Linear Algebra

for Engineers

$$\begin{array}{rcl} a_{11}x_1 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n & = & b_2 \\ & \vdots & \\ a_{mn}x_1 + \cdots + a_{mn}x_n & = & b_m \end{array} \iff \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{mn} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Parin Chaipunya, Ph.D.
Department of Mathematics
KMUTT

1. What is Linear Algebra all about?

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

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

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.

In the first half of the course, we focus on the matrix equation $Ax = b$ and go through different approaches to characterize its solution. The topics include

- Solving $Ax = b$ by Gauss-Jordan method.
- Solutions of $Ax = b$ characterized by $rank(A)$.
- Determinant and its use in the equation $Ax = b$.
- Vector spaces, subspaces, and dimension. `This part is mainly used to describe the solution set of` $Ax = b$.
- Linear transformation. `This part gives a new perspective to the matrix equation` $Ax = b$.

The second half focuses on more advanced topics that revolves around representation of matrices in different bases. One of the most applied basis is the eigenbasis, which allows useful decomposition for faster matrix computation in modern GPUs. Moreover, we speak of geometric sides of linear algebra as well as some of its applications. The second half would cover

- Canonical forms. `This allows representation of a matrix directly in the chosen basis.`
- Eigenvalues and associated eigenvectors/eigenspace. Eigenbasis.
- Diagonalization. `This is one of the decomposition methods that allows for fast computation in modern computers.`

- Similar matrices.
- Norms. Orthogonality. Quadratic forms. `We explore geometric aspects of linear algebra here.`
- Some applications.

1.1 Matrix algebra: a quick review

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

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

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

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

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 $A, B, C, \cdots$, etc.

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

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

Let us consider some special classes of matrices now.

Vectors.
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 $a, b, c, \cdots$ or more often with $x, y, u, v, \cdots$,etc.

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

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

is the same as

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

Square matrices.
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 $a_{11}, a_{22}, \cdots, a_{nn}$, as shown on the following figure:

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

↖ diagonal of A

A diagonal matrix is a square matrix whose nonzero entries are only on its diagonal. If the elements along the diagonal are $d_1, d_2, \cdots, d_n$, then we write $diag(d_1, \cdots, d_n)$ to denote such a diagonal matrix, that is

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

The identity matrix of dimension $n \times n$ is the matrix I given by

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

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.

Let us now turn to algebraic operations of matricecs.

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

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

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

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

It is clear that $(A^T)^T = A$.

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

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

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

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

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

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

- $A - B = A + (-B)$,
- and $A \pm B = B \pm A$,
- $(A \pm B)^T = A^T \pm B^T$,
- $(A \pm B) \pm C = A \pm (B \pm C) = A \pm B \pm C$,
- 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}:

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

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.

Let A, B, C be two matrices. Then we have

- $A(B \pm C) = AB \pm AC$,
- $(A \pm B)C = AC \pm BC$,
- $(AB)C = A(BC) = ABC$,
- $(AB)^T = B^T A^T$,
- $IA = A = AI$, (The two I's can be of different dimension!)
- if c is a scalar, then $c(AB) = (cA)B = A(cB)$,

whenever the above products are well-defined.

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

Exercise 1.1. 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) BC (c) ABC (d) AC

(e) BA (f) $B^T A$ (g) $C^T B A^T$ (h) $AA^T + C^T C$

Lab session 1

This lab session is an introduction to the MATLAB application and elementary syntax. The following topics will be covered:

- defining variables, matrices and vectors,
- accessing variables, entries of a matrix / vector, rows and columns,
- elementary matrix operations and broadcasting.

Assignment of the week.

Let $A = \begin{bmatrix} 1 & 3 & 0 & 0 & -1 \\ 2 & 2 & 3 & -2 & 2 \\ 3 & 4 & 9 & -2 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 2 & 2 \\ 3 & -5 & -1 \\ 0 & -2 & 1 \\ -1 & -1 & 0 \end{bmatrix}$. Find the following:

(a) The first and last rows of AB.

(b) The first and last columns of BA.

2. Linear systems and matrix equations

2.1 Transforming a linear system into a matrix equation

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 $x_1, \cdots, x_n$, 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.

> **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:

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

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

To review and familiarize linear systems, let us do some preliminary exercises.

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 $x_1 = 2, x_2 = 5$ a solution of the following linear system

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

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

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

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

Solving the previous systems, we may have observed that our calculation is dependent on the given equations. This is difficult to proceed when the number of variables and equations grow larger. Hence we would need a more disciplined strategy, and this involves changing the system into a matrix equation.

Turning a linear system 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

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

We call the 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.

Augmented matrix for $Ax = b$.

To solve $Ax = b$ by hand, it is often helpful to simplify it as an augmented matrix $[A \,|\, b]$, that is

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

Row operations.

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

- adding an equation into another,
- multiplying a constant to an equation,
- swapping order of equations.

These operations transfer to the matrix equation as (elementary) row operations, which include

- adding a row into another,
- multiplying a constant to a row,
- swapping rows.

2.2 Analyzing the equation $Ax = b$

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 *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.

2.2.1. Echelon forms

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

$$P = \begin{bmatrix} 0 & 0 & 0 & \circledast & \cdots \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & \circledast & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix},$$

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

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

$$\hat{P} = \begin{bmatrix} 0 & \circledast & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \circledast & \cdots \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

Reduced row echelon form.
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.

Exercise 2.7. 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 note with the following facts.

Fact.

- 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 2.8. Use elementary row operations to find row echelon forms of the matrices from Exercise 2.7.

2.2.2. 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 2.9. 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) $D = \begin{bmatrix} 2 & 0 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$

An observation.

- Every square matrix's row echelon form is a triangular matrix.

2.2.3. 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 2.10. 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) $B = \begin{bmatrix} -2 & 1 & 3 & 0 \\ 2 & 1 & 3 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & 1 & 3 \end{bmatrix}$

2.2.4. 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.

Nonzero rows of row echelon matrices.
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.

Rank.
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 $rank(A)$.

