Classnotes - MA-1102

Series and Matrices

Department of Mathematics
Indian Institute of Technology Madras

This classnote is only meant for academic use. It is not to be used for commercial purposes. For suggestions and improvements, contact Arindama Singh.
## Contents

### I Series

1 Series of Numbers .......................... 5
   1.1 Preliminaries ........................................... 5
   1.2 Sequences ............................................. 7
   1.3 Series ............................................... 11
   1.4 Some results on convergence ....................... 13
   1.5 Comparison tests ................................... 15
   1.6 Improper integrals .................................. 18
   1.7 Convergence tests for improper integrals .......... 20
   1.8 Tests of convergence for series .................. 24
   1.9 Alternating series ................................ 27

2 Series Representation of Functions .......... 30
   2.1 Power series ............................................. 30
   2.2 Determining radius of convergence ............... 32
   2.3 Taylor’s formulas .................................. 35
   2.4 Taylor series ......................................... 37
   2.5 Fourier series ...................................... 39

### II Matrices

3 Matrix Operations ............................ 51
   3.1 Examples of linear equations ....................... 51
   3.2 Basic matrix operations ............................. 53
   3.3 Transpose and adjoint ............................... 59
   3.4 Elementary row operations ......................... 62
   3.5 Row reduced echelon form ......................... 64
   3.6 Determinant .......................................... 66
   3.7 Computing inverse of a matrix ................. 69

4 Rank and Linear Equations ................. 73
   4.1 Matrices as linear maps ............................. 73
   4.2 Linear independence ............................... 74
4.3 Gram-Schmidt orthogonalization ........................................... 77
4.4 Determining linear independence ........................................ 79
4.5 Rank ................................................................................. 81
4.6 Solvability of linear equations ............................................. 82
4.7 Gauss-Jordan elimination .................................................. 84

5 Matrix Eigenvalue Problem ..................................................... 88
  5.1 Eigenvalues and eigenvectors ............................................. 88
  5.2 Characteristic polynomial ................................................ 88
  5.3 Special types of matrices ................................................. 91
  5.4 Diagonalization .............................................................. 94

Bibliography ........................................................................... 97

Index ..................................................................................... 98
Part I

Series
Chapter 1
Series of Numbers

1.1 Preliminaries

We use the following notation:

∅ = the empty set.

N = {1, 2, 3, ...}, the set of natural numbers.

Z = {..., −2, −1, 0, 1, 2, ...}, the set of integers.

Q = \{\frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N}\}, the set of rational numbers.

R = the set of real numbers.

R+ = the set of all positive real numbers.

N ⊊ Z ⊊ Q ⊊ R. The numbers in R − Q is the set of irrational numbers.

Examples are √2, 3.101101110111011110,... etc.

Along with the usual laws of +, ·, <, R satisfies the Archimedean property:

If a > 0 and b > 0, then there exists an n ∈ N such that na ≥ b.

Also R satisfies the completeness property:

Every nonempty subset of R having an upper bound has a least upper bound (lub) in R.

Explanation: Let A be a nonempty subset of R. A real number u is called an upper bound of A if each element of A is less than or equal to u. An upper bound ℓ of A is called a least upper bound if all upper bounds of A are greater than or equal to ℓ.

Notice that Q does not satisfy the completeness property. For example, the nonempty set A = \{x ∈ Q : x^2 < 2\} has an upper bound, say, 2. But its least upper bound is √2, which is not in Q.

Similar to lub, we have the notion of glb, the greatest lower bound of a subset of R. Let A be a nonempty subset of R. A real number v is called a lower bound of A if each element of A is greater than or equal to v. A lower bound m of A is called a greatest lower bound if all lower bounds of A are less than or equal to m. The completeness property of R implies that

Every nonempty subset of R having a lower bound has a greatest lower bound (glb) in R.
The lub acts as a maximum of a nonempty set and the glb acts as a minimum of the set. In fact, when the lub($A$) $\in A$, this lub is defined as the **maximum of** $A$ and is denoted as $\max(A)$. Similarly, if the glb($A$) $\in A$, this glb is defined as the **minimum of** $A$ and is denoted by $\min(A)$.

Moreover, both $\mathbb{Q}$ and $\mathbb{R} - \mathbb{Q}$ are **dense** in $\mathbb{R}$. That is, if $x < y$ are real numbers then there exist a rational number $a$ and an irrational number $b$ such that $x < a < y$ and $x < b < y$.

We may not explicitly use these properties of $\mathbb{R}$ but some theorems, whose proofs we will omit, can be proved using these properties. These properties allow $\mathbb{R}$ to be visualized as a number line:

From the Archemedian property it follows that the greatest integer function is well defined. That is, for each $x \in \mathbb{R}$, there corresponds, the number $\lfloor x \rfloor$, which is the greatest integer less than or equal to $x$. Moreover, the correspondence $x \mapsto \lfloor x \rfloor$ is a function.

We visualize $\mathbb{R}$ as a straight line made of expansible rubber of no thickness!

Let $a, b \in \mathbb{R}$, $a < b$.

$[a, b] = \{ x \in \mathbb{R} : a \leq x \leq b \}$, the closed interval $[a, b]$.

$(a, b) = \{ x \in \mathbb{R} : a < x < b \}$, the semi-open interval $(a, b)$.

$[a, b) = \{ x \in \mathbb{R} : a \leq x < b \}$, the semi-open interval $[a, b)$.

$(a, b] = \{ x \in \mathbb{R} : a < x \leq b \}$, the open interval $(a, b]$.

$(-\infty, b] = \{ x \in \mathbb{R} : x \leq b \}$, the closed infinite interval $(-\infty, b]$.

$(-\infty, b) = \{ x \in \mathbb{R} : x < b \}$, the open infinite interval $(-\infty, b)$.

$[a, \infty) = \{ x \in \mathbb{R} : x \geq a \}$, the closed infinite interval $[a, \infty)$.

$(a, \infty) = \{ x \in \mathbb{R} : x < \infty \}$, the open infinite interval $(a, \infty)$.

$(-\infty, \infty) = \mathbb{R}$, both open and closed infinite interval.

We also write $\mathbb{R}_+$ for $(0, \infty)$ and $\mathbb{R}_-$ for $(-\infty, 0)$. These are, respectively, the set of all positive real numbers, and the set of all negative real numbers.

A **neighborhood** of a point $c$ is an open interval $(c - \delta, c + \delta)$ for some $\delta > 0$.

The **absolute value** of $x \in \mathbb{R}$ is defined as $|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$

Thus $|x| = \sqrt{x^2}$. And $|-a| = a$ or $a \geq 0; |x - y|$ is the distance between real numbers $x$ and $y$. Moreover, if $a, b \in \mathbb{R}$, then

$$ | - a | = |a|, \quad |ab| = |a| |b|, \quad \left| \frac{a}{b} \right| = \frac{|a|}{|b|} \text{ if } b \neq 0, \quad |a + b| \leq |a| + |b|, \quad |a| - |b| \leq |a - b|.$$  

Let $x \in \mathbb{R}$ and let $a > 0$. The following are true:
1. $|x| = a$ iff $x = \pm a$.

2. $|x| < a$ iff $-a < x < a$ iff $x \in (-a, a)$.

3. $|x| \leq a$ iff $-a \leq x \leq a$ iff $x \in [-a, a]$.

4. $|x| > a$ iff $-a < x$ or $x > a$ iff $x \in (-\infty, -a) \cup (a, \infty)$ iff $x \in \mathbb{R} \setminus [-a, a]$.

5. $|x| \geq a$ iff $-a \leq x$ or $x \geq a$ iff $x \in (-\infty, -a] \cup [a, \infty)$ iff $x \in \mathbb{R} \setminus (-a, a)$.

Therefore, for $a \in \mathbb{R}, \delta > 0$,

$$|x - a| < \delta \text{ iff } a - \delta < x < a + \delta.$$ 

The following statements are useful in proving equalities from inequalities:

Let $a, b \in \mathbb{R}$.

1. If for each $\epsilon > 0$, $|a| < \epsilon$, then $a = 0$.

2. If for each $\epsilon > 0$, $a < b + \epsilon$, then $a \leq b$.

**Exercises for § 1.1**

1. Solve the following inequalities and show their solution sets in $\mathbb{R}$.
   - (a) $2x - 1 < x + 3$  
   - (b) $-\frac{x}{3} < 2x + 1$  
   - (c) $\frac{6}{x-1} \geq 5$.

2. Solve the following and show the solution sets in $\mathbb{R}$.
   - (a) $|2x - 3| = 7$  
   - (b) $|5 - \frac{2}{x}| < 1$  
   - (c) $|2x - 3| \leq 1$  
   - (d) $|2x - 3| \geq 1$.

3. Show the solution set of $x^2 - x - 2 \geq 0$ in $\mathbb{R}$.

4. Graph the inequality: $|x| + |y| \leq 1$.

### 1.2 Sequences

Since the fox runs 10 times faster than the rabbit, the fox starts 1 km behind. By the time he reaches the point where from the rabbit has started, the rabbit has moved ahead 100 m. By the time the fox reaches the second point, the rabbit has moved ahead 10 m. This way to surpass the rabbit, the fox must touch upon infinite number of points. Hence the fox can never surpass the rabbit!

The question is whether $1000 + 100 + 1 + 1/10 + 1/100 + \cdots$ is a number. We rather take the partial sums

$$1000, \ 1000 + 100, \ 1000 + 100 + 1, \ 1000 + 100 + 1 + 1/10, \ \cdots$$

which are numbers and ask whether the sequence of these numbers approximates certain real number?
For example, we may approximate $\sqrt{2}$ by the usual division procedure. We get the sequence

$$1, 1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, 1.4142135, 1.41421356, \ldots$$

Does it approximate $\sqrt{2}$?

