Linear Algebra
for Engineers

Parin Chaipunya
Department of Mathematics
KMUTT

Contents

List of Figures

Chapter 1

What is Linear Algebra all about ?

1.1 The course overview

Linear Algebra is a spin-off subject from the study of Linear Systems (or System of Linear Equations) through generalizations and re-interpretations.

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{mn}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned} \iff \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{mn} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \iff Ax = b.$$

In the above figure, we may observe a differnt view from a linear system into a simple matrix equation. This gives a new interpretation; instead of finding x that solves an equation, we find a vector x that is transformed (by a transformation A) into a new vector b.

The study of solutions of $Ax = b$ is then developed into the theory about vector space. The theory helps understanding a hidden structure that helps paving the way to computing.

Another aspect that is heavily studied revolves around matrices. Many modern applications in the digital world are the byproduct of the matrix theory. In particular, matrix is a systematic way to put data into places and matrix computations help dealing with those data — especially in the large scales.

This lecture note could be roughly divided into four parts.

Part 1. This is Chapter 2, which talks about linear systems, matrix equations and how to solve them.

Part 2. Consisting of Chapters 3–4, this part studies vector spaces, their algebraic structure and geometric structure.

Part 3. This third part includes Chapters 5–7. It starts with an interpretation of a matrix as a transformation, followed by different constants that are indications of the properties that a matrix carries. The factorizations into canonical forms are also discussed here.

Part 4. This final part is Chapter 8 which talks about some selected applications of linear algebra.

The classical lectures will also be accompanied with computer laboratory sessions to accommodate more complex calculations and to demonstrate the applications.

1.2 Matrix algebra: a quick review

1.2.1 The anatomy of a matrix

Let us give a quick summary of elementary matrix algebra. Recall first that a *matrix* is an array of numbers listed as a tableau as follows

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

This matrix is said to have *dimension* of $m \times n$, i.e. having m *rows* and n *columns*. The i^{th} row of this matrix is the array

$$\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix},$$

while the j^{th} column is the array

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

In this course, matrices are usually presented with capital roman alphabets, *i.e.*, $A, B, C, \ldots,$ *etc.*

The following illustration depicts the i^{th} row and j^{th} column of a matrix:

$$
\begin{array}{r}
\\ \\ i^{\text{th}}\text{ row} \longrightarrow \\ \\ \\
\end{array}
\begin{bmatrix}
\cdot & \cdot & \cdots & \cdot \\
\vdots & \vdots & \ddots & \vdots \\
a_{i1} & a_{i2} & \cdots & a_{in} \\
\vdots & \vdots & \ddots & \vdots \\
\cdot & \cdot & \cdots & \cdot
\end{bmatrix}
\qquad
\begin{array}{c}
\begin{bmatrix}
\cdot & \cdots & a_{1j} & \cdots & \cdot \\
\cdot & \cdots & a_{2j} & \cdots & \cdot \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\cdot & \cdot & a_{mj} & \cdot & \cdot
\end{bmatrix} \\
\uparrow \\
j^{\text{th}}\text{ column}
\end{array}
$$

1.2.2 Special classes of matrices

Vectors

Definition 1.1. A matrix that has a single column (i.e. of dimension $n \times 1$) is called a *column vector*. Likewise, a matrix that has a single row (i.e. of dimension $1 \times n$) is called a *row vector*.

In this course, we use the simple term *vector* to refer to a column vector. Vectors in this course are usually presented with lower-case letters like $a, b, c, \ldots$ or more often with $x, y, u, v, \ldots$,etc.

To save up spaces, a column vector x (with n entries) is usually written also as $x = (x_1, \ldots, x_n)$. Hence, writting

$$x = (x_1, \ldots, x_n)$$

is the same as

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Square matrices

Definition 1.2. A *square matrix* is the one having equal rows and columns.

Let $A = [a_{ij}]_{n\times n}$ be an $n \times n$ square matrix. The diagonal of A refers to the elements $a_{11}, a_{22}, \dots, a_{nn}$, as shown on the following figure:

$$\begin{bmatrix} a_{11} & \cdot & \dots & \cdot \\ \cdot & a_{22} & \dots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \dots & a_{nn} \end{bmatrix}$$

diagonal of A

Definition 1.3. A *diagonal matrix* is a square matrix whose nonzero entries are only on its diagonal.

If the elements along the diagonal are $d_1, d_2, \dots, d_n$, then we write $\operatorname{diag}(d_1, \dots, d_n)$ to denote such a diagonal matrix, that is

$$\operatorname{diag}(d_1, \dots, d_n) = \begin{bmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{bmatrix}.$$

Definition 1.4. The *identity matrix* of dimension $n \times n$ is the matrix I given by

$$I = \operatorname{diag}(\underbrace{1, \dots, 1}_{n \text{ times}}) = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}_{n\times n}$$

Definition 1.5. A matrix A is said to be an *upper triangular matrix* if all the entries below its diagonal are all 0. Likewise, it is said to be a *lower triangular matrix* if all the entries above its diagonal are all 0.

The following figure illustrates, respectively, an upper triangular matrix and a lower triangular matrix:

$$\begin{bmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{bmatrix} \qquad \begin{bmatrix} a_{11} & & 0 \\ & \ddots & \\ * & & a_{nn} \end{bmatrix}$$

We simply say that A is a *triangular* matrix if it is either an upper or a lower triangular matrix.

1.2.3 Matrix algebra

Transposition

Suppose that A is an $m \times n$ matrix given by

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = [a_{ij}]_{m \times n}.$$

The *transpose* of A, written as $A^\top$, is an $n \times m$ matrix defined by

$$A^\top = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nm} \end{bmatrix} = [a_{ji}]_{n \times m}.$$

It is clear that $(A^\top)^\top = A$.

A matrix A is said to be *symmetric* if $A^\top = A$. Of course, every symmetric matrix is a square matrix.

Scalar multiplication

We can multiply a scalar (a constant) to a matrix in a componentwise fashion. Let $c \in \mathbb{R}$ be a scalar and $A = [a_{ij}]_{m \times n}$ a matrix. Then the multiplication of c with A, written cA, is defined by

$$cA = [ca_{ij}]_{m \times n} = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{bmatrix}.$$

If $c = -1$, then we write $-A$ instead of $(-1)A$.

Matrix addition and subtraction

Adding/subtracting two or more matrices, all of them need to have the same dimension. Let $A = [a_{ij}]_{m \times n}$ and $B = [b_{ij}]_{m \times n}$ be any two matrices. Then $A \pm B$ is defined componentwise, that is

$$A \pm B = [a_{ij} \pm b_{ij}]_{m\times n} = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{22} & \dots & a_{mn} + b_{mn} \end{bmatrix}.$$

For matrices A, B, C with the same dimension, it is clear that

- $A - B = A + (-B)$,
- and $A \pm B = B \pm A$,
- $(A \pm B)^\top = A^\top \pm B^\top$,
- $(A \pm B) \pm C = A \pm (B \pm C) = A \pm B \pm C$,
- if c is a scalar, then $c(A \pm B) = cA \pm cB$.

Matrix multiplication

Now, we consider multiplying two matrices A and B, denoted AB. For this, we require the number of columns of A and the number of rows of B to be the same. Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{n\times r}$, then $AB = [c_{ij}]$ is the matrix of dimension $m \times r$ where the entry c_{ij} is defined by

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nr}.$$

The following illustration helps understanding the calculation of c_{ij}:

$$\begin{bmatrix} \cdot & \cdot & \dots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \dots & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \dots & b_{1j} & \dots & \cdot \\ \cdot & \dots & b_{2j} & \dots & \cdot \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \cdot & \cdot & b_{nj} & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} & \cdot & \\ \cdot & c_{ij} & \dots \\ & \vdots & \end{bmatrix}.$$

While m and r are not necessarily equal, BA is not even well-defined. For this reason alone, there is no reason that AB and BA are the same. In general, even for square matrices A and B, AB and BA are not equal.

Theorem 1.6. *Let A, B, C be two matrices. Then, whenever the above products are well-defined, we have*

- $A(B \pm C) = AB \pm AC$,
- $(A \pm B)C = AC \pm BC$,
- $(AB)C = A(BC) = ABC$,
- $(AB)^\top = B^\top A^\top$,
- $IA = A = AI$, (`The two` I`'s can be of different dimension!`)
- *if c is a scalar, then $c(AB) = (cA)B = A(cB)$,*

Let us practice a little bit to conclude this review session.

Exercise 1.7. Define

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -1 & -2 \\ -1 & 0 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \text{ and } C = \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix}.$$

Find

(a) AB (b) BA (c) BC

(d) AC (e) ABC (f) $B^\top A$

Chapter 2

Linear systems and matrix equations

In this chapter, we focus on systems of linear equations (or linear systems) and their reformulations into matrix equations through which all the analysis will be carried out. We shall study the classifications of the matrix equation $Ax = b$ based on its solution set being empty, singleton, or otherwise. Along the way, we study, among other things, the notions of echelon forms and rank. We shall also study the Gauss and Gauss-Jordan elimination methods which reduce any matrix into a row echelon and reduced row echelon forms, respectively. These methods are also used in finding solutions of a linear system.

Recall that a single-variable equation is said to be *linear* if it is in the form

$$ax = b.$$

This can be easily solved if $a \neq 0$ achieving $x = \frac{b}{a}$.

Having two variables, say x and y, we need two equations to fix the solution. In this way, we cannot just choose any x and y independently because they are linked by the given equations. This system of two equations are said to be a *linear system* if it is in the form

$$\begin{aligned} a_{11}x + a_{12}y &= b_1 \\ a_{21}x + a_{22}y &= b_2. \end{aligned}$$

To solve this system, we first eliminate a variable and fix the solution of the other then substitute back into one of the original equations.

The same principle applies to the case of linear systems with three variables, x, y and z, as the following form

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_1 \\ a_{21}x + a_{22}y + a_{23}z &= b_2 \\ a_{31}x + a_{32}y + a_{33}z &= b_3. \end{aligned}$$

To find a solution, we eleminate and substitute variables one by one.

We may now see the pattern: A linear system with n variables and n equations, namely $x_1, \ldots, x_n$, takes the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad\qquad & \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

Again, the same elimination-and-substitution strategy still works here, but with a lot more effort.

2.1 Solutions of a linear system

We start by asking ourselves a crucial question:

Is it necessary to have equal number of variables and equations?

The simple answer to the above question is "*No*". We may have n variables that are linked by m linear equations:

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots \qquad\qquad\qquad & \vdots & \\ \underbrace{a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n}_{n \text{ variables}} & = & b_m. \end{array}\right\} m \text{ equations}$$

The array of values $u_1, \ldots, u_n$ is called a *solution* of the above linear system if subsituting $x_1 = u_1, \ldots, x_n = u_n$ into the LHS gives RHS.

Let us consider some exercises below to refresh the idea of linear systems and how to solve them.

Exercise 2.1. Verify whether or not $x_1 = 3, x_2 = -1$ a solution of the following linear system

$$\begin{aligned} 2x_1 - 3x_2 &= 9 \\ x_1 + x_2 &= 2. \end{aligned}$$

Exercise 2.2. Verify whether or not $x_1 = 2, x_2 = 2$ a solution of the following linear system

$$\begin{aligned} x_2 - x_1 &= 0 \\ 2x_1 - 3x_2 &= -11. \end{aligned}$$

On the other hand, how do we find a solution of this system?

Now let us turn to solving a linear system instead of just verifying values.

Exercise 2.3. Find a solution to the following linear system

$$\begin{aligned} x_1 + x_2 - x_3 &= 1 \\ 2x_1 - x_2 + x_3 &= 2 \\ x_1 - x_2 - x_3 &= -1. \end{aligned}$$

Exercise 2.4. Find a solution to the following linear system

$$\begin{aligned} 3x_1 - 2x_2 + 2x_3 &= 1 \\ 2x_1 - x_2 + x_3 &= 2 \\ x_1 - x_2 &= -1. \end{aligned}$$

We should have observed that the method used to solve (and to further analyze) linear systems are different from problem to problem. To develop a consistent methodology, we first need to rewrite them as matrix equations.

2.2 Matrix equations

2.2.1 Transforming a linear system into a matrix equation

A linear system could always be transformed into a matrix equation. Given a linear system

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\vdots \qquad\qquad\qquad & \quad \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m,
\end{aligned}$$

we would define

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

We call A the *coefficient matrix*, b the *target vector*, and x an *unknown vector*. Then we have

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

and so the matrix equation

$$Ax = b$$

is equivalent to the linear system in question. When one analyze this equation, we usually construct the following *augmented matrix*

$$[A \,|\, b] = \left[\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right].$$

Note that the vertical bar separating A and b is purely notational. It is actually an $m \times (n+1)$ matrix where the first n columns are from A and the last is b.

2.2.2 Analyzing $Ax = b$

We are not only interested in solving $Ax = b$, but also to understand it.

The first and foremost behavior we would like to observe about the equation $Ax = b$ is the number of solutions it has. We may very roughly speak of *consistent systems* (those having a solution) and *inconsistent systems* (those without a solution), but we want a more detailed description in this course. This leads us to consider the three cases:

- (well-posed) $Ax = b$ has a *unique* solution,
- (under-determined) $Ax = b$ has *multiple* solutions,
- (over-determined) $Ax = b$ has *no* solution.

To convince that all the three cases could happen, let us do the following exercise.

Exercise 2.5. Let us observe that:

(a) Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$.
Then $Ax = b$ has a unique solution, which is $x^* = (1, 1, 1)$.

(b) Let $A = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$ and $b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
Then $Ax = b$ has multiple solutions. For example, $x^* = (0, 1, 0)$ and $x^{**} = (1, 2, 1)$ are both solutions for this system.

(c) Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ -1 & 4 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.
Then $Ax = b$ has no solution.

It is noticeably that the number of solutions depends on the relationship between the number of rows (equations) and the number of columns (variables). It is *casually understood* that

- a well-posed system has equal number of variables and equations,
- an under-determined system has less equations than variables,

- an over-determined system has more equations than variables.

However, to *really* use this verdict, we need to be sure that the systems is reduced into its simplest form, i.e. the *row echelon form*.

Let us do one more exercise to persuade ourselves that we need to reduce a system before judging it.

Exercise 2.6. Let $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$ and $b = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}$.
Then we have 3 variables and 3 equations, but multiple solutions. For instance, $x^* = (2, 0, 1)$ and $x^{**} = (3, 1, 0)$ are both solutions of this system. We will come back to explain this exercise again.

2.3 Echelon forms

2.3.1 Elementary row operations

Recall the operations that we did when solving linear systems in the previous exercises:

- multiplying a constant to an equation,
- adding or subtracting two equations,
- swapping order of equations.

These operations are transfered to the framework of matrix equations as *elemetary row operations* (or simply *row operations*), which include

- multiplying a constant to a row,
- adding or subtracting two rows,
- swapping rows.

2.3.2 Row echelon forms

Consider a matrix $M = [m_{ij}]_{m \times n}$. A row of M is called a *zero row* if it only consists of 0. A row that is not a zero row is called *non-zero row*. In a non-zero row, the first non-zero column in that row is called the *leading entry*.

Looking at the following matrix

$$P = \begin{bmatrix} 0 & 0 & 0 & \bigstar & \dots \\ 0 & 0 & \dots & \dots & 0 \\ 0 & \bigstar & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix},$$

we conclude that the rows #2 and #4 are zero rows of this matrix, while he circled entries ($\star$'s are any nonzero numbers) are the leading entries of their rows.

Definition 2.7. This matrix M is in a *row echelon form* if the following two conditions are satisfied:

(1) All zero rows are stacked at the bottom.

(2) In each non-zero row, its leading entry is on the left of all the leading entries below it.

One may simply observe this visually by stacking zero rows at the bottom and see that the leading entries are arranged in a staircase manner (see an example below with $\hat{P}$).

Notice that the above matrix P is not in an echelon form (why?). However, we may use row operations (swapping rows) to acheive an equivalent matrix in the following echelon form

$$\hat{P} = \begin{bmatrix} 0 & \bigstar & \dots & \dots & \dots \\ 0 & 0 & 0 & \bigstar & \dots \\ 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix}.$$

Definition 2.8. A matrix M is in a *reduced row echelon form* if it is in a row echelon form and the following additional conditions are satisfied:

(3) all leading entries are 1,

(4) all entries above a leading entry are 0.

Now let us practice checking if a given matrix is in a row echelon form or the reduced row echelon form.