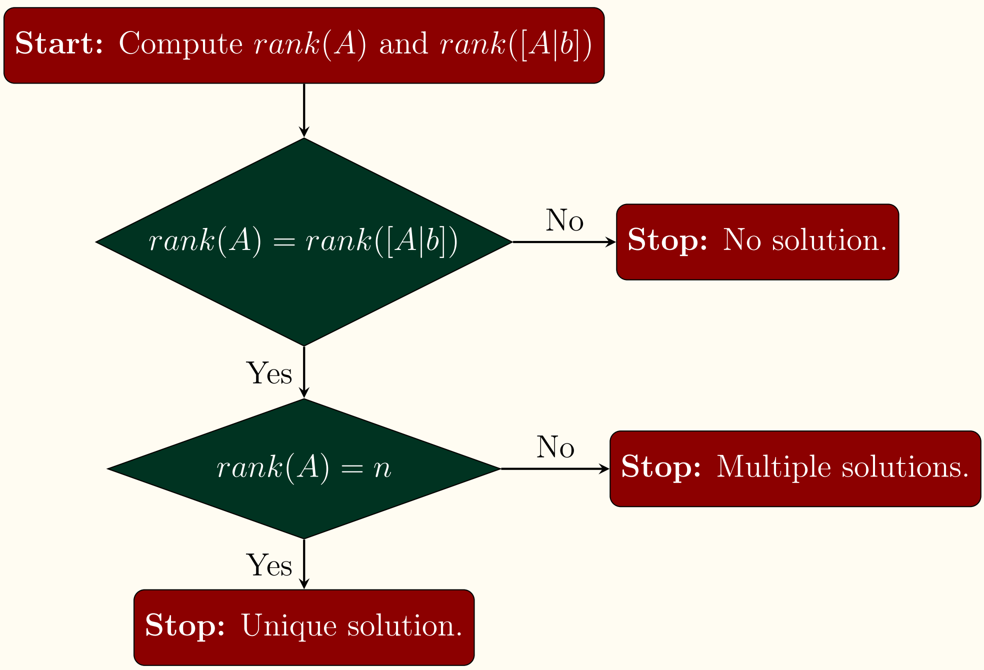
Exercise 2.11. Find the rank of the matrices from the previous exercises.

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 $[A|b]$. More precisely, the number of solutions of $Ax = b$ can be concluded from the $rank([A|b])$.

Number of solutions through the rank.
Consider the equation $Ax = b$ that corresponds to a system of n variables and m equations. Then the following conclusion is drawn:

- $rank(A) < rank([A|b]) \iff Ax = b$ has no solution;
- $rank(A) = rank([A|b]) = n \iff Ax = b$ has a unique solution;
- $rank(A) = rank([A|b]) < n \iff Ax = b$ has multiple solutions.

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



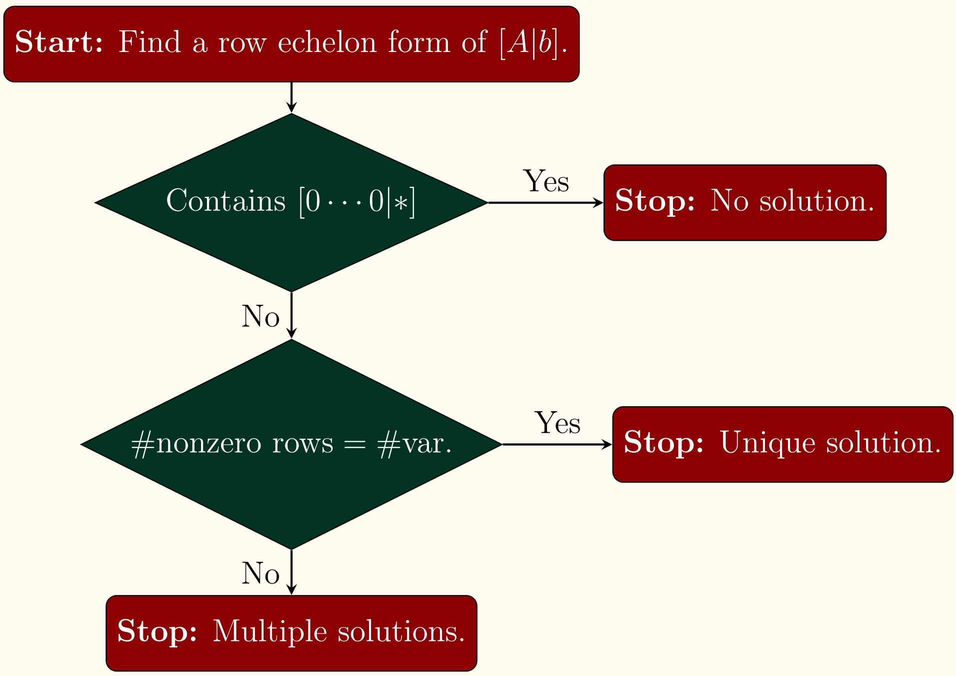
Practical aspects.
Consider again the system represented by $Ax = b$ with n variables and m equations. In practice, to see if $rank(A) < rank([A|b])$ happens, we do not

need to compute $rank(A)$ and $rank([A|b])$ separately.

Consider directly the matrix $[A|b]$. When reducing $[A|b]$ to one of its row echelon form $[\hat{A}|\hat{b}]$, 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 $[A|b]$. Then the following hold:

> If a row echelon form of $[A|b]$ consists of a row of the form $[0 \cdots 0|*]$ (where $* \neq 0$), then $rank(A) < rank([A|b])$ and so $Ax = b$ has no solution.

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



Exercise 2.12. 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 2.13. 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 2.14. 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 2.15. Determine the number of solutions of the following system:

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

2.3 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 2.16. 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 2.17. 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 2.18. Solve the following linear system

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

Exercise 2.19. Solve the following linear system

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

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 $b = 0$.

Multiple solutions... Yes, but how many?
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 $\mathcal{S}$ has a particular shape to it, e.g. a single dot, a straight line, a flat plane, a space, etc. We actually observe that all of them have *linear* shapes. 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).

Now, let us state the formal definition of a vector space.

General vector spaces.
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) $1u = u$.

Note that the scalar multiplication $\cdot$ is always omitted.