In general, we define a sequence of real numbers as a function $f : \mathbb{N} \to \mathbb{R}$. The values of the function are $f(1), f(2), f(3), \ldots$ These are called the terms of the sequence. With $f(n) = x_n$, the $n$th term of the sequence, we write the sequence in many ways such as

$$(x_n) = (x_n)_{n=1}^{\infty} = \{x_n\}_{n=1}^{\infty} = \{x_n\} = (x_1, x_2, x_3, \ldots)$$

showing explicitly its terms. For example, $x_n = n$ defines the sequence

$$f : \mathbb{N} \to \mathbb{R} \text{ with } f(n) = n,$$

that is, the sequence is $(1, 2, 3, 4, \ldots)$, the sequence of natural numbers. Informally, we say “the sequence $x_n = n$.”

The sequence $x_n = 1/n$ is the sequence $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots)$; formally, $\{1/n\}$ or $(1/n)$.

The sequence $x_n = 1/n^2$ is the sequence $(1/n^2)$, or $\{1/n^2\}$, or $(1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \ldots)$.

The constant sequence $\{c\}$ for a given real number $c$ is the constant function $f : \mathbb{N} \to \mathbb{R}$, where $f(n) = c$ for each $n \in \mathbb{N}$. It is $(c, c, c, \ldots)$.

A sequence is an infinite list of real numbers; it is ordered like natural numbers, and unlike a set of numbers.

There are sequences which approximate a real number and there are sequences which do not approximate any real number.

For example, $\{1/n\}$ approximates the real number 0, whereas $\{n\}$ approximates no real number. Also the sequence $(1, -1, 1, -1, 1, -1, \ldots)$, which may be written as $\{(-1)^n\}$, approximates no real number.

We would say that the sequence $\{1/n\}$ converges to 0 and the other two sequences diverge. The sequence $\{n\}$ diverges to $\infty$ and the sequence $\{(-1)^n\}$ diverges.

Look at the sequence $\{1/n\}$ closely. We feel that eventually, it will approximate 0, meaning that whatever tolerance I fix there is a term in the sequence, after which every term is away from 0 within that tolerance. What does it mean? Suppose I am satisfied with an approximation to 0 within the tolerance 5. Then, I see that the terms of the sequence, starting with 1 and then $1/2, 1/3, \ldots$, all of them are within 5 units away from 0. In fact, $|1/n - 0| < 5$ for all $n$. Now, you see, bigger the tolerance, it is easier to fix a tail of the sequence satisfying the tolerance condition. Suppose I fix my tolerance as 1/5. Then I see that the sixth term onwards, all the terms of the sequence are within 1/5 distance away from 0. That is, $|1/n - 0| < 1/5$ for all $n \geq 6$. If I fix my tolerance as $10^{-10}$. Then we see that $|1/n - 0| < 10^{-10}$ for all $n \geq 10^{10} + 1$. This leads to the formal definition of convergence of a sequence.

Let $\{x_n\}$ be a sequence. Let $a \in \mathbb{R}$. We say that $\{x_n\}$ converges to $a$ iff for each $\epsilon > 0$, there exists an $m \in \mathbb{N}$ such that if $n \geq m$ is any natural number, then $|x_n - a| < \epsilon$. 

8
Example 1.1. Show that the sequence \( \{1/n \} \) converges to 0.

Let \( \epsilon > 0 \). Take \( m = \lfloor 1/\epsilon \rfloor + 1 \). That is, \( m \) is the natural number such that \( m - 1 \leq 1/\epsilon < m \). Then \( m < \epsilon \). Moreover, if \( n > m \), then \( 1/n < 1/m < \epsilon \). That is, for any such given \( \epsilon > 0 \), there exists an \( m \), (we have defined it here) such that for every \( n \geq m \), we see that \( |1/n - 0| < \epsilon \). Therefore, \( \{1/n\} \) converges to 0.

Notice that in Example 1.1, we could have resorted to the Archimedian principle and chosen any natural number \( m > 1/\epsilon \).

Now that \( \{1/n\} \) converges to 0, the sequence whose first 1000 terms are like \( (n) \) and 1001st term onward, it is like \( (1/n) \) also converges to 0. Because, for any given \( \epsilon > 0 \), we choose our \( m \) as \( \lfloor 1/\epsilon \rfloor + 1001 \). Moreover, the sequence whose first 1000 terms are like \( \{n\} \) and then onwards it is \( 1,1/2,1/3,\ldots \) converges to 0 for the same reason. That is, convergence behavior of a sequence does not change if first finite number of terms are changed.

For a constant sequence \( x_n = c \), suppose \( \epsilon > 0 \) is given. We see that for each \( n \in \mathbb{N} \), \( |x_n - c| = 0 < \epsilon \). Therefore, the constant sequence \( \{c\} \) converges to \( c \).

Sometimes, it is easier to use the condition \( |x_n - a| < \epsilon \) as \( a - \epsilon < x_n < a + \epsilon \).

Such an open interval \( (a - \epsilon,a + \epsilon) \) for some \( \epsilon > 0 \) is called a neighborhood of \( a \).

A sequence thus converges to \( a \) implies the following:

1. Each neighborhood of \( a \) contains a tail of the sequence.
2. Every tail of the sequence contains numbers arbitrarily close to \( a \).

We say that a sequence \( \{x_n\} \) converges iff it converges to some \( a \). A sequence diverges iff it does not converge to any real number.

There are two special cases of divergence.

Let \( \{x_n\} \) be a sequence. We say that \( \{x_n\} \) diverges to \( \infty \) iff for every \( r > 0 \), there exists an \( m \in \mathbb{N} \) such that if \( n > m \) is any natural number, then \( x_n > r \).

We call an open interval \( (r,\infty) \) a neighborhood of \( \infty \). A sequence thus diverges to \( \infty \) implies the following:

1. Each neighborhood of \( \infty \) contains a tail of the sequence.
2. Every tail of the sequence contains arbitrarily large positive numbers.

In this case, we write \( \lim_{n \to \infty} x_n = \infty \); we also write it as “\( x_n \to \infty \) as \( n \to \infty \)” or as \( x_n \to \infty \).

We say that \( \{x_n\} \) diverges to \( -\infty \) iff for every \( r > 0 \), there exists an \( m \in \mathbb{N} \) such that if \( n > m \) is any natural number, then \( x_n < -r \).

Calling an open interval \( (-\infty,s) \) a neighborhood of \( -\infty \), we see that a sequence diverges to \( -\infty \) implies the following:
1. Each neighborhood of $-\infty$ contains a tail of the sequence.

2. Every tail of the sequence contains arbitrarily small negative numbers.

In this case, we write $\lim_{n \to \infty} x_n = -\infty$; we also write it as “$x_n \to -\infty$ as $n \to \infty$” or as $x_n \to -\infty$.

We use a unified notation for convergence to a real number and divergence to $\pm \infty$.

For $\ell \in \mathbb{R} \cup \{-\infty, \infty\}$, the notations

$$\lim_{n \to \infty} x_n = \ell, \quad x_n \to \ell \text{ as } n \to \infty, \quad x_n \to \ell$$

all stand for the phrase limit of $\{x_n\}$ is $\ell$. When $\ell \in \mathbb{R}$, the limit of $\{x_n\}$ is $\ell$ means that $\{x_n\}$ converges to $\ell$; and when $\ell = \pm \infty$, the limit of $\{x_n\}$ is $\ell$ means that $\{x_n\}$ diverges to $\pm \infty$.

**Example 1.2.** Show that (a) $\lim \sqrt{n} = \infty$; (b) $\lim \ln(1/n) = -\infty$.

(a) Let $r > 0$. Choose an $m > r^2$. Let $n > m$. Then $\sqrt{n} > \sqrt{m} > r$. Therefore, $\lim \sqrt{n} = \infty$.

(b) Let $r > 0$. Choose a natural number $m > e^r$. Let $n > m$. Then $1/n < 1/m < e^{-r}$. Consequently, $\ln(1/n) < \ln e^{-r} = -r$. Therefore, $\ln(1/n) \to -\infty$.

We state a result connecting the limit notion of a function and limit of a sequence. We use the idea of a constant sequence. A sequence $\{a_n\}$ is called a constant sequence if $a_n = \alpha$ for each $n$, where $\alpha$ is a fixed real number.

We state some results about sequences which will be helpful to us later.

**Theorem 1.1.** (Sandwich Theorem): Let $\{x\}$, $\{y_n\}$, and $\{z_n\}$ be sequences such that $x_n \leq y_n \leq z_n$ holds for all $n$ greater than some $m$. If $x_n \to \ell$ and $z_n \to \ell$, then $y_n \to \ell$.

**Theorem 1.2.** Limits of sequences to Limits of functions: Let $a < c < b$. Let $f : D \to \mathbb{R}$ be a function where $D$ contains $(a, c) \cup (c, b)$. Let $\ell \in \mathbb{R}$. Then $\lim_{x \to c} f(x) = \ell$ iff for each non-constant sequence $\{x_n\}$ converging to $c$, the sequence of functional values $\{f(x_n)\}$ converges to $\ell$.

The same way, limit of a sequence $\{a_n\}$ as $n \to \infty$ is related to the limit of a function $f(x)$ as $x \to \infty$ provided some conditions are satisfied.

**Theorem 1.3.** Let $k \in \mathbb{N}$. Let $f(x)$ be a function defined for all $x \geq k$. Let $\{a_n\}$ be a sequence of real numbers such that $a_n = f(n)$ for all $n \geq k$. If $\lim_{x \to \infty} f(x) = \ell$, then $\lim_{n \to \infty} a_n = \ell$.