Exercise 2.9. Determine whether or not the matrices given in the following are in row echelon and reduced row echelon form.

(a) $A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 1 \end{bmatrix}$ (b) $B = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 & 9 \\ 0 & 0 & -1 & 2 & -2 & 5 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$

(c) $C = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -2 & 5 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ (d) $D = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Let us state some facts that could be useful in future considerations.

Theorem 2.10. *The following facts hold true:*

- *Every matrix can be turned into a row echelon form using elementary row operations.*
- *A single matrix may be turned into several row echelon forms.*
- *every matrix can be reduced into a unique reduced row echelon form.*

Exercise 2.11. Use elementary row operations to find row echelon forms of the matrices from Exercise 2.9.

2.4 Gauss elimination

The aim of the Gauss elimination is to give a systematic procedure to simplify any matrix into a row echelon form. To do this, we eliminate as many nonzero entries as possible by the following the steps below:

1. Start with the first nonzero column. Pick a nonzero entry, swap that row to the top. Set $i = 1$.
2. Eliminate nonzero entries below the chosen entry from the previous step.
3. Find the next column with nonzero entry after the row i. Pick that nonzero entry and swap to the row $i + 1$.
4. Eliminate all the nonzero entries below the chosen entry from the previous step.

5. Set $i \leftarrow i + 1$ and repeat Steps 3–5 to complete all columns/rows.

Exercise 2.12. Use Gauss elimination to turn the following matrices into a row echelon form.

(a) $A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 1 \end{bmatrix}$ (b) $B = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 2 & 1 & 0 \\ 2 & 4 & 1 & 1 \\ 3 & 6 & 2 & 1 \end{bmatrix}$

(c) $C = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (d) $D = \begin{bmatrix} 2 & 0 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$

2.5 Gauss-Jordan elimination

The process of Gauss elimination reduces any given matrix into a row echelon form. However, if we would like to go further in obtaining a *reduced* row echelon form, we need a small extra computation in the Gauss elimination steps. This modified algorithm is known as the *Gauss-Jordan elimination* whose steps are listed as follows:

1. Start with the first nonzero column. Pick a nonzero entry, *make it into* 1 and swap that row to the top. Set $i = 1$.

2. Eliminate *all nonzero entries in that column* except the 1 from the previous step.

3. Find the next column with nonzero entry after the row i. Pick that nonzero entry, *make it into* 1 and swap to the row $i + 1$.

4. Eliminate *all the nonzero entries in that column* except the 1 from the previous step.

5. Set $i \leftarrow i + 1$ and repeat Steps 3–5 to complete all columns/rows.

Exercise 2.13. Use Gauss-Jordan elimination to obtain the reduced row echelon form of the following matrices.

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

2.6 Rank and solutions of $Ax = b$

We regard each row of a matrix as an information. A matrix is then a collection of information. Some of these information may be obtained from others and hence does not genuinely a new information. A *row echelon form* (or *reduced row echelon form*) of a matrix is viewed as a collection of *genuine* information in the sense that each row cannot be deduced from the remaining rows, hence no rows can be removed. It is natural that each matrix contains a certain amount of information, and this amount is fixed.

2.6.1 Rank

Before we define the rank of a matrix, let us point out the following fact.

Theorem 2.14. *Let A be a given matrix. Then*

- *all row echelon forms of A has the same number of nonzero rows,*
- *the reduced row echelon form of A is unique.*

With the consistence of nonzero rows in the above observation, one may define a rank of a matrix to be the number of nonzero rows in a row echelon form. That is, the rank is the number of genuine information that a matrix carries.

Definition 2.15. The *rank* of a matrix A is defined to be the number of nonzero rows of any row echelon form of A. The rank of A is denoted by $\operatorname{rank}(A)$.

Exercise 2.16. Find the rank of the matrices from the previous exercises.

2.6.2 Number of solutions

Recall that a row echelon form of a matrix reveals the minimal true information that it carries. Therefore, the number of solutions of a matrix equation $Ax = b$ can be observed from the row echelon form of the augmented matrix $[A \mid b]$. More precisely, the number of solutions of $Ax = b$ can be concluded from the rank($[A \mid b]$) as presented in the theorem below.

Theorem 2.17. *Consider the equation $Ax = b$ that corresponds to a system of n variables and m equations. Then the following conclusion is drawn:*

- $\operatorname{rank}(A) < \operatorname{rank}([A \mid b]) \iff Ax = b$ *has no solution.*
- $\operatorname{rank}(A) = \operatorname{rank}([A \mid b]) = n \iff Ax = b$ *has a unique solution.*
- $\operatorname{rank}(A) = \operatorname{rank}([A \mid b]) < n \iff Ax = b$ *has multiple solutions.*

The following flow chart helps the explaining the use of rank to determine the number of solutions of $Ax = b$.

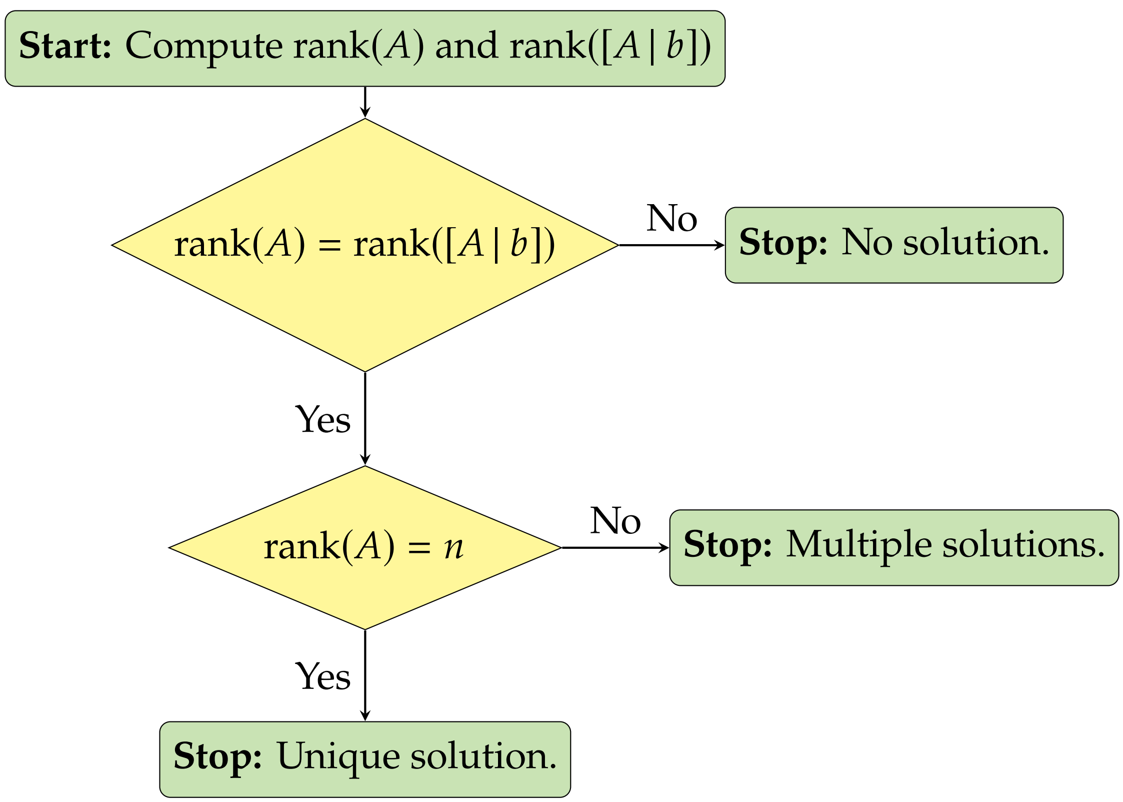

Figure 2.1: The flow chart for determining the number of solutions of $Ax = b$.

A practical aspect

Consider again the system represented by $Ax = b$ with n variables and m equations. In practice, to see if $\operatorname{rank}(A) < \operatorname{rank}([A \mid b])$ happens, we do not need to compute rank(A) and rank($[A \mid b]$) separately.

Consider directly the matrix $[A \mid b]$. When reducing $[A \mid b]$ to one of its row echelon form $[\hat{A} \mid \hat{b}]$, the $\hat{A}$ is always a row echelon form of A. Therefore, we also obtain a row echelon form of A in the process of obtaining a row echelon form of $[A \mid b]$. Then the following hold: If a row echelon form of $[A \mid b]$ consists of a row of the form $[0 \dots 0 \mid \star]$ (where $\star \neq 0$), then $\text{rank}(A) < \text{rank}([A \mid b])$ and so $Ax = b$ has no solution.

Let us rewrite the above flow chart in a more practical perspective.

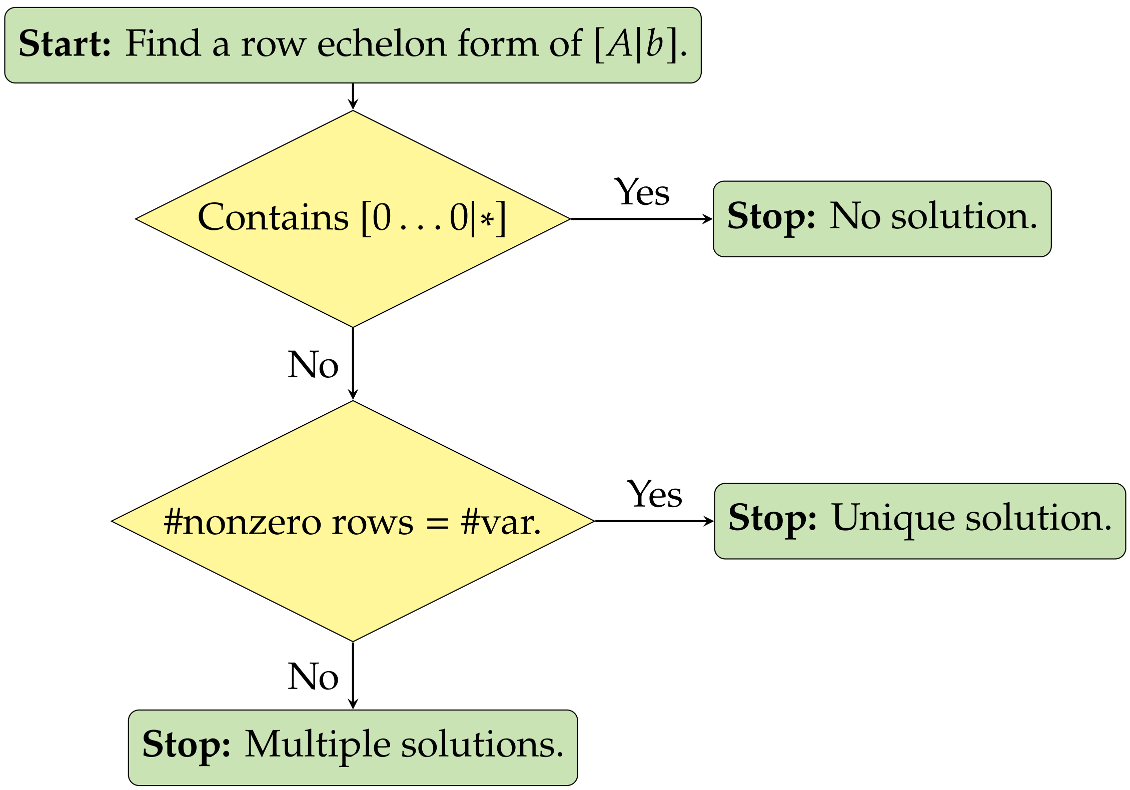

Figure 2.2: A practical flow chart for determining the number of solutions of $Ax = b$.

Exercise 2.18. Determine the number of solutions of the following system:

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

Exercise 2.19. Determine the number of solutions of the following system:

$$\begin{aligned} 2x_1 + x_2 - x_3 &= 2 \\ x_1 + x_2 &= 2 \\ x_2 + x_3 &= 0 \end{aligned}$$

Exercise 2.20. Determine the number of solutions of the following system:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 4 \\ 2x_1 + 3x_2 - x_3 - x_4 &= 3 \\ x_2 + x_3 &= 2 \\ 2x_1 - x_2 + x_3 - x_4 &= 1 \\ x_1 + x_2 - x_3 - 2x_4 &= 0 \end{aligned}$$

Exercise 2.21. Determine the number of solutions of the following system:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 4 \\ 2x_1 + 3x_2 - x_3 - x_4 &= 3 \\ x_2 + x_3 &= 2 \\ 2x_1 - x_2 + x_3 - x_4 &= 1 \\ x_1 + x_2 - x_3 - x_4 &= 0 \end{aligned}$$

2.7 Finding solutions of $Ax = b$.

Now, after being able to determine the number of solutions of $Ax = b$, it is time to actually find its solutions (if one exists). The strategy is to simplify the augmented matrix $[A|b]$ using either Gauss elimination or Gauss-Jordan elimination, and start analyzing from there.

Exercise 2.22. Solve the following linear system

$$\begin{aligned} x_1 + x_2 - x_3 &= 6 \\ x_1 - x_2 + x_3 &= -2 \\ 2x_1 - 2x_2 + 3x_3 &= -5 \end{aligned}$$

Exercise 2.23. Solve the following linear system

$$\begin{aligned} x_1 + x_2 - x_3 - x_4 &= 0 \\ x_1 - x_2 + x_3 + x_4 &= 2 \\ x_2 - 2x_3 &= -1 \\ x_1 + x_2 + 3x_3 + 2x_4 &= 7 \end{aligned}$$

Exercise 2.24. Solve the following linear system

$$\begin{aligned} 2x_1 + x_2 - 3x_3 &= 1 \\ x_1 - x_2 + x_3 &= 1 \\ 3x_1 - 2x_3 &= 2 \end{aligned}$$

Chapter 3

Vector spaces

Let us motivate the study of a vector space, again, from the equation $Ax = b$ that corresponds to a system of n variables and m equations. Let us begin with the special case where $b = 0$.

It turns out that when $Ax = 0$ has more than one solution, it has infinitely many of them. Let us demonstrate this fact. Let x^* and x^{**} both be solutions of $Ax = 0$. Take any numbers α and β. Then we get

$$A(\alpha x^* + \beta x^{**}) = \alpha A x^* + \beta A x^{**} = 0.$$

This means from two different solutions, we can generate infinitely many more by choosing different values of α and β.

Let $\mathcal{S}$ be the set that contains all the solutions of $Ax = 0$. It turns out that the shape $\mathcal{S}$ belongs to a particular class, *e.g.* a single dot, a straight line, a 2D flat plane, a 3D space, etc. We actually observe that all of them are *linear* in nature. In fact, we shall see subsequently that a solution set of the equation $Ax = 0$, if not empty, is a vector space (or a linear space).

Not limited to this, the concept of a linear space is also a backbone of modern mathematics which applies to power series expansion, Fourier series and transforms, Laplace transforms, differential equations, etc.

In this chapter, we introduce a general definition of a vector space but we will emphasize on the Euclidean space $\mathbb{R}^n$ and its subspaces. We shall also discuss notions related to a vector space including, among others, basis and dimension.

3.1 Vector spaces

Let us begine with the definition of a general vector space.

Definition 3.1 (General vector space)**.** Let V be a nonempty set (whose elements we call *vectors*) and K be a scalar field (whose elements we call *scalars* and K is usually either $\mathbb{R}$ or $\mathbb{C}$). Let $+$ and $\cdot$ be the *vector addition* (adding two vectors in V yields a vector in V) and *scalar multiplication* (multiplying a scalar in K and a vector in V yields a vector in V). Then the set V together with the operations $+$ and $\cdot$ (precisely written as the tuple $(V, +, \cdot)$) is said to be a *vector space over the scalar field* K if the following conditions are satisfied for all $u, v, w \in V$ and all $\alpha, \beta \in K$:

(V1) $u + v = v + u$;

(V2) $(u + v) + w = u + (v + w)$;

(V3) there is a *zero vector*, denoted by 0, such that $u + 0 = 0 + u = u$;

(V4) for each $u \in V$, there is a vector in V, denoted by $-u$ and called the *negative of* u, such that $u + (-u) = (-u) + u = 0$;

(V5) $\alpha(u + v) = \alpha u + \alpha v$;

(V6) $(\alpha + \beta)u = \alpha u + \beta u$;

(V7) $(\alpha\beta)u = \alpha(\beta u)$;

(V8) $1u = u$.

To clarify some frequent questions, the following remarks are in order.