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 zero vector in $\mathbb{R}^n$ is exactly the vector $0 = (\underbrace{0, \cdots, 0}_{n \text{ times}})$.
- The negative of each $u \in \mathbb{R}^n$ is unique, and it is given exactly by the formula $-u = (-1)u$.

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

Let us first note some key properties of $\mathbb{R}^n$ and its two operations, namely the *addition* '+' and *scalar multiplication* ' $\cdot$ '.

$\mathbb{R}^n$ as a vector space.
We denote $\mathbb{R}^n$ as the set of all n-dimensional real vectors $x = (x_1, \cdots, x_n)$ where $x_i \in \mathbb{R}$ for each $i = 1, \cdots, n$. 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 $\mathbb{R}^n$ is always equipped with these classical operations unless explicitly stated otherwise.

What does $\mathbb{R}^n$ carry?
Obviously, we have the following sequence:

- The space $\mathbb{R}$ 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.

Observe that $\mathbb{R}^2$ 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.

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 can use the following single-step criterion.

Single-step subspace criterion.
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.

The following double-step criteria is equally applicable. The idea is to break the single-step criterion into two simpler ones.

Double-step subspace criteria.
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).

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

Exercise 3.1. Show that $U = \{(x, y) \in \mathbb{R}^2 \,|\, x = 2y\}$ is a subspace of $\mathbb{R}^2$.

Exercise 3.2. Show that the set $U = \{(x_1, x_2) \in \mathbb{R}^2 \,|\, x_1 = x_2^2\}$ is not a subspace of $\mathbb{R}^2$.

Exercise 3.3. 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 3.4. 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 3.5. Show that the following sets are not vector subspaces of $\mathbb{R}^2$.

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

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

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

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

Exercise 3.6. Which of the following are vector spaces.

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

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

(c) $Z = \{(x_1, \cdots, x_n) \,|\, x_1 = 0 \text{ and } x_n = 0\}$

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

Exercise 3.7. 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 3.8. Let A be any $m \times n$ matrix. Then the set of all multiplications Ax is a vector subspace of $\mathbb{R}^m$. This set is actually called the image of A, or $im(A)$.

3.2 Linear combination and linear span

Linear combination.
Let $(V, +, \cdot)$ be a vector space over the field $K \in \{\mathbb{R}, \mathbb{C}\}$ and $v_1, v_2, \cdots, v_k$ 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 $\alpha_1, \cdots, \alpha_k$ are scalars.

Example. Consider the vectors $v = (4, 0, 3)$, $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$ and $v_3 = (0, 0, 1)$. Then v is a linear combination of v_1 and v_3, but not of v_1 and v_2.

Exercise 3.9. 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 3.10. 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.

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

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

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

Exercise 3.11. Let V be a vector space and $U = \{v_1, \cdots, v_k\}$ be a subset of V. Show that $span(U)$ is a vector subspace of V.

Exercise 3.12. 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 $v_1, \cdots, v_k$ 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. 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 3.13. From the previous example...

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

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

(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.

Linear independence.
Let V be a vector space. The vectors $v_1, \cdots, v_k \in V$ are said to be linearly independent if the only coefficients $\alpha_1, \cdots, \alpha_k$ that makes

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

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

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

Exercise 3.14. Show that the vectors $v_1 = (1, 0)$ and $v_2 = (0, 1)$ are linearly independent.

Exercise 3.15. Show that the vectors $v_1 = (1, 1, 1)$, $v_2 = (1, 1, 0)$ and $v_3 = (1, 0, 0)$ are linearly independent.

Exercise 3.16. Show that the vectors $v_1 = (1,1)$, $v_2 = (1,0)$ and $v_3 = (0,1)$ are linearly dependent.

Exercise 3.17. Show that if $v_1, \cdots, v_k$ 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.

3.3 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.

Basis.
Let V be a vector space. The vectors $v_1, \cdots, v_k \in V$ are called basis vectors of V if the following two conditions are satisfied

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

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

Fact. Every vector space has a basis.

Fact. All bases of a vector space have the same number of elements.

Exercise 3.18. Consider the vector space $\mathbb{R}^n$ and vectors $e_1, \cdots, e_n \in \mathbb{R}^n$ given by

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

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

Exercise 3.19. Find some other bases for $\mathbb{R}^n$.

Exercise 3.20. Find a basis for the vector space

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

Exercise 3.21. Find a basis for the vector space

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

Basis representation.
Consider any vector space V with a basis $\mathcal{B} = \{b_1, \cdots, b_n\}$. Then any element $x \in V$ can be written uniquely as

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

The shorthand notation for this is given by

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

When $V = \mathbb{R}^n$ is equipped with the standard basis $\mathcal{E} = \{e_1, \cdots, e_n\}$, we know that the basis representation $x = (x_1, \cdots, x_n)_{\mathcal{E}}$ is just the same as

> the classical expression $x = (x_1, \cdots, x_n)$. Therefore, it is agreed to write $x = (x_1, \cdots, x_n)$ without indicating the basis when the standard basis is used.

Exercise 3.22. 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 3.23. 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 thebasis $\mathcal{B}$.

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

Rank and nullity.
Let A be any given matrix. Then the rank and nullity are concluded as follows.

The rank theorem. The rank can be computed by $rank(A) = dim(im(A))$.

Nullity. The nullity of 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.

The rank-nullity theorem. If A has n columns, then $rank(A) + null(A) = n$.

Exercise 3.24. Re-assess exercises 3.20 and 3.21 in terms of dimensions using the rank-nullity theorem.

Exercise 3.25. Consider the following linear system:

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

What is the dimension of the solution set of this system ? How does it look like ?

The solution set of $Ax = b$.
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 $A\overline{x} = b$ has a solution x_0, then the solution set of $Ax = b$ is given by

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

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

Exercise 3.26. Consider the following linear system:

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

What is the dimension of the solution set of this system ? How does it look like ?

