix}$, $A_{32} = \begin{bmatrix} a_{43} & a_{44} & a_{45} & a_{46} \\ a_{53} & a_{54} & a_{55} & a_{56} \\ a_{63} & a_{64} & a_{65} & a_{66} \end{bmatrix}$, $A_{33} = \begin{bmatrix} a_{47} \\ a_{57} \\ a_{67} \end{bmatrix}$.

Fact 3.10. Suppose $A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix}$, $B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1r} \\ B_{21} & B_{22} & \cdots & B_{2r} \\ \vdots & \vdots & & \vdots \\ B_{q1} & B_{q2} & \cdots & B_{qr} \end{bmatrix}$ are partitioned matrices, and the number of columns of A_{1k} is equal to the number of rows of B_{k1} (so that all submatrices multiplications make sense). Then the usual row-column rule still WORKS!!! By treating submatrices as if they are numbers.

$$AB = C = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1r} \\ C_{21} & C_{22} & \cdots & C_{2r} \\ \vdots & \vdots & & \vdots \\ C_{p1} & C_{p2} & \cdots & C_{pr} \end{bmatrix}, \quad C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{iq}B_{qj}$$

Example 3.11. Consider $\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] = \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 2 & 2 & 1 \\ \hline 2 & 1 & 1 \end{array}\right]$, $\left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right] = \left[\begin{array}{c|cc} 1 & 1 & 1 \\ 2 & 2 & 1 \\ \hline 2 & 1 & 1 \end{array}\right]$, then

$$A_{11}B_{11} + A_{12}B_{21} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

$$A_{11}B_{12} + A_{12}B_{22} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 7 & 5 \end{bmatrix}$$

$$A_{21}B_{11} + A_{22}B_{21} = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 2 \end{bmatrix} = \begin{bmatrix} 6 \end{bmatrix}$$

$$A_{21}B_{12} + A_{22}B_{22} = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 4 \end{bmatrix}$$

$$\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right] = \left[\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hline A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array}\right] = \left[\begin{array}{c|cc} 5 & 4 & 3 \\ 8 & 7 & 5 \\ \hline 6 & 5 & 4 \end{array}\right]$$

Example 3.12. Suppose 3×3 matrix A can be partitioned into $\begin{bmatrix} R1 \\ R2 \\ R3 \end{bmatrix}$ or $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$, then Example 3.6 reads

1. $E = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $EA = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R1 \\ R2 \\ R3 \end{bmatrix} = \begin{bmatrix} R1 - 3R3 \\ R2 \\ R3 \end{bmatrix}$, EA acts as subtracting 3 times row 3 from row 1.
2. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$, $AE = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 + 2\mathbf{a}_3 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$, AE acts as adding 2 times column 3 to column 1.
3. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $EA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R1 \\ R2 \\ R3 \end{bmatrix} = \begin{bmatrix} R1 \\ 2R2 \\ R3 \end{bmatrix}$, EA acts as scaling the third row by 2.
4. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $AE = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & 3\mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$, AE acts as scaling the third column by 3.
5. $E = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $EA = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} R1 \\ R2 \\ R3 \end{bmatrix} = \begin{bmatrix} R3 \\ R2 \\ R1 \end{bmatrix}$, EA acts as interchanging row 1 and row 3.
6. $E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $AE = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_1 & \mathbf{a}_3 \end{bmatrix}$, AE acts as interchanging column 1 and column 2.

Definition 3.13. Matrices E in the previous example is called elementary matrices. They describe row and column elementary operations.

Exercise 3.14. If A is a 4 by 5 matrix, what is the elementary matrix E that acts as replacing the fourth row by adding twice of the second row.

Exercise 3.15. Suppose $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$, show that $AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_n \end{bmatrix}$

Exercise 3.16. Verify that $A^2 = I_2$ where $A = \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}$, and use partitioned matrices to show that $M^2 = I_4$, where $M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -3 & 1 \end{bmatrix}$

Exercise 3.17. Suppose $\begin{bmatrix} A & \mathbf{b} \end{bmatrix} \sim \begin{bmatrix} U & \mathbf{d} \end{bmatrix}$ is the REF/RREF, then U will also be the REF/RREF of A.

Solution. Suppose the row elementary operations applied are $E_1, E_2, \cdots, E_k$, then

$$\begin{array}{ccccccc} \begin{bmatrix} A & \mathbf{b} \end{bmatrix} \sim & E_1 \begin{bmatrix} A & \mathbf{b} \end{bmatrix} & \sim & E_2E_1 \begin{bmatrix} A & \mathbf{b} \end{bmatrix} & \sim \cdots \sim & E_k \cdots E_2E_1 \begin{bmatrix} A & \mathbf{b} \end{bmatrix} & = \begin{bmatrix} U & \mathbf{d} \end{bmatrix} \\ & \| & & \| & & \| & \\ & \begin{bmatrix} E_1A & E_1\mathbf{b} \end{bmatrix} & & \begin{bmatrix} E_2E_1A & E_2E_1\mathbf{b} \end{bmatrix} & & \begin{bmatrix} E_k \cdots E_2E_1A & E_k \cdots E_2E_1\mathbf{b} \end{bmatrix} & \end{array}$$

The same sequence of row elementary operations would reduce A to U. □

4 Lecture 4 - Matrix equations and linear independence

4.1 Vector and matrix equations

Recall a vector is a matrix with one column, the zero vector is a vector with all entries zero. For scalar (i.e. a number) c, and vectors $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$, we have

- Addition $\mathbf{a} + \mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}$
- Scalar multiplication $c\mathbf{a} = c \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{bmatrix}$
- Subtraction $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-1)\mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \\ \vdots \\ a_n - b_n \end{bmatrix}$

Remark. In handwritings, we use $\vec{v}$ or $\overrightarrow{v}$ to denote a vector, while in printing materials we often use the math bold font $\mathbf{v}$.

Definition 4.1. A vector equation is of the form

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b} \tag{4.1}$$

We can also write the vector equation (4.1) as a matrix equation with partitioned matrix

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x} = \mathbf{b}$$

Here $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Example 4.2. (1.4) can be written as a vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = x_1\begin{bmatrix}1\\2\end{bmatrix} + x_2\begin{bmatrix}1\\4\end{bmatrix} = \begin{bmatrix}x_1\\2x_1\end{bmatrix} + \begin{bmatrix}x_2\\4x_2\end{bmatrix} = \begin{bmatrix}x_1+x_2\\2x_1+4x_2\end{bmatrix} = \begin{bmatrix}10\\26\end{bmatrix} = \mathbf{b}$$

Or a matrix equation

$$A\mathbf{x} = \begin{bmatrix}1 & 1\\2 & 4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}10\\26\end{bmatrix} = \mathbf{b}$$

Example 4.3. The corresponding vector equation of $\begin{cases} x_1 + x_3 = 1 \\ 2x_2 + x_3 = 2 \end{cases}$ is

$$x_1\begin{bmatrix}1\\0\end{bmatrix} + x_2\begin{bmatrix}0\\2\end{bmatrix} + x_3\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix}$$

And the corresponding matrix equation is

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

4.2 Span

Definition 4.4.

- A linear combination of $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ is a sum $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$ for some scalars $c_1, \cdots, c_n$.
- The span of $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is the set of all its linear combinations, which we denote $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$.

Theorem 4.5. $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ has solution(s) $\iff$ $\mathbf{b}$ is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \iff \mathbf{b} \in \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$

Exercise 4.6. Let $\mathbf{a}_1 = \begin{bmatrix}1\\2\\1\end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix}-1\\0\\1\end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix}1\\-1\\-2\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}1\\1\\1\end{bmatrix}$. Is $\mathbf{b}$ in $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$?

Solution. This is equivalent of asking if whether the vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}$ has solution(s), we find an REF of its augmented matrix

$$\begin{bmatrix}1 & -1 & 1 & 1\\2 & 0 & -1 & 1\\1 & 1 & -2 & 1\end{bmatrix} \xrightarrow[R3\to R3-R1]{R2\to R2-2R1} \begin{bmatrix}1 & -1 & 1 & 1\\0 & 2 & -3 & -1\\0 & 2 & -3 & 0\end{bmatrix} \xrightarrow{R3\to R3-R2} \begin{bmatrix}1 & -1 & 1 & 1\\0 & 2 & -3 & -1\\0 & 0 & 0 & 1\end{bmatrix}$$

Since there is a pivot in the last column, by Theorem 2.18, the linear system is inconsistent, hence $\mathbf{b} \notin \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ ☐

4.3 Linear independence

Definition 4.7. $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is linearly dependent if some $\mathbf{v}_i$ is in the span of the others (so it is somewhat redundant), or equivalently, if there is a non-trivial solution $c_1, \cdots, c_n$ (i.e. not all c_i's are 0) to the vector equation

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0} \tag{4.2}$$

(4.2) is referred to as a linear dependence between $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. If (4.2) has only the trivial solution (i.e. $c_1, \cdots, c_n$ are all 0, which is of course always a solution), $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is said to be linearly independent

Remark. Equivalence between two different definitions of linear dependence

- If $\mathbf{v}_i = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, then $c_1\mathbf{v}_1 + \cdots + (-1)\mathbf{v}_i + \cdots + c_n\mathbf{v}_n = \mathbf{0}$
- If $c_1\mathbf{v}_1 + \cdots + c_i\mathbf{v}_i + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ and $c_i \neq 0$ (since not all c_i's are zero, we may assume some c_i is nonzero), then $\mathbf{v}_i = -\frac{c_1}{c_i}\mathbf{v}_1 - \cdots - \frac{c_{i-1}}{c_i}\mathbf{v}_{i-1} - \frac{c_{i+1}}{c_i}\mathbf{v}_{i+1} - \cdots - \frac{c_n}{c_i}\mathbf{v}_n$

Question. How do we determine and find linear dependence of $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$?

Answer. Let $A = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$, then non-trivial solutions to $A\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0}$ would be the linear dependences of $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. Therefore it is linearly independent if it has only the trivial(zero) solution.

Theorem 4.8. $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is linearly independent $\iff$ $A\mathbf{x} = \mathbf{0}$ has only the trivial(zero) solution $\iff$ each column of the RREF of A is a pivot column.

Proof. Consider the RREF of the augmented matrix $\begin{bmatrix}A & \mathbf{0}\end{bmatrix}$, it is necessarily $\begin{bmatrix}U & \mathbf{0}\end{bmatrix}$ for $A\mathbf{x} = \mathbf{0}$ to have only the trivial solution. ☐

Example 4.9. Suppose $\mathbf{v}_1 = \begin{bmatrix}1\\-1\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{e}_1 = \begin{bmatrix}1\\0\end{bmatrix}$, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{e}_1\}$ is linearly dependent since

$$\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{e}_1\end{bmatrix} = \begin{bmatrix}1 & 1 & 1\\-1 & 1 & 0\end{bmatrix} \sim \begin{bmatrix}1 & 0 & \frac{1}{2}\\0 & 1 & \frac{1}{2}\end{bmatrix}$$

The solution to this augmented matrix would be $\begin{cases} x_1 = -\frac{1}{2}x_3 \\ x_2 = -\frac{1}{2}x_3 \\ x_3 \text{ is free}\end{cases}$, by choosing any value nonzero value of x_3 (say 1) we get a linear dependence $-\frac{1}{2}\mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2 + \mathbf{e}_1 = \mathbf{0}$. On the other hand, $\mathbf{v}_1, \mathbf{v}_2$ are linearly independent since

$$\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix} = \begin{bmatrix}1 & 1\\-1 & 1\end{bmatrix} \sim \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

Where each column is a pivot column.

Exercise 4.10. Write the system $\begin{cases} 8x_1 - x_2 = 4 \\ 5x_1 + 4x_2 = 1 \\ x_1 - 3x_2 = 2 \end{cases}$ first as a vector equation and then as a matrix equation.

Exercise 4.11. Let $\mathbf{v}_1 = \begin{bmatrix}0\\0\\-2\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\-3\\8\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}4\\-1\\-5\end{bmatrix}$.

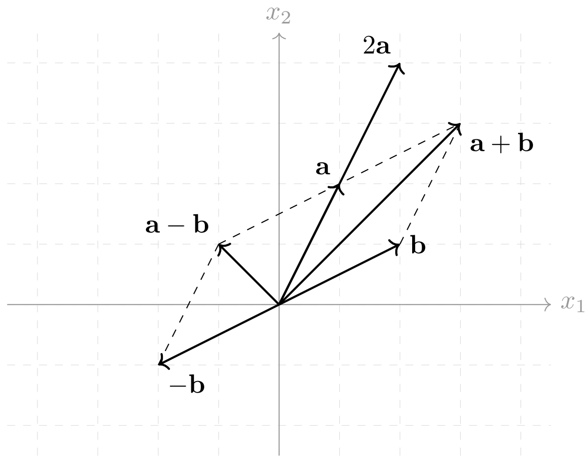
- Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^3$? Why or why not?
- Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly independent? Why or why not?

5 Lecture 5 - Geometric interpretation of solutions of linear systems

5.1 Geometric interpretation of vectors

We like to identify vectors in $M_{n\times 1}(\mathbb{R})$ with points in $\mathbb{R}^n$. And there are very nice geometric interpretation of vector additions and scalar multiplications.

Example 5.1 (Vector-point correspondence in the case of $n = 2$)**.** Let $\mathbf{a} = \begin{bmatrix}1\\2\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}2\\1\end{bmatrix}$, then $\mathbf{a} + \mathbf{b} = \begin{bmatrix}3\\3\end{bmatrix}$, $2\mathbf{a} = \begin{bmatrix}2\\4\end{bmatrix}$, $-\mathbf{b} = (-1)\mathbf{b} = \begin{bmatrix}-2\\-1\end{bmatrix}$, $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b}) = \begin{bmatrix}-1\\1\end{bmatrix}$.



5.2 Basis

Theorem 5.2. Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true statements or they are all false.

a. For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.

b. Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.

c. The columns of A span $\mathbb{R}^m$.

d. A has a pivot position in every row. (Equivalently, in the last row)

Definition 5.3. $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is said to be a basis for $\mathbb{R}^n$ if it is linearly independent and spans all of $\mathbb{R}^n$

Theorem 5.4. Let $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ forms basis of $\mathbb{R}^n \iff A \sim I_n$, in other words, each row and each column of A has a pivot.

Definition 5.5. The standard basis for $\mathbb{R}^n$ is the set of vectors $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$, where

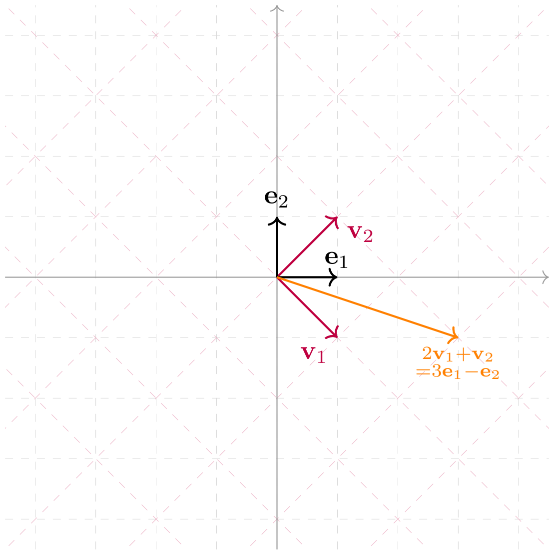
$$\mathbf{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j\text{-th entry}$$

Example 5.6. $\left\{ \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is the standard basis for $\mathbb{R}^3$, and

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$$

5.3 Geometric meaning of spans

Example 5.7. Consider Example 4.9 where $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$



Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{e}_1\}$ and Span$\{\mathbf{v}_1, \mathbf{v}_2\}$ are both the plane $\mathbb{R}^2$, $\mathbf{e}_1$ is in the span of $\{\mathbf{v}_1, \mathbf{v}_2\}$ because $\mathbf{e}_1 = \frac{1}{2}\mathbf{v}_1 + \frac{1}{2}\mathbf{v}_2$. The gray grids illustrate the span of $\{\mathbf{e}_1, \mathbf{e}_2\}$ and the purple grids illustrate the span of $\{\mathbf{v}_1, \mathbf{v}_2\}$.

Exercise 5.8. Suppose $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

- Determine whether $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ forms a basis for $\mathbb{R}^3$.
- Without performing elementary row operations, how many solutions does $A\mathbf{x} = \mathbf{b}$ have? Where $\mathbf{b} = \begin{bmatrix} 1 \\ 45 \\ -9 \end{bmatrix}$, what about $\mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

5.4 Parametric vector form

Example 5.9. In Example 2.13, the solution set can be written as parametric vector form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}x_3 + \frac{1}{2} \\ \frac{3}{2}x_3 - \frac{1}{2} \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}x_3 \\ \frac{3}{2}x_3 \\ x_3 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix} = x_3 \begin{bmatrix} \frac{1}{2} \\ \frac{3}{2} \\ 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix}$$

In Example 2.17, the solution set can be written as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2x_2 + 16 \\ x_2 \\ \frac{5}{2} \\ -\frac{9}{2} \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 16 \\ 0 \\ \frac{5}{2} \\ -\frac{9}{2} \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 16 \\ 0 \\ \frac{5}{2} \\ -\frac{9}{2} \end{bmatrix}$$

Exercise 5.10. Suppose the augmented matrix of a linear system is equivalent to the following matrix

$$\begin{bmatrix} 1 & 1 & 0 & 2 & 0 & 3 \\ 0 & 0 & 1 & -2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Write down the solution set in parametric vector form

Solution.

$$\begin{cases} x_1 + x_2 + 2x_4 = 3 \\ x_3 - 2x_4 = 2 \\ x_5 = 1 \end{cases} \Rightarrow \begin{cases} x_1 = 3 - x_2 - 2x_4 \\ x_2 \text{ is free} \\ x_3 = 2 + 2x_4 \\ x_4 \text{ is free} \\ x_5 = 1 \end{cases}$$

So the solution in parametric vector form would be

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 - x_2 - 2x_4 \\ x_2 \\ 2 + 2x_4 \\ x_4 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -x_2 \\ x_2 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -2x_4 \\ 0 \\ 2x_4 \\ x_4 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

□

5.5 Geometric interpretation of solution set to linear system

Definition 5.11. A linear system is homogeneous if it has matrix equation $A\mathbf{x} = \mathbf{0}$ (note that this always have the zero solution, called the trivial solution).

Theorem 5.12. Suppose $\begin{bmatrix} A & \mathbf{b} \end{bmatrix} \sim \begin{bmatrix} U & \mathbf{d} \end{bmatrix}$ is the RREF, then U will be the RREF of A. The solutions of $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{x} = \mathbf{b}$ differs by $\widetilde{\mathbf{d}}$ ($\widetilde{\mathbf{d}}$ is $\mathbf{d}$ with 0's inserted at the free variable positions), i.e.

$$\widetilde{\mathbf{d}} + \{\text{solutions of } A\mathbf{x} = \mathbf{0}\} = \{\text{solutions of } A\mathbf{x} = \mathbf{b}\}$$

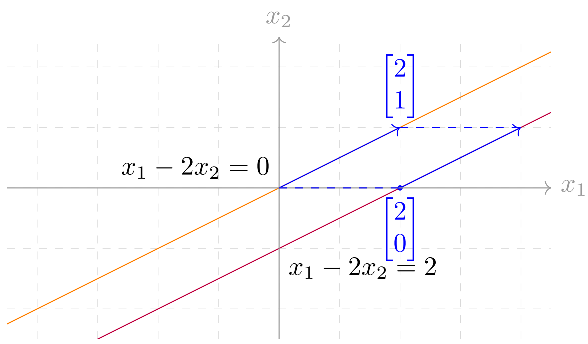
Geometrically speaking, the solution set of $A\mathbf{x} = \mathbf{b}$ is the hyperplane translated from the hyperplane through the origin (solution set of $A\mathbf{x} = \mathbf{0}$) by $\mathbf{d}$.

Example 5.13. $x_1 - 2x_2 = 2$ has augmented matrix $\begin{bmatrix} 1 & -2 & 2 \end{bmatrix}$ which is already an RREF, which has solution

$$\begin{cases} x_1 = 2 + 2x_2 \\ x_2 \text{ is free} \end{cases} \Rightarrow \begin{bmatrix} 2 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Its corresponding homogeneous equation $x_1 - 2x_2 = 0$ has solution

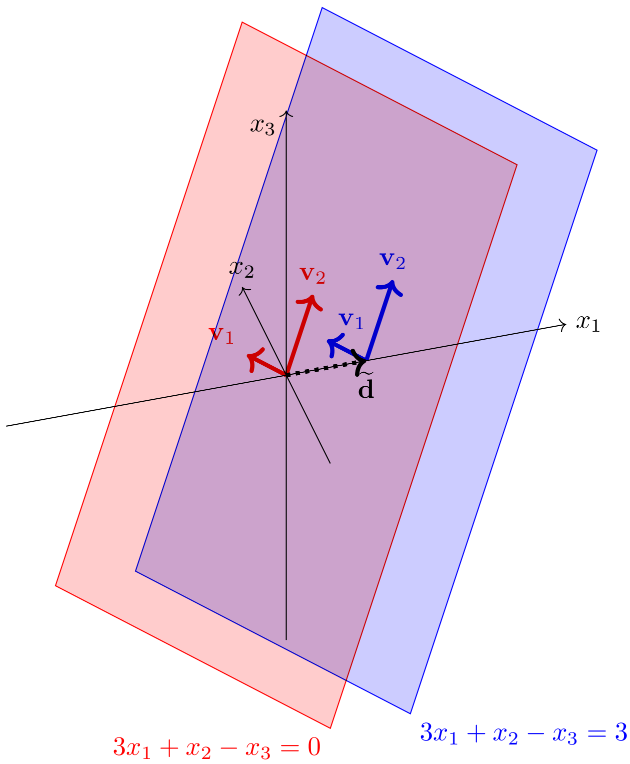
$$\begin{cases} x_1 = 2x_2 \\ x_2 \text{ is free} \end{cases} \Rightarrow x_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$



Example 5.14. Consider homogeneous linear system $3x_1 + x_2 - x_3 = 0$, and non-homogeneous linear system $3x_1 + x_2 - x_3 = 3$ .The parametric vector form of the solution sets to both are

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{1}{3} \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{1}{3} \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$



6 Lecture 6 - Linear transformations

6.1 Linear transformations and matrix transformations

Definition 6.1. A Linear transformation (or linear mapping) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a mapping satisfying

- $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for any $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$
- $T(c\mathbf{u}) = cT(\mathbf{u})$ for any scalar c and any $\mathbf{u}$ in $\mathbb{R}^n$

Example 6.2. Reflection, rotation and scaling are all linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$.