- The zero vector and the number zero are both denoted by 0. Thus the reader must be aware of the context where 0 is used.
- The negative of each $u \in \mathbb{R}^n$ is unique, and it is given exactly by the formula $-u = (-1)u$.
- We avoid to use '$\cdot$' to denote multiplication because this could be confused with the *dot product* that we shall study later.

3.2 The space $\mathbb{R}^n$ and its subspaces

We denote $\mathbb{R}^n$ as the set of all n-dimensional real vectors $x = (x_1, \dots, x_n)$ where $x_i \in \mathbb{R}$ for each $i = 1, \dots, n$. Equipped with the vector addition '+' and scalar multiplication ' $\cdot$ ' (where we always omit the symbol ' $\cdot$ '), the following properties hold for every $u, v, w \in \mathbb{R}^n$ and every $\alpha, \beta \in \mathbb{R}$:

(E1) $u + v = v + u$;

(E2) $(u + v) + w = u + (v + w)$;

(E3) there is a *zero vector*, denoted by 0, such that $u + 0 = 0 + u = u$;

(E4) for each $u \in \mathbb{R}^n$, there is a vector in $\mathbb{R}^n$, denoted by $-u$ and called the *negative of* u, such that $u + (-u) = (-u) + u = 0$;

(E5) $\alpha(u + v) = \alpha u + \alpha v$;

(E6) $(\alpha + \beta)u = \alpha u + \beta u$;

(E7) $(\alpha\beta)u = \alpha(\beta u)$;

(E8) $1u = u$.

Therefore, $\mathbb{R}^n$ (with its usual vector addition and scalar multiplication) is a vector space. In what follows, the space $\mathbb{R}^n$ (called the *Euclidean space*) is always equipped with these classical operations unless explicitly stated otherwise.

One may observe the following development as n starts off from $n = 1$. We have:

- The space $\mathbb{R}^1 = \mathbb{R}$ which looks is a straight line (a 1D object).
- The space $\mathbb{R}^2$ which looks is a flat plane (a 2D object).
- The space $\mathbb{R}^3$ which is the 3D space we live in.
- The spaces $\mathbb{R}^n$ with $n \geq 3$ are understood intuitively (adding one more axis to $\mathbb{R}^3$ and so on). They are beyond our ability to draw.

3.3 Vector subspaces

Now let's see how a space $\mathbb{R}^n$ carries within itself. The space $\mathbb{R}^2$ actually contains many straight lines. These lines behave just like the real line $\mathbb{R}$, if it contains 0. Roughly speaking, the 2D space $\mathbb{R}^2$ carries smaller 1D vector spaces. Similarly, this 3D space $\mathbb{R}^3$ contains many straight lines as well as flat planes. In the same fashion, this $\mathbb{R}^3$ carries smaller 1D and 2D vector spaces. This goes on to $\mathbb{R}^4$, $\mathbb{R}^5$, etc. This phenomenon serves as the basic intuition of a vector subspace, that is, a subset which is again a vector space in itself.

Definition 3.2 (Vector subspace)**.** Suppose that V is a vector space. Then a subset $U \subset V$ is called a *vector subspace of* V (or simply a *subspace of* V) if U is a vector space with the vector addition and scalar multiplication inherited from V.

In practice, to check whether or not a subset U is a vector subspace, we use either the single-step criterion or the double-step criterion presented as follows.

Theorem 3.3 (Single-step subspace criterion)**.** *Let V be a vector space and $U \subset V$. If the following condition holds*

(S1) $\alpha u + \beta v \in U$ *for any* $u, v \in U$ *and any* $\alpha, \beta \in K$,

then U is a vector subspace of V.

Theorem 3.4 (Double-step subspace criterion)**.** *Let V be a vector space and $U \subset V$. If the following two conditions hold*

(D1) $u + v \in U$ *for all* $u, v \in U$,

(D2) $\alpha u \in U$ *for all* $u \in U$ *and* $\alpha \in K$,

then U is a vector subspace of V.

On the contrary, to show that $U \subset V$ is not a vector subspace, we need to find those elements $u, v \in U$ and $\alpha, \beta \in U$ that violate one of the criteria (S1), (D1) or (D2). Alternatively, we could also find the vectors (and possibly scalars) in which one of (V1)–(V8) is false.

Now that we know $\mathbb{R}^n$ is a vector space, let us explore some ot its subspaces.

Exercise 3.5. Show that $U = \{(x, y) \in \mathbb{R}^2 \mid x = 2y\}$ is a subspace of $\mathbb{R}^2$.

Exercise 3.6. Show that the set $U = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2^2\}$ is not a subspace of $\mathbb{R}^2$.

Exercise 3.7. Consider a set of the form

$$X = \{(x_1, x_2) \mid x_1 = ax_2 + b\},$$

where $a, b \in \mathbb{R}$ are constants. Show the following:

(a) When $b = 0$, then X is a subspace of $\mathbb{R}^2$ for any choice of a.

(b) When $b \neq 0$, then X is not a subspace of $\mathbb{R}^2$ for any choice of a.

In particular, we may conclude from the above facts that all straight lines passing through the origin are subspaces of $\mathbb{R}^2$ and the ones not passing are not.

Exercise 3.8. Verify which of the following sets are subspaces of $\mathbb{R}^2$.

(a) $W = \{(x, -x) \mid x \in \mathbb{R}\}$.

(b) $X = \{(x, -2x) \mid x \geq 0\}$.

(c) $Y = \{(x, y) \mid y \leq x\}$.

We may furthernotice that a set involves nonlinear operations volving one or more of the variables (e.g. multiplications, powers, square roots, exponents, logarithms, etc.) is usually not a linear space.

Exercise 3.9. Show that the following sets are not vector subspaces of $\mathbb{R}^2$.

(a) $M = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 x_2 = 0\}$.

(b) $P = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + \sqrt{2x_2} = 1\}$.

(c) $X = \{(x, y) \in \mathbb{R}^2 \mid x_1(1 - x_2) = -1\}$.

(d) $Y = \{(x, x^2) \mid x \in \mathbb{R}\}$.

Exercise 3.10. Which of the following are vector spaces.

(a) $X = \{(x, -2x + 2z, z) \mid x, z \in \mathbb{R}\}$.

(b) $Y = \{(x, y, z) \in \mathbb{R}^3 \mid xy = xz\}$.

(c) $Z = \{(x_1, \ldots, x_n) \mid x_1 = 0 \text{ and } x_n = 0\}$

We next consider in the upcoming exercises the two important vector spaces that are generated by a given matrix.

Exercise 3.11. Let A be any $m \times n$ matrix. Then the solution set of $Ax = 0$ is a vector subspace of $\mathbb{R}^n$. This set is actually called the *kernel* of A, or $\ker(A)$. In some text books, the kernel is also called the *null space*.

Exercise 3.12. Let A be any $m \times n$ matrix. Then the set of all products Ax is a vector subspace of $\mathbb{R}^m$. This set is actually called the *image* of A, or $\operatorname{im}(A)$.

3.4 Linear combination and linear span

To ultimately study basis and dimension of a vector space, we start with the notions of linear combination and linear span.

Definition 3.13 (Linear combination)**.** Let $(V, +, \cdot)$ be a vector space over the field $K \in \{\mathbb{R}, \mathbb{C}\}$ and $v_1, v_2, \dots, v_k$ are vectors in V. A *linear combination* of these vectors is any vector $v \in V$ of the form

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$$

where $\alpha_1, \dots, \alpha_k$ are scalars.

Example 3.14. Consider the vectors $v = (4, 0, 3)$, $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$ and $v_3 = (0, 0, 1)$ from $\mathbb{R}^3$. Then v is a linear combination of v_1 and v_3, but not of v_1 and v_2.

Exercise 3.15. Let $v = (5, 5, -5, 3)$, $v_1 = (1, 2, 0, 0)$, $v_2 = (-1, 1, 1, 1)$ and $v_3 = (-2, 0, 3, -1)$. Is v a linear combination of the other vectors v_1, v_2, v_3 ?

Exercise 3.16. Let $v = (2, 3, -1, 1)$, $v_1 = (1, 1, 3, 2)$, $v_2 = (-1, 2, 1, 0)$ and $v_3 = (2, 0, 0, -1)$. Is v a linear combination of the other vectors v_1, v_2, v_3 ?

A nice way to generate a vector subspace from a few vectors is to use the linear space, which is nothing else but all the possible linear combinations of such given vectors.

Definition 3.17 (Linear span)**.** Let V be a vector space over the field $K \in \{\mathbb{R}, \mathbb{C}\}$ and $U = \{v_1, v_2, \dots, v_k\} \subset V$. Then the *linear span* (or simply the *span*) of U is

$$\operatorname{span}(U) = \operatorname{span}(v_1, \dots, v_k) = \left\{ \sum_{i=1}^{k} \alpha_i v_i \mid \alpha_1, \dots, \alpha_k \in K \right\},$$

i.e. the set of all linear combinations of vectors in U. We also say that the set $W = \text{span}(U)$ is *generated* by $v_1, \ldots, v_k$, or that $v_1, \ldots, v_k$ are *generating vectors* for W.

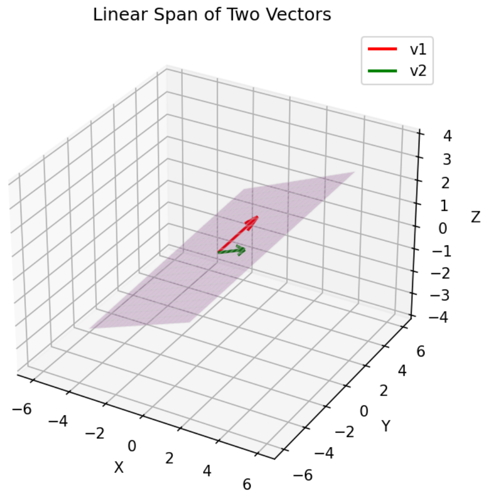

Figure 3.1: The linear span (pink plane) generated by the vectors v_1 (red) and v_2(green).

Exercise 3.18. Let V be a vector space and $U = \{v_1, \ldots, v_k\}$ be a subset of V. Show that span(U) is a vector subspace of V.

Exercise 3.19. Find span(U) where U is given by

(a) $U = \{(1,0),(0,1)\}$

(b) $U = \{(1,0),(1,1)\}$

(c) $U = \{(0,1,0),(2,0,2)\}$

Also notice that the same set may be generated by a different set of vectors.

At this point, one may also observe that a single vector v may be written as a linear combination of vectors $v_1, \ldots, v_k$ in more than one way. Hence the representation of v as

$$v = \alpha_1 v_1 + \cdots + \alpha_k v_k$$

may not be unique due to the available choices of the coefficients α_i's. Let us investigate this non-unique characteristic in the following example.

Example 3.20. Consider $v_1 = (1,1)$, $v_2 = (1,0)$ and $v_3 = (0,1)$. Then the vector $v = (2,2)$ can be obtained as a linear combination of v_1, v_2, v_3 by either

$$v = 0v_1 + 2v_2 + 2v_3$$

or

$$v = 1v_1 + 1v_2 + 1v_3$$

or

$$v = 4v_1 - 2v_2 - 2v_3$$

et cetera.

In this example, if we delete one of the vectors v_1, v_2 or v_3, then v can still be written as a linear combination of the remaining vectors, and, in a unique way. Let us do this as the next exercise.

Exercise 3.21. From the previous example...

(a) Find $\mathrm{span}(\{v_1, v_2, v_3\})$.

(b) Show that $v \in \mathrm{span}(\{v_1, v_2, v_3\})$.

(c) Show that if one of the vectors v_1, v_2, v_3 is dropped, then the new linear span equals the original one. Moreover, v can be written as a linear combination of the remaining vectors in a unique way.

(d) If we drop one more vector, then the generated span is a different one.

In this previous exercise, we see several worthy remarks. Let us list the important ones:

- $\mathbb{R}^2$ itself is a linear span.
- A subspace can be written as a linear span in several different ways.
- Some of the generating vectors are superfluous, and hence can be dropped. This means we *may* write the same linear span with less generating vectors.
- We cannot drop too many generating vectors while maintaining the same linear span. The minimal vectors left are *the most important ones*. In fact, these are *basis vectors* that we will later discuss.

3.5 Linear independence

We still need one more notion, the linear independence, to get ourselves ready for the discussion of basis vectors.

Definition 3.22 (Linear independence)**.** Let V be a vector space. The nonzero vectors $v_1, \ldots, v_k \in V$ are said to be *linearly independent* if the only coefficients $\alpha_1, \ldots, \alpha_k$ that makes

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$$

are given by $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.

If $v_1, \ldots, v_k$ are not linearly independent, we say that they are *linearly dependent*.

Exercise 3.23. Show that the vectors $v_1 = (1, 0)$ and $v_2 = (0, 1)$ are linearly independent.

Exercise 3.24. Show that the vectors $v_1 = (1, 1, 1)$, $v_2 = (1, 1, 0)$ and $v_3 = (1, 0, 0)$ are linearly independent.

Exercise 3.25. Show that the vectors $v_1 = (1, 1)$, $v_2 = (1, 0)$ and $v_3 = (0, 1)$ are linearly dependent (i.e. not linearly independent).

Exercise 3.26. Show that if $v_1, \ldots, v_k$ are linearly dependent, then at least one of these vectors can be written as a linear combination of the remaining ones with some nonzero coefficients.

3.6 Basis and dimension

The motivation to the notion of a *basis* of a vector space came from the idea that a minimal choice of vectors can be selected and form the same vector space. Notice that if the generative vectors of a vector space are linearly independent, then we have no chance dropping any vectors. In the other words, these vectors are already minimal.

3.6.1 Basis

Definition 3.27 (Basis)**.** Let V be a vector space. The vectors $v_1, \ldots, v_k \in V$ are called *basis vectors* of V if the following two conditions are satisfied

- $\operatorname{span}(\{v_1, \dots, v_k\}) = V$,
- $v_1, \dots, v_k$ are linearly independent.

The set $\{v_1, \dots, v_k\}$ is called a *basis* of V.

The following facts are noteworthy.

Theorem 3.28. *The following assertions hold true.*

- *Every vector space has a basis.*
- *All different bases of a vector space have the same number of elements.*

Exercise 3.29. Consider the vector space $\mathbb{R}^n$ and vectors $e_1, \dots, e_n \in \mathbb{R}^n$ given by

$$\begin{aligned} e_1 &= (1, \underbrace{0, 0, \dots, 0}_{n-1 \text{ times}}) \\ e_2 &= (0, 1, \underbrace{0, \dots, 0}_{n-2 \text{ times}}) \\ &\vdots \\ e_n &= (\underbrace{0, 0, \dots, 0}_{n-1 \text{ times}}, 1). \end{aligned}$$

Show that these vectors $e_1, \dots, e_n$ constitute a basis for $\mathbb{R}^n$.
This basis is known as the *standard basis* for $\mathbb{R}^n$.

Exercise 3.30. Find some other bases for $\mathbb{R}^n$.

Exercise 3.31. Find a basis for the vector space

$$V = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid 2x_1 - x_3 = 0,\ x_1 + 3x_2 = 0\}.$$

Exercise 3.32. Find a basis for the vector space

$$W = \{(x_1 + x_2 - x_3, x_3 - x_2, x_1) \in \mathbb{R}^3 \mid x_1, x_2, x_3 \in \mathbb{R}\}.$$

The next exercise gives an important insight, saying that a vector space of dimension n does not contain a proper subspace of equal dimension.

Exercise 3.33. Let V be a vector space with dimension $\dim(V) = n$. If $W \subset V$ is a subspace of V with $\dim(W) = n$, then $W = V$.

3.6.2 Basis representation

Let V be a vector space and consider a basis $\mathcal{B} = \{b_1, \ldots, b_n\}$. Then any vector $x \in V$ could be written uniquely as a linear combination of basis elements as

$$x = \alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n.$$

However, to make such an expression consistent, we need to fix the order of elements in $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ as such. The shorthand notation for this is given by

$$x = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}_{\mathcal{B}} = (\alpha_1, \ldots, \alpha_n)_{\mathcal{B}}.$$

When $V = \mathbb{R}^n$ is equipped with the standard basis $\mathcal{E} = \{e_1, \ldots, e_n\}$, we know that the basis representation $x = (x_1, \ldots, x_n)_{\mathcal{E}}$ is just the same as the classical expression $x = (x_1, \ldots, x_n)$. Therefore, it is agreed to write $x = (x_1, \ldots, x_n)$ without indicating the basis when the standard basis is used.