As an application, consider the function $\ln x$. We know that it is defined on $[1, \infty)$. Using L’Hospital’s rule, we have

$$\lim_{x \to \infty} \frac{\ln x}{x} = \lim_{x \to \infty} \frac{1}{x} = 0.$$ 

Therefore, $\lim_{n \to \infty} \frac{\ln n}{n} = \lim_{x \to \infty} \frac{\ln x}{x} = 0$.

Our main goal is to study when an infinite sum can represent a number.
Exercises for § 1.2

1. Show the following:

   (a) \( \lim_{n \to \infty} \frac{\ln n}{n} = 0 \)
   (b) \( \lim_{n \to \infty} n^{1/n} = 1 \)
   (c) \( \lim_{n \to \infty} x^{1/n} = 1 \) for \( x > 0 \)
   (d) \( \lim_{n \to \infty} x^n = 0 \) for \( |x| < 1 \)
   (e) \( \lim_{n \to \infty} \frac{x^n}{n!} = 0 \)
   (f) \( \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x \)

2. Find the limit of the sequence \( \{a_n\} \) or show that it diverges.

   (a) \( a_n = \frac{2n + 1}{1 - 3\sqrt{n}} \)
   (b) \( a_n = \frac{n + 1}{2n} \frac{n - 4}{n^{3/2}} \)
   (c) \( \sin \left(\frac{\pi}{2} + \frac{1}{n}\right) \)

1.3 Series

A series is an infinite sum of numbers. As it is, two numbers can be added; so by induction, a finite of them can also be added. For an infinite sum to be meaningful, we look at the sequence of partial sums. Let \( \{x_n\} \) be a sequence. The series \( x_1 + x_2 + \cdots + x_n + \cdots \) is meaningful when another sequence, namely,

\[
\sum_{k=1}^{n} x_k, \ldots
\]

is convergent. The infinite sum itself is denoted by \( \sum_{k=1}^{\infty} x_n \) and also by \( \sum x_n \).

We say that the series \( \sum x_n \) is convergent iff the sequence \( \{s_n\} \) is convergent, where the \( n \)th partial sum \( s_n \) is given by \( s_n = \sum_{k=1}^{n} x_k \).

Thus we may define convergence of a series as follows:

We say that the series \( \sum x_n \) converges to \( \ell \in \mathbb{R} \) iff for each \( \epsilon > 0 \), there exists an \( m \in \mathbb{N} \) such that for each \( n \geq m \), \( |\sum_{k=1}^{n} x_k - \ell| < \epsilon \). In this case, we write \( \sum x_n = \ell \).

Further, we say that the series \( \sum x_n \) converges iff the series converges to some \( \ell \in \mathbb{R} \).

The series is said to be divergent iff it is not convergent.

Similar to convergence, if the sequence of partial sums \( \{s_n\} \) diverges to \( \pm \infty \), we say that the series \( \sum x_n \) diverges to \( \pm \infty \).

That is, the series \( \sum x_n \) diverges to \( \infty \) iff for each \( r > 0 \), there exists \( m \in \mathbb{N} \) such that for each \( n \geq m \), \( \sum_{k=1}^{n} x_k > r \). We write it as \( \sum x_n = \infty \).

Similarly, the series \( \sum x_n \) diverges to \( -\infty \) iff for each \( r > 0 \), there exists \( m \in \mathbb{N} \) such that for each \( n \geq m \), \( \sum_{k=1}^{n} x_k < -r \). We write it as \( \sum x_n = -\infty \).

Notice that ‘converges to a real number’ and ‘diverges to \( \pm \infty \)’ both are written the same way.

In the unified notation, we say that a series \( \sum a_n \) sums to \( \ell \in \mathbb{R} \cup \{\infty, -\infty\} \), when either the series converges to the real number \( \ell \) or it diverges to \( \pm \infty \). In all these cases we write \( \sum a_n = \ell \).

Moreover, if a series converges to some real number \( \ell \), then the partial sums can be thought of as approximations to the number \( \ell \).
There can be series which diverge but neither to ∞ nor to −∞. For example, the series
\[ \sum_{n=0}^{\infty} (-1)^n = 1 - 1 + 1 - 1 + 1 - 1 + \cdots \]
neither diverges to ∞ nor to −∞. But it is a divergent series. Can you see why?

**Example 1.3.**
(a) The series \( \sum_{n=1}^{\infty} \frac{1}{2^n} \) converges to 1. Because, if \( \{s_n\} \) is the sequence of partial sums, then
\[
s_n = \sum_{k=1}^{n} \frac{1}{2^k} = \frac{1}{2} \cdot \frac{1 - (1/2)^n}{1 - 1/2} = 1 - \frac{1}{2^n} \to 1.
\]
(b) The series \( 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \) diverges to ∞. To see this, let \( s_n = \sum_{k=1}^{n} \frac{1}{k} \) be the partial sum up to \( n \) terms. Let \( m \) be the natural number such that \( 2^m \leq n < 2^{m+1} \). Then
\[
s_n = \sum_{k=1}^{n} \frac{1}{k} \geq 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^m - 1} \geq 1 + \left(\frac{1}{2} + \frac{1}{3}\right) + \left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right) + \cdots + \left(\sum_{k=2^{m-1}}^{2^m-1} \frac{1}{k}\right) > 1 + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \cdots + \left(\sum_{k=2^{m-1}}^{2^m-1} \frac{1}{2^m}\right) = 1 + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} = 1 + \frac{m-1}{2}.
\]
As \( n \to \infty \), we see that \( m \to \infty \). Consequently, \( s_n \to \infty \). That is, the series diverges to ∞. This is called the **harmonic series**.
(c) The series \( -1 - 2 - 3 - 4 - \cdots - n - \cdots \) diverges to −∞.
(d) The series \( 1 - 1 + 1 - 1 + \cdots \) diverges. It neither diverges to ∞ nor to −∞. Because, the sequence of partial sums here is 1, 0, 1, 0, 1, 0, 1, . . .

**Example 1.4.** Let \( a \neq 0 \). Consider the geometric series
\[ \sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + ar^3 + \cdots. \]
The \( n \)th partial sum of the geometric series is
\[
s_n = a + ar + ar^2 + ar^3 + \cdots + ar^{n-1} = \frac{a(1 - r^n)}{1 - r}.
\]
(a) If \( |r| < 1 \), then \( r^n \to 0 \). The geometric series converges to \( \lim_{n \to \infty} s_n = \frac{a}{1 - r} \).
Therefore, \( \sum_{n=0}^{\infty} ar^n = \sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1 - r} \).
(b) If \(|r| \geq 1\), then \(r^n\) diverges. The geometric series \(\sum ar^{n-1}\) diverges.

**Exercises for § 1.3**

1. You drop a ball from \(a\) meters above a flat surface. Each time the ball hits the surface after falling a distance \(h\), it rebounds a distance \(rh\), where \(r\) is positive but less than 1. Find the total distance the ball travels up and down.

2. Express \(5.232323\cdots\) in the form \(m/n\) for \(m, n \in \mathbb{N}\).

**1.4 Some results on convergence**

If a series sums to \(\ell\), then it cannot sum to \(s\) where \(s \neq \ell\).

**Theorem 1.4.** If a series \(\sum a_n\) sums to \(\ell \in \mathbb{R} \cup \{\infty, -\infty\}\), then \(\ell\) is unique.

**Proof.** Suppose the series \(\sum a_n\) sums to \(\ell\) and also to \(s\), where both \(\ell, s \in \mathbb{R} \cup \{\infty, -\infty\}\). Suppose that \(\ell \neq s\). We consider the following exhaustive cases.

**Case 1:** \(\ell, s \in \mathbb{R}\). Then \(|s - \ell| > 0\). Choose \(\epsilon = |s - \ell|/3\). We have natural numbers \(k\) and \(m\) such that for every \(n \geq k\) and \(n \geq m\),

\[
\left| \sum_{j=1}^{n} a_j - \ell \right| \leq \epsilon \quad \text{and} \quad \left| \sum_{j=1}^{n} a_j - s \right| \leq \epsilon.
\]

Fix one such \(n\), say \(M > \max\{k, m\}\). Both the above inequalities hold for \(n = M\). Then

\[
|s - \ell| = \left| s - \sum_{j=1}^{M} a_j + \sum_{j=1}^{M} a_j - \ell \right| \leq \left| \sum_{j=1}^{M} a_j - s \right| + \left| \sum_{j=1}^{M} a_j - \ell \right| \leq 2\epsilon < |s - \ell|.
\]

This is a contradiction.

**Case 2:** \(\ell \in \mathbb{R}\) and \(s = \infty\). Then there exists a natural number \(k\) such that for every \(n \geq k\), we have

\[
\left| \sum_{j=1}^{n} a_j - \ell \right| < 1.
\]

Since the series sums to \(\infty\), we have \(m \in \mathbb{N}\) such that for every \(n \geq m\),

\[
\sum_{j=1}^{n} a_j > \ell + 1.
\]

Now, fix an \(M > \max\{k, m\}\). Then both of the above hold for this \(n = M\). Therefore,

\[
\sum_{j=1}^{M} a_j < \ell + 1 \quad \text{and} \quad \sum_{j=1}^{M} a_j > \ell + 1.
\]

This is a contradiction.
Case 3: $\ell \in \mathbb{R}$ and $s = -\infty$. It is similar to Case 2. Choose “less than $\ell - 1$”.