Exercises

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.
   - $V = \{x \in \mathbb{R}^3 \mid x_1 = -x_3,\ x_2 = -x_1\}$.
   - $V = ker\left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}\right)$.
   - $V = \{(x_1 + x_3, -2x_2) \mid x_1, x_2, x_3 \in \mathbb{R}\}$.
   - $V = \{(3x_1, x_1 + x_2 - x_3, x_3) \mid x_1, x_2, x_3 \in \mathbb{R}\}$.
   - $V = \{x \in \mathbb{R}^3 \mid 2x_1 - 3x_2 = 0,\ x_1 + 2x_3 = 0\}$.
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 $dim(ker\, A) = 1 + dim(im\, A)$. Show that A has odd number of rows. (*Hint: Use the rank-nullity theorem.*)

4. Linear transformation

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 $Ax = b$.

Linear transformation.
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 $\alpha, \beta \in K$.

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 4.1. 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 4.2 (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 4.3. 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 this fact to disprove linearity of a function: If $T(0) \neq 0$, then T is not a linear transformation.

Exercise 4.4. 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 4.5. 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) $T(x_1, x_2) = \sin^2(x_1) + \cos^2(x_1) - (1 + x_2)$.

Matrix representation of a linear transformation.
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

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

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

Exercise 4.6. Find the matrix representations of all the linear transformations from Exercise 4.1.

The space $\mathscr{L}(V, W)$.
Let V and W be two vector spaces over the field $K \in \{\mathbb{R}, \mathbb{C}\}$. We define

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

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

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

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 4.7. Consider the space $\mathbb{R}^{m\times n}$. For $i = 1, \cdots, m$ and $j = 1, \cdots, n$, define E_{ij} to be the matrix whose element at the position (i, j) is 1 and all 0 elsewhere.

Kernel and range.
Given a linear transformation $T : V \to W$. The kernel (or null space) of T is given by

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

Likewise, the range of T is

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

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

- compute the kernel $ker(T)$ by solving the linear system $A_T x = 0$,
- compute the range $ran(T)$ 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 4.8. Consider the linear transformation

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

Find $ker(T)$ and $ran(T)$.

Exercises

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 $T : \mathbb{R}^n \to \mathbb{R}^m$ 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.

5. 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 $|A|$.

5.1 The 2×2 and 3×3 cases.

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*.

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

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

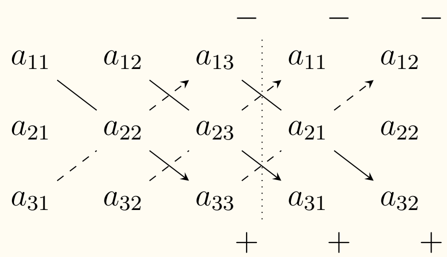
This formula is known as the 2×2 Sarrus rule.

Exercise 5.1. 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.

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

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

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



Exercise 5.2. Find $det(A)$ where A is given by

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

5.2 The $n \times n$ case.

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 may involves many more steps and concepts.

The minors, cofactors, and determinants.
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

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

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

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

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 5.3. 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 5.4. 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) $B = \begin{bmatrix} 2 & 0 & 0 & 3 & -2 \\ 1 & 1 & 0 & -3 & 1 \\ 3 & 0 & 0 & 3 & 1 \\ -1 & -1 & 0 & 0 & 2 \\ 1 & 0 & 2 & 0 & 4 \end{bmatrix}$

Properties of the determinant.
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 $det(A) = a_{11}a_{22}\cdots a_{nn}$.
- $det(AB) = det(A)det(B)$
- $det(A^T) = det(A)$
- $det(cA) = c^n det(A)$ 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, i.e. $det(A + B)$ and $det(A) + det(B)$ are independent.

Exercise 5.5. 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 5.6. Find two matrices A and B such that $det(A + B) \neq det(A) + det(B)$.

Exercise 5.7. Find two matrices A and B such that $det(A + B) = det(A) + det(B)$.

Determinant of triangular block matrices.
A matrix A can be subdivided into $p \times q$ blocks in such a way that

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

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

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

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.

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

Exercise 5.8. Find $det(A)$ where A is given by $A = \begin{bmatrix} 2 & 3 & 4 & 7 & 8 \\ -1 & 5 & 3 & 2 & 1 \\ 0 & 0 & 2 & 1 & 5 \\ 0 & 0 & 3 & -1 & 4 \\ 0 & 0 & 5 & 2 & 6 \end{bmatrix}$.

5.3 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.

Given a square $n \times n$ matrix A. We say that..

- A is of a full rank if $rank(A) = n$;
- 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.

Fact. The following conditions are equivalent:

(i) A is of a full rank.

(ii) A is invertible.

(iii) $det(A) \neq 0$.

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 $adj(A) = C(A)^T$ is called the adjoint of A.

Exercise 5.9. Let A be a 2×2 invertible matrix. Find a general formula for A^{-1}.

Exercise 5.10. Find the inverse of the following matrices.

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

Some facts concerning invertibility.
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 $x = A^{-1}b$.
- If both A and B are invertible, then AB is also invertible and

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

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

Exercise 5.11. Find a solution to the linear system

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

Exercise 5.12. Let $A = \begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} -2 & 1 \\ 2 & 0 \end{bmatrix}$. Find $(AB)^{-1}$.

Invertibility of a 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

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.

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

Exercise 5.13. 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.

Exercise 5.14. 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*.

Determinant and geometry.
If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation, then we may understand..

- how the space $\mathbb{R}^n$ is *deformed* under the application of T by looking at its range $ran(T)$.
- how some parts of $\mathbb{R}^n$ is *collapsed* to 0 by looking at its kernel $ker(T)$.

We can also note from the rank-nullity theorem that T cannot explode the space in the sense that it is impossible to gain additional dimensions by the transformation T.

Now, let A_T be the matrix representation of T. The determinant $det(A_T)$ informs us the following information:

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

Exercises

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

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

6. 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.

6.1 Eigenvalues and eigenvectors

Eigenvalues and eigenvectors.
Consider a square $n \times n$ matrix. A scalar λ (possibly a complex number even A is real) 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.

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

Exercise 6.1. 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.

Exercise 6.2. 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.

A difficulty and a practical way of finding eigenpairs.
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 unknown. 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

$$(A - \lambda I)v = 0. \tag{6.1}$$

Knowing $v \neq 0$, it is necessary to have

$$det(A - \lambda I) = 0. \tag{6.2}$$

Notice that the above equation is free of the unknown vector v. Writing down the determinant further shows that (6.2) is in fact a polynomial. Hence we may refer to (6.2) as the characteristic polynomial of A. The *fundamental theorem of algebra* then gurantees that (6.2) 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 (6.2), we return to (6.1) with the known λ's. This makes (6.1) 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 $(A - \lambda I)v = 0$.