Exercise 3.34. Let us equip $\mathbb{R}^2$ with the basis $\mathcal{B} = \{(1,0),(1,1)\}$. Then write down the vectors $x = (3,2)$ and $y = (-2,5)$ in the representation of the basis $\mathcal{B}$.

Exercise 3.35. Suppose that $\mathbb{R}^3$ is equipped with the basis

$$\mathcal{B} = \{(1,0,0),(1,1,0),(1,1,1)\}.$$

Write the vector $x = (3,2,1)$ in the representation of the basis $\mathcal{B}$.

3.6.3 Dimension

Definition 3.36. Once again that we would like to emphasize that every basis of a vector space V contains the same number of basis vectors. This consistent number is referred to as the dimension of V, or briefly as $\dim(V)$.

Example 3.37. Recall from the above exercises that $\dim(\mathbb{R}^n) = n$, $\dim(V) = 1$ and $\dim(W) = 2$.

3.7 Change of basis

In some circumstance, one may want to switch from one basis to another to gain some advantages. In this section, we first see how to find a matrix that maps a representation from one basis into another basis. This change-of-basis matrix will serve as a fundamental tool for the further studies. Let us carry out a derivation of this matrix.

Let V be an n-dimensional vector space, $\mathcal{B} = \{b_1, \cdots, b_n\}$ and $\mathcal{B}' = \{b'_1, \cdots, b'_n\}$ be two bases of V.

Suppose that a vector $v \in V$ is represented in the basis $\mathcal{B}$ as

$$v = \alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n.$$

In other words, we have $v = (\alpha_1, \cdots, \alpha_n)_{\mathcal{B}}$. To write this vector v in a new basis $\mathcal{B}'$, $v = (\alpha'_1, \cdots, \alpha'_n)_{\mathcal{B}'}$, we write the equation

$$v = \alpha'_1 b'_1 + \alpha'_2 b'_2 + \cdots + \alpha'_n b'_n$$

and solve for the unknown coefficients $\alpha'_1, \cdots, \alpha'_n$. Putting the two equations above together and let

$$B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{bmatrix} \qquad B' = \begin{bmatrix} | & | & & | \\ b'_1 & b'_2 & \cdots & b'_n \\ | & | & & | \end{bmatrix}$$

$$y = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \qquad x = \begin{bmatrix} \alpha'_1 \\ \vdots \\ \alpha'_n \end{bmatrix},$$

we arrive at the casual linear system

$$B'x = By. \tag{3.1}$$

Since the vectors $b'_1, \cdots, b'_n$ form a basis, they are linearly independent and so $\operatorname{rank}(B') = n$ which means $(B')^{-1}$ exists. Finally, we may obtain the unknown coefficients from

$$x = (B')^{-1}By.$$

We conclude the change-of-basis formula as

$$v_{\mathcal{B}'} = [\mathcal{B}' \leftarrow \mathcal{B}]v_{\mathcal{B}},$$

where

$$[\mathcal{B}' \leftarrow \mathcal{B}] = (B')^{-1}B$$

is called the *change-of-basis matrix* from $\mathcal{B}$ into $\mathcal{B}'$.

To change back from $\mathcal{B}'$ to $\mathcal{B}$, the equation (3.1) can be used again but this time x is known and y is unknown. Similar analysis yields $y = B^{-1}B'x$. We may notice here that we may derive the change-of-basis matrix from $\mathcal{B}'$ to $\mathcal{B}$ as

$$[\mathcal{B} \leftarrow \mathcal{B}'] = B^{-1}B'.$$

It is interesting to observe that

$$[\mathcal{B}' \leftarrow \mathcal{B}] = (B')^{-1}B = [B^{-1}B']^{-1} = [\mathcal{B} \leftarrow \mathcal{B}']^{-1}.$$

Exercise 3.38. Let V be a vector space and $\mathcal{B}, \mathcal{B}', \mathcal{B}''$ are three bases of V. Show that $[\mathcal{B}'' \leftarrow \mathcal{B}] = [\mathcal{B}'' \leftarrow \mathcal{B}'][\mathcal{B}' \leftarrow \mathcal{B}]$.

Exercise 3.39. Suppose that $\mathcal{S}$ is the standard basis of $\mathbb{R}^n$ and $\mathcal{B}$ be any basis of $\mathbb{R}^n$. Find the change-of-basis matrices between the two bases $\mathcal{S}$ and $\mathcal{B}$.

Exercise 3.40. Consider $\mathbb{R}^2$. Let $\mathcal{B}$ be the standard basis on $\mathbb{R}^2$ and $\mathcal{B}' = \{(1,1),(-1,1)\}$ be another basis on $\mathbb{R}^2$.

(a) Find the change-of-basis matrices $[\mathcal{B}' \leftarrow \mathcal{B}]$ and $[\mathcal{B} \leftarrow \mathcal{B}']$.

(b) Write the vector $x = (4,-1)$ in the basis $\mathcal{B}'$.

(c) Write the vector $y = (0,2)_{\mathcal{B}'}$ in the standard basis.

3.8 Rank-nullity theorem

The rank-nullity theorem is one of the connerstone in linear algebra and it provides useful information in several circumstances.

The rank of any matrix A could actually be expressed alternatively by the following result.

Theorem 3.41 (The rank theorem)**.** *If A is a matrix, then* $\operatorname{rank}(A) = \dim(\operatorname{im}(A))$.

We may define the *nullity* of a matrix following this line.

Definition 3.42. If A is a matrix, then its *nullity* of A, denoted by $null(A)$, is defined by $null(A) = \dim(\ker(A))$.

Finally, we may link the relationship between the rank and nullity in the following well-known theorem.

Theorem 3.43 (Rank-nullity theorem)**.** *If A has n columns, then*

$$\operatorname{rank}(A) + null(A) = n.$$

Exercise 3.44. Consider the following linear system:

$$\begin{aligned} 3x_1 + 2x_2 - 3x_4 &= 0 \\ 2x_2 + x_4 &= 0 \\ 2x_2 - 2x_3 - x_4 &= 0 \end{aligned}$$

Use the rank-nullity theorem to conclude the dimension of the solution set of the above system.

3.9 The geometry of the solution set of $A = b$

We have investigated already that the solution set of $Ax = b$ is a linear space if and only if $b = 0$. Moreover, the solution set of $Ax = 0$ receives a special treatment as $\ker(A)$. However, one could observe that if $A\bar{x} = b$ has a solution x_0, then the solution set of $Ax = b$ is given by

$$x_0 + \ker(A) = \{x_0 + y \mid y \in \ker(A)\}.$$

This means the solution set of $Ax = b$ is a *translation* of $\ker(A)$, so that they have the same dimension and shape.

Exercise 3.45. Consider the following linear system:

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

Use the rank-nullity theorem to draw conclusion about the geometry of the solution set of the above system.

Practice problems

1. Show that the vectors in each of the following items are linearly dependent.
    - $v_1 = (1,2,0), v_2 = (0,1,1)$.
    - $v_1 = (0,1,0), v_2 = (0,0,1), v_3 = (1,1,0)$.
    - $v_1 = (1,0,0,1), v_2 = (1,0,1,0), v_3 = (1,0,0,0), v_4 = (0,2,1,0)$.
2. Show that the vectors in each of the following items are *not* linearly independent.
    - $v_1 = (0,1,1), v_2 = (-1,-1,-1), v_3 = (1,0,0)$.
    - $v_1 = (2,2,0,0), v_2 = (0,1,0,1), v_3 = (2,1,0,3), v_4 = (-2,1,0,2)$.
3. Find a basis for the following vector spaces and determine their dimensions.
    - $V = \{x \in \mathbb{R}^3 \,|\, x_1 = -x_3,\ x_2 = -x_1\}$.
    - $V = ker\left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}\right)$.
    - $V = \{(x_1 + x_3, -2x_2) \,|\, x_1, x_2, x_3 \in \mathbb{R}\}$.
    - $V = \{(3x_1, x_1 + x_2 - x_3, x_3) \,|\, x_1, x_2, x_3 \in \mathbb{R}\}$.
    - $V = \{x \in \mathbb{R}^3 \,|\, 2x_1 - 3x_2 = 0,\ x_1 + 2x_3 = 0\}$.
4. Find dim(ker(A)) and dim(im(A)) of the following instances of a matrix A. *(Hint: Make use of the rank-nullity theorem.)*
    - $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}$
    - $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 2 & 0 & 0 \end{bmatrix}$.
    - $A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$.
    - $A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}$.

5. Determine the dimension of the solution set of the following linear system:

$$\begin{array}{rcrcrcrcl} 2x_1 & & & + & x_3 & & & = & 2 \\ & & x_2 & + & x_3 & - & x_4 & = & 2 \\ x_1 & & & & & - & x_4 & = & 0. \end{array}$$

6. Suppose that A is a matrix such that $\dim(\ker(A)) = 1 + \dim(\operatorname{im}(A))$. Show that A has odd number of rows. (*Hint: Use the rank-nullity theorem.*)

Chapter 4

Norms, dot products and orthogonality

In this chapter we concern with some geometry on vector spaces. Geometry in this context refers to *measurements*, where we mainly consider *length* and *angle*.

4.1 Norms

A *norm* is a function defined on a vector space with certain properties. A norm us usually interpreted as the *size* or *length* of an element in the underlying vector space.

Definition 4.1 (Norm)**.** Let V be a real vector space. A norm is any function $\|\cdot\| : V \to \mathbb{R}$ such that the following properties hold true for any vectors $x, y \in V$ and scalar $\alpha \in \mathbb{R}$:

(1) $\|x\| \geq 0$ and $\|x\| = 0 \iff x = 0$,

(2) $\|\alpha x\| = |\alpha|\|x\|$,

(3) $\|x + y\| \leq \|x\| + \|y\|$.

There could be several norms on a single vector space. Depending on the circumstance, a particular norm could be more appropriate than the other ones. In this note, we only focus on the norms for vectors in $\mathbb{R}^n$ and for matrices in $\mathbb{R}^{m\times n}$.

4.1.1 Norms on $\mathbb{R}^n$

Let us take a look at some common choices of norms on the vector space $\mathbb{R}^n$.

Definition 4.2 (p-Norms)**.** For $1 \leq p < \infty$, the *p-norm* of a vector $x \in \mathbb{R}^n$ is defined by

$$\|x\|_p = \sqrt[p]{|x|_1^p + \cdots + |x|_n^p}.$$

For $p = \infty$, the *∞-norm* (or the *max norm*) of x is defined by

$$\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

Among these p-norms, the case $p = 2$ is the most important and $\|x\|_2$ is also known as the *Euclidean norm* on $\mathbb{R}^n$.

Remark. Since the Euclidean norm is the most common one used in the literature, we shall write just $\|x\|$ instead of $\|x\|_2$ for simplicity.

Exercise 4.3. Show in $\mathbb{R}^2$ and $\mathbb{R}^3$ that the value $\|x\|_2$of the Euclidean norm is exactly the length of the vector x.

Exercise 4.4. Let $u = (3, 0, 1)$ and $v = (-2, 1, 2)$ be two vectors in $\mathbb{R}^3$.

(a) Find the lengths of the two vectors u and v.

(b) Find $\|u\|_1$ and $\|v\|_\infty$.

4.1.2 Norms on $\mathbb{R}^{m\times n}$

In the same way that $\mathbb{R}^n$ could be equipped with different norms, there are various choices of norms to be used in the vector space $\mathbb{R}^{m\times n}$.

The first and simplest norm is the *p-norms* which resembles the p-norms in $\mathbb{R}^n$.

Definition 4.5 (p-Norms for matrices)**.** Let $1 \leq p < \infty$. The *p-norm* of a matrix $A \in \mathbb{R}^{m\times n}$ is defined by

$$\|A\|_p = \sqrt[p]{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p}.$$

In case $p = \infty$, the *∞-norm* (also called the max-*norm*) of A is defined by

$$\|A\|_\infty = \max_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} |a_{ij}|.$$

The 2-norm is also known as the *Frobenius norm* and sometimes written as $\|A\|_F$.

Exercise 4.6. Compute the Frobenius norm of a matrix $A = \begin{bmatrix} 1 & 2 & -1 & 1 \\ -2 & -2 & 0 & -1 \end{bmatrix}$.

Definition 4.7 (Unit vectors)**.** A vector $u \in \mathbb{R}^n$ is said to be *unit* if $\|u\| = 1$.

The following process of scaling a nonzero vector into a unit vector of the same direction is called *normalization*.

Exercise 4.8 (Normalization)**.** Show that for any nonzero vector $u \in \mathbb{R}^n$, the vector $u^\circ = \frac{u}{\|u\|}$ is unit. Here, we say that the vector u is *normalized* into u°.

4.2 Dot products and orthogonality

Dor products provide a higher level of understanding of the vector space at hand. One shall see that it is also closely related to the norms. Moreover, the dot product is the key element that allows one to define and study orthogonality. Note that the orthogonality has been an important concept in modern applications concerning data and information theory.

4.2.1 Dot product

Let us start formally with the definition of the (Euclidean) dot product.

Definition 4.9. The *dot product* between two vectors $u, v \in \mathbb{R}^n$ is defined by

$$u \cdot v = u_1 v_1 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i = u^\top v.$$

The following theorem summarizes the useful properties when one computes the dot products. We also highlight the connections between the dot product and the Euclidean norm in the following theorem.

Theorem 4.10. *The following holds for any vectors $u, v, w \in \mathbb{R}^n$ and scalars $\alpha \in \mathbb{R}$:*

- $u \cdot v = v \cdot u$,
- $(\alpha u) \cdot v = u \cdot (\alpha v) = \alpha(u \cdot v)$,
- $u \cdot (v + w) = u \cdot v + u \cdot w$,

- $u \cdot u = 0 \iff u = 0$,
- $u \cdot u = \sum_{i=1}^{n} u_i^2 = \|u\|^2 \geq 0$.

Let us practice some calculations.

Exercise 4.11. Let $u = (1, -1, 2, 4)$ and $v = (2, 4, -1, a)$, where a is a constant. What constant a makes $u \cdot v = 0$?

Another important relationship between the dot product and the Euclidean norm is the following theorem, which provides an easy formula to compute the angle between two nonzero vectors.

Theorem 4.12. *Let $u, v \in \mathbb{R}^n$ be two nonzero vectors. Then*

$$u \cdot v = \|u\|\|v\| \cos\theta,$$

where θ is the angle between u and v.

In particular, we may obtain the following formula for θ:

$$\theta = \arccos\left(\frac{u \cdot v}{\|u\|\|v\|}\right).$$

Exercise 4.13. Find the angle between the two vectors $u = (1, 1, 1)$ and $v = (0, 1, 0)$.

4.2.2 Orthogonality

In many situations, we are only interested to know whether the two vectors are *orthogonal* (or *perpendicular*). In such a case, we simply observe whether the dot product vanishes. More generally, the sign of the dot product provides some rough information about the angle between two nonzero vectors.

Theorem 4.14. *If $u, v \in \mathbb{R}^n$ are two nonzero vectors and θ the angle between them, then*

- $u \cdot v = 0$ *if and only if* $\theta = 0$,
- $u \cdot v > 0$ *if and only if* $\theta \in [0, \pi/2)$,
- $u \cdot v < 0$ *if and only if* $\theta \in (-\pi/2, 0]$.

Exercise 4.15. Let $u = (1, 1)$, $v = (2, -1)$, $x = (-2, 1)$ and $y = (3, 6)$. Find out all the orthogonal pairs among these vectors.

Let us now speak of a set of pairwise orthogonal vectors.

Definition 4.16 (Orthogonal family)**.** Let $\mathcal{U} = \{v_1, \ldots, v_m\}$ be a subset of $\mathbb{R}^n$. We say that $\mathcal{U}$ is an *orthogonal family* if

$$v_i \cdot v_j = 0, \quad (i \neq j)$$

for all $i, j = 1, \ldots, m$.

Example 4.17. Show that $\mathcal{U} = \{(1, -2, 0), (4, 2, 1), (2, 1, -10)\}$ is an orthogonal family.

A more practical way to verify that $\mathcal{U}$ is an orthogonal family follows from the following theorem.

Theorem 4.18. *Let $\mathcal{U} = \{v_1, \ldots, v_m\}$ be a family of vectors in $\mathbb{R}^n$ and let U be a matrix whose columns are vectors in $\mathcal{U}$. Then $\mathcal{U}$ is an orthogonal family if and only if $U^\top U$ is a diagonal matrix.*

Proof. Suppose that $U^\top U = [w_{ij}]_{m\times m}$. Then we have $w_{ij} = v_i^\top v_j = v_i \cdot v_j$. At any $i \neq j$, we have $w_{ij} = 0$ if and only if v_i and v_j are orthogonal. ∎

It is important to point out that orthogonality implies independence.