Case 4: $\ell = \infty$, $s = -\infty$. Again choose an $M$ so that $\sum_{j=1}^{n} a_n$ is both greater than 1 and also less than $-1$ leading to a contradiction.

The results in the following theorem are sometimes helpful in ascertaining the convergence of a series without knowing what the sum of the series is.

**Theorem 1.5.** (1) (Cauchy Criterion) A series $\sum a_n$ converges iff for each $\epsilon > 0$, there exists a $k \in \mathbb{N}$ such that $|\sum_{j=m}^{n} a_j| < \epsilon$ for all $n \geq m \geq k$.

(2) (Weirstrass Criterion) Let $\sum a_n$ be a series of non-negative terms. Suppose there exists $c \in \mathbb{R}$ such that each partial sum of the series is less than $c$, i.e., for each $n$, $\sum_{j=1}^{n} a_j < c$. Then $\sum a_n$ is convergent.

The following result sometimes helps in ascertaining that a given series diverges.

**Theorem 1.6.** If a series $\sum a_n$ converges, then the sequence $\{a_n\}$ converges to 0.

*Proof:* Let $s_n$ denote the partial sum $\sum_{k=1}^{n} a_k$. Then $a_n = s_n - s_{n-1}$. If the series converges, say, to $\ell$, then $\lim s_n = \ell = \lim s_{n-1}$. It follows that $\lim a_n = 0$. □

It says that if $\lim a_n$ does not exist, or if $\lim a_n$ exists but is not equal to 0, then the series $\sum a_n$ diverges.

The series $\sum_{n=1}^{\infty} \frac{-n}{3n+1}$ diverges because $\lim_{n \to \infty} \frac{-n}{3n+1} = -\frac{1}{3} \neq 0$.

The series $\sum (-1)^n$ diverges because $\lim (-1)^n$ does not exist.

Notice what Theorem 1.6 does not say. The harmonic series diverges even though $\lim \frac{1}{n} = 0$.

Recall that for a real number $\ell$ our notation says that $\ell + \infty = \infty$, $\ell - \infty = -\infty$ and $\ell \cdot (\pm \infty) = \pm \infty$. Similarly, we accepted the convention that $\infty + \infty = \infty$ and $-\infty - \infty = -\infty$. We use these conventions in the statement of our next result.

**Theorem 1.7.** (1) If $\sum a_n$ converges to $a$ and $\sum b_n$ sums to $b$, then the series $\sum (a_n + b_n)$ sums to $a + b$; $\sum (a_n - b_n)$ sums to $a - b$; and $\sum k b_n$ sums to $k b$, where $k$ is any real number.

(2) If $\sum a_n$ converges and $\sum b_n$ diverges, then $\sum (a_n + b_n)$ and $\sum (a_n - b_n)$ diverge.

(3) If $\sum a_n$ diverges and $k \neq 0$, then $\sum k a_n$ diverges.

Proofs of the statements in Theorem 1.7 are left as exercises. However, write the first statement in the above theorem as separate statements taking $b$ as a real number, as $\infty$, and as $-\infty$.

Notice that sum of two divergent series can converge. For example, both $\sum (1/n)$ and $\sum (-1/n)$ diverge but their sum $\sum 0$ converges.

Since deleting a finite number of terms of a sequence does not alter its convergence, omitting a finite number of terms or adding a finite number of terms to a convergent (divergent) series implies
the convergence (divergence) of the new series. Of course, the sum of the convergent series will be affected. For example,

\[ \sum_{n=3}^{\infty} \left( \frac{1}{2n} \right) = \sum_{n=1}^{\infty} \left( \frac{1}{2n} \right) - \frac{1}{2} - \frac{1}{4}. \]

However,

\[ \sum_{n=3}^{\infty} \left( \frac{1}{2n-2} \right) = \sum_{n=1}^{\infty} \left( \frac{1}{2n} \right). \]

This is called re-indexing the series. As long as we preserve the order of the terms of the series, we can re-index without affecting its convergence and sum.

**Exercises for § 1.4**

1. Find the \( n \)-th partial sum of the series \( \frac{5}{2.3} + \frac{5}{3.4} + \frac{5}{4.5} + \cdots \) and then sum the series.

2. Find the limits of the following series:
   (a) \( \sum_{n=1}^{\infty} \left( \frac{1}{2^n} + \frac{(-1)^n}{5^n} \right) \)
   (b) \( \sum_{n=1}^{\infty} \frac{40n}{(2n - 1)^2(2n + 1)^2} \)

3. Find the values of \( x \) for which the series converges. And then find the limit for those values of \( x \).
   (a) \( \sum_{n=1}^{\infty} \frac{(-1)^n}{2} \left( \frac{1}{3 + \sin x} \right)^n \)
   (b) \( \sum_{n=1}^{\infty} \frac{(-1)^n}{((x - 3)/2)^n} \)

### 1.5 Comparison tests

There are various ways to determine whether a series converges or not, occasionally, some information on its sum is also obtained.

**Theorem 1.8. (Comparison Test)** Let \( \sum a_n \) and \( \sum b_n \) be series of non-negative terms. Suppose there exists \( k > 0 \) such that \( 0 \leq a_n \leq kb_n \) for each \( n \) greater than some natural number \( m \).

1. If \( \sum b_n \) converges, then \( \sum a_n \) converges.
2. If \( \sum a_n \) diverges to \( \infty \), then \( \sum b_n \) diverges to \( \infty \).

**Proof:** (1) Consider all partial sums of the series having more than \( m \) terms. We see that

\[ a_1 + \cdots + a_m + a_{m+1} + \cdots + a_n \leq a_1 + \cdots + a_m + k \sum_{j=m+1}^{n} b_j. \]

Since \( \sum b_n \) converges, so does \( \sum_{j=m+1}^{n} b_j \). By Weirstrass criterion, \( \sum a_n \) converges.

(2) Similar to (1).

**Caution:** The comparison test holds for series of non-negative terms.
Theorem 1.9. (Ratio Comparison Test) Let $\sum a_n$ and $\sum b_n$ be series of non-negative terms. Suppose there exists $m \in \mathbb{N}$ such that for each $n > m$, $a_n > 0$, $b_n > 0$, and $\frac{a_{n+1}}{a_n} \leq \frac{b_{n+1}}{b_n}$.

1. If $\sum b_n$ converges, then $\sum a_n$ converges.

2. If $\sum a_n$ diverges to $\infty$, then $\sum b_n$ diverges to $\infty$.

Proof: For $n > m$,

$$a_n = \frac{a_n}{a_{n-1}} \cdot \frac{a_{n-1}}{a_{n-2}} \cdots \frac{a_{m+1}}{a_m} a_m \leq \frac{b_n}{b_{n-1}} \cdot \frac{b_{n-1}}{b_{n-2}} \cdots \frac{b_{m+1}}{b_m} b_m = \frac{a_m}{b_m} b_n.$$  

By the Comparison test, if $\sum b_n$ converges, then $\sum a_n$ converges. This proves (1). And, (2) follows from (1) by contradiction. $\square$

Theorem 1.10. (Limit Comparison Test) Let $\sum a_n$ and $\sum b_n$ be series of non-negative terms. Suppose that there exists $m \in \mathbb{N}$ such that for each $n > m$, $a_n > 0$, $b_n > 0$, and $\lim_{n \to \infty} \frac{a_n}{b_n} = k$.

1. If $k > 0$ then $\sum b_n$ and $\sum a_n$ converge or diverge to $\infty$, together.

2. If $k = 0$ and $\sum b_n$ converges, then $\sum a_n$ converges.

3. If $k = \infty$ and $\sum b_n$ diverges to $\infty$ then $\sum a_n$ diverges to $\infty$.

Proof: (1) Let $\epsilon = k/2 > 0$. The limit condition implies that there exists $m \in \mathbb{N}$ such that

$$\frac{k}{2} < \frac{a_n}{b_n} < \frac{3k}{2} \quad \text{for each } n > m.$$  

By the Comparison test, the conclusion is obtained.

(2) Let $\epsilon = 1$. The limit condition implies that there exists $m \in \mathbb{N}$ such that

$$-1 < \frac{a_n}{b_n} < 1 \quad \text{for each } n > m.$$  

Using the right hand inequality and the Comparison test we conclude that convergence of $\sum b_n$ implies the convergence of $\sum a_n$.

(3) If $k > 0$, $\lim(b_n/a_n) = 1/k$. Use (1). If $k = \infty$, $\lim(b_n/a_n) = 0$. Use (2). $\square$

Example 1.5. For each $n \in \mathbb{N}$, $n! \geq 2^{n-1}$. That is, $\frac{1}{n!} \leq \frac{1}{2^{n-1}}$.

Since $\sum_{n=1}^{\infty} \frac{1}{2^{n-1}}$ is convergent, $\sum_{n=1}^{\infty} \frac{1}{n!}$ is convergent. Therefore, adding 1 to it, the series

$$1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \cdots$$

is convergent. In fact, this series converges to $e$. To see this, consider

$$s_n = 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!}, \quad t_n = \left(1 + \frac{1}{n}\right)^n.$$
By the Binomial theorem,
\[ t_n = 1 + \frac{1}{2!}(1 - \frac{1}{n}) + \cdots + \frac{1}{n!} \left[ (1 - \frac{1}{n})(1 - \frac{2}{n}) \cdots (1 - \frac{n-1}{n}) \right] \leq s_n. \]

Thus taking limit as \( n \to \infty \), we have
\[ e = \lim_{n \to \infty} t_n \leq \lim_{n \to \infty} s_n. \]

Also, for \( n > m \), where \( m \) is any fixed natural number,
\[ t_n \geq 1 + \frac{1}{2!}(1 - \frac{1}{n}) + \cdots + \frac{1}{m!} \left[ (1 - \frac{1}{n})(1 - \frac{2}{n}) \cdots (1 - \frac{m-1}{n}) \right] \]

Taking limit as \( n \to \infty \) we have
\[ e = \lim_{n \to \infty} t_n \geq s_m. \]

Since \( m \) is arbitrary, taking the limit as \( m \to \infty \), we have
\[ e \geq \lim_{m \to \infty} s_m. \]

Therefore, \( \lim_{m \to \infty} s_m = e \). That is, the series \( \sum_{n=0}^{\infty} \frac{1}{n!} = e \).