Exercise 6.3. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then A has no real eigenvalues.

Exercise 6.4. Find all eigenvalues and eigenvectors of a matrix $A = \begin{bmatrix} -1 & 0 \\ 2 & 0 \end{bmatrix}$.

Exercise 6.5. 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) $B = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$

Some properties of eigenvalues-eigenvectors.
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^T 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

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

Exercise 6.6. Check the last facts (sum and products of eigenvalues) with matrices in Exercise 6.5.

6.2 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.

Eigenbasis.
Let A be a square matrix of dimension $n \times n$. If A has linearly independent *real* eigenvectors $v_1, \cdots, v_n$, then these eigenvectors form a basis $\mathcal{S} = \{v_1, \cdots, v_n\}$. This special basis is called the eigenbasis generated by A. 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 6.7. Find the eigenbases associated to each of the following matrices.

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

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

(c) $C = \begin{bmatrix} 1 & 11 & 111 \\ 0 & 22 & 222 \\ 0 & 0 & 333 \end{bmatrix}$

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

(e) $E = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

7. Change of basis revisited

In this chapter, we revisit bases of a vector space. In particular, suppose that a basis $\mathcal{B}$ is initially chosen for a vector space V and for some reason we migrate to using another basis $\mathcal{B}'$. In such case, we need to recalculate the vector x in the representation of $\mathcal{B}$ into that of $\mathcal{B}'$. This process turns out to be a linear transformation.

On the other hand, we used to represent a linear operator $T : \mathbb{R}^n \to \mathbb{R}^m$ with an $m \times n$ matrix A_T. This matrix corresponds to the situation where both $\mathbb{R}^n$ and $\mathbb{R}^m$ are equipped with the standard bases. However, it is also possible to tailor-made this matrix representation that would accept and return vectors promptly in other bases.

7.1 Change-of-basis matrix

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 especially in this chapter.

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

Let $v \in V$ be a vector which 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 $\mathcal{B}'$, i.e. $v = (\alpha'_1, \cdots, \alpha'_n)_{\mathcal{B}'}$, we solve the equation

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

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

$$B'x = By. \tag{7.1}$$

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

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

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 (7.1) 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 7.1. 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 7.2. 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.

Exercise 7.3. 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 7.4. Consider $\mathbb{R}^3$. Let $\mathcal{B} = \{(1,0,1),(0,1,0),(1,0,0)\}$ and $\mathcal{B}' = \{(1,0,0),(1,0,1),(0,1,0))\}$ be two bases of $\mathbb{R}^3$.

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

(b) Write the vector $x = (-1,1,2)_{\mathcal{B}}$ in the basis $\mathcal{B}'$.

(c) Write the vector $y = (2,-1,1)_{\mathcal{B}'}$ in the standard basis.

7.2 Matrices and linear transformations under different bases

Suppose that we work with $\mathbb{R}^n$ with basis $\mathcal{B}$, $\mathbb{R}^m$ with basis $\mathcal{B}'$ and an $m \times n$ matrix that links $\mathbb{R}^n$ to $\mathbb{R}^m$. Here, we would like to modify an $m \times n$ matrix so that it multiplies with a vector $x_{\mathcal{B}}$ and returns a vector $y_{\mathcal{B}'}$. The same is asked to linear transformations. We would like to pass a vector of a chosen basis of $\mathbb{R}^n$ into the transformation and instantly get a resulting vector, again, in a chosen basis of $\mathbb{R}^m$.

Representing a matrix in a chosen bases.
Assume that $\mathbb{R}^n$ is equipped with a basis $\mathcal{B}$ and $\mathbb{R}^m$ is equipped with a basis $\mathcal{B}'$. Let A be an $m \times n$ matrix. Normally, a matrix A multiplies with a vector $x \in \mathbb{R}^n$ and returns a vector $y \in \mathbb{R}^m$, i.e. $y = Ax$, where both spaces are assumably equipped with the standard bases.

If we wants to multiply $x_{\mathcal{B}}$ to A, we would need to change $x_{\mathcal{B}}$ into the standard basis representation first, then multiply it to A. If we further need the resulting vector y in a basis $\mathcal{B}'$, we would need to change the resulting vector into such a basis. The full process can be concluded in the following

$$y_{\mathcal{B}'} = [\mathcal{B}' \leftarrow \mathcal{S}']A[\mathcal{S} \leftarrow \mathcal{B}]x_{\mathcal{B}},$$

where $\mathcal{S}$ and $\mathcal{S}'$ denote the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. From this equation, we may see that the matrix

$$A^{\mathcal{B}}_{\mathcal{B}'} = [\mathcal{B}' \leftarrow \mathcal{S}']A[\mathcal{S} \leftarrow \mathcal{B}]$$

provides a version of the matrix A that multiplies $x_{\mathcal{B}}$ and returns $y_{\mathcal{B}'}$ in the respective bases, as required.

Exercise 7.5. Consider the matrix $A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$. Let $\mathcal{B} = \{(1, -1), (0, 1)\}$ and $\mathcal{B}' = \{(0, 1), (1, 0)\}$ be two bases of $\mathbb{R}^2$. Find the representation of A that multiplies a vector in the basis $\mathcal{B}$ and results in the basis $\mathcal{B}'$.

Linear transformations with respect to bases.
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Assume that $\mathbb{R}^n$ is equipped with the basis $\mathcal{B}$ and $\mathbb{R}^m$ is equipped with the basis $\mathcal{B}'$.