Exercise 4.19. Suppose that $\mathcal{U}$ is an orthogonal family of nonzero vectors. Show that its elements are linearly independent.

4.3 Orthogonal projection

If we look at two vectors, one may have a feeling that the first vector have a hint of the information that the second one provides. For example, a vector $(1, 1)$ is going slightly *right*, which means it has a flavor of another vector $(1, 0)$. In the same way, it also has a flavor of $(0, 1)$. On the other hand, the two vectors $(1, 0)$ and $(0, 1)$ has no flavors of one another. This is because they are orthogonal. The flavor we are observing right now is formally called the *orthogonal projection*.

4.3.1 Orthogonal projection onto a vector

We begin with the orthogonal projection of a vector onto another vector.

Definition 4.20. Let $v \in \mathbb{R}^n$ be a nonzero vector. The *orthogonal projection of a vector $u \in \mathbb{R}^n$ onto v* is defined by

$$\mathrm{proj}_v(u) = (u \cdot v)\frac{v}{\|v\|^2} = \frac{u \cdot v}{v \cdot v}v.$$

Exercise 4.21. Let $u = (1, 0)$ and $v = (1, 1)$. Find $\mathrm{proj}_v(u)$.

Exercise 4.22. Let $u = (-1, 1, 0)$ and $v = (1, -1, 2)$. Verify that both $\mathrm{proj}_v(u)$ and $\mathrm{proj}_u(v)$ are both the zero vector.

The following fact provides the instrumental idea for the development of the Gram-Schmidt orthogonalization process in the next section.

Exercise 4.23. If $u, v \in \mathbb{R}^n$ are nonzero vectors, show that $u - \mathrm{proj}_v(u)$ is orthogonal to v.

4.3.2 Orthogonal projection onto a subspace

Sometimes we need to make a projection onto a subspace rather than just a single vector (like in principal component analysis). This is achieved by taking the sum of orthogonal projections onto each basis elements of the subspace in question.

Theorem 4.24. *If V is a subspace of a vector space $\mathbb{R}^n$, then the orthogonal projection of $u \in \mathbb{R}^n$ onto V is obtained by the formula*

$$\mathrm{proj}_V(u) = \sum_{j=1}^{m} \mathrm{proj}_{v_j}(u),$$

where $\mathcal{V} = \{v_1, \ldots, v_m\}$ is a basis of V.

Exercise 4.25. Let $V = \mathrm{span}((1, 1, -1), (1, 2, 1))$. Find the orthogonal projection of the vector $u = (1, 1, 1)$ onto V.

4.4 Gram-Schmidt orthogonalization process

Motivated by the observation made in the previous section, we present now a process called *Gram-Schmidt orthogonalization process* which creates an orthogonal basis out of any basis by removing the *flavor* of all the previous basis elements.

Recall first that a basis for a vector space is said to be an *orthogonal basis* if it is also an orthogonal family.

Theorem 4.26 (Gram-Schmidt orthogonalization process)**.** *Let V be a vector space with dot product and let $\mathcal{U} = \{v_1, \cdots, v_n\}$ be a basis of V. We define*

$$\begin{aligned}
v_1' &= v_1 \\
v_2' &= v_2 - \mathrm{proj}_{v_1'}(v_2) \\
v_3' &= v_3 - \mathrm{proj}_{v_2'}(v_3) - \mathrm{proj}_{v_1'}(v_3) \\
&\vdots \\
v_i' &= v_i - \mathrm{proj}_{v_{i-1}'}(v_i) - \cdots - \mathrm{proj}_{v_1'}(v_i) \\
&\vdots \\
v_n' &= v_n - \mathrm{proj}_{v_{n-1}'}(v_n) - \cdots - \mathrm{proj}_{v_1'}(v_n).
\end{aligned}$$

Then the family $\mathcal{U}' = \{v_1', \cdots, v_n'\}$ is an orthogonal basis for V.

Remark. The Gram-Schmidt orthogonalization process is order-sensitive. If we re-index the basis elements, then the resulting orthogonal basis could be a different one.

Exercise 4.27. Consider $\mathbb{R}^2$ with the basis $\mathcal{U} = \{(1,-3),(-1,0)\}$. Use the Gram-Schmidt process to create an orthogonal basis from $\mathcal{U}$.

Exercise 4.28. Consider $\mathbb{R}^3$ with the basis $\mathcal{B} = \{(1,0,0),(1,0,1),(0,-2,0)\}$. Orthogonalize $\mathcal{B}$ using the Gram-Schmidt process.

Practice problems

1. Find the lengthiest and shortest vectors among the following:
   - $v_1 = (3, 2, -1)$,
   - $v_2 = (1, -3, 4)$,
   - $v_3 = (-2, 3, 1)$,
   - $v_4 = (1, 0, 4)$.
2. Let $u = (1, -2, 0)$ and $v = (0, 1, 1)$. Find the angle between the two vectors u and v.
3. Consider the following bases of $\mathbb{R}^3$:
   - $\mathcal{B}_1 = \{(1, 0, 0), (0, 0, 1), (0, -1, 1)\}$
   - $\mathcal{B}_2 = \{(0, -1, 1), (1, 1, 0), (0, -1, 0)\}$
   - $\mathcal{B}_3 = \{(1, 0, 1), (0, 1, 2), (1, 1, 0)\}$

   Which of these are orthogonal ?
4. Use the Gram-Schmidt process to orthogonalize the non-orthogonal bases from the previous question.

Chapter 5

Linear transformations

There is a special class of functions that maps between two vector spaces, called *linear transformation*. We will discover later that this class is special because it can be fully described with matrices in the Euclidean setting. This characterization also brings about a new interpretation to the equation $Ax = b$.

Definition 5.1. Let V and W be two vector spaces over the field $K \in \{\mathbb{R}, \mathbb{C}\}$. A function $T : V \to W$ is said to be a *linear transformation* if the condition

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

holds true for all $x, y \in V$ and $\alpha, \beta \in K$.

In this definition, it should be remarked that the additions and multiplications appeared on both sides of the equation, i.e. $\alpha x + \beta y$ and $\alpha T(x) + \beta T(y)$, are additions and multiplications on different spaces V and W, respectively.

Exercise 5.2. Show that the following functions are linear transformations.

(a) $T : \mathbb{R}^2 \to \mathbb{R}$ defined by $T(x) = T(x_1, x_2) = 2x_1 - x_2$.

(b) $T : \mathbb{R} \to \mathbb{R}^2$ defined by $T(x) = (2x, x)$.

(c) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x) = T(x_1, x_2, x_3) = (2x_1 - x_2, x_2 - 2x_3)$.

(d) $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x) = (x_2, x_1)$.

Exercise 5.3 (Matrices as linear transformations.)**.** Let A be an $m \times n$ matrix. Show that the function $T : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$T(x) = Ax$$

is a linear transformation. With this fact, we can now view the matrix A as a function that transforms $x \in \mathbb{R}^n$ into $Ax \in \mathbb{R}^m$.

Exercise 5.4. Let V and W be two vector spaces and $T : V \to W$ be a linear transformation. Show that $T(0) = 0$. Again, one should be cautious that 0 on the LHS and RHS of the equation belongs to different vector spaces V and W, respectively.

We may use the converse of the above fact to disprove linearity of a function: If $T(0) \neq 0$, then T is not a linear transformation.

Exercise 5.5. Show that the following functions are not linear transformations.

(a) $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = 3x + 1$.

(b) $T : \mathbb{R} \to \mathbb{R}^2$ defined by $T(x) = (x, x^2)$.

(c) $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x) = (x_1 + x_2, x_1 x_2)$.

(d) $T : \mathbb{R}^2 \to \mathbb{R}$ defined by $T(x) = \sin(x_1) + \cos(x_2)$.

Exercise 5.6. Determine whether or not the following functions are linear transformations.

(a) $T(x_1, x_2) = \sin^2(x_1) + \cos^2(x_2)$.

(b) $T(x) = \sin^2(x) + \cos^2(x)$.

(c) $T(x_1, x_2) = \sin^2(x_1) + \cos^2(x_1) - (1 + x_2)$.

5.1 Matrix representation

We have already seen that an $m \times n$ matrix defines a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$. In this section, we show that the converse is also true. That is, any linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$ defines an $m \times n$ matrix. This matrix representation plays a very important role in the analysis of a linear transformation itself.

5.1.1 The formula for the matrix representation

We have seen that every matrix can be seen as a linear transformation. Now, we show that every linear transformation can also be seen as a matrix. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then there is an $m \times n$ matrix A_T in which

$$T(x) = A_T x \qquad (\forall x \in \mathbb{R}^n).$$

This matrix A_T is called the *matrix representation of* T (or shortly the *matrix of* T) and can be formulated by

$$A_T = \begin{bmatrix} | & | & & | \\ T(e_1) & T(e_2) & \dots & T(e_n) \\ | & | & & | \end{bmatrix}$$

where $e_1, \dots, x_n$ are vectors in the standard basis of $\mathbb{R}^n$.

Exercise 5.7. Find the matrix representations of all the linear transformations from Exercise 5.2.

5.1.2 Kernel and range

Definition 5.8. Given a linear transformation $T : V \to W$. The *kernel* (or *null space*) of T is given by

$$\ker(T) = \{x \in V \mid T(x) = 0\}.$$

Likewise, the *range* of T is

$$\operatorname{ran}(T) = \{y \in W \mid \exists x \in V : y = T(x)\} = \{T(x) \mid x \in V\}.$$

It can be noticed that $\ker(T) = \ker(A_T)$ and $\operatorname{ran}(T) = \operatorname{im}(A_T)$, hence they both are linear subspaces. Moreover, we may

- compute the kernel $\ker(T)$ by solving the linear system $A_T x = 0$,
- compute the range $\operatorname{ran}(T)$ as the linear span of the columns of A_T.

This relationship allows the application of the rank-nullity theorem in the context of linear transformations.

Exercise 5.9. Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T(x_1, x_2, x_3) = (2x_1 - x_2, x_1 + x_2).$$

Find $\ker(T)$ and $\operatorname{ran}(T)$.

5.2 Vector space of all linear transformations

Following the following result, we deduce that the set of all linear trasformations between two vector spaces is actually a vector space.

Theorem 5.10. *Let V and W be two vector spaces over the field $K \in \{\mathbb{R}, \mathbb{C}\}$. We define*

$$\mathscr{L}(V, W) = \{\textit{All linear transformations } T : V \to W\}.$$

For any $S, T \in \mathscr{L}$ and scalar $\alpha \in K$, we define addition $S + T$ and scalar multiplication αT by

$$(S+T)(x) = S(x) + T(x) \quad \textit{and} \quad (\alpha T)(x) = \alpha T(x).$$

Then $\mathscr{L}(V, W)$ is a vector space over the field K under the above albegra.

If $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, then each linear transformation $T : V \to W$ is equivalent to a matrix A_T. Hence the vector space $\mathscr{L}(V, W)$ is then reduced to

$$\mathbb{R}^{m \times n} = \{\text{All matrices of dimension } m \times n\}.$$

Hence we write $\mathscr{L}(V, W) \approx \mathbb{R}^{m \times n}$ to denote this equivalence.

Exercise 5.11. Consider the space $\mathbb{R}^{m \times n}$. For $i = 1, \ldots, m$ and $j = 1, \ldots, n$, define E_{ij} to be the matrix whose element at the position (i, j) is 1 and all 0 elsewhere. Show that these matrices form a basis for $\mathbb{R}^{m \times n}$.

5.3 Composite linear transformation

Now, let us consider compositions of linear transformations.

Exercise 5.12. Suppose that U, V, W are three vector spaces and let $T : U \to V$ and $S : V \to W$ be two linear transformation. Show that the composition $S \circ T : U \to W$, defined by

$$S \circ T(x) = S(T(x)),$$

is also a linear transformation.

Exercise 5.13. Following the previous exercise:

(a) Show that $\ker(T) \subset \ker(S \circ T)$.

(b) Give an example where we have $\ker(T) \neq \ker(S \circ T)$.

(c) Give an example where we have $\ker(T) = \ker(S \circ T)$.

Exercise 5.14. Follwing the above:

(a) Show that $\operatorname{ran}(S \circ T) \subset \operatorname{ran}(S)$.

(b) Give an example where we have $\operatorname{ran}(S \circ T) \neq \operatorname{ran}(S)$.

(c) Give an example where we have $\operatorname{ran}(S \circ T) = \operatorname{ran}(S)$.

5.4 Linear transformation in computer graphics

Many of the transformation in computer graphics are actually linear transformation. The most important and fundamental ones are the flips and rotations.

5.4.1 2D Flippings

Flipping refers to the situation where a point is mirrored across a given linear plane. The most basic ones are the flippings across the x- and y-axis, respectively.

Flipping across x-axis

When a point $(x, y) \in \mathbb{R}^2$ is flipped across the x-axis, it is moved to $(x, -y)$. Hence we define the flipping across the x-axis (the plane $y = 0$) by a function $\mathcal{F}_{(y=0)} : \mathbb{R}^2 \to \mathbb{R}^2$, with

$$\mathcal{F}_x(x, y) = (x, -y).$$

This is clearly a linear transformation. The matrix representation of $\mathcal{F}_{(y=0)}$ will be denoted by $F_{(y=0)}$ and is given by

$$F_{(y=0)} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Flipping across y-axis

Similarly, a point $(x, y) \in \mathbb{R}^2$ is flipped across the y-axis (the plane $x = 0$) means it is moved to $(-x, y)$. Then we define this flip by a linear transformation $\mathcal{F}_{(x=0)} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ with

$$\mathcal{F}_{(x=0)}(x, y) = (-x, y).$$

Its matrix representation is then derived as

$$F_{(x=0)} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Flipping across any plane $ax + by = 0$

One could also construct a linear transformation of flipping a point across an arbitrary plane given by an equation $ax + by = 0$. The process involves the use of plane geometry, and the resulting flipping matrix is

$$F_{(ax+by=0)} = \frac{1}{a^2 + b^2} \begin{bmatrix} b^2 - a^2 & -2ab \\ -2ab & a^2 - b^2 \end{bmatrix}.$$

The following figure illustrates the flipping of points in $\mathbb{R}^2$ across the plane $x + y = 0$.

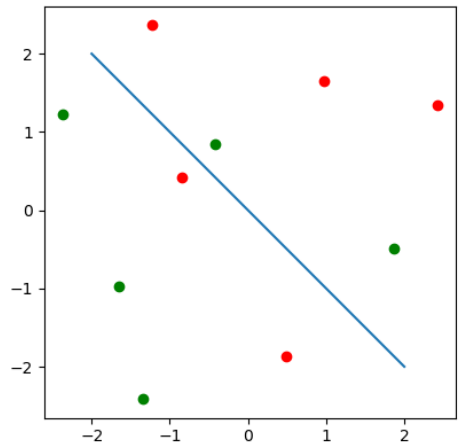

Figure 5.1: Green points flipped across the plane $x + y = 0$ into red points.

5.4.2 2D rotations

Here, we consider an operator that rotates a point $(x, y) \in \mathbb{R}^2$ about the origin by a certain angle θ in the counter-clowise direction. Let us write $\mathcal{R}_\theta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ to denote the 2D rotation. To derive the formula of $\mathcal{R}_\theta$, we write (x, y) in the polar coordinate so that

$$x = r\cos(\theta_0) \quad \text{and} \quad y = r\sin(\theta_0)$$

for some $r \geq 0$ and $\theta_0 \in [0, 2\pi)$. Then we have

$$\mathcal{R}_\theta(x, y) = (r\cos(\theta_0 + \theta), r\sin(\theta_0 + \theta)).$$

Next, we derive that

$$\begin{aligned} r\cos(\theta_0 + \theta) &= r[\cos(\theta_0)\cos(\theta) - \sin(\theta_0)\sin(\theta)] \\ &= [r\cos(\theta_0)]\cos(\theta) - [r\sin(\theta_0)]\sin(\theta) \\ &= x\cos(\theta) - y\sin(\theta) \end{aligned}$$

and also

$$\begin{aligned} r\sin(\theta_0 + \theta) &= r[\sin(\theta_0)\cos(\theta) + \cos(\theta_0)\sin(\theta)] \\ &= [r\sin(\theta_0)]\cos(\theta) + [r\cos(\theta_0)]\sin(\theta) \\ &= x\sin(\theta) + y\cos(\theta). \end{aligned}$$

Hence we have

$$\begin{aligned}\mathcal{R}_\theta(x, y) &= (x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta)) \\ &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.\end{aligned}$$

This shows that $\mathcal{R}_\theta$ is a linear transformation and its matrix representation is given by

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

Practice problems

1. Determine whether the following functions are linear transformations.

   ◦ $T(x_1, x_2) = (2x_1, 3x_1 - x_2)$.
   ◦ $T(x_1, x_2, x_3) = (0, 1, x_1 + x_2 + x_3)$.
   ◦ $T(x_1, x_2) = (x_1 - x_2^2 + x_1x_2, 2 + x_2)$.
   ◦ $T(x_1, x_2, x_3) = \cos\left(\frac{3\pi}{2}\right)(x_1 + x_2 + x_3)$.
   ◦ $T(x_1, x_2) = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$.