**Example 1.6.** Determine when the series \( \sum_{n=1}^{\infty} \frac{n+7}{n(n+3)\sqrt{n+5}} \) converges.

Let \( a_n = \frac{n+7}{n(n+3)\sqrt{n+5}} \) and \( b_n = \frac{1}{n^{3/2}} \). Then
\[ \frac{a_n}{b_n} = \frac{\sqrt{n(n+7)}}{(n+3)\sqrt{n+5}} \to 1 \text{ as } n \to \infty. \]

Since \( \frac{1}{n^{3/2}} \) is convergent, Limit comparison test says that the given series is convergent.

**Exercises for § 1.5**

1. Determine whether the following series converge or diverge:
   (a) \( \sum_{n=1}^{\infty} \frac{1}{2\sqrt{n} + \sqrt{n}} \)
   (b) \( \sum_{n=1}^{\infty} \frac{1 + \cos n}{n^2} \)
   (c) \( \sum_{n=1}^{\infty} \frac{2 + (-1)^n}{(5/4)^n} \)
   (d) \( \sum_{n=3}^{\infty} \frac{1}{\ln \ln n} \)
   (e) \( \sum_{n=1}^{\infty} \left( \frac{1 - 3}{n} \right)^n \)
   (f) \( \sum_{n=1}^{\infty} \sin \frac{1}{n} \)
   (g) \( \sum_{n=1}^{\infty} a_n \), where \( a_1 = 1/2 \), \( a_{n+1} = \frac{n + \ln n}{n + 10} a_n \).
   (h) \( \sum_{n=1}^{\infty} a_n \), where \( a_1 = 3 \), \( a_{n+1} = \frac{n}{n + 1} a_n \).
   (i) \( \sum_{n=1}^{\infty} a_n \), where \( a_n = n/2^n \) if \( n \) is prime; else, \( a_n = 1/2^n \).
1.6 Improper integrals

There is a nice connection between integrals and series. To see this connection, we consider the so called improper integrals.

In the definite integral \( \int_a^b f(x) \, dx \) we required that both \( a, b \) are finite and also the range of \( f(x) \) is a subset of some finite interval. However, there are functions which violate one or both of these requirements, and yet, the area under the curves and above the \( x \)-axis remain bounded.

Such integrals are called Improper Integrals. Suppose \( f(x) \) is continuous on \([0, \infty)\). It makes sense to write
\[
\int_0^\infty f(x) \, dx = \lim_{b \to \infty} \int_0^b f(x) \, dx
\]
provided that the limit exists. In such a case, we say that the improper integral \( \int_0^\infty f(x) \, dx \) converges and its value is given by the limit. We say that the improper integral diverges iff it is not convergent.

Obviously, we are interested in computing the value of an improper integral, in which case, the integral is required to converge. Integrals of the type \( \int_a^b f(x) \, dx \) can become improper when \( f(x) \) is not continuous at a point in the interval \([a, b]\). Here are the possible types of improper integrals.

1. If \( f(x) \) is continuous on \([a, \infty)\), then
\[
\int_a^\infty f(x) \, dx = \lim_{b \to \infty} \int_a^b f(x) \, dx.
\]
2. If \( f(x) \) is continuous on \((-\infty, b]\), then
\[
\int_{-\infty}^b f(x) \, dx = \lim_{a \to -\infty} \int_a^b f(x) \, dx.
\]
3. If \( f(x) \) is continuous on \((-\infty, \infty)\), then
\[
\int_{-\infty}^\infty f(x) \, dx = \int_{-\infty}^c f(x) \, dx + \int_c^\infty f(x) \, dx, \text{ for any } c \in \mathbb{R}.
\]
4. If \( f(x) \) is continuous on \((a, b]\) and discontinuous at \( x = a \), then
\[
\int_a^b f(x) \, dx = \lim_{t \to a^+} \int_t^b f(x) \, dx.
\]
5. If \( f(x) \) is continuous on \([a, b)\) and discontinuous at \( x = b \), then
\[
\int_a^b f(x) \, dx = \lim_{t \to b^-} \int_a^t f(x) \, dx.
\]
6. If \( f(x) \) is continuous on \([a, c) \cup (c, b]\) and discontinuous at \( x = c \), then
\[
\int_a^b f(x) \, dx = \int_a^c f(x) \, dx + \int_c^b f(x) \, dx.
\]
In each case, if the limit of the concerned integral is finite, then we say that the improper integral (on the left) **converges**, else, the improper integral **diverges**; the finite value as obtained from the limit is the **value** of the improper integral. A convergent improper integral converges to its value.

Two important sub-cases of divergent improper integrals are when the limit of the concerned integral is \( \infty \) or \(-\infty\). In these cases, we say that the improper integral **diverges to** \( \infty \) or to \(-\infty\) as is the case.

**Example 1.7.** For what values of \( p \in \mathbb{R} \), the improper integral \( \int_{1}^{\infty} \frac{dx}{x^p} \) converges? What is its value, when it converges?

**Case 1:** \( p = 1 \).

\[
\int_{1}^{b} \frac{dx}{x^p} = \int_{1}^{b} \frac{dx}{x} = \ln b - \ln 1 = \ln b.
\]

Since \( \lim_{b \to \infty} \ln b = \infty \), the improper integral diverges to \( \infty \).

**Case 2:** \( p < 1 \).

\[
\int_{1}^{b} \frac{dx}{x^p} = \frac{-x^{-p+1}}{-p+1} \bigg|_{1}^{b} = \frac{1}{1-p} (b^{1-p} - 1).
\]

Since \( \lim_{b \to \infty} b^{1-p} = \infty \), the improper integral diverges to \( \infty \).

**Case 3:** \( p > 1 \).

\[
\int_{1}^{b} \frac{dx}{x^p} = \frac{1}{1-p} (b^{1-p} - 1) = \frac{1}{1-p} \left( \frac{1}{b^{p-1}} - 1 \right).
\]

Since \( \lim_{b \to \infty} \frac{1}{b^{p-1}} = 0 \), we have

\[
\int_{1}^{\infty} \frac{dx}{x^p} = \lim_{b \to \infty} \int_{1}^{b} \frac{dx}{x^p} = \lim_{b \to \infty} \frac{1}{1-p} \left( \frac{1}{b^{p-1}} - 1 \right) = \frac{1}{p-1}.
\]

Hence, the improper integral \( \int_{1}^{\infty} \frac{dx}{x^p} \) converges to \( \frac{1}{p-1} \) for \( p > 1 \) and diverges to \( \infty \) for \( p \leq 1 \).

**Example 1.8.** For what values of \( p \in \mathbb{R} \), the improper integral \( \int_{0}^{1} \frac{dx}{x^p} \) converges?

**Case 1:** \( p = 1 \).

\[
\int_{0}^{1} \frac{dx}{x^p} = \lim_{a \to 0+} \int_{a}^{1} \frac{dx}{x} = \lim_{a \to 0+} [\ln 1 - \ln a] = \infty.
\]

Therefore, the improper integral diverges to \( \infty \).

**Case 2:** \( p < 1 \).

\[
\int_{0}^{1} \frac{dx}{x^p} = \lim_{a \to 0+} \int_{a}^{1} \frac{dx}{x^p} = \lim_{a \to 0+} \frac{1 - a^{1-p}}{1-p} = \frac{1}{1-p}.
\]

Therefore, the improper integral converges to \( 1/(1-p) \).
Case 3: $p > 1$.

$$\int_0^1 \frac{dx}{x^p} = \lim_{a \to 0^+} \frac{1 - a^{1-p}}{1 - p} = \lim_{a \to 0^+} \frac{1}{p - 1} \left( \frac{1}{a^{p-1}} - 1 \right) = \infty.$$ 

Hence the improper integral diverges to $\infty$.

The improper integral $\int_0^1 \frac{dx}{x^p}$ converges to $\frac{1}{1 - p}$ for $p < 1$ and diverges to $\infty$ for $p \geq 1$.

Exercises for § 1.6

1. Evaluate the following improper integrals:
   
   (a) $\int_2^\infty \frac{2}{x^2 - x} \, dx$
   
   (b) $\int_0^1 \frac{x + 1}{\sqrt{x^2 + 2x}} \, dx$
   
   (c) $\int_0^\infty \frac{dx}{(1 + x^2)(1 + \tan^{-1} x)}$
   
   (d) $\int_0^\infty 2e^{-x} \sin x \, dx$
   
   (e) $\int_{-1}^4 \frac{dx}{\sqrt{|x|}}$
   
   (f) $\int_0^2 \frac{dx}{\sqrt{|x - 1|}}$

1.7 Convergence tests for improper integrals

Sometimes it is helpful to be sure that an improper integral converges, even if we are unable to evaluate it.

**Theorem 1.11. (Comparison Test)** Let $f(x)$ and $g(x)$ be continuous functions on $[a, \infty)$. Suppose that $0 \leq f(x) \leq g(x)$ for all $x \geq a$.