Exercise 6.3. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y \end{bmatrix}$. Is T is a linear mapping?

Solution. Suppose $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, then

- $$\begin{aligned} T(\mathbf{u}+\mathbf{v}) = T\left(\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \end{bmatrix}\right) &= \begin{bmatrix} (u_1+v_1)+(u_2+v_2) \\ u_2+v_2 \end{bmatrix} \\ &= \begin{bmatrix} (u_1+u_2)+(v_1+v_2) \\ u_2+v_2 \end{bmatrix} = \begin{bmatrix} u_1+u_2 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1+v_2 \\ v_2 \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v}) \end{aligned}$$

- $$T(c\mathbf{u}) = T\left(\begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}\right) = \begin{bmatrix} cu_1+cu_2 \\ cu_2 \end{bmatrix} = c\begin{bmatrix} u_1+u_2 \\ u_2 \end{bmatrix} = cT(\mathbf{u})$$

$\square$

Definition 6.4. A matrix transformation is the mapping defined via matrix multiplication, i.e. $T(\mathbf{x}) = A\mathbf{x}$ for some $m \times n$ matrix A. It is a linear transformation thanks to Fact 3.8 c),d) since

- $T(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$

- $T(c\mathbf{u}) = A(c\mathbf{u}) = c(A\mathbf{u}) = cT(\mathbf{u})$

Example 6.5. In fact, T in Exercise 6.3 is a matrix transformation

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\mathbf{x}$$

Definition 6.6. In general, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ is the standard basis, then any $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, and

$$T(\mathbf{x}) = x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n) = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{6.1}$$

Denote $A = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$ (which is called the standard matrix for the linear transformation T), then $T(\mathbf{x}) = A\mathbf{x}$, so every linear transformation T from $\mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation.

Example 6.7. In Exercise 6.3

$$T(\mathbf{e}_1) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1+0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad T(\mathbf{e}_2) = T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0+1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Therefore the standard matrix for T is $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

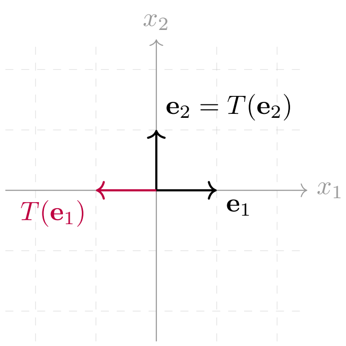
Exercise 6.8. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 + 1 \\ x_2 \end{bmatrix}$. Is T is a linear mapping?

Solution. T is not a linear transformation since $T(\mathbf{x}) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = A\mathbf{x} + \mathbf{p}$ is not a matrix transformation (6.1) ($\mathbf{p} \neq 0$) □

Example 6.9. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation

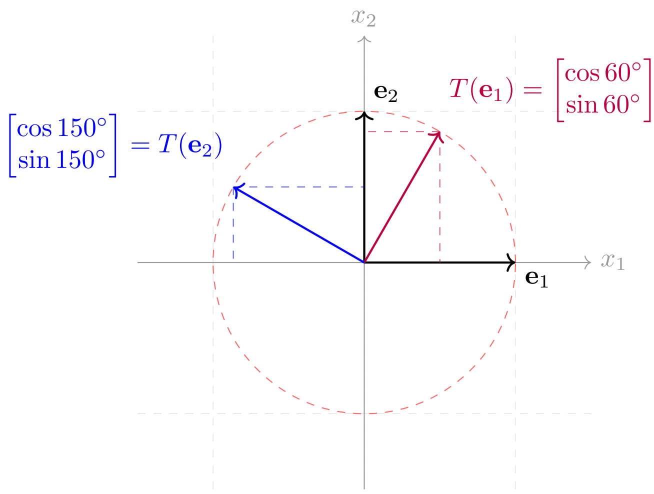
a) Suppose T is the reflection over x_2-axis, then the standard matrix for T is

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$



b) Suppose T is the rotation by 60° counter-clockwise, then the standard matrix for T is

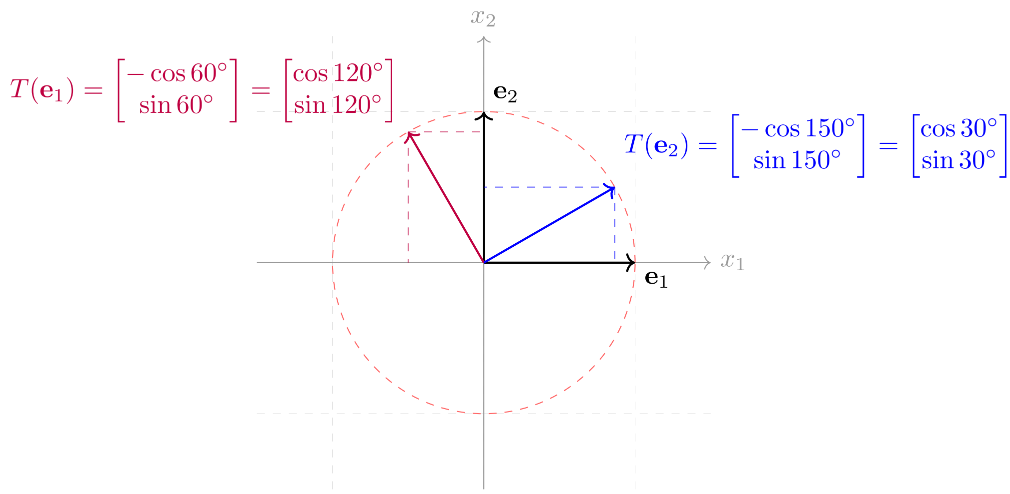
$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$



Exercise 6.10. $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear transformation that rotate 60° counter-clockwise and then reflects over x_2-axis, what is its standard matrix?

Solution. The standard matrix for T is

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$



□

6.2 Properties of linear transformations

Definition 6.11. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a mapping. We call

- $\mathbb{R}^n$ the domain of T
- $\mathbb{R}^m$ the codomain of T
- $T(\mathbf{x})$ the image of $\mathbf{x}$ under T
- $T^{-1}(\mathbf{b}) = \{\mathbf{x} | T(\mathbf{x}) = \mathbf{b}\}$ the set of preimages of $\mathbf{b}$ under T
- the set of images $\{T\mathbf{x} | \mathbf{x} \in \mathbb{R}^n\}$ the range of T

Exercise 6.12. Suppose the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ is defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 + x_3 \\ 2x_1 - x_3 \\ x_1 + x_2 + x_3 \end{bmatrix}$, what is the standard matrix of T? What is the image $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$, what is the set of vectors with image $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$, what is the range?

Solution. The standard matrix is $A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}$, the image $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$ under T is

$$T\left(\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

The set of vectors with image $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ under T is the solution set to $T(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$ (this is Example 2.7), which is $\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$. And since there is a pivot in each row, by Theorem 5.2, the range of T is $\mathbb{R}^3$ □

Definition 6.13. A mapping T is said to be onto $\mathbb{R}^m$ if each $b \in \mathbb{R}^m$ is the image of at least one $x \in \mathbb{R}^n$.

Codomain is larger than the range if T is not onto

Definition 6.14. A mapping T is said to be one-to-one if each $b \in \mathbb{R}^m$ is the image of at most one $x \in \mathbb{R}^n$.

Theorem 6.15. Suppose A is the standard matrix for linear transformation T (i.e. $T(\mathbf{x}) = A\mathbf{x}$), then

- T is one-to-one $\iff$ $A\mathbf{x} = \mathbf{b}$ has at most one solution $\iff$ $A\mathbf{x} = \mathbf{0}$ has the unique trivial solution $\iff$ RREF of A has a pivot in each column $\iff$ columns of A are linearly independent.
- T is onto $\iff$ $A\mathbf{x} = \mathbf{b}$ always has solution $\iff$ the columns of A span $\mathbb{R}^m$ $\iff$ RREF of A has a pivot in each row.

Exercise 6.16. Suppose the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ is defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 + x_3 \\ 2x_1 - x_3 \end{bmatrix}$, Is T onto? Is T one-to-one?

Solution. This is the Example 2.13. The standard matrix for T is $A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \end{bmatrix} \xrightarrow{R2 \to R2 - 2R1} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 2 & -3 \end{bmatrix}$, since there is a pivot in each row but not in each column, it is onto but not one-to-one □

Exercise 6.17.

- If $T : \mathbb{R}^3 \to \mathbb{R}^2$ is a linear transformation, could it be one-to-one?
- If $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation, could it be onto?

Solution. Both no! Due to Theorem 6.15

- Since A is a 2×3 matrix, there will be at most 2 pivots (only 2 rows), so there won't be enough pivots to fill all columns.
- Since A is a 3×2 matrix, there will be at most 2 pivots (only 2 columns), so there won't be enough pivots to fill all rows.

$\square$

Exercise 6.18. Suppose $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 + x_3 \\ 2x_1 - x_3 \\ x_1 + x_2 - 2x_3 \end{bmatrix}$.

- What is the domain of T?
- What is the codomain of T?
- What is the standard matrix of T?
- What is the image of $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$?
- What is the set of vectors with image being $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$?
- What is the set of vectors with image being $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$?
- Is T onto?
- Is T one-to-one?

6.3 Composition of linear transformations

Definition 6.19. Suppose

- $T_1 : \mathbb{R}^p \to \mathbb{R}^n$ is a linear transformation with standard matrix A_1 (which should be $n \times p$)
- $T_2 : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with standard matrix A_2 (which should be $m \times n$)

Define the composition $T_2 \circ T_1$ of T_1 and T_2 as $(T_2 \circ T_1)(\mathbf{x}) = T_2(T_1(\mathbf{x}))$

$$\begin{tikzcd} \mathbb{R}^p \arrow[r, "T_1"] \arrow[rr, bend right, "T_2 \circ T_1"'] & \mathbb{R}^n \arrow[r, "T_2"] & \mathbb{R}^m \end{tikzcd}$$

Then $T_2 \circ T_1 : \mathbb{R}^p \to \mathbb{R}^m$ is also a linear transformation (Why? Verify this). For $\mathbf{x} \in \mathbb{R}^p$,

$$(T_2 \circ T_1)(\mathbf{x}) = T_2(T_1(\mathbf{x})) = A_2(T_1(\mathbf{x})) = A_2(A_1\mathbf{x}) = (A_2A_1)\mathbf{x}$$

So we have concluded that the standard matrix for $T_2 \circ T_1$ is the $m \times p$ matrix A_2A_1.

Note. You should compose maps to the left.

Example 6.20. Consider Example 6.9., If we let $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ to denote the rotation by $60°$ counter-clockwise, $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ to denote reflection over x_2-axis, and their standard matrices are

$$\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}, \qquad \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then look at Exercise 6.10, this is the composition $T_2 \circ T_1$, which has the standard matrix

$$A_2A_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

Question. Suppose $A = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$ is the standard matrix for the linear transformation of rotating $60°$ counter-clockwise (Example 6.9). What is A^7?

Answer. $A^7 = AAAAAAA$ is the standard matrix for composition of linear transformations $T \circ T \circ T \circ T \circ T \circ T \circ T$ which is rotate $7 \times 60° = 420°$, but that is the same as rotating $420° - 360° = 60°$ which is the same linear transformation as T, so $A^7 = A$

$$\begin{tikzcd}
A^7 = AAAAAAA \arrow[r, equal] & A \\
T^{\circ 7} = T \circ T \circ T \circ T \circ T \circ T \circ T \arrow[u, squiggly, "\text{Standard matrix}"] \arrow[r, equal, "\text{same effect}"'] & T \arrow[u, squiggly, "\text{Standard matrix}"']
\end{tikzcd}$$

Exercise 6.21. Let T be the linear transformation that rotate $\mathbb{R}^2$ counter-clockwise of angle θ, find the standard matrix for T. What about the standard matrix for $T^{\circ 100}$

7 Lecture 7 - Matrix transpose and matrix inverse

7.1 Matrix transpose

Definition 7.1. Suppose $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ is a $m \times n$ matrix, we define its transpose by flipping it over the diagonal, and this is the $n \times m$ matrix

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

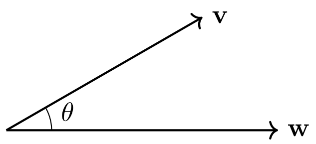
Example 7.2. Suppose $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

Theorem 7.3. Here are some properties of matrix transpose

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(cA)^T = cA^T$
- $(AB)^T = B^T A^T$

Definition 7.4. For any $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^n$, we can define the dot product to be $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = v_1 w_1 + \cdots + v_n w_n$. $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2}$ is the length $\mathbf{v}$

Remark. There is a nice geometric interpretation of dot product. Suppose the angle between $\mathbf{v}$ and $\mathbf{w}$ is θ, then $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta$.



Exercise 7.5. Let $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$.

- What is the length of $\mathbf{v}$?
- What is $\mathbf{v} \cdot \mathbf{w}$?
- what is the angle between $\mathbf{v}$ and $\mathbf{w}$?

Question. Have you wondered about how linear algebra would look like if we make the identification of row vectors $M_{1\times n}(\mathbb{R}^n)$ and $\mathbb{R}^n$ instead of column vectors.

Answer. The standard basis for $\mathbb{R}^n$ would be $\{\mathbf{e}_1^T, \mathbf{e}_2^T, \cdots, \mathbf{e}_n^T\}$, where $\mathbf{e}_j^T = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix}$, where every entry is 0 except the j-th entry being 1. For any $\mathbf{x}^T \in \mathbb{R}^n$ and any linear transformation $T : \mathbb{R}^n \to \mathbf{R}^m$, we have

$$\begin{aligned} T(\mathbf{x}^T) &= T(\mathbf{x})^T = (x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + \cdots + x_n T(\mathbf{e}_n))^T \\ &= x_1 T(\mathbf{e}_1)^T + x_2 T(\mathbf{e}_2)^T + \cdots + x_n T(\mathbf{e}_n)^T \\ &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} T(\mathbf{e}_1)^T \\ T(\mathbf{e}_2)^T \\ \vdots \\ T(\mathbf{e}_n)^T \end{bmatrix} = \mathbf{x}^T A^T \end{aligned}$$

Here $A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$

If you consider composition $T_2 \circ T_1$ where T_1, T_2 are linear transformations with standard matrices A, B, you would get

$$(T_2 \circ T_1)(\mathbf{x}^T) = T_2(T_1(\mathbf{x}^T)) = T_1(\mathbf{x}^T)B^T = (\mathbf{x}^T A^T)B^T = \mathbf{x}^T (A^T B^T)$$

On the other hand, we should know that $(T_2 \circ T_1)(\mathbf{x}^T) = \mathbf{x}^T (A^T B^T)$, so this implies $A^T B^T = (BA)^T$.

7.2 Matrix inverse

Definition 7.6. Suppose linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is both onto and one-to-one (i.e. for each vector $\mathbf{b}$ in the codomain $\mathbb{R}^m$ there is a unique preimage, which we denote as $T^{-1}(\mathbf{b})$). Suppose A is the standard matrix for T, then m necessarily equal n as shown in Exercise 6.17, so A must be a square matrix. We know $T(\mathbf{x}) = \mathbf{b}$ always has a unique solution which should be $T^{-1}(\mathbf{b})$, it can be shown that $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ as mapping is actually also a linear transformation (Why? See if you can figure this out). Then the standard matrix of T^{-1} is defined to be A^{-1} (the inverse matrix of A). Note that

$$\begin{aligned} (T \circ T^{-1})(\mathbf{b}) &= T(T^{-1}(\mathbf{b})) = T(\mathbf{x}) = \mathbf{b} \\ (T^{-1} \circ T)(\mathbf{x}) &= T^{-1}(T(\mathbf{x})) = T^{-1}(\mathbf{b}) = \mathbf{x} \end{aligned}$$

Since $T \circ T^{-1}$, $T^{-1} \circ T$ work like the identity mappings, so $AA^{-1} = A^{-1}A = I$. In this case, we see that A is equivalent to the identity matrix (because of Theorem 4.9, A has a pivot in each row and column).

Remark. Because we can write elementary row operations as left elementary matrix multiplications, so we know there are elementary matrices $E_1, E_2, \cdots, E_k$ such that

$$\begin{aligned} A &\sim E_1 A \sim E_2 E_1 A \sim E_3 E_2 E_1 A \sim \cdots \\ &\sim E_k E_{k-1} \cdots E_2 E_1 A = I \end{aligned}$$

If we multiply A^{-1} on the right on both sides, we get $E_k E_{k-1} \cdots E_2 E_1 = A^{-1}$. Thanks to this observation, we introduce an algorithm for computing matrix inverses. Let's consider the RREF of the following partitioned matrix

$$\begin{aligned} \left[\begin{array}{c|c} A & I \end{array}\right] &\sim \left[\begin{array}{c|c} E_1 A & E_1 \end{array}\right] \sim \left[\begin{array}{c|c} E_2 E_1 A & E_2 E_1 \end{array}\right] \sim \cdots \\ &\sim \left[\begin{array}{c|c} E_k E_{k-1} \cdots E_2 E_1 A & E_k E_{k-1} \cdots E_2 E_1 \end{array}\right] = \left[\begin{array}{c|c} I & A^{-1} \end{array}\right] \end{aligned}$$

Exercise 7.7. Find the inverse of the following matrices.

a) $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ b) $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ c) $\begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}$

Solution.

a)

$$\left[\begin{array}{cc|cc} -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \xrightarrow{R1\to(-1)R1} \left[\begin{array}{cc|cc} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right]$$

Hence $A^{-1} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

b)

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 5 & 0 & 1 \end{array}\right] \xrightarrow{R2\to R2-3R1} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -1 & -3 & 1 \end{array}\right]$$
$$\xrightarrow{R1\to R1+2R2} \left[\begin{array}{cc|cc} 1 & 0 & -5 & 2 \\ 0 & -1 & -3 & 1 \end{array}\right] \xrightarrow{R2\to(-1)R2} \left[\begin{array}{cc|cc} 1 & 0 & -5 & 2 \\ 0 & 1 & 3 & -1 \end{array}\right]$$

Hence $A^{-1} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$

c)

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 2 & 0 & -1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow[R3\to R3-R1]{R2\to R2-2R1} \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 2 & -3 & -2 & 1 & 0 \\ 0 & 2 & 0 & -1 & 0 & 1 \end{array}\right]$$
$$\xrightarrow{R3\to R3-R2} \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 2 & -3 & -2 & 1 & 0 \\ 0 & 0 & 3 & 1 & -1 & 1 \end{array}\right] \xrightarrow{R2\to R2+R3} \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 1 \\ 0 & 0 & 3 & 1 & -1 & 1 \end{array}\right]$$
$$\xrightarrow[R3\to R3/3]{R2\to R2/2} \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{array}\right] \xrightarrow{R1\to R1+R2-R3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ 0 & 1 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{array}\right]$$

Hence $A^{-1} = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}$

□

7.3 Properties of matrix transposes and inverses

Definition 7.8. A square matrix A is invertible (or non-singular) if it has an inverse A^{-1} such that $AA^{-1} = A^{-1}A = I$. A is called singular if A is not invertible.

Theorem 7.9. Suppose T is a linear transformation with standard matrix A, then

$$T \text{ is invertible with inverse } T^{-1} \iff A \text{ is invertible with inverse } A^{-1} \iff A \sim I$$

Theorem 7.10. $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $A^{-1} = \dfrac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, here $\det A = ad - bc$

Example 7.11. If $A = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$, then

$$A^{-1} = \frac{1}{\frac{1}{2}\frac{1}{2} - \frac{\sqrt{3}}{2}\left(-\frac{\sqrt{3}}{2}\right)} \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

Theorem 7.12. If A is invertible, then the linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$

Proof. Multiply A^{-1} on the left on both sides $\square$

Example 7.13. Let's consider (1.4), in which case $A = \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 10 \\ 26 \end{bmatrix}$, then $A^{-1} = \dfrac{1}{1 \cdot 4 - 1 \cdot 2} \begin{bmatrix} 4 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -\frac{1}{2} \\ -1 & \frac{1}{2} \end{bmatrix}$, and

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} 2 & -\frac{1}{2} \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 10 \\ 26 \end{bmatrix} = \begin{bmatrix} 2 \cdot 10 - \frac{1}{2} \cdot 26 \\ -10 + \frac{1}{2} \cdot 26 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

Theorem 7.14. Here are some properties of matrix inverse

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$

Exercise 7.15. What is $(A^T)^{-1}$ in Exercise 7.7, c)?

Solution. Use Theorem 7.14, we know

$$(A^T)^{-1} = (A^{-1})^T = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}^T = \begin{bmatrix} \frac{1}{6} & -\frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

$\square$

Exercise 7.16. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix A.