Suppose that A_T is the matrix representation of T, then A_T multiplies and results respectively in the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. Using a similar observation, one may see that $(A_T)^{\mathcal{B}}_{\mathcal{B}'}$ provides a way to input and output vectors with T in required bases. To conclude, we may define the matrix of T that accepts $x_{\mathcal{B}}$ and returns $y = T(x)$ in the basis $\mathcal{B}'$ with

$$T^{\mathcal{B}}_{\mathcal{B}'}(x_{\mathcal{B}}) = (A_T)^{\mathcal{B}}_{\mathcal{B}} x_{\mathcal{B}} = [\mathcal{B}' \leftarrow \mathcal{S}'] A_T [\mathcal{S} \leftarrow \mathcal{B}] x_{\mathcal{B}} = y_{\mathcal{B}'}.$$

Exercise 7.6. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation defined by $T(x_1, x_2) = (x_1, 0, x_1 + x_2)$. Assume that we equip $\mathbb{R}^2$ with the basis $\mathcal{B} = \{(0,1), (1,0)\}$ and $\mathbb{R}^3$ with $\{(1,0,0), (1,1,0), (0,0,1)\}$. How to define $T^{\mathcal{B}}_{\mathcal{B}'}$?

Exercise 7.7. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation defined by

$$T(x_1, x_2) = \begin{bmatrix} 1 & -2 \\ -3 & 1 \end{bmatrix}.$$

Suppose that $\mathbb{R}^2$ is equipped with the basis $\mathcal{B} = \{(3,1),(1,1)\}$. Find the representation of T that accepts and returns vectors in $\mathbb{R}^2$ under the basis $\mathcal{B}$.

Exercise 7.8. Consider a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T(x) = Ax = \begin{bmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{bmatrix} \begin{bmatrix} | \\ x \\ | \end{bmatrix}.$$

Find the representation of T with respect to the eigenbasis of A.

We may observe from the previous exercise that A is a diagonal matrix in the eigenbasis representation. Moreover, the diagonal elements are the eigenvalues. This is not only by chance, but happens to all matrices that admit eigenbases. This process is called diagonalization.

Diagonalization.
Let A be an $n \times n$ matrix. Then A is said to be diagonalizable if there is an $n \times n$ invertible matrix U and an $n \times n$ diagonal matrix D such that

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

It is not clear when a matrix is diagonalizable since we need to find the two unknown matrices U and D. However, we can utilize eigenvalues/eigenvectors knowledge to diagoalize a matrix.

Diagonalization by eigenvalues/eigenvectors.
Suppose that an $n \times n$ matrix A has n independent real eigenvectors $v_1, \cdots, v_n$ which are associated respectively to eigenvalues $\lambda_1, \cdots, \lambda_n$. Then A is diagonalizable and we have

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

where V and Λ are given by

$$V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

Exercise 7.9. Diagonalize the following matrices.

(a) $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$

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

Being able to diagonalize a matrix allows an easier way to compute matrix power. Observe first in the next exercise how a diagonal matrix powers up.

Exercise 7.10. Let $D = diag(d_1, \cdots, d_n)$. Show that

$$D^k = \underbrace{DD \cdots D}_{k \text{ times}} = diag(d_1^k, \cdots, d_n^k).$$

Knowing the power formula for diagonal matrices, we may observe the following: If A is diagonalizable into $A = UDU^{-1}$, then

$$\begin{aligned} A^k &= \underbrace{(UDU^{-1})(UDU^{-1}) \cdots (UDU^{-1})}_{k \text{ times}} \\ &= UD \underbrace{[(U^{-1}U)D][(U^{-1}U)D] \cdots [(U^{-1}U)D]}_{k-1 \text{ times}} U^{-1} \\ &= UD^kU^{-1}. \end{aligned}$$

Exercise 7.11. Let $A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. Find A^{10}.

8. Quadratic forms

In a multivariate polynomial of variables $x_1, \cdots, x_n$, the k^{th} degree term is the one that has products consisting of k variables out of $x_1, \cdots, x_n$. Among these, the terms with degree $= 2$ is of the most importance. These second-degree terms are exactly the *quadratic forms*, as we call them in linear algebra.

Quadratic forms.
Consider a vector space $\mathbb{R}^n$. Then a function $q : \mathbb{R}^n \to \mathbb{R}$ is called a quadratic form on $\mathbb{R}^n$ if

$$q(x) = q(x_1, \cdots, x_n) = \sum_{i,j=1}^{n} c_{ij} x_i x_j = \sum_{i=1}^{n} c_{ii} x_i^2 + \sum_{i \neq j} c_{ij} x_i x_j,$$

for some $c_{ij} \in \mathbb{R}$.

Note here that we do not exclude the possibility that c_{ii} and d_{ij} are all 0, however we shall focus mainly on the case where at least one of them is nonzero.

Example.

- On the real line $\mathbb{R}$ (one variable case), a quadratic form reduces to $q(x) = ax^2$ where $a \in \mathbb{R}$.
- In the two-variable case, we may name the two variables x and y. Then a quadratic form looks like $q(x, y) = ax^2 + bxy + cx^2$, where $a, b, c \in \mathbb{R}$.

Exercise 8.1. Let q be a quadratic form on $\mathbb{R}^n$. Show that there exists a symmetric $n \times n$ matrix A such that $q(x) = x^T Ax$.

Exercise 8.2. Consider the quadratic form

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

Find the symmetric matrix A in which $q(X) = X^T AX$ where $X = (x, y)$.

Exercise 8.3. Write down the quadratic form

$$q(x_1, x_2, x_3) = 2x_1^2 - x_2x_3 + x_1x_3 + x_3^2$$

in the vector-matrix form.

As we seen above, every matrix gives rise to a quadratic form and vice versa. The sign of the associated quadratic form then defines the *sign* or *positivity* of that matrix.

Positivity (or sign) of a matrix.
The positivity of a matrix plays a fundamental role of how a quadratic form behavee geometrically. Let A be a symmetric $n \times n$ matrix. We say that A is

- positive semidefinite if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$;
- positive definite if $x^T Ax > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$;
- negative semidefinite if $-A$ is positive semidefinite;
- negative definite if $-A$ is positive definite;
- indefinite if it is neither one from above.

It is not ideal for to verify positivity of a matrix using the above definition since the sign of $x^T Ax$ have to be tested at all x. This is where eigenvalues can be helpful, as they fully characterized the positivity of A.