1. If $\int_a^\infty g(x) \, dx$ converges, then $\int_a^\infty f(x) \, dx$ converges.

2. If $\int_a^\infty f(x) \, dx$ diverges to $\infty$, then $\int_a^\infty g(x) \, dx$ diverges to $\infty$.

**Proof:** Since $0 \leq f(x) \leq g(x)$ for all $x \geq a$,

$$\int_a^b f(x) \, dx \leq \int_a^b g(x) \, dx.$$ 

As $\lim_{b \to \infty} \int_a^b g(x) \, dx = \ell$ for some $\ell \in \mathbb{R}$, $\lim_{b \to \infty} \int_a^b f(x) \, dx$ exists and the limit is less than or equal to $\ell$. This proves (1). Proof of (2) is similar to that of (1). \qed

We also use a similar result stated below, without proof.

**Theorem 1.12. (Limit Comparison Test)** Let $f(x)$ and $g(x)$ be positive continuous functions on $[a, \infty)$. If $\lim_{x \to \infty} \frac{f(x)}{g(x)} = L$, where $0 < L < \infty$, then $\int_a^\infty f(x) \, dx$ and $\int_a^\infty g(x) \, dx$ either both converge, or both diverge.

Theorems 1.11 and 1.12 talk about non-negative functions. The reason is the following result, which we will not prove:

**Theorem 1.13.** Let $f(x)$ be a continuous function on $[a, b)$, for $b \in \mathbb{R}$ or $b = \infty$. If the improper integral $\int_a^b |f(x)| \, dx$ converges, then the improper integral $\int_a^b f(x) \, dx$ also converges.
Example 1.9.

(a) \( \int_1^\infty \frac{\sin^2 x}{x^2} \, dx \) converges because \( \frac{\sin^2 x}{x^2} \leq \frac{1}{x^2} \) for all \( x \geq 1 \) and \( \int_1^\infty \frac{dx}{x^2} \) converges.

(b) \( \int_2^\infty \frac{dx}{\sqrt{x^2 - 1}} \) diverges to \( \infty \) because \( \frac{1}{\sqrt{x^2 - 1}} \geq \frac{1}{x} \) for all \( x \geq 2 \) and \( \int_2^\infty \frac{dx}{x} \) diverges to \( \infty \).

(c) \( \int_1^\infty \frac{dx}{1 + x^2} \) converges or diverges?

Since \( \lim_{x \to \infty} \left[ \frac{1}{1 + x^2} / \frac{1}{x^2} \right] = \lim_{x \to \infty} \frac{x^2}{1 + x^2} = 1 \), the limit comparison test says that the given improper integral and \( \int_1^\infty \frac{dx}{x^2} \) both converge or diverge together. The latter converges, so does the former. However, they may converge to different values.

\[
\int_1^\infty \frac{dx}{1 + x^2} = \lim_{b \to \infty} [\tan^{-1} b - \tan^{-1} 1] = \frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}.
\]

\[
\int_1^\infty \frac{dx}{x^2} = \lim_{b \to \infty} \left( \frac{-1}{b} - \frac{-1}{1} \right) = 1.
\]

(d) Does the improper integral \( \int_1^\infty \frac{10^{10} \, dx}{e^x + 1} \) converge?

\[
\lim_{x \to \infty} \frac{10^{10}}{e^x + 1} / \frac{1}{e^x} = \lim_{x \to \infty} \frac{10^{10} e^x}{e^x + 1} = 10^{10}.
\]

Also, \( e \geq 2 \) implies that for all \( x \geq 1 \), \( e^x \geq x^2 \). So, \( e^{-x} \leq x^{-2} \). Since \( \int_1^\infty \frac{dx}{x^2} \) converges, \( \int_1^\infty \frac{dx}{e^x} \) also converges. By limit comparison test, the given improper integral converges.

Example 1.10. Show that \( \Gamma(x) = \int_0^\infty e^{-t} t^{x-1} \, dt \) converges for each \( x > 0 \).

Fix \( x > 0 \). Since \( \lim_{t \to \infty} e^{-t} t^{x+1} = 0 \), there exists \( t_0 \geq 1 \) such that \( 0 < e^{-t} t^{x+1} < 1 \) for \( t > t_0 \). That is,

\[
0 < e^{-t} t^{x-1} < t^{-2} \quad \text{for} \quad t > t_0.
\]

Since \( \int_1^\infty t^{-2} \, dt \) is convergent, \( \int_{t_0}^\infty t^{-2} \, dt \) is also convergent. By the comparison test,

\[
\int_{t_0}^\infty e^{-t} t^{x-1} \, dt \quad \text{is convergent}.
\]

The integral \( \int_1^{t_0} e^{-t} t^{x-1} \, dt \) exists and is not an improper integral.

Next, we consider the improper integral \( \int_0^1 e^{-t} t^{x-1} \, dt \). Let \( 0 < a < 1 \).
For $a \leq t \leq 1$, we have $0 < e^{-t}t^{x-1} < t^{x-1}$. So,

$$
\int_a^1 e^{-t}t^{x-1} \, dt < \int_a^1 t^{x-1} \, dt = \frac{1 - a^x}{x} < \frac{1}{x}.
$$

Taking the limit as $a \to 0^+$, we see that the

$$
\int_0^1 e^{-t}t^{x-1} \, dt \text{ is convergent,}
$$

and its value is less than or equal to $1/x$. Therefore,

$$
\int_0^\infty e^{-t}t^{x-1} \, dt = \int_0^1 e^{-t}t^{x-1} \, dt + \int_1^\infty e^{-t}t^{x-1} \, dt
$$

is convergent.

The function $\Gamma(x)$ is defined on $(0, \infty)$. For $x > 0$, using integration by parts,

$$
\Gamma(x + 1) = \int_0^\infty t^{x-1}e^{-t} \, dt = \left[ -t^x e^{-t} \right]_0^\infty - \int_0^\infty xt^{x-1}(-e^{-t}) \, dt = x\Gamma(x).
$$

It thus follows that $\Gamma(n + 1) = n!$ for any non-negative integer $n$. We take $0! = 1$.

**Example 1.11.** Test the convergence of $\int_{-\infty}^\infty e^{-t^2} \, dt$.

Since $e^{-t^2}$ is continuous on $[-1, 1]$, $\int_{-1}^1 e^{-t^2} \, dt$ exists.

For $t > 1$, we have $t < t^2$. So, $0 < e^{-t^2} < e^{-t}$. Since $\int_1^\infty e^{-t} \, dt$ is convergent, by the comparison test, $\int_1^\infty e^{-t^2} \, dt$ is convergent.

Now, $\int_{-a}^{-1} e^{-t^2} \, dt = \int_a^1 e^{-t^2} \, dt(-t) = \int_a^1 e^{-t^2} \, dt$. Taking limit as $a \to \infty$, we see that $\int_{-\infty}^1 e^{-t^2} \, dt$ is convergent and its value is equal to $\int_1^\infty e^{-t^2} \, dt$.

Combining the three integrals above, we conclude that $\int_{-\infty}^\infty e^{-t^2} \, dt$ converges.

The Gamma function takes other forms by substitution of the variable of integration. Substituting $t$ by $rt$ we have

$$
\Gamma(x) = r^x \int_0^\infty e^{-rt}t^{x-1} \, dt \quad \text{for } 0 < r, 0 < x.
$$

Substituting $t$ by $t^2$, we have

$$
\Gamma(x) = 2 \int_0^\infty e^{-t^2}t^{2x-1} \, dt \quad \text{for } 0 < x.
$$

Using multiple integrals it can be shown that $\Gamma(1/2) = 2 \int_0^\infty e^{-t^2} \, dt = \sqrt{\pi}$.

**Example 1.12.** Show that $\Gamma(1/2) = 2 \int_0^\infty e^{-t^2} \, dt = \sqrt{\pi}$.

$$
\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-x}x^{-1/2} \, dx = 2 \int_0^\infty e^{-t^2} \, dt \quad (x = t^2)
$$
To evaluate this integral, consider the double integral of $e^{-x^2-y^2}$ over two circular sectors $D_1$ and $D_2$, and the square $S$ as indicated below.

Since the integrand is positive, we have $\int\int_{D_1} < \int\int_S < \int\int_{D_2}$.

Now, evaluate these integrals by converting them to iterated integrals as follows:

$$
\int e^{-r^2} r \, dr \int_0^{\pi/2} d\theta < \int_0^R e^{-x^2} \, dx \int_0^R e^{-y^2} \, dy < \int_0^{\sqrt{2} R} e^{-r^2} r \, dr \int_0^{\pi/2} d\theta
$$

$$
\frac{\pi}{4} (1 - e^{-R^2}) < \left( \int_0^R e^{-x^2} \, dx \right)^2 < \frac{\pi}{4} (1 - e^{-2R^2})
$$

Take the limit as $R \to \infty$ to obtain

$$
\left( \int_0^\infty e^{-x^2} \, dx \right)^2 = \frac{\pi}{4}
$$

From this, the result follows.

**Example 1.13.** Prove: $B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} \, dt$ converges for $x > 0, y > 0$.

We write the integral as a sum of two integrals:

$$
B(x, y) = \int_0^{1/2} t^{x-1} (1-t)^{y-1} \, dt + \int_{1/2}^1 t^{x-1} (1-t)^{y-1} \, dt
$$

Setting $u = 1 - t$, the second integral looks like

$$
\int_{1/2}^1 t^{x-1} (1-t)^{y-1} \, dt = \int_0^{1/2} u^{y-1} (1-u)^{x-1} \, dt
$$

Therefore, it is enough to show that the first integral converges. Notice that here, $0 < t \leq 1/2$.