- If A is invertible, then A has n pivots. ✓
- If T is one-to-one, then A is invertible. ✓
- If columns of A span $\mathbb{R}^n$, then A is invertible. ✓
- If A is invertible, $A\mathbf{x} = \mathbf{0}$ only has the trivial solution. ✓
- If T is onto, then T is one-to-one. ✓
- If T is one-to-one, then T is onto. ✓

Exercise 7.17. Suppose $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation with standard matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$. Is A^{-1} invertible? Is T one-to-one? Is A^T invertible? If so, what is $(A^T)^{-1}$? If so, what is $(A^{-1})^{-1}$. Is T invertible (i.e. does T^{-1} exist)? What is the standard matrix of T^{-1}? Is T onto?

Solution.

$$[\ A \mid I\ ] = \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow[R2 \to R2+2R3]{R1 \to R1-3R3} \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & -3 \\ 0 & -1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R2 \to (-1)R2} \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & -3 \\ 0 & 1 & 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R1 \to R1-2R2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 2 & 1 \\ 0 & 1 & 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] = [\ I \mid A^{-1}\ ]$$

So $A^{-1} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$. By Theorem 7.14, we know

$$(A^{-1})^{-1} = A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$(A^T)^{-1} = (A^{-1})^T = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$$

Therefore we know A^{-1} and A^T are invertible.
In general, if T is invertible, then A is invertible, so A^{-1} will be the standard matrix for T^{-1} as $T^{-1}(\mathbf{x}) = A^{-1}\mathbf{x}$, in more explict terms, we have

$$T^{-1}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 + x_3 \\ -x_2 - 2x_2 \\ x_3 \end{bmatrix}$$

□

8 Lecture 8 - Determinant

Definition 8.1. We say a square matrix A is upper triangular if it only has zeros to the left of the diagonal

$$\begin{bmatrix} * & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & * \end{bmatrix}$$

A is lower triangular if it only has zeros to the right of the diagonal

$$\begin{bmatrix} * & 0 & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 \\ * & * & * & * & * & 0 \\ * & * & * & * & * & * \end{bmatrix}$$

A is diagonal if A only has nonzero entries on the diagonal

$$\begin{bmatrix} * & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & * \end{bmatrix}$$

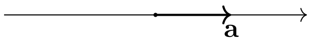
A diagonal matrix is both upper triangular and lower triangular

8.1 Geometric definition of determninants

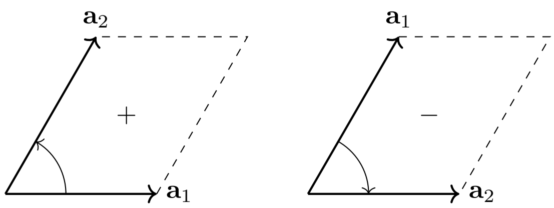
Now let's introduce determinants (ONLY for square matrices!!!). Consider the parallelepiped P with edges $\mathbf{a}_1, \cdots, \mathbf{a}_n$ in $\mathbb{R}^n$. We would like the following geometric definition of determinants.

Definition 8.2. The determinant of $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ (Usually denoted $\det A$ or $|A| = \begin{vmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{vmatrix}$, replacing brackets with vertical lines) as *signed* volumes of P . Therefore we have $\mathrm{Vol}(P) = |\det A|$, i.e. actual volume is the absolute value of the determinant.

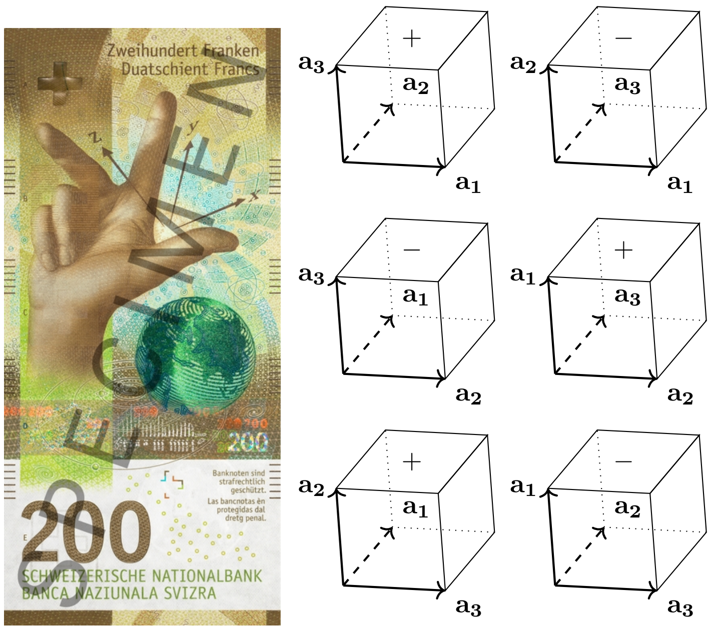
Example 8.3 ($n = 1$)**.** Suppose $A = [a]$ is a 1 by 1 matrix, then $\det A$ is the signed length of $a \in \mathbb{R}^1$, which is a itself! Namely $\det A = a$.



Example 8.4 ($n = 2$)**.** Suppose $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$ is a 2 by 2 matrix. $\det A$ is the actual positive area of the parallelogram if $\mathbf{a}_1$ turns counter-clockwise to $\mathbf{a}_2$, otherwise the negative area.

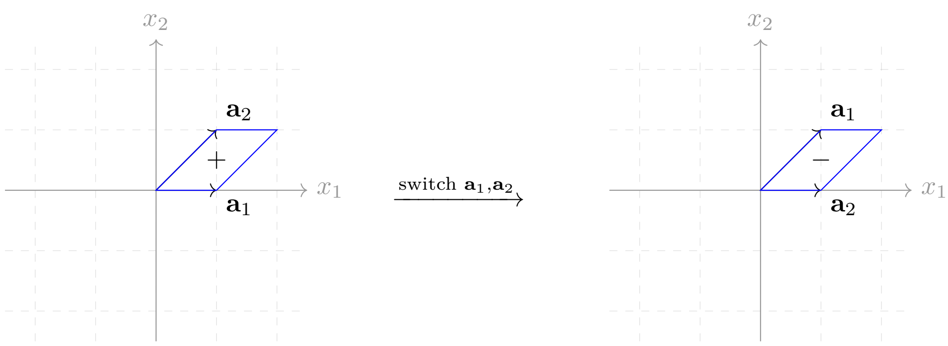


Example 8.5 ($n = 3$)**.** Suppose $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$ is a 3 by 3 matrix. To decide the sign of the volume of the parallelepiped, we follow the right-hand rule.

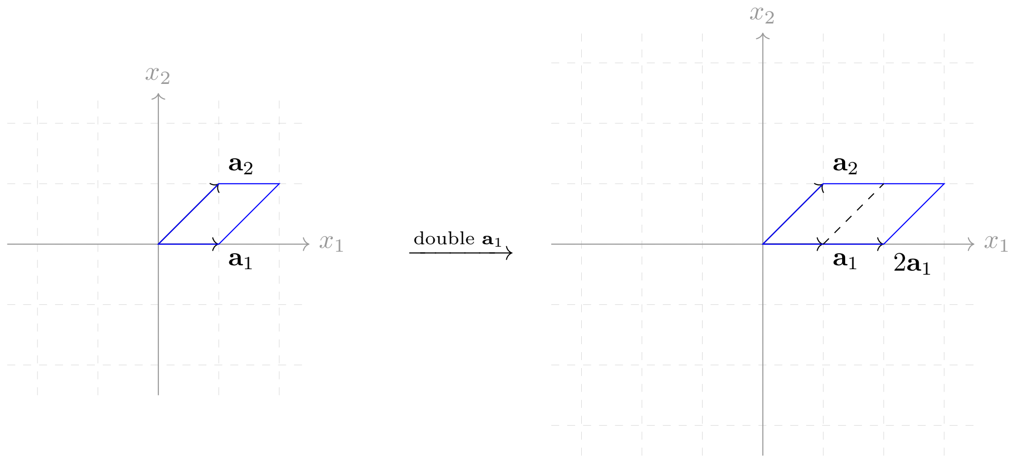


Determinant has following three properties:

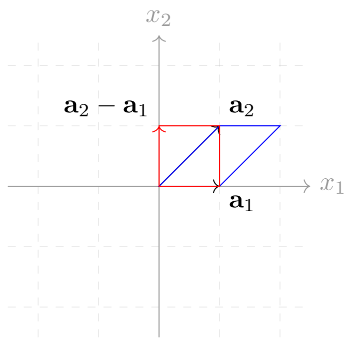
1. Interchanging $\mathbf{a}_1, \mathbf{a}_2$ changes the sign of determinant. Namely $\det \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_1 \end{bmatrix} = -\det \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$



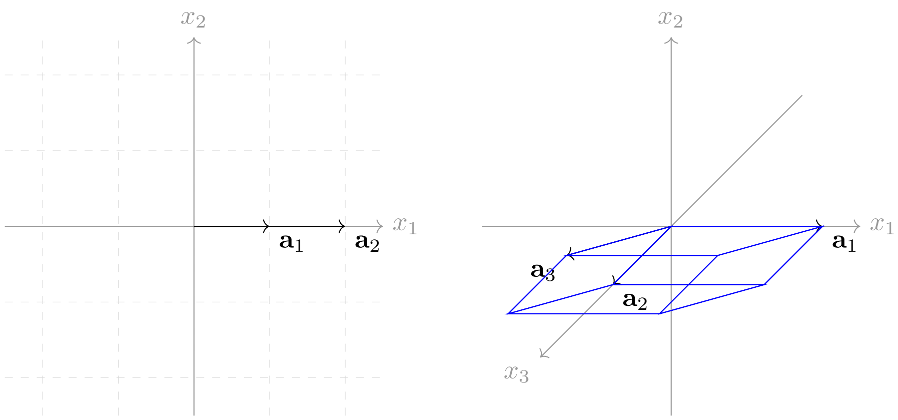
2. Scaling $\mathbf{a}_1$ scales the determinant. Namely $\det \begin{bmatrix} c\mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = c \det \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$



3. Adding a multiple of $\mathbf{a}_1$ to $\mathbf{a}_2$ doesn't change the determinant. Namely $\det\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 + c\mathbf{a}_1\end{bmatrix} = \det\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}$



Remark. If $\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$ is linearly dependent, then A is singular, i.e. not invertible, then the determinant will be zero, since the parallelepiped will be constraint in a hyperplane which has zero volume. Take $n = 2$ and 3 for examples



8.2 Properties of determinants

Theorem 8.6. Suppose $A = \begin{bmatrix} h & 0 \\ * & B \end{bmatrix}$ or $\begin{bmatrix} h & * \\ \mathbf{0} & B \end{bmatrix}$ where h is a scalar, B is a $(n-1) \times (n-1)$ submatrix, then $\det A = h \cdot \det B$

Corollary 8.7. Determinant of a triangular matrix is the product of the diagonal elements.

Proof. Apply Theorem 8.6 repeatedly. $\square$

Theorem 8.8. $\det(A) = \det(A^T)$

Thanks to Theorem 8.8, we can compute determinants via elementary row and column operations.

Example 8.9. Use elementary row operations to evaluate the following (Note that we omit the backward phase (which are replacements) since it doesn't change the determinants)

i.

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 2 & 1 \end{vmatrix} \xlongequal[R3\to R3-R1]{R2\to R2-2R1} \begin{vmatrix} 1 & -1 & 1 \\ 0 & 2 & -3 \\ 0 & 3 & 0 \end{vmatrix} \xlongequal{\text{factor } R3} 3\begin{vmatrix} 1 & -1 & 1 \\ 0 & 2 & -3 \\ 0 & 1 & 0 \end{vmatrix} \xlongequal{R2\to R2-2R3} 3\begin{vmatrix} 1 & -1 & 1 \\ 0 & 0 & -3 \\ 0 & 1 & 0 \end{vmatrix}$$

$$\xlongequal{R2\leftrightarrow R3} (-1)\cdot 3\begin{vmatrix} 1 & -1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{vmatrix} = (-1)\cdot 3\cdot 1\cdot 1\cdot(-3) = 9$$

ii.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 0 & -1 \\ -1 & 2 & 2 \end{vmatrix} \xlongequal[R3\to R3+R1]{R2\to R2-2R1} \begin{vmatrix} 1 & 2 & 1 \\ 0 & -4 & -3 \\ 0 & 4 & 3 \end{vmatrix} \xlongequal{R3\to R3+R2} \begin{vmatrix} 1 & 2 & 1 \\ 0 & -4 & -3 \\ 0 & 0 & 0 \end{vmatrix} = 1\cdot(-4)\cdot 0 = 0$$

iii.

$$\begin{vmatrix} 1 & 2 & 0 & 0 \\ 2 & 8 & 0 & 0 \\ 1 & 3 & 2 & 0 \\ -4 & 5 & 7 & 1 \end{vmatrix} \xlongequal{C2\to C2-2C1} \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 1 & 2 & 0 \\ -4 & 13 & 7 & 1 \end{vmatrix} \xlongequal{\text{transpose}} \begin{vmatrix} 1 & 2 & 1 & -4 \\ 0 & 4 & 1 & 13 \\ 0 & 0 & 2 & 7 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1\cdot 4\cdot 2\cdot 1 = 8$$

Exercise 8.10. Suppose I is the $n\times n$ identity matrix, what is $\det I$, $\det(-I)$, $\det(2I)$ and $\det(aI)$?

Solution. Note that I is a diagonal matrix. $\det I = 1$, $\det(-I) = (-1)^n$, $\det(2I) = 2^n$, and in general $\det(aI) = a^n$ by factoring each row. $\square$

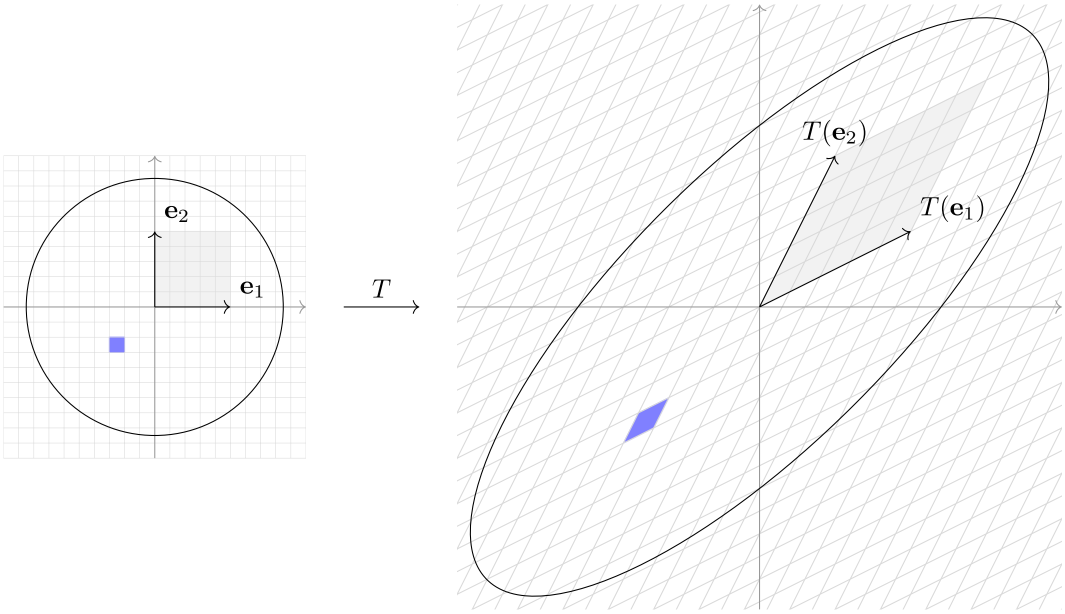
Example 8.11. Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \xlongequal{R2\to R2-\frac{c}{a}R1} \begin{vmatrix} a & b \\ 0 & d-\frac{bc}{a} \end{vmatrix} = a\left(d - \frac{bc}{a}\right) = ad - bc$$

Definition 8.12. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation with standard matrix A, the determinant of T is defined to be $\det T = \det A$.

Question. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation with standard matrix $A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. What is the area of the image of the unit circle under T?

Answer. We sketch the unit circle and its image



Since every small blue squares has been deformed into small parallelograms which are congruent to the gray square, parallelogram respectively and are of the same (signed) ratio $\det A$, so any area gets scaled by $\det A = 2 \cdot 2 - 1 \cdot 1 = 3$ under T.

Remark. In multi-variable calculus, this is known as the Jacobian.

Theorem 8.13. Suppose A, B are $n \times n$ matrices, then $\det(AB) = (\det A)(\det B)$

Proof. Suppose T_1, T_2 are linear transformations with standard matrices A, B, respectively. Consider $T_2 \circ T_1$ which should scale area by $\det(BA)$. On the other hand, this should scale area by $\det A$ and then $\det B$ □

Theorem 8.14. A is invertible $\iff$ $\det A \neq 0$. In addition, $\det(A^{-1}) = \dfrac{1}{\det A}$.

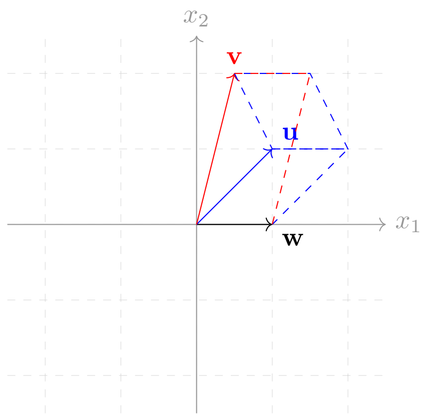
Proof. If A is invertible, then A^{-1} is well-defined, then $1 = \det I = \det(AA^{-1}) = (\det A)(\det(A^{-1})) \Rightarrow \det(A^{-1}) = \dfrac{1}{\det A}$, so $\det A \neq 0$. Conversely, if $\det A \neq 0$, A would have n pivots, so a pivot in each row and column, thus A will be invertible. □

Exercise 8.15. Compute the determinants of the following matrices

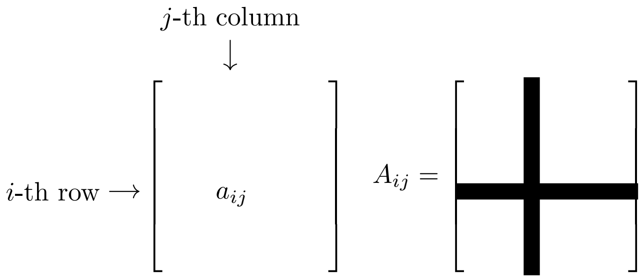
- $\begin{bmatrix} 1 & -2 \\ 4 & 3 \end{bmatrix}$
- $\begin{bmatrix} 1 & 5 & -4 \\ -1 & -4 & 5 \\ -2 & -8 & 7 \end{bmatrix}$
- $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 3 \\ 0 & 7 & -1 & -8 \\ 2 & 2 & -5 & 1 \end{bmatrix}$

8.3 Cofactor expansion

There is one more property of determinants. $\det \begin{bmatrix} \mathbf{w} & \mathbf{u}+\mathbf{v} \end{bmatrix} = \det \begin{bmatrix} \mathbf{w} & \mathbf{u} \end{bmatrix} + \det \begin{bmatrix} \mathbf{w} & \mathbf{v} \end{bmatrix}$



Definition 8.16. We use a_{ij} to be denote the (i, j)-th entry of the matrix A, and A_{ij} to denote the submatrix of A by deleting the i-th row and the j-th column



We define the (i, j)-cofactor to be $C_{ij} = (-1)^{i+j} \det A_{ij}$ (we also call $\det A_{ij}$ a minor).

For $n \geq 2$, with the help of Lemma **??**, we derive cofactor expansion formula.
The cofactor expansion across the i-th row is

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

The cofactor expansion across the j-th column is

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

$$\begin{bmatrix} \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \\ \end{bmatrix} \qquad \begin{bmatrix} & a_{1j} & \\ & a_{2j} & \\ & \vdots & \\ & a_{nj} & \end{bmatrix}$$

Proof of Theorem 8.8. Note that row cofactor expansion of A is the same as column cofactor expansion of A^T. Hence we can prove this inductively on the size of the matrix. $\square$

Example 8.17.

$$\begin{vmatrix} 1 & 2 & 3 & 0 \\ 0 & 3 & -1 & 0 \\ -1 & 2 & 1 & 2 \\ 2 & -3 & 1 & 0 \end{vmatrix} \xlongequal{\text{cofactor expansion across last column}} 2(-1)^{3+4} \begin{vmatrix} 1 & 2 & 3 \\ 0 & 3 & -1 \\ 2 & -3 & 1 \end{vmatrix}$$

$$\xlongequal{R3 \to R3 - 2R1} (-2) \begin{vmatrix} 1 & 2 & 3 \\ 0 & 3 & -1 \\ 0 & -7 & -5 \end{vmatrix} \xlongequal{\text{cofactor expansion across first column}} (-2) \cdot 1(-1)^{1+1} \begin{vmatrix} 3 & -1 \\ -7 & -5 \end{vmatrix}$$

$$= (-2)(3(-5) - (-1)(-7)) = 44$$

Remark. When use the cofactor expansion, we want to apply it to rows/columns with more 0's

Exercise 8.18. Suppose $A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}$. Please find the cofactor expansion of A across the

a) 1st row

b) 2nd column

And evaluate determinant of A.