Eigenvalue criteria for positivity.
Let A be a symmetric $n \times n$ matrix. Then

- A is positive semidefinite $\iff$ all eigenvalues of A are ≥ 0;
- A is positive definite $\iff$ all eigenvalues of A are > 0;
- A is negative semidefinite $\iff$ all eigenvalues of A are ≤ 0;
- A is negative definite $\iff$ all eigenvalues of A are < 0;
- A is indefinite $\iff$ A has both positive and negative eigenvalues.

Exercise 8.4. Determine the positivity of the following matrices.

(a) $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$

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

(c) $C = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(d) $D = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 9 \\ 1 & 0 & 0 \end{bmatrix}$

Convexity of quadratic functions and extremities.
A quadratic function of n variables is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x_1, \cdots, x_n) = f(x) = x^T Ax + b^T x + c$$

for some $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The quadratic part of this function is actually the quadratic form $x^T Ax$. The positivity of this term actually determines how the graph of this quadratic function would look like. To be precise, we have the following conclusions:

- f is convex if A is positive semidefinite;
- f is concave if A is negative semidefinite;
- f is a saddle surface if A is indefinite;

Extremities of quadratic functions.
If f is a quadratic function given by the formula $f(x) = x^T Ax + b^T x + c$, then the following criteria hold:

- If f is convex, then f can only have minimizers which can be found by solving the linear system $Ax = -\frac{b}{2}$.
- If f is concave, then f can only have maximizers which can be found be solving the linear system $Ax = -\frac{b}{2}$.
- If f is a saddle surface the it has neither a minimizer or a maximizer.

Note that none of the above criteria guarantees an existence.

Exercise 8.5. Let $A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ and $c = -2$ and consider the quadratic function $f(x) = x^T Ax + b^T x + c$.

(a) Classify the type of extreme points of f.

(b) Find all extreme points of f.

(c) Replace b with $b_0 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and repeat (a) and (b).

9. Dot product, norm 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*.

9.1 Dot products and norms.

Dot product.
Consider the space $\mathbb{R}^n$. The dot product on $\mathbb{R}^n$, denoted by $\cdot$, is defined between two vectors $u, v \in \mathbb{R}^n$ by

$$u \cdot v = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

This dot product enjoys the following fundamental properties for all vectors $x, y, z \in \mathbb{R}^n$ and all scalars $\alpha, \beta \in \mathbb{R}$:

(1) $x \cdot x \geq 0$ and $x \cdot x = 0 \iff x = 0$;

(2) $x \cdot y = y \cdot x$;

(3) $(\alpha x + \beta y) \cdot z = \alpha(x \cdot z) + \beta(y \cdot z)$.

Note that the above three conditions are the defining properties of the more general notion, which can defined on any (real) vector space, called *inner product.* Actually, any positive semidefinite matrix Q can be used to define an inner product by $x \cdot_Q y = x^T Q y$.

Norm.
Consider the space $\mathbb{R}^n$. The norm of any vector $x \in \mathbb{R}^n$ is defined by

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

This norm enjoys the following properties for all vectors $x, y \in \mathbb{R}^n$ and all scalars $\alpha \in \mathbb{R}$:

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

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

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

Again, these three conditions are the defining properties of a more general notion, also called a norm. Any function enjoying these three properties are called a norm, and this can be defined on any vector spaces. Moreover, if V is equipped with an inner product $x \cdot y$, then the norm is always defined by $\|x\| = \sqrt{x \cdot x}$.

In general, we interpret $\|x\|$ as the *length* or *magnitude* of the vector x. We can also use the norm to indirectly measure the distance between two vectors, namely x and y, by measuring the length of $x - y$. That is, the distance between x and y is $d(x, y) = \|x - y\|$.

Exercise 9.1. Consider two vectors $u = (3, 0, 1)$ and $v = (-2, 1, 2)$.

(a) Find $u \cdot v$.

(b) Find $\|u\|$ and $\|v\|$.

(c) Find how far is u from v.

(d) Verify that $\|u \pm v\| \leq \|u\| + \|v\|$.

We have discussed that a norm can be defined on any vector space. In fact, there can even be more than one norm on the same space depending on which norm is appropriate with the context. We hereby provide a few examples of norms on $\mathbb{R}^n$ itself as well as on the space $\mathbb{R}^{m \times n}$ of $m \times n$ matrices.

Exercise 9.2. Consider $\mathbb{R}^n$ and let $p \in [1, +\infty)$. The p-norm on $\mathbb{R}^n$ is defined by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

Moreover, the ∞-norm on $\mathbb{R}^n$ is defined by

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i|.$$

Let $u = (1, -1)$ and $v = (3, -2)$. Find $\|u\|_1$, $\|v\|_3$ and $\|u + v\|_\infty$.

Another important instant can be found in the space of matrices. The following examples are typical norms that are used frequently in practice.

Example. Let $\mathbb{R}^{m \times n}$ be the vector space of all $m \times n$ matrices.

- For $p \in [1, +\infty)$, the Frobenius p-norm is defined by

$$\|A\|_p = \left(\sum_{\substack{i=1,\cdots,m \\ j=1,\cdots,n}} |a_{ij}|^p \right)^{1/p}.$$

- The Frobenius ∞-norm if defined by

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

- Suppose that $\mathbb{R}^n$ and $\mathbb{R}^m$ are equipped with the norm $\|\cdot\|_{\mathbb{R}^n}$ and $\|\cdot\|_{\mathbb{R}^m}$, respectively. Then the relative norm for any matrix $A \in \mathbb{R}^{m\times n}$ is given by

$$\|A\| = \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_{\mathbb{R}^m}}{\|x\|_{\mathbb{R}^n}}.$$

9.2 Angle between two vectors and orthogonality

It is more interesting to see how dot products could help in geometric understanding of a vector space.

Angle between two vectors.
Suppose that V is a real vector space that has an inner product $x \cdot y$ and norm $\|x\| = \sqrt{x \cdot x}$. Then the two notions are linked by the equation

$$x \cdot y = \|x\|\|y\| \cos\theta,$$

where θ is the angle between the two vectors x and y in V. Note that the angle θ is only defined when both x and y are nonzero vectors. Rearranging this equation gives

$$\theta = \arccos\left(\frac{x \cdot y}{\|x\|\|y\|}\right).$$

With the given formula, we observe that $x \cdot y = 0$ if and only if either:

- one of the vectors is zero, or
- both vectors are nonzero and they are othogonal, i.e. $\theta = \frac{\pi}{2}$.