2. Suppose that A is an $m \times n$ matrix and $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is defined by

$$T(x) = Ax.$$

   Show that the matrix representation of T is $A_T = A$.

3. Find a matrix representation of the following linear transformations.

   ◦ $T(x_1, x_2) = (x_1, x_2, x_1, x_2)$.
   ◦ $T(x_1, x_2, x_3) = x_1 + x_2 + x_3$.
   ◦ $T(x_1, x_3, x_3, x_4) = (x_1 + x_3, -x_2 - 3x_4)$.

4. Find $\dim(ker\,T)$ and $\dim(ran\,T)$ of T from Exercise 3 by using the rank-nullity theorem.

5. Find $ker\,T$ and $ran\,T$ of T from Exercise 3.

Chapter 6

Determinant and invertibility

The determinant is a scalar value assigned to a *square* matrix. Many properties of a matrix is encoded withing its *determinant*. The most important one would be the *invertibility* of a matrix (or of the corresponding linear transformation). The determinant can also be used to explain how a transformation actually *transforms* the space. The determinant of a matrix A is denoted by $\det(A)$ or $|A|$.

6.1 Sarrus rules

The determinant can be defined for any $n \times n$ matrices, but let us first focus on the 2×2 and 3×3 cases as they can be easily computed using the *Sarrus rule*.

6.1.1 2×2 matrices

Let A be a 2×2 matrix. Then its determinant can be computed by

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

This formula is known as the 2×2 *Sarrus rule*.

Exercise 6.1. Find the determinant of $A = \begin{bmatrix} 1 & 3 \\ -2 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix}$. Note that $\det(A)$ can be positive, negative and 0, and this is independent of the sign of elements in a matrix.

6.1.2 3×3 matrices

Let A be a 3×3 matrix. Then its determinant can be computed by

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

This formula is known as the 3×3 Sarrus rule. The following mnemonic device is useful for this rule:

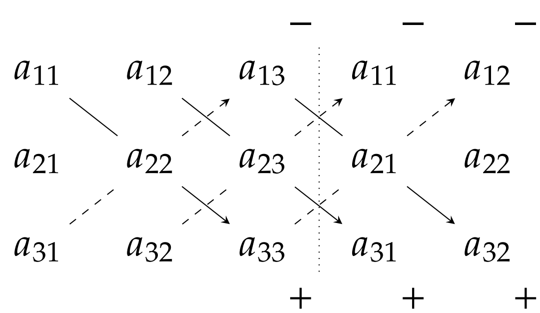

Figure 6.1: The mnemonic diagram for 3×3 Sarrus rule.

Exercise 6.2. Find $\det(A)$ where A is given by

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 3 \\ 0 & 1 & 0 \end{bmatrix}.$$

6.2 General definition

The situation is quite different when it comes to determinant of matrices or larger than 3×3. There is no mnemonic device for this and it requires the immediate notions of minor and cofactor matrices.

Definition 6.3 (Minor matrix)**.** Let A be an $n \times n$ matrix. We write A_{-ij} to denote the $(n-1) \times (n-1)$ matrix obtained by removing Row#i and Col#j from A. The *minor matrix* of A, denoted by $M(A)$, is an $n \times n$ matrix

$$M(A) = [M_{ij}]_{n \times n}$$

whose element M_{ij} is computed by

$$M_{ij} = \det(A_{-ij}).$$

Definition 6.4 (Cofactor matrix)**.** The *cofactor matrix* of A, denoted by $C(A)$, is an $n \times n$ matrix

$$C(A) = [C_{ij}]_{n\times n}$$

whose element C_{ij} is computed by

$$C_{ij} = (-1)^{i+j} M_{ij} = (-1)^{i+j} \det(A_{-ij}).$$

We are now ready for the formal definition of the determinant of any $n \times n$ matrix.

Definition 6.5 (Determinant)**.** The *determinant* of A can be computed by fixing any row i and calculate

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in},$$

or equivalently by fixing any column j and calculate

$$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij} = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

One should notice now the *inductive* nature of minors and cofactors of A. For instance, to compute the minor (and the cofactor) of a 4×4 matrix, we need to compute determinants of 3×3 matrices. Similarly, computing the minor (and the cofactor) of a 5×5 matrix requires the determinants of 4×4 matrices, etc.

Let us warm up with a 3×3 matrix wihtout using the Sarrus rule.

Exercise 6.6. Let

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

Find $\det(A)$ using different rows/columns.

Let us move on to larger matrices.

Exercise 6.7. Find the determinant of the following matrices.

(a) $A = \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -3 \\ 3 & -1 & -2 & 3 \\ 1 & 1 & 2 & 0 \end{bmatrix}$ (b) $B = \begin{bmatrix} 2 & 0 & 0 & 3 & -2 \\ 1 & 1 & 0 & -3 & 1 \\ 3 & 0 & 0 & 3 & 1 \\ -1 & -1 & 0 & 0 & 2 \\ 1 & 0 & 2 & 0 & 4 \end{bmatrix}$

Let us now state some useful properties of the determinant.

Theorem 6.8. *Suppose that A and B are square $n \times n$ matrices, then..*

- *If A contains a zero row or a zero column, then* $\det(A) = 0$.
- *If A contains two identical rows or columns, then* $\det(A) = 0$.
- *If A is a triangular matrix, then* $\det(A) = a_{11}a_{22}\dots a_{nn}$.
- $\det(AB) = \det(A)\det(B)$
- $\det(A^\top) = \det(A)$
- $\det(cA) = c^n \det(A)$ *for any scalar $c \in \mathbb{R}$.*

From the last property, it implies that $\det(-A) = (-1)^n \det(A)$ so that we have $\det(-A) = -\det(A)$ when n is odd and $\det(-A) = \det(A)$ when n is even.

Also note that the determinant is not distributed under summation, that is, $\det(A + B)$ and $\det(A) + \det(B)$ are independent.

Exercise 6.9. Let $A = \begin{bmatrix} 1 & 0 & 3 & 1 \\ 0 & 11 & 3 & -4 \\ 0 & 0 & 11 & 385 \\ 0 & 0 & 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} -3 & 0 & 0 & 0 \\ 2 & 28 & 0 & 0 \\ -98 & 42 & 11 & 0 \\ 13 & 0 & -56 & 2 \end{bmatrix}$. Find $\det(2AB)$.

Exercise 6.10. Find two matrices A and B such that $\det(A + B) \neq \det(A) + \det(B)$.

Exercise 6.11. Find two matrices A and B such that $\det(A + B) = \det(A) + \det(B)$.

6.3 Block triangular matrices

We may extend the technique known for triangular matrices to a more general cae of block triangular matrices.

Any matrix A can be subdivided into $p \times q$ blocks in such a way that

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1q} \\ A_{21} & A_{22} & \dots & A_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \dots & A_{pq} \end{bmatrix},$$

with appropriate dimensions for each submatrices A_{kl}'s.

We say that A is in an *upper block triangular form* if A can be subdivided into $p \times p$ blocks with each block A_{kl} being a square matrix and

$$A = \begin{bmatrix} A_{11} & & & * \\ 0 & A_{22} & & \\ \vdots & & \ddots & \\ 0 & \dots & 0 & A_{pp} \end{bmatrix},$$

where $*$ means any values are acceptable. Likewise, we can define a *lower block triangular matrix* in the same way. We simply say that A is a *block trinangular matrix* if it is either upper or lower block traingular.

The following theorem simplifies the calculation the determinant of a block triangular.

Theorem 6.12. *If A is a block triangular matrix with all diagonals $A_{11}, \dots, A_{pp}$ being square matrices, then* $\det(A) = \det(A_{11})\det(A_{22})\dots\det(A_{nn})$.

Exercise 6.13. Find $\det(A)$ where $A = \begin{bmatrix} 2 & 3 & 4 & 7 & 8 \\ -1 & 5 & 3 & 2 & 1 \\ 0 & 0 & 2 & 1 & 5 \\ 0 & 0 & 3 & -1 & 4 \\ 0 & 0 & 5 & 2 & 6 \end{bmatrix}$.

6.4 Determinant, rank and invertibility.

It is interesting to see how the determinant is related more to the equation $Ax = b$ or even to the matrix A itself.

Definition 6.14 (Invertibility)**.** Given a square $n \times n$ matrix A. We say that

- A has a *full rank* if $\operatorname{rank}(A) = n$;
- A is *invertible* if there exists an $n \times n$ matrix, denoted with A^{-1}, satisfying $AA^{-1} = A^{-1}A = I$. In this case, A^{-1} is called the *inverse matrix* of A.

One may suspect an existence of a matrix B in which $AB = I$ but $BA \neq I$ (or vice verse). It turns out for a square matrix that such exceptional situation would not happen.

The following result shows a useful equivalence between concepts we have seen.

Theorem 6.15. *The following conditions are equivalent:*

- *A has a full rank.*
- *A is invertible.*
- $\det(A) \neq 0$.

In the case where A is invertible (which is actually where $\det(A) \neq 0$*), the inverse matrix of A can be calculated by*

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

where $adj(A) = C(A)^\top$ is called the adjoint *of A.*

Exercise 6.16. Let A be a 2×2 invertible matrix. Find a general formula for A^{-1}.

Exercise 6.17. Find the inverse of the following matrices.

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

We have the following properties concerning inverse matrices.

Theorem 6.18. *Let A and B be two $n \times n$ matrices.*

- *If A is invertible, then the equation $Ax = b$ has a unique solution for all $b \in \mathbb{R}^n$. Moreover, the solution is given by $x = A^{-1}b$.*

- *If both A and B are invertible, then the product AB is also invertible and*

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

- *If one of A or B is not invertible, then the product AB is not invertible.*

Exercise 6.19. Find a solution of the following linear system by using the inverse of the coefficient matrix.

$$\begin{aligned} 2x_1 &= 4 \\ x_1 - 2x_3 &= -3 \\ x_2 &= 1. \end{aligned}$$

Exercise 6.20. Let $A = \begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} -2 & 1 \\ 2 & 0 \end{bmatrix}$. Find $(AB)^{-1}$.

6.5 Inverse linear transformation

As we now know that linear transformations (on $\mathbb{R}^n$) and matrices are *equivalent* concepts. Their invertibility can be concluded from the matrix representations

Definition 6.21 (Invertible transformation)**.** A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if there is a function $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ in which $T \circ T^{-1}(x) = T^{-1} \circ T(x) = x$ for all $x \in \mathbb{R}^n$. The function T^{-1} here is called the inverse transformation (or briefly the inverse) of T.

The most important property that we need to address is that the inverse T^{-1} is also a linear transformation. This fact follows from the following result.

Theorem 6.22. *If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation whose matrix representation is A_T, then the inverse T^{-1} of T is given by the inverse matrix A_T^{-1}. From this, we know that T^{-1} is also a linear transformation.*

Exercise 6.23. Let $T(x) = T(x_1, x_2, x_3) = (x_1, x_1 + x_2, x_1 + x_3)$ be a linear transformation. Find the inverse transformation of T.

Beyond invertibility, the determinant could also provide other information about a linear transformation. We conclude the section with the following characterizations of nonzero determinant.

Theorem 6.24. *Let T be a linear transformation whose matrix representation is A_T. Then*

- $\det(A_T) \neq 0$ *if and only if* $\ker(T) = \{0\}$.
- T *is one-to-one if and only if* $\ker(T) = \{0\}$.
- $\det(A_T) \neq 0$ *if and only if T is* one-to-one *and* onto.

6.6 Geometry of determinant

Let A be a matrix that induces a linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation. Then $\det(A)$ encodes how the space $\mathbb{R}^n$ is deformed under the application of T_A. We can also note from the rank-nullity theorem that T_A cannot explode the space in the sense that it is impossible to gain additional dimensions by the transformation T_A.

More precisely, the determinant $\det(A)$ informs us the following information:

- The absolute value $|\det(A)|$ is the area/volume of the parallelotope generated by columns of A. Since columns of A are $T_A(e_i)$'s, we obtain the information of how the parallellotope generated by $e_1, \ldots, e_n$ are enlarged/shrinked through the transformation T_A.
- The sign of $\det(A)$ tells about the change of orientation of the transformation. If $\det(A)$ is negative, then the transformation does not preserve the orientation of the axes.

Exercise 6.25. Consider a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ whose matrix representation is

$$A_T = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}.$$

Find $\det(A_T)$ and explains how T transform the box bounded by the standard basis of $\mathbb{R}^2$.

Practice problems

1. Compute the determinants of the following matrices.

   ◦ $A = \begin{bmatrix} 2 & 7 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

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

   ◦ $C = \begin{bmatrix} 1 & 12 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ -1 & 5 & 0 & 1 & 2 \\ -11 & 2 & 1 & 1 & 0 \\ 1 & -4 & 1 & 1 & 1 \end{bmatrix}$.

2. Which of the above matrices are invertible?

3. Consider if the linear transformation

$$T(x_1, x_2, x_3) = (x_1 - x_2, x_2 + 2x_3, 2x_1 - x_3)$$

   is invertible. If T is invertible, find T^{-1}.

4. Show that the linear transformation

$$T(x_1, x_2) = (2x_1 - x_2, x_2 - 2x_1)$$

   is not invertible.

5. Find T^{-1} of the following linear transformation

$$T(x_1, x_2, x_3, x_4) = (x_4, x_2, x_3, x_1).$$

Chapter 7

Eigenvalues and eigenvectors

Eigenvalues and eigenvectors are important concepts that has long list of applications. For example, it can be used to transform a differential equation into a polynomial. In data analytics, eigenvectors represents important information that is extracted from a square matrix. This technique is known as principal component analysis, which is one of the main tool to extract information from a large pool of data. In this course, we will see only an immediate use of eigenvectors in the construction of basis that can be used for storing large matrix with limited memory.

7.1 Eigenvalues and eigenvectors

Definition 7.1 (Eigenvalues and eigenvectors)**.** Consider a square $n \times n$ matrix. A scalar λ is called an eigenvalue of A if there exists a nonzero vector v (again, possibly complex) for which

$$Av = \lambda v.$$

For an eigenvalue λ of A, all nonzero vectors v in which $Av = \lambda v$ are called eigenvectors of A associated with λ. A pair (λ, v) is called an eigenpair if λ is an eigenvalue of A and v is an eigenvector associated to λ.

Geometrically, if we regard A as a transformation, then (λ, v) is a pair of eigenvalue-eigenvector of A if and only if the transformed vector Av lies on the same line as the original vector v.

In many situtations, the eigenvalues λ's are allowed to be complex numbers even though the matrix is real. However, this complex eigenvalues fall outside the scope of this lecture note. In this course, we shall only focus on real eigenpairs.

Before we proceed to finding eigenvalues and the associated eigenvectors of a matrix, let us show that eigenvectors associated to each eigenvalue, together with the zero vector, form a vector space.

Exercise 7.2. Let A be an $n \times n$ matrix and $\lambda \in \mathbb{R}$ be an eigenvalue of A. Then the set

$$E_\lambda = \{0\} \cup \{\text{All eigenvectors of } \lambda.\} = \{v \in \mathbb{R}^n \mid Av = \lambda v\}$$

is a vector space, which is called the *eigenspace* associated to λ.

Exercise 7.3. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then $\lambda = 1$ is an eigenvalue of A and the associated eigenvectors are any vectors along the x-axis. Also describe this geometrically.