Solution.

$$\begin{aligned}
\det A &= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \\
&= a_{11}(-1)^{1+1}\det A_{11} + a_{12}(-1)^{1+2}\det A_{12} + a_{13}(-1)^{1+3}\det A_{13} \\
&= 1 \cdot (-1)^{1+1}\begin{vmatrix} 0 & -1 \\ 1 & 1 \end{vmatrix} + (-1)\cdot(-1)^{1+2}\begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} + 1 \cdot (-1)^{1+3}\begin{vmatrix} 2 & 0 \\ 1 & 1 \end{vmatrix} \\
&= 1 \cdot (0 \cdot 1 - (-1)\cdot 1) + (-1)\cdot(2 \cdot 1 - (-1)\cdot 1) + 1 \cdot (-1)(2 \cdot 1 - (-1)\cdot 1) \\
&= 6
\end{aligned}$$

$$\begin{aligned}
\det A &= a_{12}C_{12} + a_{22}C_{22} + a_{32}C_{32} \\
&= a_{11}(-1)^{1+1}\det A_{11} + a_{12}(-1)^{1+2}\det A_{12} + a_{13}(-1)^{1+3}\det A_{13} \\
&= (-1)\cdot(-1)^{1+2}\begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} + 0 \cdot (-1)^{2+2}\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} + 1 \cdot (-1)^{3+2}\begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} \\
&= (-1)\cdot(-1)(2 \cdot 1 - (-1)\cdot 1) + 0 \cdot (1 \cdot 1 - 1 \cdot 1) + 1 \cdot (-1)(1 \cdot (-1) - 1 \cdot 2) \\
&= 6
\end{aligned}$$

□

Exercise 8.19. Suppose $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & -1 \\ -1 & 2 & 1 \end{bmatrix}$. Write out the cofactor expansion of A across the second row, and evaluate the determinant $\det A$.

Solution.

$$\begin{aligned}
\det A &= a_{21}C_{21} + a_{22}C_{22} + a_{23}C_{23} \\
&= a_{21}(-1)^{2+1}\det A_{21} + a_{22}(-1)^{2+2}\det A_{22} + a_{23}(-1)^{2+3}\det A_{23} \\
&= 1 \cdot (-1)^{2+1}\begin{vmatrix} 2 & 1 \\ 2 & 1 \end{vmatrix} + 0 \cdot (-1)^{2+2}\begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} + (-1)\cdot(-1)^{2+3}\begin{vmatrix} 2 & 1 \\ 2 & 1 \end{vmatrix} \\
&= 1 \cdot (-1)(2 \cdot 1 - 1 \cdot 2) + 0 \cdot (1 \cdot 1 - 1 \cdot (-1)) + (-1)\cdot(-1)(1 \cdot 2 - 2 \cdot (-1)) \\
&= 4
\end{aligned}$$

□

Remark. The REF of a square matrix A is upper triangular, and $\det A = 0$ if A has less then n pivots.

Exercise 8.20. Suppose $A = \begin{bmatrix} 2 & -1 & 3 & 1 \\ 0 & -2 & 0 & -1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Please find use cofactor expansion to find the $\det A$

Solution. Note that A is upper triangular, so we could do cofactor expansions across first columns multiple times

$$\begin{vmatrix} 2 & -1 & 3 & 1 \\ 0 & -2 & 0 & -1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 2(-1)^{1+1} \begin{vmatrix} -2 & 0 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \end{vmatrix} = 2 \cdot (-2)(-1)^{1+1} \begin{vmatrix} 3 & 1 \\ 0 & 1 \end{vmatrix} = 2 \cdot (-2) \cdot 3 \cdot (-1)^{1+1} 1 = -12$$

□

Exercise 8.21. Consider $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. What is $A_{11}, A_{12}, A_{21}, A_{22}$? What is $C_{11}, C_{12}, C_{21}, C_{22}$. Write down the cofactor expansion of A across the

- 1st row
- 2nd row
- 1st column
- 2nd column

Solution. $A_{11} = [d], A_{12} = [c], A_{21} = [b], A_{22} = [a]$ are all 1 by 1 matrices. $C_{11} = (-1)^{1+1} \det A_{11} = d, C_{12} = (-1)^{1+2} \det A_{21} = -c, C_{21} = (-1)^{2+1} \det A_{21} = -b, C_{22} = (-1)^{2+2} \det A_{22} = a$. So the cofactor expansions are

- $\det A = aC_{11} + bC_{12} = ad - bc$
- $\det A = cC_{21} + dC_{22} = -bc + ad$
- $\det A = aC_{11} + cC_{21} = ad - bc$
- $\det A = bC_{12} + dC_{22} = -bc + ad$

Note that all of the above calculations show that $\det A = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$. □

9 Lecture 9 - Vector spaces and subspaces

9.1 Vector space

To motivate the definition of a vector space, let's consider the following example

Example 9.1. Let $\mathbb{P}_n$ denote the set of (real) polynomials of degree less or equal to n. For example $\mathbb{P}_0 = \mathbb{R}$ is just the set of real numbers, and

$$\begin{aligned} \mathbb{P}_1 &= \{a_0 + a_1 t | a_0, a_1 \in \mathbb{R}\} \\ \mathbb{P}_2 &= \{a_0 + a_1 t + a_2 t^2 | a_0, a_1, a_2 \in \mathbb{R}\} \\ \mathbb{P}_3 &= \{a_0 + a_1 t + a_2 t^2 + a_3 t^3 | a_0, a_1, a_2, a_3 \in \mathbb{R}\} \\ &\vdots \\ \mathbb{P}_n &= \{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n | a_0, a_1, a_2, \cdots, a_n \in \mathbb{R}\}. \end{aligned}$$

You may soon realize that $\mathbb{P}_n$ can be identified with $\mathbb{R}^{n+1}$.

$$\begin{tikzcd}
\mathbb{P}_1 \arrow[r, equal, "\sim"] \arrow[d, equal] & \mathbb{R}^2 \arrow[d, equal] \\
\{a_0 + a_1 t\} \arrow[r, leftrightsquigarrow] & \left\{\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}\right\}
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{P}_2 \arrow[r, equal, "\sim"] \arrow[d, equal] & \mathbb{R}^3 \arrow[d, equal] \\
\{a_0 + a_1 t + a_2 t^2\} \arrow[r, leftrightsquigarrow] & \left\{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}\right\}
\end{tikzcd}$$

$$\begin{tikzcd}
\mathbb{P}_n \arrow[r, equal, "\sim"] \arrow[d, equal] & \mathbb{R}^{n+1} \arrow[d, equal] \\
\{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n\} \arrow[r, leftrightsquigarrow] & \left\{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}\right\}
\end{tikzcd}$$

More concrete examples could be

1. For $\mathbb{P}_1 \cong \mathbb{R}^2$, $1 + 2t \leftrightsquigarrow \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
2. For $\mathbb{P}_2 \cong \mathbb{R}^3$, $3t^2 - 1 \leftrightsquigarrow \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}$

If we consider addition and scalar multiplication, we have

$$\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \qquad 2 \cdot \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}$$

$$(1 + 2t^2) + (2 + t) = 3 + t + 2t^2 \qquad 2 \cdot (1 + 2t^2) = 2 + 4t^2$$

So we may conclude that addition and scalar multiplication in $\mathbb{P}_n$ can be identically translated to addition and scalar multiplication in $\mathbb{R}^{n+1}$

Remark. We call $\{1, t, t^2, \cdots, t^n\}$ the *standard basis* of $\mathbb{P}_n$, corresponding to the standard basis for $\mathbb{R}^{n+1}$

Example 9.2. $\left\{1, t, t^2\right\}$ is the standard basis for $\mathbb{P}_2$, and

$$p(t) = a_0 + a_1 t + a_2 t^2 = a_0 \cdot 1 + a_1 \cdot t + a_2 \cdot t^2$$

Example 9.3. Let's denote $M_{m \times n}(\mathbb{R})$ the set of $m \times n$ matrices. For example

$$M_{2\times 2}(\mathbb{R}) = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \middle| a_{11}, a_{12}, a_{21}, a_{22} \in \mathbb{R} \right\}$$

$$M_{3\times 2}(\mathbb{R}) = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \middle| a_{11}, a_{12}, a_{21}, a_{22}, a_{31}, a_{32} \in \mathbb{R} \right\}$$

$$M_{2\times 3}(\mathbb{R}) = \left\{ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \middle| a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23} \in \mathbb{R} \right\}$$

$$\vdots$$

$$M_{m\times n}(\mathbb{R}) = \left\{ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \middle| a_{11}, \cdots, a_{1n}, \cdots, a_{m1}, \cdots, a_{mn} \in \mathbb{R} \right\}.$$

You may realize that $M_{m\times n}(\mathbb{R})$ can be identified with $\mathbb{R}^{mn}$

$$\begin{tikzcd}
M_{2\times 2}(\mathbb{R}) \arrow[r, equal, "\sim"] \arrow[d, equal] & \mathbb{R}^4 \arrow[d, equal] \\
\left\{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right\} \arrow[r, leftrightsquigarrow] & \left\{\begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}\right\}
\end{tikzcd}
\qquad
\begin{tikzcd}
M_{3\times 2}(\mathbb{R}) \arrow[r, equal, "\sim"] \arrow[d, equal] & \mathbb{R}^6 \arrow[d, equal] \\
\left\{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right\} \arrow[r, leftrightsquigarrow] & \left\{\begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \\ a_{31} \\ a_{32} \end{bmatrix}\right\}
\end{tikzcd}$$

$$\begin{tikzcd}
M_{2\times 3}(\mathbb{R}) \arrow[r, equal, "\sim"] \arrow[d, equal] & \mathbb{R}^6 \arrow[d, equal] \\
\left\{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\right\} \arrow[r, leftrightsquigarrow] & \left\{\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix}\right\}
\end{tikzcd}$$

$$\begin{tikzcd}
M_{m\times n}(\mathbb{R}) \arrow[r, equal, "\sim"] \arrow[d, equal] & \mathbb{R}^{mn} \arrow[d, equal] \\
\left\{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}\right\} \arrow[r, leftrightsquigarrow] & \left\{\begin{bmatrix} a_{11} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{m1} \\ \vdots \\ a_{mn} \end{bmatrix}\right\}
\end{tikzcd}$$

In more concrete terms, addition and scalar multiplication can be identified as the following

$$\begin{tikzcd}[column sep=small]
\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \arrow[d, leftrightsquigarrow] & + & \begin{bmatrix} -1 & 0 \\ 2 & 3 \end{bmatrix} \arrow[d, leftrightsquigarrow] & = & \begin{bmatrix} 0 & 0 \\ 4 & 6 \end{bmatrix} \arrow[d, leftrightsquigarrow] \\
\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} & + & \begin{bmatrix} -1 \\ 0 \\ 2 \\ 3 \end{bmatrix} & = & \begin{bmatrix} 0 \\ 0 \\ 4 \\ 6 \end{bmatrix}
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small]
2 \cdot & \begin{bmatrix} -1 & 0 \\ 2 & 3 \end{bmatrix} \arrow[d, leftrightsquigarrow] & = & \begin{bmatrix} -2 & 0 \\ 4 & 6 \end{bmatrix} \arrow[d, leftrightsquigarrow] \\
2 \cdot & \begin{bmatrix} -1 \\ 0 \\ 2 \\ 3 \end{bmatrix} & = & \begin{bmatrix} -2 \\ 0 \\ 4 \\ 6 \end{bmatrix}
\end{tikzcd}$$

So we may conclude that addition and scalar multiplication in $M_{m\times n}(\mathbb{R})$ can be identically translated to addition and scalar multiplication in $\mathbb{R}^{mn}$

Remark. We call $\{E_{ij}\}$ the *standard basis* of $M_{m\times n}(\mathbb{R})$, corresponding to the standard basis for $\mathbb{R}^{mn}$. Here E_{ij} is the $m \times n$ matrix that only has a single 1 in the (i, j)-th spot, but 0's elsewhere.

Example 9.4. $\left\{ E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is the standard basis for $M_{2\times 2}(\mathbb{R})$, and

$$\begin{aligned}
\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} &= \begin{bmatrix} a_{11} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & a_{12} \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ a_{21} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & a_{22} \end{bmatrix} \\
&= a_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_{12} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_{21} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\
&= a_{11}E_{11} + a_{12}E_{12} + a_{21}E_{21} + a_{22}E_{22}
\end{aligned}$$

Definition 9.5. A (real) vector space is a set V of objects, called *vectors*, on which are defined two operations, called *addition* $\boldsymbol{+}$ and *(left) scalar multiplication* $\bullet$, subject to axioms

0. $\boldsymbol{u} \boldsymbol{+} \boldsymbol{v}$ and $c \bullet \boldsymbol{v}$ are still in V
1. $\boldsymbol{u} \boldsymbol{+} \boldsymbol{v} = \boldsymbol{v} \boldsymbol{+} \boldsymbol{u}$
2. $(\boldsymbol{u} \boldsymbol{+} \boldsymbol{v}) \boldsymbol{+} \boldsymbol{w} = \boldsymbol{u} \boldsymbol{+} (\boldsymbol{v} \boldsymbol{+} \boldsymbol{w})$
3. There is a *zero vector* $\mathbf{0}$ such that $\boldsymbol{u} \boldsymbol{+} \mathbf{0} = \boldsymbol{u}$
4. For each $\boldsymbol{u}$ in V, there is a vector $\boldsymbol{-u}$ in V such that $\boldsymbol{u} \boldsymbol{+} (\boldsymbol{-u}) = \mathbf{0}$
5. $c \bullet (\boldsymbol{u} \boldsymbol{+} \boldsymbol{v}) = c \bullet \boldsymbol{u} \boldsymbol{+} c \bullet \boldsymbol{v}$
6. $(c + d) \bullet \boldsymbol{u} = c \bullet \boldsymbol{u} \boldsymbol{+} d \bullet \boldsymbol{u}$
7. $c \bullet (d \bullet \boldsymbol{u}) = (cd) \bullet \boldsymbol{u}$
8. $1 \bullet \boldsymbol{u} = \boldsymbol{u}$

Example 9.6. Set V to be $\mathbb{R}^n$, $\boldsymbol{+}$ to be addition $+$ for vectors, $\bullet$ to be scalar multiplication $\cdot$ for vectors, then this is a vector space

Example 9.7 (non-example)**.** Suppose $V = \mathbb{R}$, $a \boldsymbol{+} b = a + b + 1$, $c \bullet a = c \cdot a = ca$, we can check

0. $a \boldsymbol{+} b = a + b + 1 \in \mathbb{R}$, $c \bullet a = ca \in \mathbb{R}$
1. $a \boldsymbol{+} b = a + b + 1 = b + a + 1 = b \boldsymbol{+} a$
2. $(a \boldsymbol{+} b) \boldsymbol{+} c = (a + b + 1) + c + 1 = a + (b + c + 1) + 1 = a \boldsymbol{+} (b \boldsymbol{+} c)$
3. There is a *zero vector* $\mathbf{0} = -1$ such that $a \boldsymbol{+} \mathbf{0} = a + (-1) + 1 = a$
4. For each a, we have $\boldsymbol{-}a = -a - 2$ such that $a \boldsymbol{+} (\boldsymbol{-}a) = a + (-a - 2) + 1 = -1 = \mathbf{0}$

However $2 \bullet (a \boldsymbol{+} b) = 2(a + b + 1) \neq 2a + 2b + 1 = 2 \bullet a \boldsymbol{+} 2 \bullet b$. Therefore, this is not a vector space

9.2 Subspace

Definition 9.8. Suppose V is a vector space with addition $\boldsymbol{+}$ and scalar multiplication $\bullet$. A subspace H is a non-empty subset which closed under addition and scalar multiplication, i.e. for any $\boldsymbol{u}, \boldsymbol{v} \in H$, $c \in \mathbb{R}$, $\boldsymbol{u} \boldsymbol{+} \boldsymbol{v}, c \bullet \boldsymbol{u} \in H$

Remark. It is easy to check that a subspace H is again a vector space.

Exercise 9.9. Recall that $M_{2\times 2}(\mathbb{R})$ is the set of 2 by 2 matrices, and that a square matrix A is *symmetric* if $A^T = A$. Consider a subset V consists of 2 by 2 symmetric matrices, i.e. $V = \left\{A \in M_{2\times 2}(\mathbb{R}) \middle| A^T = A\right\}$

1. Show that V is a vector space.
2. Find a basis for V.

Solution.

1. For any $A, B \in V$, $c \in \mathbb{R}$, by definition we know that $A^T = A$, $B^T = B$, we want to show that $A + B \in V$, $cA \in V$ (condition for subspace), i.e. $(A + B)^T = A^T + B^T$, $(cA)^T = cA$. This is true because

$$(A + B)^T = A^T + B^T = A + B, \qquad (cA)^T = cA^T = cA$$

Therefore V is a subspace of $M_{2\times 2}(\mathbb{R})$, and thus a vector space

2. Suppose $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$, then $a_{12} = a_{21}$, so we may conclude that

$$V = \left\{ \begin{bmatrix} a & b \\ b & c \end{bmatrix} \in M_{2\times 2}(\mathbb{R}) \middle| a, b, c \in \mathbb{R} \right\}$$

Note that

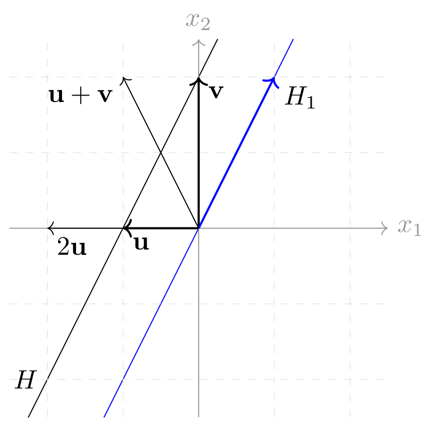
$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \\ b & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{9.1}$$

And that linear combination (9.1) is the zero matrix $\iff$ $a = b = c = 0$, thus $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for V

□

Exercise 9.10. Suppose $H = \{p(t) = a_0 + a_1 t + a_2 t^2 \in \mathbb{P}_2 | a_0 + a_1 + a_2 = 0\}$ is the set of polynomials of degree ≤ 2 and sum of coefficients zero. Show that H is a vector space.

Example 9.11. Consider the vector space $V = \mathbb{R}^2$, and H is the set of solutions to the linear equation $2x_1 - x_2 + 2 = 0$, then H is not a subspace. For example, if we choose $\mathbf{u} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, then $\mathbf{u} + \mathbf{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ is not in H, nor is $2\mathbf{u} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$



The reason is that H is not homogeneous. If we consider H_1 to be solution set of the homogeneous equation $2x_1 - x_2 = 0$, we see that H_1 is a subspace as it is the span of a single vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

10 Lecture 10 - Null spaces, Column spaces, Row spaces, Rank and Nullity

10.1 Null space, Column Space and Row space

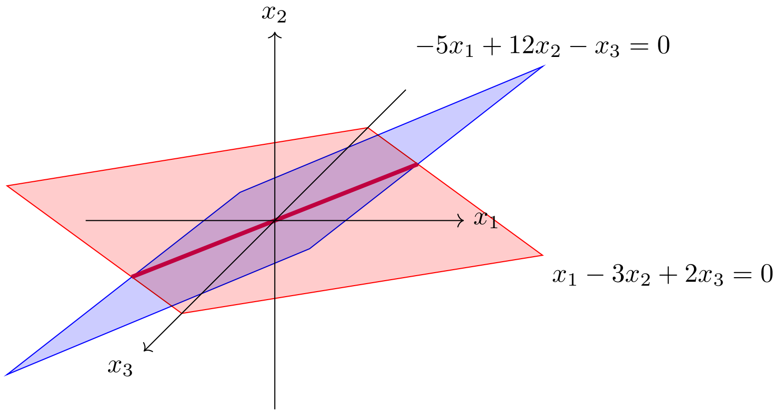
Definition 10.1. Suppose V is a vector space, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\} \subseteq V$ is a set of linearly independent vectors that span V, we define the dimension of V to be $\dim V = n$.

Definition 10.2. Suppose A is a $m \times n$ matrix, we define the null space of A to be $\operatorname{Nul} A = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{0}\}$. Note that the solution set to linear system $A\mathbf{x} = \mathbf{0}$ is the intersection of m hyperplanes (one for each homogeneous equation) that pass through the origin.

Example 10.3. $A = \begin{bmatrix} 1 & -3 & 2 \\ -5 & 12 & -1 \end{bmatrix}$, the to find the $\operatorname{Nul} A$ is equivalent to solve $A\mathbf{x} = \mathbf{0}$

$$\begin{bmatrix} A & \mathbf{0} \end{bmatrix} \xrightarrow{R2 \to R2 + 5R1} \begin{bmatrix} 1 & -3 & 2 & 0 \\ 0 & -3 & 9 & 0 \end{bmatrix} \xrightarrow{R2/(-3)} \begin{bmatrix} 1 & -3 & 2 & 0 \\ 0 & 1 & -3 & 0 \end{bmatrix} \xrightarrow{R1 \to R1 + 3R2} \begin{bmatrix} 1 & 0 & -7 & 0 \\ 0 & 1 & -3 & 0 \end{bmatrix}$$

Hence the solution set is $\begin{cases} x_1 = 7x_3 \\ x_2 = 3x_3 \\ x_3 \text{ is free} \end{cases}$, in parametric form, $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix}$, which describes a line in $\mathbb{R}^3$ that passes through the origin, and this line is the intersection of planes $x_1 - 3x_2 + 2x_3 = 0$ and $-5x_1 + 12x_2 - x_3 = 0$



Remark. As discussed in Example 9.11, in general, the solution set of $A\mathbf{x} = \mathbf{b}$ is not a subspace of $\mathbb{R}^n$ unless $\mathbf{b} = \mathbf{0}$. And in fact, any subspace of $\mathbb{R}^n$ is the null space for some $m \times n$ matrix A, i.e. the intersection of hyperplanes passing through the origin

Definition 10.4. Suppose $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ is an $m \times n$ matrix, then the column space (denote as $\operatorname{Col} A$) is the subspace $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$ in $\mathbb{R}^m$. Suppose $A = \begin{bmatrix} R1 \\ R2 \\ \vdots \\ Rm \end{bmatrix}$, then the row space (denote as $\operatorname{Row} A$) is the subspace spaned by row vectors $\operatorname{Span}\{R1, R2, \cdots, Rm\}$ in $\mathbb{R}^n$ written horizontally.