Hence, we conclude the orthogonality condition as follows.

Orthogonality criterion.
Two nonzero vectors x and y are orthogonal if and only if $x \cdot y = 0$.

Exercise 9.3. Let $x = (0, 0, 1)$, $y = (1, 1, 1)$ and $z = (0, 1, 0)$.

(a) Find the angle between x and y.

(b) Find the angle between x and z.

(c) Find the angle between y and z.

Orthogonal family.
Let $U = \{v_1, v_2, \cdots, v_k\}$ be a set containing vectors in a vector space V with inner product. We say that U is an orthogonal family (or pairwise orthogonal) if $v_i \cdot v_j = 0$ for all i, j with $i \neq j$.

Exercise 9.4.
Let $U = \{(1,0,1),(0,1,0),(-1,0,1)\}$ and $V = \{(1,0,1),(0,0,1),(0,1,0)\}$.

- Show that U is an orthogonal family.
- Show that V is not an orthogonal family.

Exercise 9.5. Show that if $U = \{v_1, \cdots, v_k\}$ is an orthogonal family, then it is linearly independent.

9.3 Orthogonal projection

Flavor of another vector – Orthogonal projection.
Let u and v be two vectors in a vector space V with dot product. Looking at them geometrically, the vector u may have some *hint* of the information given by the other vector v. We illustrate this idea by setting $u = (1, 0)$ and $v = (1, 1)$. Eventhough they are linearly independent, the vector v still has a flavor of leaning to the *east* which is the direction of u. On the other hand, if we take $w = (0, 1)$ as the vector pointing to the north, then v also leans as well towards the direction of w. Intuitively, the *flavor* of u that is contained in the vector v is called the *projection* of v onto u.

The orthogonal projection of u onto v is defined by

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

With the above description, we observe that u has no flavor of v means the two vectors are orthogonal. If $\operatorname{proj}_v u$ describes the flavor of v contained in u, removing this part from u should leave the result free of v. This could be observed by

$$\begin{aligned}
(u - \operatorname{proj}_v u) \cdot v &= u \cdot v - \operatorname{proj}_v u \cdot v \\
&= u \cdot v - (u \cdot v) \frac{v}{\|v\|^2} \cdot v \\
&= u \cdot v - (u \cdot v) \frac{v \cdot v}{\|v\|^2} \\
&= u \cdot v - u \cdot v \\
&= 0.
\end{aligned}$$

This observation will serve as the fundamental for the Gram-Schmidt process in the next section, where we orthogonalize a given basis.

Exercise 9.6. Let $u = (1, 0)$ and $v = (1, 1)$. Find $\operatorname{proj}_v u$ and $u - \operatorname{proj}_v u$.

Exercise 9.7. Let $u = (1, -1, 2)$, $v = (0, 2, 0)$ and $w = (1, 0, -1)$.

(a) Compute $y = u - \operatorname{proj}_v u$.

(b) Compute $z = u - \operatorname{proj}_v u - \operatorname{proj}_w u$.

(c) Show that y is orthogonal to v but not to w.

(d) Show that z is orthogonal to both v and w.

9.4 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.

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

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

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

Exercise 9.8. 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 9.9. 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.

10. Some applications

In this chapter, we discuss some applications of linear algebra, particularly on solving least squares problems that are solved through linear systems. To be more precise, we shall be talking about (1) least-square solution of an inconsistent linear system, and (2) linear regression approached from least mean-square-error perspective.

10.1 Least-sqaure solution of an inconsistent linear system

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 $\overline{x}$ in the sense that $A\overline{x}$ best approximates b.

Least-square solution.
A vector $\overline{x}$ is called the least-square solution of the linear system $Ax = b$ if $\overline{x}$ minimizes the squared norm $\|Ax - b\|^2$.

We shall now verify that

- The least-square solution is valid for any linear system.

  This could be addressed by observing that

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

  which shows that the minimum-norm solution is a finding a minimum of a quadratic function. Moreover, we know that

$$x^T (A^T A)x = (x^T A^T)Ax = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$$

  for all x, hence $A^T A$ is positive semidefinite and that underlying quadratic function is convex and the context of minimizing $\|Ax - b\|$

is relevant. Finally, finding the least-square solution of $Ax = b$ is equivalent to solving the following linear system

$$A^T Ax = A^T b.$$

- If $Ax = b$ is consistent, then the least-square solution and the traditional solution are the same.

  This can be simply deduced as any traditional solution $\overline{x}$ gives $\|Ax - b\| = 0$, minimizing the least squares.

Exercise 10.1. 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-square solution.

Exercise 10.2. Show that the following system has no traditional solution and, instead, find the least-square solution

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

10.2 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 an *affine* function to the observed data.

1-Dimensional linear regression.
Let us first consider the 1-dimensional case. This means that the observed data y_i depends *linearly* on a *single* controlled variable x_i, i.e.

$$y_i \approx ax_i + b$$

for some *good choice* of $a \in \mathbb{R}$ and $b \in \mathbb{R}$.

Suppose that we make m observations 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

For any $a, b \in \mathbb{R}$, we expect to have the approximation $y_i \approx ax_i + b$ with the least *mean-square error* (*MSE*). The MSE between the actual observation y_i and the predicted value $ax_i - b$ can be computed by

$$MSE = \sqrt{(y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \cdots + (y_m - (ax_m + b))^2}$$

Putting $y = (y_1, \cdots, y_m)$, $x = (x_1, \cdots, x_m)$, $\mathbf{1} = (\underbrace{1, \cdots, 1}_{m \text{ times}})$, $A = [x; \mathbf{1}]$ and $u = (a, b)$, we can further simplify the MSE formula above into

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

This implies the MSE reduces to a squared norm with a linear system inside. Therefore, selecting $u = (a, b)$ that minimizes the MSE is finding a least-square solution to the system $Au = y$. Finally, the parameter $u = (a, b)$ can be determined by solving the linear system

$$A^T A u = A^T y.$$

Exercise 10.3. 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.