Now, let us turn to a practical aspect of finding eigenvalues and their associated eigenvectors. When one tries to solve the eigenvalue-eigenvector equation

$$Av = \lambda v,$$

a difficulty arises from the fact that both scalar λ and nonzero vector v are unknowns. Hence the right-hand-side product λv implies that the above equation itself is *nonlinear*. However, we can rearrange the said equation into the following form

$$(A - \lambda I)v = 0. \tag{7.1}$$

In order to obtain $v \neq 0$, it is necessary to have

$$\det(A - \lambda I) = 0. \tag{7.2}$$

Notice that the above equation is free of the unknown vector v. Writing down the determinant further shows that (7.2) is in fact a polynomial, known as the *characteristic polynomial* of A. The *fundamental theorem of algebra* then gurantees that (7.2) always has repeatable n complex solutions (provided that A is of dimension $n \times n$).

Once eigenvalues λ are sorted out from the characteristic polynomial (7.2), we return to (7.1) with the known λ's. This makes (7.1) simply a linear system for each eigenvalues λ.

Finally, let us write down a summary of steps to find all eigenvalues and eigenvectors of a matrix A:

Step 1. **Find all eigenvalues λ's** by solving $\det(\lambda I - A) = 0$.

Step 2. **Finding eigenvectors v's associated to each** λ by solving the linear systems $(A - \lambda I)v = 0$.

Let us practice finding the eigenvalues and eigenvectors using the above procedures.

Exercise 7.4. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then A has no real eigenvalues.

Exercise 7.5. Find all eigenvalues and eigenvectors of a matrix $A = \begin{bmatrix} -1 & 0 \\ 2 & 0 \end{bmatrix}$.

The characteristic equation of an $n \times n$ matrix always has the degree of n. It is often that the (real) root of the equation is repeated, at which case we have *repeated* eigenvalues. The following are two examples where we have repeated eigenvalues.

Exercise 7.6. Find all eigenvalues and eigenvectors of the following matrices.

(a) $A = \begin{bmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{bmatrix}$ (b) $B = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$

We end the section with a list of fundamental properties of the eigenvalues and eigenvectors that could be useful for future use.

Theorem 7.7. *Consider a square matrix A.*

- *If (λ_1, v_1) and (λ_2, v_2) are eigenpairs of A with $\lambda_1 \neq \lambda_2$, then v_1 and v_2 are linearly independent.*
- *If $\det(A) = 0$, then $\lambda = 0$ is one of the eigenvalues of A.*
- *If λ is an eigenvalue of A, then $k\lambda$ is an eigenvalue of kA.*
- *If λ is an eigenvalue of A and $n \in \mathbb{N}$, then λ^n is an eigenvalue of A^n.*
- *A and $A^\top$ have the same eigenvalues.*
- *If A is invertible and λ is an eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}.*

- *If A is a triangular matrix, then eigenvalues of A are the elements in the diagonal of A.*
- *If A is symmetric, then all eigenvalues of A are real.*
- *Let $\lambda_1, \cdots, \lambda_n$ be all the eigenvalues of A (repeatable), then*

$$\sum_{i=1}^{n} \lambda_i = tr(A) = \sum_{i=1}^{n} a_{ii} \quad \textit{and} \quad \prod_{i=1}^{n} \lambda_i = \det(A).$$

7.2 Eigenbasis

One of the most important features from the eigenpairs is the possiblity to construct a new meaningful basis, called the *eigenbasis*, that simplifies the computation related to a given matrix.

Definition 7.8 (Eigenbasis)**.** Let A be a square matrix of dimension $n \times n$. If A has linearly independent *real* eigenvectors $v_1, \ldots, v_n$, then these eigenvectors form a basis $\mathcal{E} = \{v_1, \ldots, v_n\}$. This special basis is called the *eigenbasis* generated by A.

Remark. According to the previous section, if A has n distinct *real* eigenvalues, then the eigenvectors associated to each eigenvalues will automatically be linearly independent.

Exercise 7.9. Find the eigenbases associated to each of the following matrices.

(a) $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ (b) $B = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & -1 \end{bmatrix}$ (c) $C = \begin{bmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{bmatrix}$

A stronger result could be drawn when A is symmetric, and this is known as the spectral theorem.

Theorem 7.10 (Spectral theorem)**.** *If A is an $n \times n$ symmetric matrix, then A has eigenvectors $v_1, \ldots, v_n$ that form an orthonormal basis on $\mathbb{R}^n$. In simpler words, the matrix A has an orthonormal eigenbasis.*

Let us confirm the spectral theorem with one of the matrices above.

Exercise 7.11. Verify that the eigenbasis for $B = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & -1 \end{bmatrix}$ is orthonormal.

7.3 Diagonalization

Diagonalization is the process that changes the basis under which a square matrix becomes a diagonal matrix. While it is not always possible to diagonalize every matrix, there could be more than one way to diagonalize a matrix. However, it turns out that a matrix is always diagonalized under its own eigenbasis. Let us formally state these ideas below using a more tractable approach.

Definition 7.12. A $n \times n$ matrix A is *diagonalizable* if there is an $n \times n$ invertible matrix U and an $n \times n$ diagonal matrix D such that

$$A = UDU^{-1}.$$

Here, since U is invertible, the columns of U forms a basis $\mathcal{U}$ on $\mathbb{R}^n$. Moreover, the two matrices U and U^{-1} could be interpreted, respectively, as the change-of-basis matrices $[\mathcal{S} \leftarrow \mathcal{U}]$ and $[\mathcal{U} \leftarrow \mathcal{S}]$. With this view, the diagonal matrix D is the representation of A in the basis $\mathcal{U}$, that is

$$D = [\mathcal{U} \leftarrow \mathcal{S}]A[\mathcal{S} \leftarrow \mathcal{U}] = A_{\mathcal{U}}^{\mathcal{U}}.$$

Theorem 7.13. *If an $n \times n$ matrix A has n independent real eigenvectors $v_1, \dots, v_n$ which are associated respectively to eigenvalues $\lambda_1, \dots, \lambda_n$, then A is diagonalizable and we have*

$$A = V\Lambda V^{-1},$$

where V is the matrix whose i^{th} column is v_i, and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$.

Exercise 7.14. Diagonalize the following matrices.

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

Diagonalization has some uses in matrix computation. The first use case is when one wants to compute a matrix power A^k when k is large. Instead of computing the matrix power A^k, if A is diagonalized into $A = V\Lambda V^1$, then

$$\begin{aligned} A^k &= \underbrace{(V\Lambda V^{-1})(V\Lambda V^{-1})\cdots(V\Lambda V^{-1})}_{k \text{ times}} \\ &= V\Lambda \underbrace{[(V^{-1}V)\Lambda][(V^{-1}V)\Lambda]\cdots[(V^{-1}V)\Lambda]}_{k-1 \text{ times}} V^{-1} \\ &= V\Lambda^k V^{-1}, \end{aligned}$$

and recall that $\Lambda^k = \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)$. Note that matrix power is widely used to approximate the matrix exponential that appears frequently in system of ordinary differential equations.

Exercise 7.15. Let $A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. Find A^{10}.

7.4 Singular value decomposition

We could in fact decompose any $m \times n$ matrix A in a similar method as diagonalization, called *singular value decomposition*.

Definition 7.16. Let A be an $m \times n$ matrix. A *singular value decomposition* (or *SVD*) of A is the factorization of the form

$$A = U\Sigma V^\top,$$

where

- $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)_{m\times n}$, with $\sigma_i = \sqrt{\lambda_i}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ are the eigenvalues of $A^\top A$,
- V is an $n \times n$ matrix whose i^{th} column, v_i, is the normalized eigenvector of $A^\top A$ associated to λ_i,
- U is an $m \times m$ matrix whose i^{th} column, u_i, is given by $u_i = \frac{1}{\sigma_i} v_i$ at i in which $\sigma_i > 0$, and $u_i = 0$ at i in which $\sigma_i = 0$.

Remark. Some remarks are in order.

- The number of nonzero eigenvalues of $A^\top A$ equals the rank of A.

- In some source, the matrix U is required that $U^\top U = I$. Here we may replace the u_i at i in which $\sigma_i = 0$ with anything that fullfil $U^\top U = I$. In fact, the values at which $\sigma_i = 0$ does not participate in the factorization.

The following fact highlights the reason why SVD is considered an important factorization technique.

Theorem 7.17. *Any matrix can be factorized using SVD.*

If A is regarded as a transformation, then the SVD says that every transformation is achieved by making a rotation with $V^\top$, followed by rescale along the axes with Σ, and finally followed by another rotation U.

Exercise 7.18. Compute a SVD of the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix}$

7.5 Quadratic forms

A *quadratic form* is a generalization of the square term in a polynomial of one variable. In the multi-variable setting, a quadratic form refers to those terms involving the product of any two variables. To give a better picture, we list the quadratic forms of up to 3 variables below.

- 1 **variable.** $q(x) = ax^2$, where $a \in \mathbb{R}$ is a constant.
- 2 **variables.** $q(x, y) = ax^2 + by^2 + cxy$, where $a, b, c \in \mathbb{R}$.
- 3 **variables.** $q(x, y, z) = ax^2 + by^2 + cy^2 + dxy + exz + fyz$.

Following this line, we now state the formal definition of a quadratic form of n variables.

Definition 7.19 (Quadratic forms)**.** A function $q : \mathbb{R}^n \to \mathbb{R}$ is called a *quadratic form* on $\mathbb{R}^n$ if

$$q(x) = q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=i}^{n} c_{ij} x_i x_j,$$

where c_{ij}'s are real coefficients.

7.5.1 Matrix-vector form

The most important result regarding quadratic forms is about their matrix-vector representation.

Theorem 7.20. *Let $q(x_1, \dots, x_n) = \sum_{i=1}^{n} \sum_{j=i}^{n} c_{ij} x_i x_j$ be a quadratic form on $\mathbb{R}^n$. Then there exists a symmetric $n \times n$ matrix A such that*

$$q(x) = x^\top A x.$$

In particular, this matrix is given by

$$A = \frac{1}{2} \begin{bmatrix} 2c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{12} & 2c_{22} & c_{23} & \dots & c_{2n} \\ c_{13} & c_{23} & 2c_{33} & \dots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & c_{3n} & \dots & 2c_{nn} \end{bmatrix}. \tag{7.3}$$

Exercise 7.21. Consider the quadratic form

$$q(x, y) = 2x^2 + 3xy + y^2.$$

Find a matrix A in which $q(u) = u^\top A u$, where $y = (x, y)$.

Exercise 7.22. Write down the quadratic form

$$q(x_1, x_2, x_3) = 2x_1^2 - x_2 x_3 + x_1 x_3 + x_3^2$$

in the vector-matrix form.

7.5.2 Classification of quadratic forms

It is useful in applications to identify the *convexity* of a quadratic function. The following Figure 7.1 illustrates respectively a convex, a concave, and an indefinite quadratic forms.

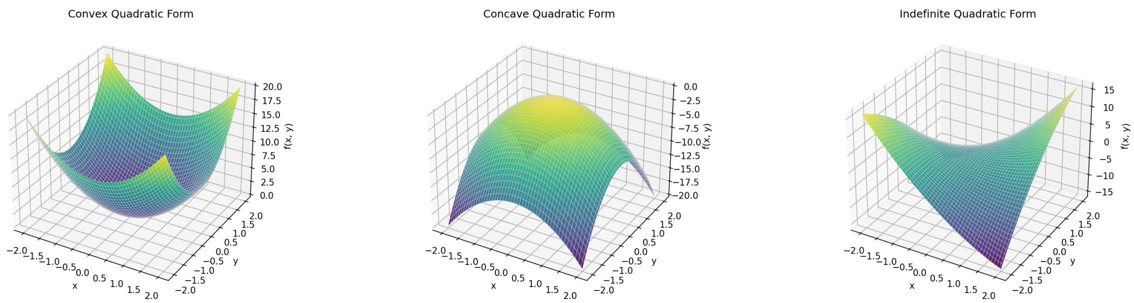

Figure 7.1: (Left) Convex quadratic form.
(Middle) Concave quadratic form.
(Right) Indefinite quadratic form.

Looking at Figure 7.1, one may make infer that a convex quadratic form is suitable for minimization, a concave quadratic form is suitable for maximization, and an indefinite one does not work well with both.

When a quadratic form $q(x)$ is converted into its matrix-vector form $q(x) = x^\top Ax$ using the formula (7.3), such convexity could be observed through eigenvalues of A. Before we go to the main result, we first define the *signs* of a symmetric matrix using the signs of its eigenvalues.

Definition 7.23 (Matrix classification)**.** A symmetric matrix A is said to be

- *positive semidefinite* if all eigenvalues of A are ≥ 0,
- *positive definite* if all eigenvalues of A are > 0,
- *negative semidefinite* if all eigenvalues of A are ≤ 0,
- *negative definite* if all eigenvalues of A are < 0,
- *indefinite* if A has both positive and negative eigenvalues.

Remark. From the definition, one may observe that positive definiteness implies positive semidefiniteness and similar to the negative counterpart.

The following result characterizes the convexity of a quadratic form using the sign of its underlying matrix.

Theorem 7.24. *A quadratic form $q(x) = x^\top Ax$ is*

- ∘ convex $\iff$ *A is positive semidefinite,*
- ∘ concave $\iff$ *A is negative semidefinite,*
- ∘ indefinite $\iff$ *A is indefinite.*

Exercise 7.25. Classify the quadratic form $q(x) = x^\top Ax$, where A is given as follows.

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

7.5.3 Quadratic forms reformulation without cross terms

One may notice that a complication of a quadratic form is rooted in its *cross terms*. Opposing to the *pure terms* that is the true squares $c_{ii}x_i^2$'s, the cross terms are the those $c_{ij}x_ix_j$ involving two different variables. In general, we may write a quadratic form as

$$q(x) = \sum_{i=1}^{n}\sum_{j=i}^{n} c_{ij}x_ix_j = \underbrace{\sum_{i=1}^{n} c_{ii}x_i^2}_{\text{pure terms}} + \underbrace{\sum_{i=1}^{n}\sum_{j=i+1}^{n} c_{ij}x_ix_j}_{\text{cross terms}}.$$

When writing $q(x)$ as a matrix-vector form $q(x) = x^\top Ax$, the diagonal entries are responsible for the pure terms' coefficients and the off-diagonal entries are corresponding to the cross terms.

At this stage, one may apply the diagonalization technique studied in the Section 7.3 to the matrix A. Recall that A is symmetric, so the spectral theorem ensures that it has an orthonormal eigenbasis $\{v_1, \ldots, v_n\}$, where v_i is a normalized eigenvector of A associated to an eigenvalue λ_i. Let V and Λ be defined accordingly to the diagonalization $A = V\Lambda V^{-1}$. From the orthonormality, we have

$$VV^\top = I = V^\top V$$

which means $V^{-1} = V^\top$. Therefore, the diagonalization becomes

$$A = V\Lambda V^\top$$

and finally we obtain

$$q(x) = x^\top V\Lambda V^\top x = (V^\top x)^\top \Lambda(V^\top x) = y^\top \Lambda y.$$

If we put $\hat{q}(y) = y^\top \Lambda y$, then $\hat{q}$ is a quadratic form that has no cross terms.

We have just showed that the diagonalization technique could always be applied to find a new basis (in this case the orthonormal eigenbasis) under which the cross terms are eliminated.

Exercise 7.26. Use the diagonalization technique to eliminate the cross terms of the quadratic form $q(x) = x^\top Ax$, where A is given below.

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

Practice problems

1. Find all eigenvalues and eigenvectors of the following matrices:

   - $A = \begin{bmatrix} 3 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix}$
   - $B = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \\ 3 & 0 & -1 \end{bmatrix}$
   - $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 1 \end{bmatrix}$
   - $D = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & 3 & 2 \end{bmatrix}$

2. Find the eigenbasis, whenever possible, from the matrices A, B, C, D above.

Chapter 8

Applications

In this chapter, we discuss several applications of linear algebra using the concepts and tools of earlier chapters.

8.1 Minimizing convex quadratic functions

Let us start with the definition.

Definition 8.1 (Quadratic functions)**.** A quadratic function is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x) = x^\top Ax + b^\top x + c,$$

where $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix, $b \in \mathbb{R}^n$ is a vector, and $c \in \mathbb{R}$ is a scalar.

Remark. A quadratic function is *not* the same as a quadratic form. Rather, it is a quadratic form $x^\top Ax$ added to an affine function $b^\top x + c$.