Remark. Suppose column vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ in A has some linear dependence $A\mathbf{x} = \mathbf{0}$, then, after elementary row reduction $A \sim EA$, $(EA)\mathbf{x} = E(A\mathbf{x}) = E\mathbf{0} = \mathbf{0}$ again has the same linear dependence. In other words, the linear dependence of columns of A is preserved by row equivalence.

Remark. Since row elementary operations can be reversed, so $\operatorname{Row} A$ is preserved under row equivalence

We may conclude the following algorithm for finding basis for $\operatorname{Nul} A$, $\operatorname{Col} A$, $\operatorname{Row} A$ by simply use elementary row reductions!!!

Theorem 10.5. Suppose A is a $m \times n$ matrix, $A \sim U$ is of RREF form

- The solution set of $\begin{bmatrix} U & \mathbf{0} \end{bmatrix}$ in parametric vector form gives a basis for $\operatorname{Nul} A$. Note that $\dim \operatorname{Nul} A =$ the number of free variables.
- A basis for $\operatorname{Col} A$ could be the set of pivot columns in A. Note that $\dim \operatorname{Col} A =$ the number of pivots
- A basis for $\operatorname{Row} A$ could be the set of non-zero row vectors in U (Or any REF of A actually). Note that $\dim \operatorname{Row} A =$ the number of pivots

10.2 Rank and Nullity

Definition 10.6. $\dim \operatorname{Nul} A$ is also name the nullity of A. The number of pivots of A (which is equal to both $\dim \operatorname{Col} A$ and $\dim \operatorname{Row} A$) is called the rank of A

Theorem 10.7 (Rank-Nullity theorem)**.** Notice that the number of columns in A (say a $m \times n$ matrix) is equal to the number of free variables and the number of pivot columns, thus we have

$$n = \text{nullity} + \text{rank}$$

Example 10.8. $A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 1 \\ 0 & 2 & -3 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 \end{bmatrix}$, which is an REF and an RREF respectively. There is only one free variable x_3, so the nullity is 1, and the 1st, 2nd columns are pivot columns, so the rank is 2. We see that Theorem 10.7 holds as $3 = 1 + 2$, and

$$\text{Nul}\, A = \text{Span}\left\{ \begin{bmatrix} \frac{1}{2} \\ \frac{3}{2} \\ 1 \end{bmatrix} \right\}$$

$$\text{Col}\, A = \text{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$\text{Row}\, A = \text{Span}\left\{ \begin{bmatrix} 1 & 0 & -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} 0 & 1 & -\frac{3}{2} \end{bmatrix} \right\} \text{ or } \text{Span}\left\{ \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & -3 \end{bmatrix} \right\}$$

Exercise 10.9. $A = \begin{bmatrix} 1 & 2 & 0 & 4 & 5 \\ 0 & 0 & 5 & -7 & 8 \\ 0 & 0 & 0 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & 4 & 0 \\ 0 & 0 & 1 & -\frac{7}{5} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ Note that here we have 2 free variables x_2, x_4, so the nullity is 2, and the 1st, 3rd, 5th columns are pivot columns, so the rank is 3. We see that Theorem 10.7 holds as $5 = 2 + 3$, and

$$\text{Nul}\, A = \text{Span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ \frac{7}{5} \\ 1 \end{bmatrix} \right\}$$

$$\text{Col}\, A = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 8 \\ -9 \\ 0 \end{bmatrix} \right\}$$

$$\text{Row}\, A = \text{Span}\left\{ \begin{bmatrix} 1 & 2 & 0 & 4 & 5 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 5 & -7 & 8 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 & -9 \end{bmatrix} \right\}$$

Question. If you have a set $\mathcal{S}$ of vectors in $\mathbb{R}^m$, how do you find a subset of $\mathcal{S}$ that is a basis for $\text{Span}\{\mathcal{S}\}$ (i.e. remove linear dependences)?

Answer. Collect these vectors as the column vectors of a matrix, and then find its columns space.

Exercise 10.10. Suppose $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$. Find a basis for $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

Exercise 10.11. Suppose $A = \begin{bmatrix} 1 & -7 & 0 & 6 & 5 \\ 0 & 0 & 1 & -2 & -3 \\ -1 & 7 & -4 & 2 & 7 \end{bmatrix}$. Find $\text{Nul}\, A, \text{Row}\, A, \text{Col}\, A$, and the nullity, rank of A.

11 Lecture 11 - Linear transformations in general

11.1 Linear transformation

Definition 11.1. Suppose V, W are vector spaces, a *linear transformation* is a mapping $T : V \to W$ such that

- $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$

- $T(c \bullet u) = c \bullet T(\boldsymbol{u})$

Just as before, we call V the *domain* of T, W the *codomain* of V, the *image* of $\boldsymbol{u}$ under T is $T(\boldsymbol{u})$, the set of images $\{T(\boldsymbol{u}) | \boldsymbol{u} \in V\}$ the *range* (denoted as $\operatorname{Range} T$), and the set $\{\boldsymbol{u} | T(\boldsymbol{w}) = \boldsymbol{u}\}$ the preimage of $\boldsymbol{w}$ under T. We still say that T is *one-to-one* if any $\boldsymbol{w} \in W$, there is at most one $\boldsymbol{u} \in V$ such that $T(\boldsymbol{u}) = \boldsymbol{w}$. T is *onto* the range is the codomain. T is said to be *invertible* if T has a inverse (this happens if and only if T is both one-to-one and onto), in this case we also call T an isomorphism.

Definition 11.2. We call $\{\boldsymbol{u} | T(\boldsymbol{u}) = \boldsymbol{0}\}$ the kernel (or null space) of T

Remark. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(\mathbf{x}) = A\mathbf{x}$ is a matrix transformation, then $\ker T = \operatorname{Nul} A$, $\operatorname{Range} T = \operatorname{Col} A$

Example 11.3. The identification

$$T : \mathbb{P}_2 \to \mathbb{R}^3, \qquad T(a_0 + a_1 t + a_2 t^2) = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

in Eaxmple 9.1 is an invertible linear transformation with inverse linear transformation

$$T^{-1} : \mathbb{R}^3 \to \mathbb{P}_2, \qquad T^{-1}\left(\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}\right) = a_0 + a_1 t + a_2 t^2$$

Example 11.4. The identification

$$T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}^4, \qquad T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

in Eaxmple 9.3 is an invertible linear transformation with inverse linear transformation

$$T^{-1} : \mathbb{R}^4 \to M_{2\times 2}(\mathbb{R}), \qquad T^{-1}\left(\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Theorem 11.5. Suppose $T : V \to W$ is a linear transformation between vector spaces, then

- $\ker T$ is a subspace of V.
- $\operatorname{Range} T$ is a subspace of W.

Proof.

- Suppose $\boldsymbol{u}, \boldsymbol{v} \in \ker T$, then by definition $T(\boldsymbol{u}) = \boldsymbol{0}$, $T(\boldsymbol{v}) = \boldsymbol{0}$, so $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v}) = \boldsymbol{0}$. And for any $c \in \mathbb{R}$, $T(c \bullet \boldsymbol{u}) = c \bullet T(\boldsymbol{u}) = \boldsymbol{0}$. In other words, we have shown that $\boldsymbol{u} + \boldsymbol{v}, c \bullet \boldsymbol{u} \in \ker T$, so $\ker T$ is a subspace.
- For any $T(\boldsymbol{u}), T(\boldsymbol{v}) \in \operatorname{Range} T$, $T(\boldsymbol{u}) + T(\boldsymbol{v}) = T(\boldsymbol{u} + \boldsymbol{v}) \in \operatorname{Range} T$, and for any $c \in \mathbb{R}$, $c \bullet T(\boldsymbol{u}) = T(c \bullet \boldsymbol{u}) \in \operatorname{Range} T$, so $\operatorname{Range} T$ is a subspace

□

Exercise 11.6. Suppose $T : \mathbb{P}_2 \to \mathbb{R}$ takes the sum of coefficients, i.e. $T(a_0 + a_1 t + a_2 t^2) = a_0 + a_1 + a_2$. Show that T is a linear transformation, and H in Exercise 9.10 is a vector space.

Solution. since for any $p(t) = a_0 + a_1 + a_2, q(t) = b_0 + b_1 t + b_2 t^2 \in \mathbb{P}_2$, $c \in \mathbb{R}$, we have

$$\begin{aligned} T(p+q) &= T((a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2) = (a_0 + b_0) + (a_1 + b_1) + (a_2 + b_2) \\ &= (a_0 + a_1 + a_2) + (b_0 + b_1 + b_2) = T(p) + T(q) \end{aligned}$$

$$T(cp) = T((ca_0) + (ca_1)t + (ca_2)t^2) = (ca_0) + (ca_1) + (ca_2) = c(a_0 + a_1 + a_2) = cT(p)$$

So $V = \ker T$ is a subspace of V by Theorem 11.5 □

Exercise 11.7. Suppose $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}$ takes the sum of diagonal, i.e. $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a+d$. Show that T is a linear transformation, and the set $H = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a+d=0 \right\}$ of 2×2 matrix with trace(sum of diagonal element) 0 is a vector space.

11.2 Matrices of linear transformations

Question. Can we realize a general linear transformation T as a matrix transformation?

Answer. Just need to know its effect on any basis! Suppose $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is a basis of $\mathbb{R}^n$, then any vector $\mathbf{v}$ can be written as $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, and by linearality, we have

$$T(\mathbf{v}) = T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n) \tag{11.1}$$

Theorem 11.8 (Unique representation theorem)**.** Suppose $\mathcal{B} = \{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ is a basis of a vector sapce V, then any vector $\boldsymbol{v} \in V$ can be uniquely represented as a linear combination $x_1 \bullet \boldsymbol{b}_1 \boldsymbol{+} \cdots \boldsymbol{+} x_n \bullet \boldsymbol{b}_n$

Remark. Here $[\boldsymbol{v}]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is called the $\mathcal{B}$-coordinate vector (or coordinate vector relative to the basis $\mathcal{B}$) of $\boldsymbol{v}$

Proof. Suppose $\boldsymbol{v} = c_1 \bullet \boldsymbol{b}_1 \boldsymbol{+} \cdots \boldsymbol{+} c_n \bullet \boldsymbol{b}_n = d_1 \bullet \boldsymbol{b}_1 \boldsymbol{+} \cdots \boldsymbol{+} d_n \bullet \boldsymbol{b}_n$ are two linear combinations that express $\boldsymbol{v}$, then we have $(c_1 - d_1) \bullet \boldsymbol{b}_1 \boldsymbol{+} \cdots \boldsymbol{+} (c_n - d_n) \bullet \boldsymbol{b}_n = \mathbf{0}$, since $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ is linearly independent, we have $c_1 = d_1, \cdots, c_n = d_n$. Therefore the expression is unique. ☐

Definition 11.9. We call $[\quad]_{\mathcal{B}} : V \to \mathbb{R}^n$ the coordinate mapping

Theorem 11.10. The coordinate mapping $[\quad]_{\mathcal{B}} : V \to \mathbb{R}^n$ is an isomorphism

Example 11.11. The identifications in Example 9.1 and Example 9.3 are coordinate mappings $[\quad]_{\mathcal{E}} : \mathbb{P}_2 \to \mathbb{R}^3$ with standard basis $\mathcal{E} = \{1, t, t^2\}$ and $[\quad]_{\mathcal{E}} : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}^4$ with standard basis $\mathcal{E} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$, respectively

Example 11.12. $\mathcal{B} = \left\{ \mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$, $\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$. Solve linear system $\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ we get $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$

Remark. $\mathbf{x} = [\mathbf{x}]_{\mathcal{E}}$ as $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$.

Exercise 11.13.

Question. How should we study linear transformations via matrices in general?

Assume $T : V \to W$ is a linear transformation between vector spaces, $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ is a basis for V and $\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_m\}$ is a basis for W, then for any

$$\boldsymbol{v} = x_1 \bullet \boldsymbol{b}_1 \boldsymbol{+} \cdots \boldsymbol{+} x_n \bullet \boldsymbol{b}_n = \begin{bmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_n \end{bmatrix} [\boldsymbol{v}]_{\mathcal{B}}$$

We have

$$T(\boldsymbol{v}) = T(x_1 \bullet \boldsymbol{b}_1 \boldsymbol{+} \cdots \boldsymbol{+} x_n \bullet \boldsymbol{b}_n) = x_1 \bullet T(\boldsymbol{b}_1) \boldsymbol{+} \cdots \boldsymbol{+} x_n \bullet T(\boldsymbol{b}_n) = x_1 \bullet T(\boldsymbol{b}_1) \boldsymbol{+} \cdots \boldsymbol{+} x_n \bullet T(\boldsymbol{b}_n)$$

$$= \begin{bmatrix} T(\boldsymbol{b}_1) & \cdots & T(\boldsymbol{b}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} T(\boldsymbol{b}_1) & \cdots & T(\boldsymbol{b}_n) \end{bmatrix} [\boldsymbol{v}]_{\mathcal{B}}$$

By Theorem 11.8, we can write

$$\begin{aligned}
T(\boldsymbol{b}_1) &= a_{11} \bullet \boldsymbol{c}_1 + a_{21} \bullet \boldsymbol{c}_2 + \cdots + a_{m1} \bullet \boldsymbol{c}_m \\
T(\boldsymbol{b}_2) &= a_{12} \bullet \boldsymbol{c}_1 + a_{22} \bullet \boldsymbol{c}_2 + \cdots + a_{m2} \bullet \boldsymbol{c}_m \\
&\vdots \\
T(\boldsymbol{b}_n) &= a_{1n} \bullet \boldsymbol{c}_1 + a_{2n} \bullet \boldsymbol{c}_2 + \cdots + a_{mn} \bullet \boldsymbol{c}_m
\end{aligned}$$

Therefore we have

$$\begin{bmatrix} T(\boldsymbol{b}_1) & \cdots & T(\boldsymbol{b}_n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m \end{bmatrix} A \tag{11.2}$$

Where

$$A = \begin{bmatrix} [T(\boldsymbol{b}_1)]_{\mathcal{C}} & \cdots & [T(\boldsymbol{b}_n)]_{\mathcal{C}} \end{bmatrix} \tag{11.3}$$

is called the matrix of T relative to bases $\mathcal{B}$ and $\mathcal{C}$, thus

$$T(\boldsymbol{v}) = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m \end{bmatrix} A[\boldsymbol{v}]_{\mathcal{B}}$$

On the other hand, we should have

$$T(\boldsymbol{v}) = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m \end{bmatrix} [T(\boldsymbol{v})]_{\mathcal{C}}$$

So we may also conclude that

$$[T(\boldsymbol{v})]_{\mathcal{C}} = A[\boldsymbol{v}]_{\mathcal{B}} \tag{11.4}$$

The above discussion can be summerized by the following commutative diagram

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "{[\ \ ]_{\mathcal{B}}}"'] & W \arrow[d, "{[\ \ ]_{\mathcal{C}}}"] \\ \mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^m \end{tikzcd} \qquad \begin{tikzcd} \boldsymbol{v} \arrow[r, mapsto] \arrow[d, mapsto] & T(\boldsymbol{v}) \arrow[d, mapsto] \\ {[\boldsymbol{v}]_{\mathcal{B}}} \arrow[r, mapsto] & A[\boldsymbol{v}]_{\mathcal{B}} = [T(\boldsymbol{v})]_{\mathcal{C}} \end{tikzcd} \tag{11.5}$$

Remark. The philosophy here is that any statement about a general linear transformation can be converted to a corresponding statement about matrix transformation.

Example 11.14. Suppose $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ with bases $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$, $\{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$, then (11.2) reads $A = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$ is the standard matrix

Example 11.15. Consider linear transformation

$$T : \mathbb{P}_2 \to \mathbb{R}^3, \qquad T(a_0 + a_1 t + a_2 t^2) = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

and $\mathcal{B} = \{1 + t, t + t^2, 1 + t^2\}$ is a basis for $\mathbb{P}_2$, $\mathcal{E}$ is the standard basis for $\mathbb{R}^3$, then matrix of T relative to bases $\mathcal{B}$ and $\mathcal{E}$ can be read from (11.3) as

$$A = \begin{bmatrix} T(1+t) & T(t+t^2) & T(1+t^2) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Example 11.16. Consider linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ x_2 \end{bmatrix}$, $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$, $\mathcal{C} = \left\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}\right\}$ are bases for $\mathbb{R}^2$. Then the matrix of T relative to bases $\mathcal{B}$ and $\mathcal{C}$ can be read from (11.2)

$$\begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} A = \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix} \Rightarrow A = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ -2 & 1 \end{bmatrix}$$

Exercise 11.17.

Remark. Suppose $T : V \to W$ and $S : W \to U$ are linear transformations between vector spaces, $\mathcal{B} = \{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ is a basis for V, $\mathcal{C} = \{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_n\}$ is a basis for W and $\mathcal{D} = \{\boldsymbol{d}_1, \cdots, \boldsymbol{d}_n\}$ is a basis for U, then the matrix of $S \circ T$ relative to bases $\mathcal{B}$, $\mathcal{D}$ is BA.

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "{[\ \ ]_{\mathcal{B}}}"'] \arrow[rr, bend left, "S\circ T"] & W \arrow[r, "S"] \arrow[d, "{[\ \ ]_{\mathcal{C}}}"] & U \arrow[d, "{[\ \ ]_{\mathcal{D}}}"] \\
\mathbb{R}^n \arrow[r, "A"] \arrow[rr, bend right, "BA"'] & \mathbb{R}^m \arrow[r, "B"] & \mathbb{R}^p
\end{tikzcd}
\qquad
\begin{tikzcd}
\boldsymbol{v} \arrow[r, mapsto] \arrow[d, mapsto] & T(\boldsymbol{v}) \arrow[r] \arrow[d, mapsto] & S(T(\boldsymbol{v})) = (S \circ T)(\boldsymbol{v}) \arrow[d, mapsto] \\
{[\boldsymbol{v}]_{\mathcal{B}}} \arrow[r, mapsto] & A[\boldsymbol{v}]_{\mathcal{B}} = [T(\boldsymbol{v})]_{\mathcal{C}} \arrow[r, mapsto] & BA[\boldsymbol{v}]_{\mathcal{B}} = [S(T(\boldsymbol{v}))]_{\mathcal{D}}
\end{tikzcd}$$

If T is invertible, then the matrix of T^{-1} relative to bases $\mathcal{C}$, $\mathcal{B}$ is A^{-1}

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "{[\ \ ]_{\mathcal{B}}}"'] & W \arrow[l, bend left, "T_{-1}"] \arrow[d, "{[\ \ ]_{\mathcal{C}}}"] \\
\mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^m \arrow[l, bend left, "A^{-1}"]
\end{tikzcd}$$

Exercise 11.18. Suppose $T : \mathbb{P}_2 \to \mathbb{R}_3$ is evaluation at 2, i.e. $T(p(t)) = p(2)$ for $p(t) \in \mathbb{P}_2$. Show that T is a linear transformation. Is T onto? Is T one-to-one? Is T invertible? Find a basis for $\ker T$. Find a basis for $\operatorname{Range} T$.

Exercise 11.19. Suppose H is the set of 2 by 2 matrices that are symmetric with the sum of diagonal being zero. Show H is a vector space. Find a basis of H.