**Case 1:** $x \geq 1$.

For $0 < t < 1/2$, $1-t > 0$. Therefore, for all $y > 0$, the function $(1-t)^{y-1}$ is well defined, continuous, and bounded on $(0, 1/2]$. So is the function $t^{x-1}$. Therefore, the integral $\int_0^{1/2} t^{x-1} (1-t)^{y-1} \, dt$ exists and is not an improper integral.

**Case 2:** $0 < x < 1$.

Here, the function $t^{x-1}$ is well defined and continuous on $(0, 1/2]$. By Example 1.8, the integral $\int_0^{1/2} t^{x-1} \, dt$ converges. Since $t^{x-1}(1-t)^{y-1} \leq t^{x-1}$ for $0 < t \leq 1/2$, we conclude that
\[
\int_0^{1/2} t^{x-1} (1 - t)^{y-1} \, dt \text{ converges.}
\]
By setting \( t = 1 - t \), we see that \( B(x, y) = B(y, x) \).

By substituting \( t \) with \( \sin^2 t \), the Beta function can be written as

\[
B(x, y) = 2 \int_0^{\pi/2} (\sin t)^{2x-1} (\cos t)^{2y-1} \, dt, \quad \text{for } x > 0, y > 0.
\]

Changing the variable \( t \) to \( t/(1 + t) \), the Beta function can be written as

\[
B(x, y) = \int_0^\infty \frac{t^{x+1}}{(1 + t)^{x+y}} \, dt \quad \text{for } x > 0, y > 0.
\]

Again, using multiple integrals it can be shown that

\[
B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \quad \text{for } x > 0, y > 0.
\]

**Exercises for § 1.7**

1. Test for convergence, the following improper integrals:
   
   \( a \) \( \int_{-\pi/2}^{\pi/2} \frac{\cos x}{(\pi - 2x)^{1/3}} \, dx \)  \( b \) \( \int_0^1 \frac{dx}{x - \sin x} \)  \( c \) \( \int_{-1}^1 \ln |x| \, dx \)
   
   \( d \) \( \int_0^\infty \frac{dx}{\sqrt{x-1}} \)  \( e \) \( \int_1^\infty \frac{dx}{\sqrt{e^x - x}} \)  \( f \) \( \int_{-\infty}^\infty \frac{dx}{e^x + e^{-x}} \)
   
   \( g \) \( \int_1^2 \frac{dx}{x(\ln x)^p} \)  \( h \) \( \int_2^\infty \frac{dx}{x(\ln x)^p} \)  Hint: \( p < 1, p = 1, p > 1 \). Put \( x = e^t \).

**1.8 Tests of convergence for series**

**Theorem 1.14. (Integral Test)** Let \( \sum a_n \) be a series of positive terms. Let \( f : [1, \infty) \to \mathbb{R} \) be a continuous, positive and non-increasing function such that \( a_n = f(n) \) for each \( n \in \mathbb{N} \).

1. If \( \int_1^\infty f(t) \, dt \) is convergent, then \( \sum a_n \) is convergent.
2. If \( \int_1^\infty f(t) \, dt \) diverges to \( \infty \), then \( \sum a_n \) diverges to \( \infty \).

**Proof:** Since \( f \) is a positive and non-increasing, the integrals and the partial sums have a certain relation.
\[
\int_1^{n+1} f(t) \, dt \leq a_1 + a_2 + \cdots + a_n \leq a_1 + \int_1^n f(t) \, dt.
\]

If \( \int_1^n f(t) \, dt \) is finite, then the right hand inequality shows that \( \sum a_n \) is convergent.
If \( \int_1^n f(t) \, dt = \infty \), then the left hand inequality shows that \( \sum a_n \) diverges to \( \infty \). \qed

Notice that when the series converges, the value of the integral can be different from the sum of the series. Moreover, Integral test assumes implicitly that \( \{a_n\} \) is a monotonically decreasing sequence. Further, the integral test is also applicable when the interval of integration is \([m, \infty)\) instead of \([1, \infty)\).

**Example 1.14.** Show that \( \sum_{n=1}^{\infty} \frac{1}{n^p} \) converges for \( p > 1 \) and diverges for \( p \leq 1 \).

For \( p = 1 \), the series is the harmonic series; and it diverges. Suppose \( p \neq 1 \). Consider the function \( f(t) = \frac{1}{t^p} \) from \([1, \infty)\) to \( \mathbb{R} \). This is a continuous, positive and decreasing function.

\[
\int_1^\infty \frac{1}{t^p} \, dt = \lim_{b \to \infty} \frac{t^{-p+1}}{-p+1}\bigg|_1^b = \frac{1}{1-p} \lim_{b \to \infty} \left(\frac{1}{b^{p-1}} - 1\right) = \begin{cases} 
\frac{1}{p-1} & \text{if } p > 1 \\
\infty & \text{if } p < 1.
\end{cases}
\]

Then the Integral test proves the statement. We note that for \( p > 1 \), the sum of the series \( \sum n^{-p} \) need not be equal to \((p-1)^{-1}\).

There are simple tests which are applicable to series of positive terms, whether the terms are decreasing or not. We discuss those next.

**Theorem 1.15. (D’ Alembert Ratio Test)**

Let \( \sum a_n \) be a series of positive terms. Suppose \( \lim_{n \to \infty} \frac{a_{n+1}}{a_n} = \ell \).

1. If \( \ell < 1 \), then \( \sum a_n \) converges.
2. If \( \ell > 1 \) or \( \ell = \infty \), then \( \sum a_n \) diverges to \( \infty \).
3. If \( \ell = 1 \), then no conclusion is obtained.

**Proof:** (1) Given that \( \lim(a_{n+1}/a_n) = \ell < 1 \). Choose \( \delta \) such that \( \ell < \delta < 1 \). There exists \( m \in \mathbb{N} \) such that for each \( n > m \), \( a_{n+1}/a_n < \delta \). Then

\[
\frac{a_n}{a_{m+1}} = \frac{a_n}{a_{n-1}} \cdot \frac{a_{n-1}}{a_{n-2}} \cdots \frac{a_{m+2}}{a_{m+1}} < \delta^{n-m}.
\]

Thus, \( a_n < \delta^{n-m} a_{m+1} \). Consequently,

\[
a_{m+1} + a_{m+2} + \cdots + a_n < a_{m+1}(1 + \delta + \delta^2 + \cdots \delta^{n-m}).
\]

Since \( \delta < 1 \), this approaches a limit as \( n \to \infty \). Therefore, the series

\[
a_{m+1} + a_{m+2} + \cdots a_n + \cdots
\]
converges. In that case, the series \( \sum a_n = (a_1 + \cdots + a_m) + a_{m+1} + a_{m+2} + \cdots \) converges.

(2) Given that \( \lim(a_{n+1}/a_n) = \ell > 1 \). Then there exists \( m \in \mathbb{N} \) such that for each \( n > m \), \( a_{n+1} > a_n \). Then

\[
    a_{m+1} + a_{m+2} + \cdots + a_n > a_{m+1}(n - m).
\]

Since \( a_{m+1} > 0 \), this approaches \( \infty \) as \( n \to \infty \). Therefore, the series

\[
    a_{m+1} + a_{m+2} + \cdots
\]

diverges to \( \infty \). In that case, the series \( \sum a_n = (a_1 + \cdots + a_m) + a_{m+1} + a_{m+2} + \cdots \) diverges to \( \infty \). The other case of \( \ell = \infty \) is similar.

(3) For the series \( \sum (1/n) \), \( \lim(a_{n+1}/a_n) = \lim(n/(n+1)) = 1 \). This series, as we know is divergent to \( \infty \).

But the series \( \sum (1/n^2) \) is convergent although \( \lim(a_{n+1}/a_n) = 1 \).

\( \square \)

**Example 1.15.** Does the series \( \sum_{n=1}^{\infty} \frac{n!}{n^n} \) converge?

Write \( a_n = n!/(n^n) \). Then

\[
    \frac{a_{n+1}}{a_n} = \frac{(n+1)!n^n}{(n+1)n!} = \left( \frac{n}{n+1} \right)^n \to \frac{1}{e} < 1 \text{ as } n \to \infty.
\]

By D’Alembert’s ratio test, the series converges.

Then it follows that the sequence \( \{\frac{n!}{n^n}\} \) converges to 0.

**Theorem 1.16. (Cauchy Root Test)**

*Let \( \sum a_n \) be a series of positive terms. Suppose \( \lim_{n \to \infty} (a_n)^{1/n} = \ell. \)*

1. If \( \ell < 1 \), then \( \sum a_n \) converges.

2. If \( \ell > 1 \) or \( \ell = \infty \), then \( \sum a_n \) diverges to \( \infty \).

3. If \( \ell = 1 \), then no conclusion is obtained.

**Proof:**

(1) Suppose \( \ell < 1 \). Choose \( \delta \) such that \( \ell < \delta < 1 \). Due to the limit condition, there exists an \( m \in \mathbb{N} \) such that for each \( n > m \), \( (a_n)^{1/n} < \delta \). That is, \( a_n < \delta^n \). Since \( 0 < \delta < 1 \), \( \sum \delta^n \) converges. By Comparison test, \( \sum a_n \) converges.

(2) Given that \( \ell > 1 \) or \( \ell = \infty \), we see that \( (a_n)^{1/n} > 1 \) for infinitely many values of \( n \). That is, the sequence \( \{(a_n)^{1/n}\} \) does not converge to 0. Therefore, \( \sum a_n \) is divergent. It diverges to \( \infty \) since it is a series of positive terms.