Theorem 8.2. *A quadratic function $f(x) = x^\top Ax + b^\top x + c$ is convex (concave) if and only if A is positive semidefinite (negative semidefinite). Moreover, its minimizers (maximizers) can be found by solving the linear system*

$$Ax = -\frac{b}{2}.$$

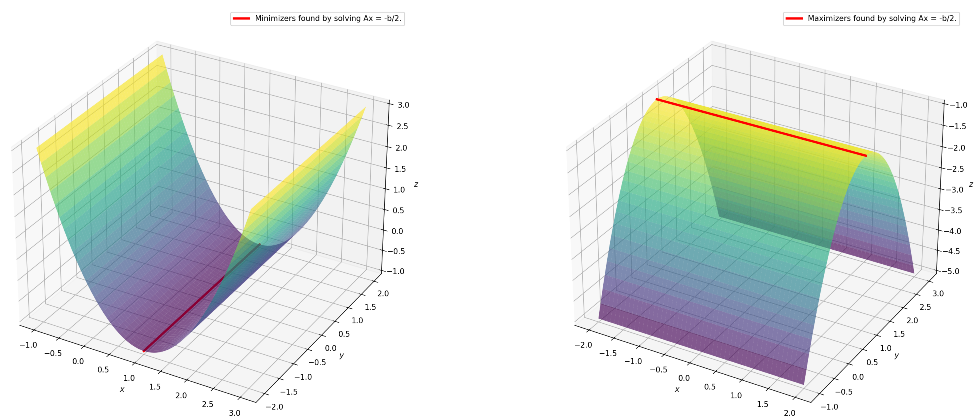

Figure 8.1: A convex and concave quadratic function with their extreme points.

Exercise 8.3. Find all minimizers of the convex quadratic function $f(x) = x^\top Ax + b^\top + c$ with

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad c = -1.$$

Exercise 8.4. Consider a quadratic function $f(x) = x^\top Ax + b^\top x + c$ where

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \quad c = 2.$$

Show that $f(x)$ is convex and find all minimizers of $f(x)$.

8.2 Least-squares solutions of linear systems

In this section, we apply the minimization of a convex quadratic function to achieve a practical aspect of inconsistent linear systems. Recall that a linear system $Ax = b$ is called *inconsistent* if it has no solution. This happens if and only if $rank(A) < rank([A\ b])$. When a linear system is inconsistent, we usually seek a *relaxed* solution $\bar{x}$ in the sense that $A\bar{x}$ is the closest to b.

Definition 8.5. A vector $\bar{x}$ is called a *least-squares solution* of a linear system $Ax = b$ if $\bar{x}$ minimizes the *residual* function $R(x) = \|Ax - b\|^2$.

The first important fact is that we can always find a least-squares solution even if the linear system has no solution in the classical sense.

Theorem 8.6. *A linear system always has a least-squares solution.*

The next result is also very important in the practical aspect.

Theorem 8.7. *The residual function is a convex quadratic function.*

Proof. From the definition of a residual function, we have

$$\begin{aligned}
R(x) &= \|Ax - b\|^2 \\
&= (Ax - b)^\top (Ax - b) \\
&= (x^\top A^\top - b^\top)(Ax - b) \\
&= x^\top A^\top Ax - b^\top Ax - x^\top A^\top b + b^\top b \\
&= x^\top (A^\top A)x - 2(A^\top b)^\top x + b^\top b.
\end{aligned}$$

This shows that R is a quadratic function. Since the matrix $Q = A^\top A$ is positive semidefinite, it follows that $R(x)$ is convex. ∎

Now that we know $R(x)$ is a convex quadratic function, the least-squares solutions of $Ax = b$ are obtained by solving the new linear system

$$A^\top Ax = A^\top b.$$

Let us first show that the least-squares solutions coincide with the classical solutions for consistent systems.

Theorem 8.8. *If a linear system $Ax = b$ is consistent, then the least-squares solutions are exactly the classical solutions.*

Proof. First notice that $R(x) = \|Ax - b\|^2 \geq 0$ for any x. Moreover we know that $R(x) = 0$ if and only if $Ax = b$. This means the classical solutions of $Ax = b$ are the least-squares solutions, and vice versa. ∎

We now look at some examples.

Exercise 8.9. Consider the linear system

$$\begin{aligned}
2x_1 + 3x_2 &= 2 \\
x_1 - x_2 &= 1 \\
x_1 + 4x_2 &= 0
\end{aligned}$$

(a) Show that this system has no solution.

(b) Find the least-squares solution.

Exercise 8.10. Show that the following system has no classical solution and find the least-squares solution

$$\begin{aligned} x_1 + x_2 &= 0 \\ 2x_1 + x_3 &= 2 \\ x_2 - x_3 &= -1 \\ x_1 + x_3 &= -1. \end{aligned}$$

8.3 Linear regression

Regression is a statistical tool that is used to find the function of a particular class that fits best to the observed data. In this course, we only focus on the *linear* regression which means that we would like to fit a *straight line* to the observed data. To keep things simple, we only consider the 1-dimensional case.

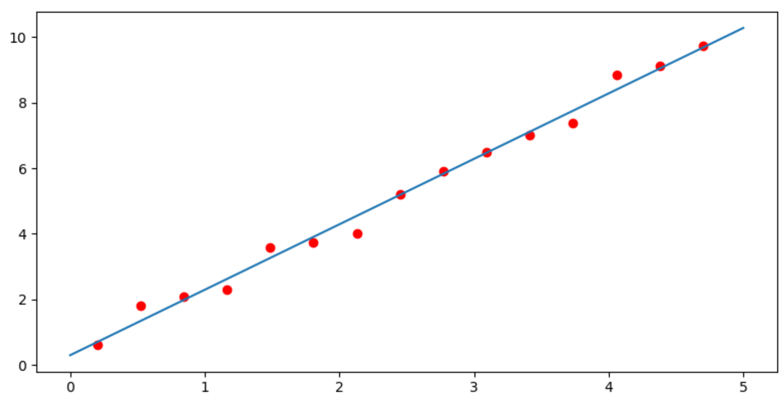

Figure 8.2: Linear regression

Suppose that we take m samples at controlled variable $x_1, \cdots, x_m$ where we observed the value $y_1, \cdots, y_m$, respectively. This could be presented as a table

Controlled variable (x)	Observed data (y)
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_m	y_m

In this 1-dimensional linear regression paradigm, we make an assumption that the observed data y_i depends *linearly* on a *single* controlled variable x_i. That is, we believe that, with a *good choice* of *parameters* $a, b \in \mathbb{R}$, we will get

$$\underbrace{y_i}_{\text{observed value}} \approx \underbrace{ax_i + b}_{\text{predicted value}}$$

for all samples $i = 1, \ldots, m$.

For any $u = (a, b)$, we measure how good the approximation $y_i \approx ax_i + b$ is using the *squares error*[1]. The squares error between the observed value y_i and the predicted value $ax_i + b$ can be computed by the formula

$$\mathrm{Err}(u) = (y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \cdots + (y_m - (ax_m + b))^2.$$

Putting

$$\begin{aligned} y &= (y_1, \ldots, y_m) \\ x &= (x_1, \ldots, x_m) \\ \mathbf{1} &= (1, \ldots, 1)_{m\times 1} \\ A &= [x; \mathbf{1}] \\ u &= (a, b), \end{aligned}$$

we can further simplify the error formula $\mathrm{Err}(u)$ into

$$\begin{aligned} \mathrm{Err}(u) &= \|y - (ax + b\mathbf{1})\|^2 \\ &= \|y - [x\ \mathbf{1}]u\|^2 \\ &= \|Au - y\|^2. \end{aligned}$$

[1] A more classical approach is to use the *root mean-squares error* (*RMSE*) or the *mean-squares error* (*MSE*). However, these approach end up with the same solutions as the squares error approach that we used.

This implies the function Err reduces to a squared norm with a linear system inside. Therefore, selecting $u = (a, b)$ that minimizes the squares error $\mathrm{Err}(u)$ is finding a least-square solution to the system $Au = y$. Finally, the *best* parameter $u = (a, b)$ can be determined by solving the linear system

$$A^\top A u = A^\top y.$$

Let us now do a practice exercise for linear regression.

Exercise 8.11. A scientist is experimenting a new drug that slows down the growth of certain species of bacteria under a controlled environment. The procedures are as follows: The initial colony of 1 CFU (colony-forming unit) is deployed in an agar plate. For 60 minutes, the scientist will observe the number of bacterial cells every 5 minutes. The following is the observed data

Observed time (min.)	Number of bacterial cells (CFU)
0	1
5	1.45
10	2.02
15	2.5
20	2.97
25	3.49
30	3.92
35	4.55
40	5.11
45	5.60
50	6.13
55	6.64
60	7.23

Write the linear system for the linear regression that estimate the data in the above table.

8.4 Principal component analysis

Principal component analysis (briefly *PCA*) is another statistical tool that is used to reduce data dimension while maintaining the most information possible. The data dimension corresponds to the number of features of each sample.

Given a dataset of m points in $\mathbb{R}^n$:

$$\hat{\mathcal{X}} = \{\hat{x}_1, \ldots, \hat{x}_m\} \subset \mathbb{R}^n.$$

One of the naïve methods to reduce the data dimension is to delete some of the features from the dataset. However, this does not retain the most information possible. Instead, we seek an orthogonal projection onto a lower-dimensional subspace of designated dimension $k \leq n$ such that the *information*, measured using the *variance*, is maximized. The basis of this optimal subspace is called the *principal components*, and it happens to be the (normalized) eigenvectors corresponded to the k largest eigenvalues of the covariance matrix of the *centered* dataset.

To make an optimal dimension reduction using the PCA, we first need to center the data:

$$\begin{aligned}
&\mathcal{X} = \{x_1, \ldots, x_m\}, \\
&x_i = \hat{x}_i - \bar{x}, \quad \bar{x} = \frac{1}{m}\sum_{i=1}^{m} \hat{x}_i.
\end{aligned}$$

The covariance matrix of the centered dataset $\mathcal{X}$ is defined by

$$S = \frac{1}{m}\sum_{i=1}^{m} x_i x_i^\top = \frac{1}{m} X X^\top,$$

where X is the matrix whose i^{th} column is x_i.

The k *principal components* are the orthonormal basis $\mathcal{V} = \{v_1, \ldots, v_k\} \subset \mathbb{R}^n$ such that the *information* measured with the variance, $\sum_{j=1}^{k} v_j^\top S v_j$, is maximized.

Theorem 8.12. *The k principal components of the centered dataset $\mathcal{X} = \{x_1, \ldots, x_m\} \subset \mathbb{R}^n$, with $k \leq n$, is obtained by the normalized eigenvectors corresponded to the k largest eigenvalues of $S = \frac{1}{m} X X^\top$. That is, if*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

are then eigenvalues of S, then the k principal components $\mathcal{V} = \{v_1, \ldots, v_k\}$ satisfy

$$(S - \lambda_j I) v_j = 0 \quad \textit{and} \quad \|v_j\| = 1$$

for all $j = 1, \ldots, k$.

Once the k principal components $\mathcal{V} = \{v_1, \ldots, v_k\}$ is obtained, the centered dataset $\mathcal{X}$ is projected onto $V = \mathrm{span}(\mathcal{V})$ to obtain the projected dataset

$$\begin{aligned} \mathcal{X}^* &= \{x_1^*, \ldots, x_m^*\} \subset V, \\ x_i^* &= \mathrm{proj}_V(x_i), \qquad i = 1, \ldots, m. \end{aligned}$$

Recall that $\dim(V) = k \leq n$, so that $\mathcal{X}^*$ is a dataset in a lower-dimensional space.

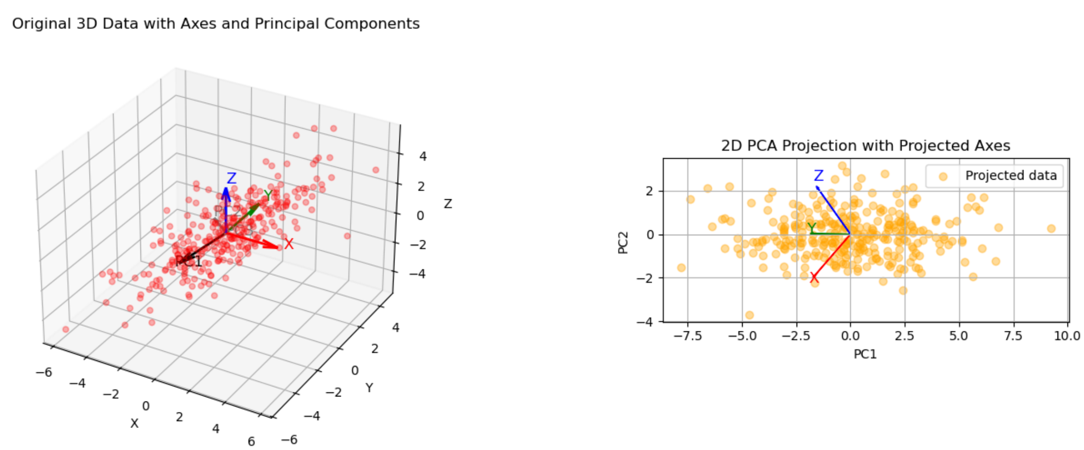

Figure 8.3: The orthogonal projection of 3-dimensional data onto the first two principal components.

8.5 Image compression

In this subsection, we demonstrate how a digital image file could be compressed into a smaller file size without losing significant details using singular value decomposition (SVD). One should be clear that compression and resizing are two different things. Any compression method should not reduce the pixel size of an image but rather the memory used to store it.

A grayscale digital image is nothing but an $m \times n$ matrix A whose pixel entry a_{ij} is a gray value between 0 (representing black) and 1 (representing white). In a similar way, an rgb color digital image is representing with three layers of the respective color's intensity matrix.

We shall only demonstrate a grayscale case for simplicity. Suppose that A is an $m \times n$ matrix representing a grayscal digital image. Suppose that A is factorized with SVD into

$$A = U\Sigma V^\top,$$

where we recall that $U \in \mathbb{R}^{m\times m}$, $\Sigma \in \mathbb{R}^{m\times n}$ and $V \in \mathbb{R}^{n\times n}$. Also, remember that the information is stored in U and V at columns i's such that $\sigma_i > 0$. Moreover, since $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, the information values are sorted in a decreasing order. That means the information values in the early columns of U and V are higher than the later columns. Such an observation motivates us to *discard* the later columns and only keep the first k columns that matter the most in our image matrix:

$$A \approx A_k = U_k\Sigma_k V_k^\top,$$

where $k \leq \min\{m, n\}$ and

- $U_k = [u_1\ u_2\ \ldots\ u_k]_{m\times k}$,
- $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)_{k\times k}$,
- $V_k = [v_1\ v_2\ \ldots\ v_k]_{n\times k}$.

This method is sometimes called the *low-rank approximation* or *rank k approximation*, due to the fact that the full image is approximated up to the rank k matrix A_k. Note that A_k maintains the dimension $m \times n$ of the original data.

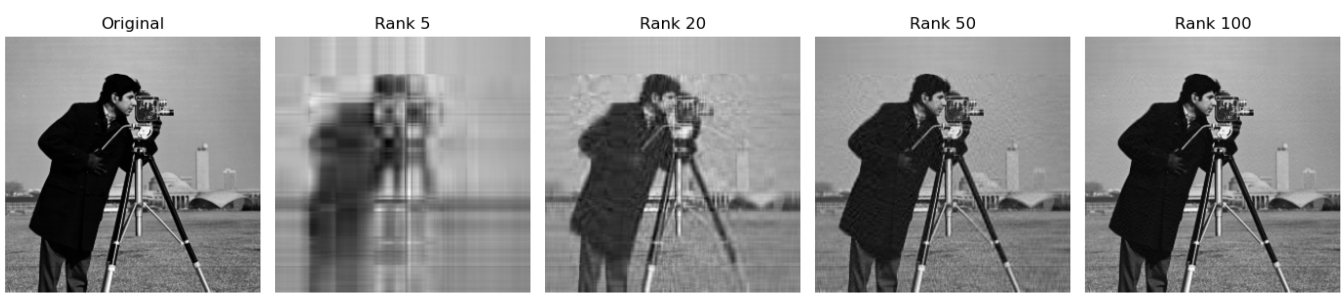

Figure 8.4: The original image and its rank k approximations for $k = 5, 20, 50, 100$.

You’re now on your own, kid!

Outroduction.

I truly hope you did learn some boring mathematics but, in the back of your head, aware how creative stuffs are made out of it.

Linear algebra, together with many other branches of mathematics, are milestones in bleeding-edge technologies around us.

Good engineers know how things work.

Great engineers know how to make them better!

All the best.

Parin Chaipunya