11.3 Change of basis

Now let's talk about change of basis. Suppose V is a vector space with two basis $\mathcal{B} = \{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ and $\mathcal{C} = \{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_n\}$, and $\mathrm{id}_V : V \to V$, $\mathrm{id}_V(\boldsymbol{v}) = \boldsymbol{v}$ is the identity mapping. Diagram (11.5) becomes

$$\begin{tikzcd}
V \arrow[r, equal, "\mathrm{id}_V"] \arrow[d, "{[\ \ ]_{\mathcal{B}}}"'] & V \arrow[d, "{[\ \ ]_{\mathcal{C}}}"] \\
\mathbb{R}^n \arrow[r, "\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}"] & \mathbb{R}^n
\end{tikzcd} \tag{11.6}$$

Where equation (11.3) and equation (11.4) become

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [\boldsymbol{b}_1]_{\mathcal{C}} & \cdots & [\boldsymbol{b}_n]_{\mathcal{C}} \end{bmatrix}$$

$$[\boldsymbol{v}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\boldsymbol{v}]_{\mathcal{B}}$$

which is the matrix of id_V relative to basis $\mathcal{B}$ and $\mathcal{C}$, and we call this the change of coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$. Also remark 11.2 gives us

$$\underset{\mathcal{D}\leftarrow\mathcal{B}}{P} = \left(\underset{\mathcal{D}\leftarrow\mathcal{C}}{P}\right)\left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right), \quad \left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)^{-1} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}$$

Example 11.20. Continue Example 9.9, we have shown that $\mathcal{B} = \left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\} = \{B_1, B_2, B_3\}$ is a basis for V. Let's show that $\mathcal{C} = \left\{\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\1 & 1\end{bmatrix}\right\} = \{C_1, C_2, C_3\}$ is another basis for V. First note that

$$C_1 = \begin{bmatrix}1 & 1\\1 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix} = B_1 + B_2$$

$$C_2 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} = B_1 + B_3$$

$$C_3 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = B_2 + B_3$$

So $\{[C_1]_{\mathcal{B}}, [C_2]_{\mathcal{B}}, [C_3]_{\mathcal{B}}\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$, and it is easy to show that this is a basis for $\mathbb{R}^3$, hence $\mathcal{C}$ is a basis for V. Also we know that

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} [C_1]_{\mathcal{B}} & [C_2]_{\mathcal{B}} & [C_3]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

And

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \left(\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}\right)^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Note that here the diagram (11.6) becomes

$$\begin{tikzcd} V \arrow[r, equal, "\mathrm{id}_V"] \arrow[d, "{[\ \ ]_{\mathcal{B}}}"'] & V \arrow[d, "{[\ \ ]_{\mathcal{C}}}"] \\ \mathbb{R}^3 \arrow[r, "\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}"] & \mathbb{R}^3 \end{tikzcd}$$

Example 11.21. Continue Example 11.6, let's find a basis for $V = \ker T$ which is supposed to correspond to $\operatorname{Nul} A$, where A is the matrix relative to both standard bases $\mathcal{E} = \{1, t, t^2\}$ for $\mathbb{P}_2$ and $\mathcal{E} = \left\{ \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ which can be read from (11.4)

$$A = \begin{bmatrix} [T(1)]_{\mathcal{E}} & [T(t)]_{\mathcal{E}} & [T(t^2)]_{\mathcal{E}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

So we get

$$\operatorname{Nul} A = \operatorname{Span} \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Which gives a basis $\{-1 + t, -1 + t^2\}$ for V. Note that here the diagram (11.5) becomes

$$\begin{tikzcd} \mathbb{P}_2 \arrow[r, "T"] \arrow[d, "{[\ \ ]_{\mathcal{E}}}"'] & \mathbb{R} \arrow[d, equal, "{[\ \ ]_{\mathcal{E}}}"] \\ \mathbb{R}^3 \arrow[r, "A"] & \mathbb{R} \end{tikzcd}$$

An algorithm for computing $A^{-1}B$

$$\left[\ A \mid B\ \right] \sim \left[\ I \mid A^{-1}B\ \right]$$

Exercise 11.22. Suppose $\mathcal{B} = \{2t^2 - 1, 3t + 1 - t^2, 3 - t\}$ and $\mathcal{C} = \{1 + t, t^2, -t\}$ are both bases for $\mathbb{P}_2$. Please find the change of basis matrix form $\mathcal{B}$ to $\mathcal{C}$

Solution. First let's find the change of basis matrices from $\mathcal{B}$ and $\mathcal{C}$ to the standard basis $\mathcal{E} = \{1, t, t^2\}$. We have

$$\underset{\mathcal{E}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [-1 + 2t^2]_{\mathcal{E}} & [1 + 3t - t^2]_{\mathcal{E}} & [3 - t]_{\mathcal{E}} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 3 \\ 0 & 3 & -1 \\ 2 & -1 & 0 \end{bmatrix}$$

And

$$\underset{\mathcal{E}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} [1 - t]_{\mathcal{E}} & [t]_{\mathcal{E}} & [t^2]_{\mathcal{E}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

Hence we have

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \left(\underset{\mathcal{C}\leftarrow\mathcal{E}}{P}\right)\left(\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}\right) = \left(\underset{\mathcal{E}\leftarrow\mathcal{C}}{P}\right)^{-1}\left(\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}\right)$$

Which can computed via the above algorithm

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & 3 \\ 1 & 0 & -1 & 0 & 3 & -1 \\ 0 & 1 & 0 & 2 & -1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & 3 \\ 0 & 1 & 0 & 2 & -1 & 0 \\ 0 & 0 & 1 & -1 & -2 & 4 \end{array}\right]$$

$$\begin{tikzcd} \mathbb{P}_2 \arrow[r, equal, "\mathrm{id}_V"] \arrow[d, "{[\ \ ]_\mathcal{B}}"'] & \mathbb{P}_2 \arrow[r, equal, "\mathrm{id}_V"] \arrow[d, "{[\ \ ]_\mathcal{E}}"'] & \mathbb{P}_2 \arrow[d, "{[\ \ ]_\mathcal{C}}"] \\ \mathbb{R}^3 \arrow[r, "\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}"'] \arrow[rr, bend right, "\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}"'] & \mathbb{R}^3 \arrow[r, bend left, "\underset{\mathcal{C}\leftarrow\mathcal{E}}{P}"] & \mathbb{R}^3 \arrow[l, "\underset{\mathcal{E}\leftarrow\mathcal{C}}{P}"] \end{tikzcd}$$

□

Example 11.23. Suppose $\mathbb{R}^2$ has bases $\mathcal{B} = \left\{\mathbf{b}_1 = \begin{bmatrix} 7 \\ 5 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -3 \\ -1 \end{bmatrix}\right\}$, $\mathcal{C} = \left\{\mathbf{c}_1 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}\right\}$. $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \left(\underset{\mathcal{E}\leftarrow\mathcal{C}}{P}\right)^{-1}\left(\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}\right)$, which can be computed via

$$\left[\begin{array}{cc|cc} 1 & -2 & 7 & -3 \\ -5 & 2 & 5 & -1 \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & -3 & 1 \\ 0 & 1 & -5 & 2 \end{array}\right]$$

So $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} -3 & 1 \\ -5 & 2 \end{bmatrix}$

$$\begin{tikzcd} \mathbb{R}^2 \arrow[r, equal] \arrow[d, "{[\ \ ]_\mathcal{B}}"'] & \mathbb{R}^2 \arrow[r, equal] \arrow[d, "{[\ \ ]_\mathcal{E}}"'] & \mathbb{R}^2 \arrow[d, "{[\ \ ]_\mathcal{C}}"] \\ \mathbb{R}^2 \arrow[r, "\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}"'] \arrow[rr, bend right, "\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}"'] & \mathbb{R}^2 \arrow[r, bend left, "\underset{\mathcal{C}\leftarrow\mathcal{E}}{P}"] & \mathbb{R}^2 \arrow[l, "\underset{\mathcal{E}\leftarrow\mathcal{C}}{P}"] \end{tikzcd}$$

Theorem 11.24. Let's generalize the rank-nullity theorem 10.7, with the remark 11.1, we have

$$\dim V = \dim \operatorname{Range} T + \dim \ker T$$

Definition 11.25. Suppose $T : V \to V$ is a linear transformation, and $\mathcal{B} = \{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ is a basis for V, we can the matrix relative to $\mathcal{B}$ and $\mathcal{B}$ the $\mathcal{B}$-matrix of T (denoted $[T]_\mathcal{B}$), i.e. (11.5) becomes

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "{[\ \ ]_\mathcal{B}}"'] & V \arrow[d, "{[\ \ ]_\mathcal{B}}"] \\ \mathbb{R}^n \arrow[r, "{[T]_\mathcal{B}}"] & \mathbb{R}^n \end{tikzcd} \qquad \begin{tikzcd} \boldsymbol{v} \arrow[r, mapsto] \arrow[d, mapsto] & T(\boldsymbol{v}) \arrow[d, mapsto] \\ {[\boldsymbol{v}]_\mathcal{B}} \arrow[r, mapsto] & {[T]_\mathcal{B}[\boldsymbol{v}]_\mathcal{B} = [T(\boldsymbol{v})]_\mathcal{B}} \end{tikzcd}$$

And (11.3) gives $[T]_\mathcal{B} = \begin{bmatrix} [T(\boldsymbol{b}_1)]_\mathcal{B} & \cdots & [T(\boldsymbol{b}_n)]_\mathcal{B} \end{bmatrix}$ Suppose both $\mathcal{B} = \{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ and $\mathcal{C} = \{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_n\}$ form bases for V, then

$$[T]_\mathcal{B} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}[T]_\mathcal{C}\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}[T]_\mathcal{C}\left(\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}\right)^{-1}$$

is similar via

$$
\begin{tikzcd}
V \arrow[rrr, "T"] \arrow[dr, "{[\ ]_{\mathcal{B}}}"] \arrow[ddd, "\mathrm{id}"'] & & & V \arrow[dl, "{[\ ]_{\mathcal{B}}}"'] \arrow[ddd, "\mathrm{id}"] \\
& \mathbb{R}^n \arrow[r, "{[T]_{\mathcal{B}}}"] & \mathbb{R}^n \arrow[d, "\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}"'] & \\
& \mathbb{R}^n \arrow[u, "\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}"'] \arrow[r, "{[T]_{\mathcal{C}}}"] & \mathbb{R}^n & \\
V \arrow[ur, "{[\ ]_{\mathcal{C}}}"'] \arrow[rrr, "T"'] & & & V \arrow[ul, "{[\ ]_{\mathcal{C}}}"]
\end{tikzcd}
$$

If $T : \mathbb{R}^n \to \mathbb{R}^n$, $T(\mathbf{x}) = A\mathbf{x}$ is a matrix transformation, $\mathcal{E}$ is the standard basis, and $\mathcal{C} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is basis consists of eigenvectors, then $[T]_{\mathcal{C}} = D$ is diagonal with eigenvalues, as shown in Theorem 12.18.

Remark. In fact, any two similar matrix can be realized as matrices of the same linear transformation in different basis.

12 Lecture 15 - Eigendecomposition

12.1 Eigenvalues, eigenvectors and eigenspaces

Definition 12.1. Suppose A is a square $n \times n$ matrix. A λ-eigenvector of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an eigenvalue of A has λ-eigenvectors.

Definition 12.2. We say that a vector space V is trivial if $V = \{0\}$ is the zero vector space.

Question. How to decide whether a nonzero vector $\mathbf{x}$ is an eigenvector?

Answer. We can evaluate $A\mathbf{x}$ and see if it is a multiple of $\mathbf{x}$

Example 12.3. $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, determine whether $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ are eigenvectors of A.

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = (-4)\mathbf{u}$$

Hence $\mathbf{u}$ is a (-4)-eigenvector of A.

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix}$$

Which is not a multiple of $\mathbf{v}$, so $\mathbf{v}$ is not an eigenvector.

Question. How to decide if λ is an eigenvalue of A?

Answer. By definition, we know that λ is an eigenvalue of $A \iff A\mathbf{x} = \lambda\mathbf{x}$ has a nontrivial solution $\iff \lambda\mathbf{x} - A\mathbf{x} = \lambda I\mathbf{x} - A\mathbf{x} = (\lambda I - A)\mathbf{x} = \mathbf{0}$ has a nontrivial solution $\iff \mathrm{Nul}(\lambda I - A)$ is nontrivial $\iff \lambda I - A$ is invertible $\iff \det(\lambda I - A) = 0$

Example 12.4. $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. Determine whether $\lambda = 2$ is an eigenvalue of A

$$\det(2I - A) = \det\left(2\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}\right) = \begin{vmatrix} 2 & -1 & -6 \\ 2 & -1 & -6 \\ 2 & -1 & -6 \end{vmatrix} = 0$$

Definition 12.5. From above criterion, we see that the set of λ-vectors is $\mathrm{Nul}(\lambda I - A)$ which is a subspace of $\mathbb{R}^n$, we call this the λ-eigenspace.

Example 12.6. Continue Example 12.4. Let's find a basis for the 2-eigenspace of A, which is equivalent of finding a basis for $\text{Nul}(2I - A)$, since $\left[\begin{array}{c|c} 2I - A & \mathbf{0} \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & -\frac{1}{2} & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$ Hence a basis could be $\left\{ \begin{bmatrix} \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$

12.2 Characteristic polynomials

Question. How could we find all the eigenvalues of A?

Definition 12.7. From above discussion, we see that λ is an eigenvalue of $A \iff \det(\lambda I - A) = 0$, this motivates the following definition. We call $\det(tI - A)$ the characteristic polynomial of A, and $\det(tI - A) = 0$ the characteristic equation. And the roots of the characteristic polynomial are the eigenvalues of A.

Definition 12.8. We call the dimension of the λ-eigenspace ($\dim \text{Nul}(\lambda I - A)$) the geometric multiplicity of λ, and the multiplicity of λ as a root in the charateristic polynomial $\det(tI - A)$ the algebraic multiplicity of λ.

Example 12.9. Suppose $A = \begin{bmatrix} 1 & -4 \\ 4 & 2 \end{bmatrix}$, then the characteristic polynomial of A would be

$$\det(tI - A) = \det\left(t\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & -4 \\ 4 & 2 \end{bmatrix}\right) = \begin{vmatrix} t-1 & 4 \\ -4 & t-2 \end{vmatrix} = (t-1)(t-2) - 4\cdot(-4) = t^2 - 3t + 18$$

And the characteristic equation is $t^2 - 3t + 18 = 0$. Since $\Delta = (-3)^2 - 4 \cdot 1 \cdot 18 = -63 < 0$, this equation has no (real) solutions, A doesn't have (real) eigenvalues

Note. Recall that the quadratic equation $ax^2 + bx + c = 0$ has no (real) solutions $\iff$ the discriminant $\Delta = b^2 - 4ac < 0$.

Example 12.10. Suppose $A = \begin{bmatrix} 3 & -2 \\ 2 & -1 \end{bmatrix}$, then the characteristic polynomial of A would be

$$\det(tI - A) = \det\left(t\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 3 & -2 \\ 2 & -1 \end{bmatrix}\right) = \begin{vmatrix} t-3 & 2 \\ -2 & t+1 \end{vmatrix} = (t-3)(t+1) - 2\cdot(-2) = t^2 - 2t + 1$$

And the characteristic equation is $t^2 - 2t + 1 = (t-1)^2 = 0$. Hence the eigenvalues of A is 2 with algebraic multiplicity 2

Example 12.11. Suppose $A = \begin{bmatrix} 3 & -2 \\ 2 & -1 \end{bmatrix}$, then the characteristic polynomial of A would be

$$\det(tI - A) = \begin{vmatrix} t-3 & 2 & -3 \\ 0 & t+1 & 0 \\ -6 & -7 & t+4 \end{vmatrix} \xlongequal{\text{cofactor expansion across the 2nd row}} (t+1)(-1)^{2+2}\begin{vmatrix} t-3 & -3 \\ -6 & t+4 \end{vmatrix}$$
$$= (t+1)((t-3)(t+4) - (-3)\cdot(-6)) = (t+1)(t^2 + t - 30) = (t+1)(t+6)(t-5)$$

So we see that the eigenvalues of A are $-1, -6, -5$, all with algebraic multiplicities 1,1,1.

Example 12.12. Suppose $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$, then characteristic polynomial is

$$\det(tI - A) = \begin{vmatrix} t & -1 & 1 \\ -1 & t & -1 \\ -1 & -1 & t \end{vmatrix} \xlongequal[R2\to R2-R3]{R2\to R1+t\cdot R2} \begin{vmatrix} 0 & -1-t & 1+t^2 \\ 0 & t+1 & -1-t \\ -1 & -1 & t \end{vmatrix}$$
$$\xlongequal{\text{cofactor expansion across the 1st column}} (-1)(-1)^{3+1}\begin{vmatrix} -1-t & 1+t^2 \\ t+1 & -1-t \end{vmatrix}$$
$$= (-1)((-1-t)^2 - (1+t^2)(t+1)) = (-1)(t - t^3) = t(t+1)(t-1)$$

Thus the eigenvalues are $0, 1, -1$, with algebraic multiplicities $1, 1, 1$.

Example 12.13. Suppose

$$A = \begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

is an $n \times n$ triangular matrix, then the characteristic polynomial of A is

$$\det(tI - A) = \begin{vmatrix} t-\lambda_1 & * & \cdots & * \\ 0 & t-\lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t-\lambda_n \end{vmatrix} = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n)$$

so the eigenvalues are the diagonal elements $\lambda_1, \lambda_2, \cdots, \lambda_n$. Note that λ_i's might not be distinct, so there might be some multiplicities.

Example 12.14. Suppose $A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, then the characteristic polynomial is

$$\det(tI - A) = \begin{vmatrix} t-1 & -2 & 0 & 1 \\ 0 & t-2 & -1 & -3 \\ 0 & 0 & t-3 & -4 \\ 0 & 0 & 0 & t-1 \end{vmatrix} = (t-1)^2(t-2)(t-3)$$

And the eigenvalues are $1, 2, 3$, with multiplicities $2, 1, 1$ respectively.

12.3 Similarity

Definition 12.15. A and B are said to be similar if there exists an invertible matrix P such that $A = PBP^{-1}$. Note that this definition is *symmetric* in the sense that we also have $B = P^{-1}AP = (P^{-1})A(P^{-1})^{-1}$. Similarity is *transitive* in the sense that if A, B are similar, B, C are similar, then so do A, C. The reason is that suppose $A = PBP^{-1}$, $B = RCR^{-1}$, we would have $A = PBP^{-1} = PRCR^{-1}P^{-1} = (PR)C(PR)^{-1}$.

Definition 12.16. We can the mapping $A \mapsto PAP^{-1}$ a similar transformation

Theorem 12.17. Similar matrices have the same

- determinant
- characteristic polynomial

Proof. Suppose A, B are similar, $A = PBP^{-1}$, then

- $\det(A) = \det(PBP^{-1}) = \det(P)\det(B)\det(P^{-1}) = \det(P)\det(B)\det(P)^{-1} = \det(B)$
- Note that $tI - A = tPIP^{-1} - PBP^{-1} = P(tI - A)P^{-1}$, so $tI - A$ and $tI - B$ are similar, so they have the same determinant which is the characteristic polynomial.

□

12.4 Eigendecomposition and diagonalization

Theorem 12.18 (diagonalization theorem)**.** Suppose $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are eigenvectors of A with eigenvalues $\lambda_1, \cdots, \lambda_n$, then we have

$$A \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

Therefore we get a diagonalization of A, which reads $A = PDP^{-1}$, where $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$

Theorem 12.19. If $\mathbf{v}_1, \cdots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \cdots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ is linearly independent.

Proof. Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is linearly independent, up to reordering the indices, we may assume $\mathbf{v}_r = c_1\mathbf{v}_1 + \cdots + c_{r-1}\mathbf{v}_{r-1}$ □

Exercise 12.20. Diagonalize $A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$

Solution. First we want to find all eigenvalues of A

$$\det(tI - A) = \begin{vmatrix} t-1 & -3 & -3 \\ 3 & t+5 & 3 \\ -3 & -3 & t-1 \end{vmatrix} \xlongequal{R3 \to R3+R2} \begin{vmatrix} t-1 & -3 & -3 \\ 3 & t+5 & 3 \\ 0 & t+2 & t+2 \end{vmatrix}$$
$$\xlongequal{\text{factor out } (t+2) \text{ from row 3}} (t+2)\begin{vmatrix} t-1 & -3 & -3 \\ 3 & t+5 & 3 \\ 0 & 1 & 1 \end{vmatrix}$$
$$\xlongequal[R2 \to R2-3R3]{R1 \to R1+3R3} (t+2)\begin{vmatrix} t-1 & 0 & 0 \\ 3 & t+2 & 0 \\ 0 & 1 & 1 \end{vmatrix} = (t+2)(t-1)(t+2)(1) = (t+2)^2(t-1)$$

So the eigenvalues of A are $\lambda_1 = 1$ with algebraic multiplicity 1 and $\lambda_2 = \lambda_3 = -2$ which is of algebraic multiplicity 2. For a basis of the 1-eigenspace $\mathrm{Nul}(I - A)$, consider

$$\left[\begin{array}{c|c} I - A & \mathbf{0} \end{array}\right] = \left[\begin{array}{ccc|c} 0 & -3 & -3 & 0 \\ 3 & 6 & 3 & 0 \\ -3 & -3 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

So it could be $\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$ For a basis of the (-2)-eigenspace $\mathrm{Nul}(-2I - A)$, consider

$$\left[\begin{array}{c|c} -2I - A & \mathbf{0} \end{array}\right] = \left[\begin{array}{ccc|c} -3 & -3 & -3 & 0 \\ 3 & 3 & 3 & 0 \\ -3 & -3 & -3 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

So it could be $\left\{ \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$. By Theorem 12.19, we know that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ are linearly independent, thus forming a basis for $\mathbb{R}^3$. And we have a diagonalization

$$A = PDP^{-1}, \quad P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

□

Example 12.21. $A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$, the characteristic polynomial of A is

$$\det(tI - A) = \begin{vmatrix} t-2 & -4 & -3 \\ 4 & t+6 & 3 \\ -3 & -3 & t-1 \end{vmatrix} \xlongequal{R1 \to R1+R2} \begin{vmatrix} t+2 & t+2 & 0 \\ 4 & t+6 & 3 \\ -3 & -3 & t-1 \end{vmatrix}$$

$$\xlongequal{\text{factor } (t+2) \text{ from row } 1} (t+2) \begin{vmatrix} 1 & 1 & 0 \\ 4 & t+6 & 3 \\ -3 & -3 & t-1 \end{vmatrix} \xlongequal[R3 \to R3+3R1]{R2 \to R2-4R1} (t+2) \begin{vmatrix} 1 & 1 & 0 \\ 0 & t+2 & 3 \\ 0 & 0 & t-1 \end{vmatrix}$$

$$= (t-1)(t+2)^2$$

So the eigenvalues of A are $\lambda_1 = 1$ with algebraic multiplicity 1 and $\lambda_2 = \lambda_3 = -2$ which is of algebraic multiplicity 2. To find a basis for the 1-eigenspace, consider

$$\left[\begin{array}{c|c} I - A & \mathbf{0} \end{array}\right] = \left[\begin{array}{ccc|c} -1 & -4 & -3 & 0 \\ 4 & 7 & 3 & 0 \\ -3 & -3 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} -1 & -4 & -3 & 0 \\ 0 & -9 & -9 & 0 \\ 0 & 9 & 9 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} -1 & -4 & -3 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

It could be $\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$. To find a basis for the (-2)-eigenspace, consider

$$\left[\begin{array}{c|c} -2I - A & \mathbf{0} \end{array}\right] = \left[\begin{array}{ccc|c} -4 & -4 & -3 & 0 \\ 4 & 4 & 3 & 0 \\ -3 & -3 & -3 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} -4 & -4 & -3 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right]$$

$$\sim \left[\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

It could be $\left\{ \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$. Note that there is not enough eigenvectors ($2 < 3$!!!) for a basis for $\mathbb{R}^3$. The real reason that this failed to be a basis was that the geometric multiplicity of -2 (which is equal to 1) is less that the algebraic multiplicity of -2 (which is equal to 2)

Theorem 12.22. Supoose A is an $n \times n$ matrix. Then the geometric multiplicity of an eigenvalue λ is less or equal to the algebraic multiplicity of λ. A is diagonalizable $\iff$ A has n linearly independent eigenvectors $\iff$ geometric multiplicities are always the same as algebraic multiplicities.

Theorem 12.23. Combining Theorem 12.19, we know that an $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Definition 12.24. By Theorem 12.17 we know that similar matrices have the same characteristic polynomials (hence the same eigenvalues) and determinants. So we may define notions like *eigenvalues*, *eigenvectors*, *characteristic polynomials* and *determinants* for a linear transformation $T : V \to V$.

Example 12.25. Suppose $T = \dfrac{d}{dt} : \mathbb{P}_2 \to \mathbb{P}_2$, $T(a_0 + a_1 t + a_2 t^2) = a_1 + 2a_2 t$ is a linear transformation (verify this!), $\mathcal{E} = \{1, t, t^2\}$ is the standard basis, we have

$$D = [T]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

so the characteristic polynomial of T is $\det(tI - [T]_{\mathcal{E}}) = t(t-1)(t-2)$, so the eigenvalues for T will be $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 2$, and the corresponding eigenvectors for $[T]_{\mathcal{E}}$ could be $\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$, which in turn corresponds to eigenvectors $\{1, t, t^2\}$ for T. However if we choose basis $\mathcal{C} = \{1+t, t, t^2\}$ is the standard basis, we have

$$P = \underset{\mathcal{E}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P^{-1} = \underset{\mathcal{C}\leftarrow\mathcal{E}}{P} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and thus

$$A = [T]_{\mathcal{C}} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}[T]_{\mathcal{C}}\left(\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}\right)^{-1} = PDP^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The determinant of T is equal to either $\det([T]_{\mathcal{B}})$ or $\det([T]_{\mathcal{C}})$, which is zero.

Example 12.26. $A = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$, let's compute A^{50}. First we find the eigenvalues $\lambda_1 = 1$, $\lambda_2 = 4$. and we get corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, so we have $D = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ and $P = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}$, so $P^{-1} = \frac{1}{3}\begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$. So we have

$$\begin{aligned} A^{50} &= \overbrace{(PDP^{-1})(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})}^{50} \\ &= \overbrace{(PD\cancel{P^{-1}})(\cancel{P}D\cancel{P^{-1}})(\cancel{P}D\cancel{P^{-1}})\cdots(\cancel{P}DP^{-1})}^{50} \\ &= PD^{50}P^{-1} = \frac{1}{3}\begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 4^{50} \end{bmatrix}\begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

13 Lecture 17 - Orthogonalization

Recall for $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$, we can define the *dot product* $\mathbf{u} \cdot \mathbf{v}$ to be $\mathbf{u}^T\mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots u_nv_n$. The length(or norm) of a vector $\mathbf{u}$ can be expressed as $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + \cdots + u_n^2}$. Geometrically, the dot product can interpreted as $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$, here θ is the angle between $\mathbf{u}$ and $\mathbf{v}$ (which could be between 0 and π)

- If $\mathbf{u} \cdot \mathbf{v} < 0$, then $\cos\theta < 0$, θ is obtuse, if in addition $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|$, then $\mathbf{u}, \mathbf{v}$ are of the opposite direction.
- If $\mathbf{u} \cdot \mathbf{v} = 0$, then $\cos\theta = 0$, $\theta = \frac{\pi}{2}$, or $\mathbf{u} = \mathbf{0}$, or $\mathbf{v} = \mathbf{0}$. In this case, we say that $\mathbf{u}, \mathbf{v}$ are orthogonal (denoted $\mathbf{u} \perp \mathbf{v}$, $\perp$ stands for perpendicular).
- If $\mathbf{u} \cdot \mathbf{v} > 0$, then $\cos\theta > 0$, θ is acute, if in addition $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|$, then $\mathbf{u}, \mathbf{v}$ are of the same direction.

13.1 Orthogonal and orthonormal basis

Definition 13.1. We say a vector $\mathbf{u}$ is of unit length (or a unit vector, or a normalized vector) if $\|\mathbf{u}\| = 1$

Definition 13.2. $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is called an orthogonal set if $\mathbf{v}_i, \mathbf{v}_j$ $(i \neq j)$ are orthogonal. $\mathcal{B}$ is called an orthogonal basis if $\mathcal{B}$ is in addition a basis. $\mathcal{B}$ is called an orthonormal basis if the basis vectors are in addition normalized.

Remark. Suppose $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, to test if $\mathcal{B}$ is an orthogonal (or orthonormal) set, we just need to write $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$, and test if

$$A^T A = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T\mathbf{v}_1 & \mathbf{v}_1^T\mathbf{v}_2 & \cdots & \mathbf{v}_1^T\mathbf{v}_n \\ \mathbf{v}_2^T\mathbf{v}_1 & \mathbf{v}_2^T\mathbf{v}_2 & \cdots & \mathbf{v}_2^T\mathbf{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_n^T\mathbf{v}_1 & \mathbf{v}_n^T\mathbf{v}_2 & \cdots & \mathbf{v}_n^T\mathbf{v}_n \end{bmatrix}$$

is diagonal (or if $A^T A = I$)

Theorem 13.3. Suppose $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is an orthogonal set of non-zero vectors, then $\mathcal{B}$ is linearly independent

Proof. Suppose not, we may assume $\mathbf{0} \neq \mathbf{v}_p = c_1\mathbf{v}_1 + \cdots + r_{p-1}\mathbf{v}_{p-1}$, but then

$$0 < \|\mathbf{v}_p\|^2 = \mathbf{v}_p \cdot (c_1\mathbf{v}_1 + \cdots + r_{p-1}\mathbf{v}_{p-1}) = 0$$

Which is a contradiction. $\square$

13.2 Gram-Schmidt process

Question. Suppose you have two vectors $\mathbf{u}, \mathbf{v}$ (here $\mathbf{v} \neq \mathbf{0}$), what is the orthogonal projection of $\mathbf{u}$ onto $\mathbf{v}$ (Which we denote as $\operatorname{Proj}_{\mathbf{v}} \mathbf{u}$)?

Answer. First you realize that $\operatorname{Proj}_{\mathbf{v}} \mathbf{u}$ is parallel to $\mathbf{v}$, so we write it as $\lambda\mathbf{v}$, and we know $\|\operatorname{Proj}_{\mathbf{v}} \mathbf{u}\| = \lambda\|\mathbf{v}\| = \|\mathbf{u}\| \cos\theta$, so we may conclude that

$$\lambda = \frac{\|\mathbf{u}\| \cos\theta}{\|\mathbf{v}\|} = \frac{\|\mathbf{u}\|\|\mathbf{v}\| \cos\theta}{\|\mathbf{v}\|^2} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}$$

Therefore we have derived the equation

$$\operatorname{Proj}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}\mathbf{v} \tag{13.1}$$

Question. If W is a subspace of $\mathbb{R}^n$ with an orthogonal basis $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$, how could you express the orthogonal projection of a vector $\mathbf{y} \in \mathbb{R}^n$ onto W (denoted by $\operatorname{Proj}_W \mathbf{y}$).

Suppose $\operatorname{Proj}_W \mathbf{y} = \lambda_1\mathbf{u}_1 + \cdots + \lambda_p\mathbf{u}_p$, then we should have for any i

$$0 = (\mathbf{y} - \operatorname{Proj}_W \mathbf{y}) \cdot \mathbf{u}_i = (\mathbf{y} - \lambda_1\mathbf{u}_1 + \cdots + \lambda_p\mathbf{u}_p) \cdot \mathbf{u}_i = \mathbf{y} \cdot \mathbf{u} - \lambda_i\mathbf{u}_1 \cdot \mathbf{u}_1 \Rightarrow \lambda_i = \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}$$

There we have the equation

$$\begin{aligned} \operatorname{Proj}_W \mathbf{y} &= \operatorname{Proj}_{\mathbf{u}_1} \mathbf{y} + \cdots + \operatorname{Proj}_{\mathbf{u}_p} \mathbf{y} \\ &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p \end{aligned} \tag{13.2}$$

If we further assume that $\mathcal{B}$ is an orthonormal basis, then $\mathbf{u}_1 \cdot \mathbf{u}_1 = \|\mathbf{u}_i\|^2 = 1$. Let's write $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{bmatrix}$. (13.2) becomes

$$\operatorname{Proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p = UU^T\mathbf{y} \tag{13.3}$$

Example 13.4. Consider $\mathcal{B} = \left\{ \mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Let $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 0 & 0 \end{bmatrix}$, then $A^T A = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$ is diagonal, hence $\mathcal{B}$ is an orthogonal set, the orthogonal projection of $\mathbf{y}$ onto W is

$$\begin{aligned} \text{Proj}_W \mathbf{y} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 \\ &= \frac{1 \cdot 2 + 1 \cdot 1 + 1 \cdot 0}{2^2 + 1^2 + 0^2}\mathbf{u}_1 + \frac{1 \cdot (-1) + 1 \cdot 2 + 1 \cdot 0}{(-1)^2 + 2^2 + 0^2}\mathbf{u}_2 \\ &= \frac{3}{5}\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{5}\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \end{aligned}$$

Question. Suppose we are given arbitrary basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ for a subspace W of $\mathbb{R}^n$, how could we get a orthogonal (or orthonormal) basis $\mathcal{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ from it

Answer. We apply the Gram-Schmidt process

- $\mathbf{u}_1 = \mathbf{v}_1$
- $\begin{aligned} \mathbf{u}_2 &= \mathbf{v}_2 - \text{Proj}_{\mathbf{u}_1} \mathbf{v}_2 \\ &= \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 \end{aligned}$
- $\begin{aligned} \mathbf{u}_3 &= \mathbf{v}_3 - \text{Proj}_{\mathbf{u}_1} \mathbf{v}_3 - \text{Proj}_{\mathbf{u}_2} \mathbf{v}_3 \\ &= \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 \end{aligned}$
- $\vdots$
- $\begin{aligned} \mathbf{u}_p &= \mathbf{v}_p - \text{Proj}_{\mathbf{u}_1} \mathbf{v}_p - \cdots - \text{Proj}_{\mathbf{u}_{p-1}} \mathbf{v}_p \\ &= \mathbf{v}_p - \frac{\mathbf{v}_p \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 - \cdots - \frac{\mathbf{v}_p \cdot \mathbf{u}_{p-1}}{\mathbf{u}_{p-1} \cdot \mathbf{u}_{p-1}}\mathbf{u}_{p-1} \end{aligned}$

Then $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ would be an orthogonal basis. To get an orthonormal basis, just normalize these vectors.

Example 13.5. Consider $\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \right\}$. Let's use Gram-Schmidt process to find an orthogonal (and an orthonormal) basis from it.

- First we choose $\mathbf{u}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- $\mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{5}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$. Let's instead take a multiple of this to be our $\mathbf{u}_2$, namely we set $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$
- $\mathbf{u}_3 = \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$

Thus $\left\{ \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right\}$ is an orthogonal basis for $\mathbb{R}^3$. If we further normalize it, we have

$$\mathbf{w}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \mathbf{w}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} = \frac{1}{\sqrt{6}} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \mathbf{w}_3 = \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

is an orthonormal basis.

Exercise 13.6. Use Gram-Schmidt process to find an orthogonal and orthonormal basis of the following basis

1. $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$
2. $\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix} \right\}$

13.3 Orthogonal matrix and QR factorization

Definition 13.7. A square matrix U is called an orthogonal matrix if the columns of U form an orthonormal basis $\iff U^T U = U U^T = I$. Note that in this particular case, $U^T = U^{-1}$

Remark. The rows of U also form an orthonormal basis.

Note. The name *orthogonal matrix* is unfortunate since it is made out of orthonormal basis. It should really be called the orthonormal matrix. But since mathematicians has been using this for a long time. So we will stick to this usage.

The Gram-Schmidt process gives us the so-called QR-factorization. First we can rewrite the Gram-Schmidt process as

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1 \\
\mathbf{v}_2 &= \frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \mathbf{u}_2 \\
\mathbf{v}_3 &= \frac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 + \mathbf{u}_3 \\
&\vdots \\
\mathbf{v}_n &= \frac{\mathbf{v}_n \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{v}_n \cdot \mathbf{u}_{n-1}}{\mathbf{u}_{n-1} \cdot \mathbf{u}_{n-1}}\mathbf{u}_{n-1} + \mathbf{u}_n
\end{aligned}$$

After normalization (let's set $\mathbf{u}_i := \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$ to be unit vectors), we may write the above as the following set of equations with coefficients

$$\begin{aligned}
\mathbf{v}_1 &= r_{11}\mathbf{u}_1 \\
\mathbf{v}_2 &= r_{12}\mathbf{u}_1 + r_{22}\mathbf{u}_2 \\
\mathbf{v}_3 &= r_{13}\mathbf{u}_1 + r_{23}\mathbf{u}_2 + r_{33}\mathbf{u}_3 \\
&\vdots \\
\mathbf{v}_n &= r_{1n}\mathbf{u}_1 + \cdots + r_{n-1,n}\mathbf{u}_{n-1} + r_{nn}\mathbf{u}_n
\end{aligned}$$

If we write $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$, $Q = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ (this is an orthogonal matrix), then have

$$A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_{nn} \end{bmatrix} = QR$$

It can be shown that QR-factorization always exists, even A is not invertible (columns does not form a basis)

Example 13.8. Continuing Example 13.5, we consider the QR-factorization of $A = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} = \begin{bmatrix}1 & 1 & 3\\ 1 & 2 & -1\\ 1 & 2 & 1\end{bmatrix}$, we know that $Q = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}}\end{bmatrix}$, so we know that

$$R = Q^{-1}A = Q^TA = \begin{bmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}}\\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1 & 1 & 3\\ 1 & 2 & -1\\ 1 & 2 & 1\end{bmatrix} = \begin{bmatrix}\sqrt{3} & \frac{5}{\sqrt{3}} & \sqrt{3}\\ 0 & \frac{2}{\sqrt{6}} & -\sqrt{6}\\ 0 & 0 & \sqrt{2}\end{bmatrix}$$

Note that here we didn't actually use Gram-Schmidt process to evaluate R, instead we only used the existence of QR-factorization.

Definition 13.9. Suppose W is a subspace of $\mathbb{R}^n$, the orthogonal complement $W^\perp = \{\mathbf{v}|\mathbf{v} \perp W\}$ is the set of vectors that are orthogonal to W

Remark. It is easy to show that W is a subspace of $\mathbb{R}^n$ and $(W^\perp)^\perp = W$. This is kind of a *duality*, take $\mathbb{R}^3$ for an example, the orthogonal complement of a line would be the perpendicular plane, the orthogonal complement of a plane would be the perpendicular line, the orthogonal complement of the origin (a point) would be the whole space ($\mathbb{R}^3$) and the orthogonal complement of $\mathbb{R}^3$ would be the origin $\{\mathbf{0}\}$. From this we also see that $\dim W + \dim W^\perp = n$.

We make the following crucial observation: $\mathbf{x}$ is in $\operatorname{Nul} A \iff \mathbf{x}$ is orthogonal all row vectors of $A \iff \mathbf{x} \perp \operatorname{Row} A$, so we may conclude that $(\operatorname{Row} A)^\perp = \operatorname{Nul} A$, and similarly we know that $(\operatorname{Col} A)^\perp = (\operatorname{Row}(A^T))^\perp = \operatorname{Nul}(A^T)$.

Exercise 13.10. Suppose $W = \operatorname{Span}\left\{\begin{bmatrix}1\\1\\0\\2\\0\\3\end{bmatrix}, \begin{bmatrix}0\\0\\1\\-2\\0\\2\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\\1\end{bmatrix}\right\}$, find $(\operatorname{Col} A)^\perp$

14 Lecture 18 - Least-square problem

14.1 Least-square solutions

Question. linear system $A\mathbf{x} = \mathbf{b}$ may not always be solvable, nonetheless, we still want to find the *best possible* solution $\hat{\mathbf{x}}$

Answer. According to Gauss, by best possible we mean the least-square solutions. In concrete terms, such $\hat{\mathbf{x}}$ stisfies that $\|A\hat{\mathbf{x}} - \mathbf{b}\| \le \|A\mathbf{x} - \mathbf{b}\|$ for any other choice of $\mathbf{x}$.

By a simple argument we can show that $A\hat{\mathbf{x}}$ must be the projection of $\mathbf{b}$ onto $\operatorname{Col} A$, so we know that

$$(A\hat{\mathbf{x}} - \mathbf{b}) \perp \operatorname{Col} A \iff (A\hat{\mathbf{x}} - \mathbf{b}) \in (\operatorname{Col} A)^\perp = \operatorname{Nul}(A^T) \iff A^T(A\hat{\mathbf{x}} - \mathbf{b}) = \mathbf{0} \iff A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

We call equation

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{14.1}$$

the normal equation of $A\mathbf{x} = \mathbf{b}$. And we have shown that the least square solutions of the linear system $A\mathbf{x} = \mathbf{b}$ is the solution set to the normal equation (14.1). The solution to the least square problem might not be unique.

Theorem 14.1. $A^TA\mathbf{x} = A^T\mathbf{b}$ has a unique solution $\iff$ the columns of A is linearly independent $\iff A^TA$ is invertible. And the unique solution would of course be $(A^TA)^{-1}A^T\mathbf{b}$

Example 14.2. Suppoes $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$. It is easy to verify that the linear system $A\mathbf{x} = \mathbf{b}$ is not linearly consistent. To solve for the least-square solution, we look at its normal equation

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^T A\mathbf{x} = A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

And solve for $\mathbf{x}(A^TA)^{-1}A^T\mathbf{b} = (A^TA)^{-1}A^T\mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

Exercise 14.3. Suppose $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$, solve for the least square solution of $A\mathbf{x} = \mathbf{b}$.

14.2 Machine Learning and approximation

Question. Suppose you know a curve has the form

$$y = \beta_1 f_1(x_1, \cdots, x_k) + \beta_2 f_2(x_1, \cdots, x_k) + \cdots + \beta_n f_n(x_1, \cdots, x_k)$$

And you know what functions $f_1, f_2, \cdots, f_n$ should be (they could be linear functions, quadratic functions, polynomial functions, exponential functions, logarithmic function, trigonometry functions, etc.). You also have collected a bunch of data points $(x_1^{(1)}, \cdots, x_k^{(1)}, y^{(1)}), (x_1^{(2)}, \cdots, x_k^{(2)}, y^{(2)}), \cdots, (x_1^{(m)}, \cdots, x_k^{(m)}, y^{(m)})$. How do you find best fitting coefficients $\beta_1, \beta_2, \cdots, \beta_n$? First let's define the residual

$$\epsilon^{(i)} = y^{(i)} - \beta_1 f_1(x_1^{(i)}, \cdots, x_k^{(i)}) - \beta_2 f_2(x_1^{(i)}, \cdots, x_k^{(i)}) - \cdots - \beta_n f_n(x_1^{(i)}, \cdots, x_k^{(i)})$$

Our goal to minimize the residual. One might want to minimize the quantity $\sum_{i=1}^m |\epsilon^{(i)}|$, however, this is cumbersome, difficult to work with, and mathematically unsatisfactory. So Gauss instead considered the square sum $\sum_{i=1}^m |\epsilon^{(i)}|^2$, posed and solved the least square problem!!!

Answer. We use normal equations.

Let's write

$$\mathbf{y} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon^{(1)} \\ \vdots \\ \epsilon^{(m)} \end{bmatrix}, X = \begin{bmatrix} f_1(x_1^{(1)}, \cdots, x_k^{(1)}) & \cdots & f_n(x_1^{(1)}, \cdots, x_k^{(1)}) \\ \vdots & \ddots & \vdots \\ f_1(x_1^{(m)}, \cdots, x_k^{(m)}) & \cdots & f_n(x_1^{(m)}, \cdots, x_k^{(m)}) \end{bmatrix}$$

Then have $\mathbf{y} = X\beta = \epsilon$. To minimize $\|\epsilon\|^2$ will be the same as solving the normal equation $X^TX\beta = X^T\mathbf{y}$ of $X\beta = \mathbf{y}$.

Example 14.4. Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2, 1), (5, 2), (7, 3)$ and $(8, 3)$.

15 Lecture 19 - Complex eigenvalues

15.1 Complex numbers

Theorem 15.1. Consider quadratic equation $at^2 + bt + c = 0$, the quadratic formula reads $t = \frac{-b \pm \sqrt{\Delta}}{2a}$, where $\Delta = b^2 - 4ac$ is the discriminant.

Let's review some general stuff about complex numbers.

Definition 15.2. The set of complex numbers $\mathbb{C}$ is the defined to be $\{z = a + bi | a, b \in \mathbb{R}\}$, with $i = \sqrt{-1}$ so that $i^2 = -1$. We call $\operatorname{Re} z = a$ the real part of z and $\operatorname{Im} z = b$ to be the imaginary part of z. We call $r = |z| = \sqrt{a^2 + b^2}$ the modulus (or absolute value) of z, and the angle φ between z and the real axis the argument of z.

Definition 15.3. There is a natural identification between the complex numbers and the plane $\mathbb{R}^2$ via $z = a + bi \leftrightsquigarrow (a, b)$ (this is why the set of complex numbers is often called the complex plane). We can define

- Addition via $(a + bi) + (c + di) = (a + c) + (b + d)i$
- Multiplication via $(a + bi)(c + di) = ac + adi + bci + bdi^2 = (ac - bd) + (ad + bc)i$.
- Conjugate as $\overline{z} = a - bi$.

Remark. Through this identification it is easy to see that $a = r \cos \varphi$, $b = r \sin \varphi$, and we have Euler's formula $z = re^{i\varphi} = r \cos \varphi = i \sin \varphi$.

Remark. If we have a complex-valued matrix A, then the conjugation is defined entrywise (Note that column vectors are matrices). If we write a complex valued matrix $A = \operatorname{Re} A + i \operatorname{Im} A$, then the conjugation would be $\overline{A} = \operatorname{Re} A - i \operatorname{Im} A$. For example, if $A = \begin{bmatrix} 1+2i & 3 \\ -1-i & i \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -1 & 0 \end{bmatrix} + i \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}$, then $\overline{A} = \begin{bmatrix} 1-2i & 3 \\ -1+i & -i \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -1 & 0 \end{bmatrix} - i \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}$. It is easy to check that $\overline{A+B} = \overline{A} + \overline{B}$, $\overline{cA} = \overline{c}\overline{A}$, $\overline{AB} = \overline{A}\overline{B}$.

15.2 Eigen-decomposition over the complex numbers

Question. $A = \begin{bmatrix} 1 & -2 \\ 1 & 3 \end{bmatrix}$, can we diagonalize it?

Answer. First we compute its characteristic polynomial $t^2 - 4t + 5$, an then we can solve the quadratic equation as follows

$$\Delta = (-4)^2 - 4 \cdot 1 \cdot 5 = -4, t = \frac{-(-4) \pm \sqrt{-4}}{2} = \frac{-(-4) \pm 2i}{2} = 2 \pm i$$

Hence the eigenvalues are $\lambda_1 = 2 + i, \lambda_2 = 2 - i$. The eigenvector for λ_1 can be computed via

$$\left[\begin{array}{c|c} \lambda_1 I - A & \mathbf{0} \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1-i & 0 \\ 0 & 0 & 0 \end{array}\right]$$

So we get $\mathbf{v}_1 = \begin{bmatrix} -1+i \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} i \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} i = \operatorname{Re} \mathbf{v}_1 + i \operatorname{Im} \mathbf{v}_1$. The eigenvector for λ_2 can be computed via

$$\left[\begin{array}{c|c} \lambda_2 I - A & \mathbf{0} \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1+i & 0 \\ 0 & 0 & 0 \end{array}\right]$$

So we get $\mathbf{v}_2 = \begin{bmatrix} -1-i \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} -i \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \end{bmatrix} i = \operatorname{Re} \mathbf{v}_2 + i \operatorname{Im} \mathbf{v}_2$. Now we may realize that we have λ_1 and λ_2 are conjugate of each other. $\mathbf{v}_1$ and $\mathbf{v}_2$ are conjugate of each other.

Theorem 15.4. In general, if A is a 2 by 2 real-valued matrix with complex eigenvalues (characteristic polynomial have complex roots, no real roots), then they are conjugate of each other, we can write them as $\lambda = a - bi$ and $\overline{\lambda} = a + bi$, suppose the eigenvector for λ is $\mathbf{v} = \operatorname{Re} \mathbf{v} + i \operatorname{Im} \mathbf{v}$, then the eigenvector for $\overline{\lambda}$ will be $\overline{\mathbf{v}} = \operatorname{Re} \mathbf{v} - i \operatorname{Im} \mathbf{v}$. If we write $P = \begin{bmatrix} \operatorname{Re} \mathbf{v} & \operatorname{Im} \mathbf{v} \end{bmatrix}$, and $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, then $AP = PC$, therefore we get decomposition $A = PCP^{-1}$.

Proof. We have $A\mathbf{v} = \lambda\mathbf{v}$ (and hence by conjugation we have $A\overline{\mathbf{v}} = A\overline{\mathbf{v}} = \overline{A}\overline{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$, note that here A is real-valued, so $\overline{A} = A$), we can rewrite this as

$$(A \operatorname{Re}\mathbf{v})+i(A \operatorname{Im}\mathbf{v}) = A(\operatorname{Re}\mathbf{v}+i \operatorname{Im}\mathbf{v}) = A\mathbf{v} = \lambda\mathbf{v} = (a-bi)(\operatorname{Re}\mathbf{v}+i \operatorname{Im}\mathbf{v}) = (a \operatorname{Re}\mathbf{v}+b \operatorname{Im}\mathbf{v})+i(a \operatorname{Im}\mathbf{v}-b \operatorname{Re}\mathbf{v})$$

Looking at its real and imaginary parts we conclude that

$$A \operatorname{Re}\mathbf{v} = a \operatorname{Re}\mathbf{v} + b \operatorname{Im}\mathbf{v}, \qquad A \operatorname{Im}\mathbf{v} = a \operatorname{Im}\mathbf{v} - b \operatorname{Re}\mathbf{v}$$

This is precisely $AP = PC$ □

Remark. Matrix C is special in the following sense (it can be decomposed as a composition of rotation and scaling)

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} r\cos\varphi & -r\sin\varphi \\ r\sin\varphi & r\cos\varphi \end{bmatrix} = r\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

Here we suppose $\overline{\lambda} = a + bi = re^{i\varphi} = r\cos\varphi + ir\sin\varphi$ so that $a = r\cos\varphi$, $b = r\sin\varphi$.

Example 15.5. In the previous example, we could take $\lambda = 2 - i$ so that $a = 2, b = 1$, $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + i\begin{bmatrix} -1 \\ 0 \end{bmatrix}$ so that $\operatorname{Re}\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\operatorname{Im}\mathbf{v} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$, and we have the decomposition

$$A = \begin{bmatrix} 1 & -2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} = PCP^{-1}$$

Exercise 15.6. Suppose $A = \begin{bmatrix} -1 & -1 \\ 5 & -5 \end{bmatrix}$. Find an invertible matrix P and a matrix C of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ such that $A = PCP^{-1}$

16 Lecture 20 - Orthogonal diagonalization

16.1 Orthogonal diagonalization

Theorem 16.1. Suppose A is a symmetric ($A^T = A$) real-valued matrix, and $\mathbf{v}_1$, $\mathbf{v}_2$ are λ_1-eigenvector, λ_2-eigenvectors respectively. Then $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$

Proof. Consider

$$\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\lambda_1\mathbf{v}_1)^T\mathbf{v}_2 = (A\mathbf{v}_1)^T\mathbf{v}_2 = \mathbf{v}_1^T A^T\mathbf{v}_2 = \mathbf{v}_1^T A\mathbf{v}_2 = \mathbf{v}_1 \cdot (\lambda_2\mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2)$$

We get $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$, since $\lambda_1 - \lambda_2 \neq 0$, $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ □

Theorem 16.2. Suppose A is a symmetric real-valued matrix, then the eigenvalues are also real.

Proof. Suppose λ is an eigenvalue of A, then there exists some λ-eigenvector such that $A\mathbf{v} = \lambda\mathbf{v}$, and that the length of the complex vector $\mathbf{v}$ is $\|\mathbf{v}\|^2 = \overline{\mathbf{v}} \cdot \mathbf{v}$ is a positive real number (Note that for a complex number $z = a + bi$, $|z|^2 = a^2 + b^2 = (a + bi)(a - bi) = z\overline{z}$). Since A is symmetric and real-valued, $\overline{A}^T = A$. We have

$$\overline{\lambda}\|\mathbf{v}\|^2 = \overline{(A\mathbf{v})}^T\mathbf{v} = \overline{\mathbf{v}}^T\overline{A}^T\mathbf{v} = \overline{\mathbf{v}}^T A\mathbf{v} = \lambda\|\mathbf{v}\|^2$$

Which implies that $(\lambda - \overline{\lambda})\|\mathbf{v}\|^2 = 0$, so $\lambda = \overline{\lambda}$, i.e. λ is real-valued. □

Fact 16.3. A symmetric real-valued matrix is diagonalizable.

Theorem 16.4. Suppose A is a symmetric real-valued matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$ (maybe repeated). A can be orthogonal diagonalized as $A = PDP^T$, where D is the diagonal matrix $\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$, and P is an orthogonal matrix.

Proof. To get an orthonormal basis for each eigenspace $\mathrm{Nul}(\lambda I - A)$, you just need to find an arbitrary basis, and then apply Gram-Schmidt process to get an orthonormal basis, then by Theorem 16.1, the set of eigenvectors form an orthonormal basis for $\mathbb{R}^n$, assume they are $\mathbf{u}_1, \cdots, \mathbf{u}_n$, in corresponds to eigenvalues $\lambda_1, \cdots, \lambda_n$, then we get orthogonal diagonalization

$$A = PDP^{-1} = PDP^T$$

Here $P = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ is an orthogonal basis and thus $P^{-1} = P^T$. $\square$

Example 16.5. Suppose $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$, the characteristic polynomial is $(t-3)(t-8)$, so we have eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 8$, we can then find the eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, we may realized that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, i.e. $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2$, we can further normalize them into $\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$ and $\mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$. So we have orthogonal diagonalization

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = PDP^T$$

Example 16.6. Suppose $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$, then characteristic polynomial is

$$\begin{aligned}
\det(tI - A) &= \begin{vmatrix} t-3 & 2 & -4 \\ 2 & t-6 & -2 \\ -4 & -2 & t-3 \end{vmatrix} \overset{R3 \to R3 + 2R2}{=\!=\!=} \begin{vmatrix} t-3 & 2 & -4 \\ 2 & t-6 & -2 \\ 0 & 2t-14 & t-7 \end{vmatrix} \\
&\overset{\text{factor } (t-7) \text{ from 3rd row}}{=\!=\!=\!=\!=} (t-7) \begin{vmatrix} t-3 & 2 & -4 \\ 2 & t-6 & -2 \\ 0 & 2 & 1 \end{vmatrix} \overset{C2 \to C2 - C3}{=\!=\!=} (t-7) \begin{vmatrix} t-3 & 10 & -4 \\ 2 & t-2 & -2 \\ 0 & 0 & 1 \end{vmatrix} \\
&\overset{\text{cofactor expansion on the 3rd row}}{=\!=\!=\!=\!=\!=} (t-7) \cdot 1 \cdot (-1)^{3+3} \begin{vmatrix} t-3 & 10 \\ 2 & t-2 \end{vmatrix} \\
&= (t-7)(t^2 - 5t - 14) = (t-7)^2(t+2)
\end{aligned}$$

Hence the eigenvalues are $\lambda_1 = \lambda_2 = 7$, $\lambda_3 = -2$

$$\left[\begin{array}{c|c} 7I - A & \mathbf{0} \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

so we may choose

$$\mathbf{v}_1 = 2 \begin{bmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

as eigenvectors for $\lambda_1 = \lambda_2 = 7$. We can use Gram-Schmidt process to get an orthogonal set: $\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 - \dfrac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} \mathbf{w}_1 = \begin{bmatrix} \frac{4}{5} \\ \frac{2}{5} \\ 1 \end{bmatrix}$, we can choose $\mathbf{w}_2 = \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}$. We can normalize them into

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{3\sqrt{5}} \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}$$

$$\left[\begin{array}{c|c} -2I - A & \mathbf{0} \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

so we may choose

$$\mathbf{v}_3 = 2\begin{bmatrix}-1\\-\frac{1}{2}\\1\end{bmatrix} = \begin{bmatrix}-2\\-1\\2\end{bmatrix}$$

as the eigenvector for $\lambda_3 = -2$. We can normalize it into

$$\mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{1}{3}\begin{bmatrix}-2\\-1\\2\end{bmatrix}$$

Now we get the orthogonal decomposition

$$A = \begin{bmatrix}3 & -2 & 4\\-2 & 6 & 2\\4 & 2 & 3\end{bmatrix} = \begin{bmatrix}-\frac{1}{\sqrt{5}} & \frac{4}{3\sqrt{5}} & -\frac{2}{3}\\ \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} & -\frac{1}{3}\\ 0 & \frac{5}{3\sqrt{5}} & \frac{2}{3}\end{bmatrix}\begin{bmatrix}7 & 0 & 0\\0 & 7 & 0\\0 & 0 & -2\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0\\ \frac{4}{3\sqrt{5}} & \frac{2}{3\sqrt{5}} & \frac{5}{3\sqrt{5}}\\ -\frac{2}{3} & -\frac{1}{3} & \frac{2}{3}\end{bmatrix} = PDP^T$$

16.2 Spectral decomposition

Theorem 16.7. Suppose A is a symmetric real-valued matrix, and $A = PDP^T$ is its orthogonal diagonalization, then we have the so-called spectral decomposition

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T$$

Proof.

$$A = PDP^T = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\begin{bmatrix}\mathbf{u}_1^T\\ \vdots\\ \mathbf{u}_n^T\end{bmatrix} = \begin{bmatrix}\lambda_1\mathbf{u}_1 & & \\ & \ddots & \\ & & \lambda_n\mathbf{u}_n\end{bmatrix}\begin{bmatrix}\mathbf{u}_1^T\\ \vdots\\ \mathbf{u}_n^T\end{bmatrix} = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T$$

□

Remark. Recall that if $\mathbf{u}$ is of unit length, then $\mathrm{Proj}_{\mathbf{u}}\,\mathbf{x} = \mathbf{u}\mathbf{u}^T\mathbf{x}$, so

$$A\mathbf{x} = \lambda_1\mathbf{u}_1\mathbf{u}_1^T\mathbf{x} + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T\mathbf{x}$$

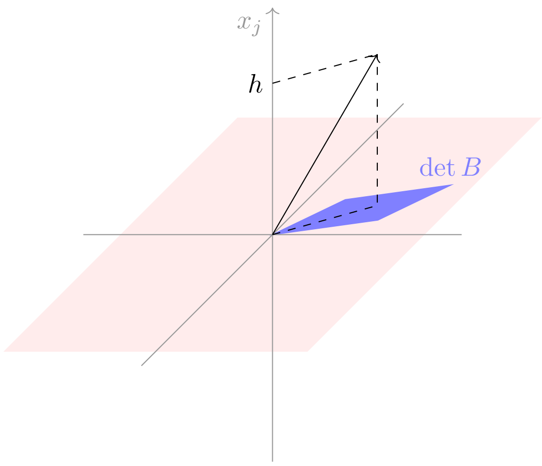
You can think of this as decompose the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$ into the sum of scaled orthogonal projections.

17 Appendix

Theorem 17.1. The RREF of a matrix A is unique.

Proof. The linear dependences of columns of A is presevered under row elementary operations,thus the pivot columns are unique. □

Lemma 17.2. Suppose $A = \begin{bmatrix}B_1 & * & B_2\\0 & h & 0\\B_3 & * & B_4\end{bmatrix}$, where h is the (i, j)-th entry. If we write $B = \begin{bmatrix}B_1 & B_2\\B_3 & B_4\end{bmatrix}$ then $\det A = (-1)^{i+j}h \cdot \det B$



18 Online Assignments

18.1 Online Assignment 1

Problem 18.1. Rewrite the following linear systems as augmented matrices and then solve them, show all your work

1. $\begin{cases} 5x_1 + x_2 = 2 \\ 3x_1 - x_2 = 6 \end{cases}$

2. $\begin{cases} x_1 + x_2 + x_3 = 6 \\ x_1 - x_2 + x_3 = 2 \\ -x_1 + x_2 + x_3 = 4 \end{cases}$

Solution.

1. The augmented matrix is

$$\begin{bmatrix} 5 & 1 & 2 \\ 3 & -1 & 6 \end{bmatrix} \xrightarrow{5R2} \begin{bmatrix} 5 & 1 & 2 \\ 15 & -5 & 30 \end{bmatrix} \xrightarrow{R2 \to R2 - 3R1} \begin{bmatrix} 5 & 1 & 2 \\ 0 & -8 & 24 \end{bmatrix} \xrightarrow{R2/(-8)} \begin{bmatrix} 5 & 1 & 2 \\ 0 & 1 & -3 \end{bmatrix}$$
$$\xrightarrow{R1 \to R1 - R2} \begin{bmatrix} 5 & 0 & 5 \\ 0 & 1 & -3 \end{bmatrix} \xrightarrow{R1/5} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -3 \end{bmatrix}$$

Thus the solution to this linear system is $\begin{cases} x_1 = 1 \\ x_2 = -3 \end{cases}$

2. The augmented matrix is

$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 1 & -1 & 1 & 2 \\ -1 & 1 & 1 & 4 \end{bmatrix} \xrightarrow[R3 \to R3 + R1]{R2 \to R2 - R1} \begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & -2 & 0 & -4 \\ 0 & 2 & 2 & 10 \end{bmatrix} \xrightarrow{R3 \to R3 + R2} \begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & -2 & 0 & -4 \\ 0 & 0 & 2 & 6 \end{bmatrix}$$
$$\xrightarrow[R3/3]{R2/(-2)} \begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix} \xrightarrow{R1 \to R1 - R2 - R3} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Thus the solution to this linear system is $\begin{cases} x_1 = 1 \\ x_2 = 2 \\ x_3 = 3 \end{cases}$

□

Problem 18.2. How many solutions does the following linear systems of equations have

1. $\begin{cases} 5x_1 + 7x_2 = 3 \\ -10x_1 - 14x_2 = -3 \end{cases}$

2. $\begin{cases} 2x_1 - x_2 = 4 \\ x_1 - \frac{1}{2}x_2 = 2 \end{cases}$

Solution.

1. The augmented matrix is

$$\begin{bmatrix} 5 & 7 & 3 \\ -10 & -14 & -3 \end{bmatrix} \xrightarrow{R2 \to R2+2R1} \begin{bmatrix} 5 & 7 & 3 \\ 0 & 0 & 3 \end{bmatrix}$$

Since the last column has pivot, by Theorem 2.18, this linear system has no solutions

2.

$$\begin{bmatrix} 2 & -1 & 4 \\ 1 & -\frac{1}{2} & 2 \end{bmatrix} \xrightarrow{R2 \to R2-\frac{1}{2}R1} \begin{bmatrix} 2 & -1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

Since the last column has no pivot and there is a free variable x_2, by Theorem 2.18, this linear system has infinitely solutions

□

Problem 18.3. Consider the following matrix

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

1. Which columns are the pivot columns of A?
2. Write down the RREF of the this matrix

Solution.

1. The pivot columns of A are columns 1,3,4.
2.

$$A \xrightarrow[R3/2]{R2/(-1)} \begin{bmatrix} 1 & 2 & 2 & 3 & 1 \\ 0 & 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow[R2 \to R2+2R3]{R1 \to R1-3R3} \begin{bmatrix} 1 & 2 & 2 & 0 & -5 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{R1 \to R1-2R2} \begin{bmatrix} 1 & 2 & 0 & 0 & -11 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

□

Problem 18.4. Determine which of the following statements are true

1. The following matrix is of row reduced echelon form

$$\begin{bmatrix} 1 & 2 & 2 & 3 & 1 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & -1 & 2 & 1 \end{bmatrix}$$

2. The following two matrices are equivalent

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 2 \\ 2 & 4 & 1 & 2 & 4 \\ 1 & 1 & -3 & 2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 2 & 3 & 1 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & -1 & 2 & 1 \end{bmatrix}$$

Solution.

1. False. $(3,3)$, $(2,4)$-th entries are pivots which breaks the "staircase shape" $\begin{bmatrix} 1 & 2 & 2 & 3 & 1 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & -1 & 2 & 1 \end{bmatrix}$

2. False. Because

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 2 \\ 2 & 4 & 1 & 2 & 4 \\ 1 & 1 & -3 & 2 & 1 \end{bmatrix} \xrightarrow[R3-R1]{R2\to R2-2R1} \begin{bmatrix} 1 & -1 & 1 & 3 & 2 \\ 0 & 6 & -1 & -4 & 0 \\ 0 & 2 & -4 & -1 & -1 \end{bmatrix}$$

$$\xrightarrow{R2\to R2-3R3} \begin{bmatrix} 1 & -1 & 1 & 3 & 2 \\ 0 & 0 & 11 & 14 & 3 \\ 0 & 2 & -4 & -1 & -1 \end{bmatrix} \xrightarrow{R2\leftrightarrow R3} \begin{bmatrix} 1 & -1 & 1 & 3 & 2 \\ 0 & 2 & -4 & -1 & -1 \\ 0 & 0 & 11 & 14 & 3 \end{bmatrix}$$

has pivot columns 1,2,3, and

$$\begin{bmatrix} 1 & 2 & 2 & 3 & 1 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & -1 & 2 & 1 \end{bmatrix} \xrightarrow{R2\leftrightarrow R3} \begin{bmatrix} 1 & 2 & 2 & 3 & 1 \\ 0 & 0 & -1 & 2 & 1 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix}$$

has pivot columns 1,3,4. By Theorem 2.10, they are not equivalent, otherwise they would have the same RREF, which implies same pivot columns.

□

Problem 18.5. Determine the value(s) of h such that the matrix is the augmented matrix of a consistent linear system $\begin{bmatrix} 1 & h & 1 \\ 2 & 4 & 4 \end{bmatrix}$

Solution. First consider

$$\begin{bmatrix} 1 & h & 1 \\ 2 & 4 & 4 \end{bmatrix} \xrightarrow{R2\to R2-2R1} \begin{bmatrix} 1 & h & 1 \\ 0 & 4-2h & 2 \end{bmatrix}$$

By Theorem 2.18, the linear system has solutions $\iff$ the last column is not a pivot column $\iff 4-2h \neq 0 \iff h \neq 2$ □

Problem 18.6. Do the three lines $x_1 - 4x_2 = 1$, $2x_1 - x_2 = -3$, and $-x_1 - 3x_2 = 4$ have a common point of intersection? Explain.

Solution. Note that a common point of intersection would be a solution to the linear system $\begin{cases} x_1 - 4x_2 = 1 \\ 2x_1 - x_2 = -3 \\ -x_1 - 3x_2 = 4 \end{cases}$, consider its augmented matrix

$$\begin{bmatrix} 1 & -4 & 1 \\ 2 & -1 & -3 \\ -1 & -3 & 4 \end{bmatrix} \xrightarrow[R3\to R3+R1]{R2\to R2-2R1} \begin{bmatrix} 1 & -4 & 1 \\ 0 & 7 & -5 \\ 0 & -7 & 5 \end{bmatrix} \xrightarrow{R3\to R3+R2} \begin{bmatrix} 1 & -4 & 1 \\ 0 & 7 & -5 \\ 0 & 0 & 0 \end{bmatrix}$$

By Theorem 2.18, since the last column is not a pivot column, the linear system is consistent, i.e. these three lines has comon point(s) of intersection. □

References

Index