(3) Once again, for both the series \( \sum (1/n) \) and \( \sum (1/n^2) \), we see that \( (a_n)^{1/n} \) has the limit 1. But one is divergent, the other is convergent. \( \square \)

**Remark:** In fact, for a sequence \( \{a_n\} \) of positive terms if \( \lim_{n \to \infty} \frac{a_{n+1}}{a_n} \) exists, then \( \lim_{n \to \infty} (a_n)^{1/n} \) exists and the two limits are equal.
To see this, suppose $\lim_{n \to \infty} \frac{a_{n+1}}{a_n} = \ell$. Let $\epsilon > 0$. Then we have an $m \in \mathbb{N}$ such that for all $n > m$, $\ell - \epsilon < \frac{a_{n+1}}{a_n} < \ell + \epsilon$. Use the right side inequality first. For all such $n$, $a_n < (\ell + \epsilon)^{n-m}a_m$. Then
\[
(a_n)^{1/n} < (\ell + \epsilon)((\ell + \epsilon)^{-m}a_m)^{1/n} \to \ell + \epsilon \text{ as } n \to \infty.
\]
Therefore, $\lim(a_n)^{1/n} \leq \ell + \epsilon$ for every $\epsilon > 0$. That is, $\lim(a_n)^{1/n} \leq \ell$.

Notice that this gives an alternative proof of Theorem 1.16.

Example 1.16. Does the series $\sum_{n=0}^{\infty} 2^{(-1)^n-n} = 2 + \frac{1}{4} + \frac{1}{2} + \frac{1}{16} + \cdots$ converge?

Let $a_n = 2^{(-1)^n-n}$. Then
\[
\frac{a_{n+1}}{a_n} = \begin{cases} 
1/8 & \text{if } n \text{ even} \\
2 & \text{if } n \text{ odd}.
\end{cases}
\]
Clearly, its limit does not exist. But
\[
(a_n)^{1/n} = \begin{cases} 
2^{1/n-1} & \text{if } n \text{ even} \\
2^{-1/n-1} & \text{if } n \text{ odd}
\end{cases}
\]
This has limit $1/2 < 1$. Therefore, by Cauchy root test, the series converges.

Exercises for § 1.8

1. Test for convergence the following series:
   (a) $\sum_{n=1}^{\infty} \frac{n}{10^{100}n^2 + 2^{100}}$
   (b) $\sum_{n=1}^{\infty} \frac{5}{n + 1}$
   (c) $\sum_{n=2}^{\infty} \frac{\ln n}{\sqrt{n}}$
   (d) $\sum_{n=2}^{\infty} \frac{\sqrt{n}}{\ln n}$
   (e) $\sum_{n=1}^{\infty} \frac{1}{n + \sqrt{n}}$
   (f) $\sum_{n=3}^{\infty} \frac{1/n}{(\ln n) \sqrt{(\ln n)^2 - 1}}$
   (g) $\sum_{n=1}^{\infty} \frac{(\ln n)^n}{n^2}$
   (h) $\sum_{n=1}^{\infty} \frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots 2n(3^n + 1)}$
   (i) $\sum_{n=1}^{\infty} n!e^{-n}$

2. Show that neither the ratio test nor the root test determine convergence of the series $\sum_{n=2}^{\infty} \frac{1}{(\ln n)^p}$.

1.9 Alternating series

If the terms of a series have alternating signs, then these tests are not applicable. For example, the methods discussed so far fail on deciding whether the series $\sum (-1)^n/n$ converges or not.

Theorem 1.17. (Leibniz Alternating Series Test)
Let $\{a_n\}$ be a sequence of positive terms decreasing to 0; that is, for each $n$, $a_n \geq a_{n+1} > 0$, and $\lim_{n \to \infty} a_n = 0$. Then the series $\sum_{n=1}^{\infty} (-1)^n a_n$ converges, and its sum lies between $a_1 - a_2$ and $a_1$.
Proof: The partial sum up to $2n$ terms is

$$s_{2n} = (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2n-1} - a_{2n}) = a_1 - [(a_2 - a_3) + \cdots + (a_{2n-2} - a_{2n-1})] - a_{2n}.$$  

It is a sum of $n$ positive terms bounded above by $a_1$ and below by $a_1 - a_2$. Hence $s_{2n}$ converges to some $s$ such that $a_1 - a_2 \leq s \leq a_1$.

The partial sum up to $2n+1$ terms is $s_{2n+1} = s_{2n} + a_{2n+1}$. It converges to $s$ as $\lim a_{2n+1} = 0$. Hence the series converges to some $s$ with $a_1 - a_2 \leq s \leq a_1$. \[\square\]

The bounds for $s$ can be sharpened by taking $s_{2n} \leq s \leq s_{2n-1}$ for each $n > 1$.

Leibniz test now implies that the series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \cdots$ is convergent to some $s$ with $1/2 \leq s \leq 1$. By taking more terms, we can have different bounds such as

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} = \frac{7}{12} \leq s \leq 1 - \frac{1}{2} + \frac{1}{3} = \frac{10}{12}$$

In contrast, the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots$ diverges to $\infty$.

We say that the series $\sum a_n$ is absolutely convergent iff the series $\sum |a_n|$ is convergent.

An alternating series $\sum a_n$ is said to be conditionally convergent iff it is convergent but it is not absolutely convergent.

Thus for a series of non-negative terms, convergence and absolute convergence coincide. As we just saw, an alternating series may be convergent but not absolutely convergent.

The series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \cdots$ is a conditionally convergent series. It shows that the converse of the following theorem is not true.

**Theorem 1.18.** An absolutely convergent series is convergent.

Proof: Let $\sum a_n$ be an absolutely convergent series. Then $\sum |a_n|$ is convergent. Let $\epsilon > 0$. By Cauchy criterion, there exists an $n_0 \in \mathbb{N}$ such that for all $n > m > n_0$, we have

$$|a_m| + |a_{m+1}| + \cdots + |a_n| < \epsilon.$$  

Now,

$$|a_m + a_{m+1} + \cdots + a_n| \leq |a_m| + |a_{m+1}| + \cdots + |a_n| < \epsilon.$$  

Again, by Cauchy criterion, the series $\sum a_n$ is convergent. \[\square\]

An absolutely convergent series can be rearranged in any way we like, but the sum remains the same. Whereas a rearrangement of the terms of a conditionally convergent series may lead to divergence or convergence to any other number. In fact, a conditionally convergent series can always be rearranged in a way so that the rearranged series converges to any desired number; we will not prove this fact.
Example 1.17. Do the series (a) \( \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{2^n} \) and (b) \( \sum_{n=1}^{\infty} \frac{\cos n}{n^2} \) converge?

(a) \( \sum_{n=1}^{\infty} \frac{1}{2^n} \) converges. Therefore, the given series converges absolutely; hence it converges.

(b) \( \left| \frac{\cos n}{n^2} \right| \leq \frac{1}{n^2} \); and \( \sum (n^{-2}) \) converges. By comparison test, the given series converges absolutely; and hence it converges.

Example 1.18. Discuss the convergence of the series \( \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^p} \).

For \( p > 1 \), the series \( \sum n^{-p} \) converges. Therefore, the given series converges absolutely for \( p > 1 \).

For \( 0 < p \leq 1 \), by Leibniz test, the series converges. But \( \sum n^{-p} \) does not converge. Therefore, the given series converges conditionally for \( 0 < p \leq 1 \).

For \( p \leq 0 \), \( \lim \frac{(-1)^{n+1}}{n^p} \neq 0 \). Therefore, the given series diverges in this case.

Exercises for § 1.9

1. Which of the following series converge absolutely, converge conditionally, and diverge?

(a) \( \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^{3/2}} \)  
(b) \( \sum_{n=1}^{\infty} \frac{(-1)^{n+1}n^{10}}{10^n} \)  
(c) \( \sum_{n=1}^{\infty} \frac{(-1)^{n+1}\ln n}{n} \)

(d) \( \sum_{n=1}^{\infty} \frac{(-1)^n}{1 + \sqrt{n}} \)  
(e) \( \sum_{n=2}^{\infty} \frac{(-1)^n}{\ln(n^3)} \)  
(f) \( \sum_{n=1}^{\infty} (-2/3)^n n^2 \)

(g) \( \sum_{n=1}^{\infty} \frac{(-1)^n\tan^{-1} n}{1 + n^2} \)  
(h) \( \sum_{n=1}^{\infty} \frac{(-1)^n}{n \ln n} \)  
(i) \( \sum_{n=1}^{\infty} (-1)^n(\sqrt{n+1} - \sqrt{n}) \)
Chapter 2

Series Representation of Functions

2.1 Power series

A power series apparently is a generalization of a polynomial. A polynomial in $x$ looks like

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n.$$ 

A power series is an infinite sum of the same form. The question is though a polynomial defines a function when $x \in \mathbb{R}$, when does a power series define a function? That is, for what values of $x$, a power series sums to a number?

Let $a \in \mathbb{R}$. A **power series about** $x = a$ is a series of the form

$$\sum_{n=0}^{\infty} c_n(x - a)^n = c_0 + c_1(x - a) + c_2(x - a)^2 + \cdots$$

The point $a$ is called the **center** of the power series and the real numbers $c_0, c_1, \cdots, c_n, \cdots$ are its **co-efficients**.

If the power series converges to $f(x)$ for all $x \in D$, for some subset $D$ of $\mathbb{R}$, then we say that the power series **sums to** the function $f(x)$, whose domain is $D$.

In such a case, we also say that the power series **represents** the function $f(x)$.

For example, the geometric series

$$1 + x + x^2 + \cdots + x^n + \cdots$$

is a power series about $x = 0$ with each co-efficient as $1$. We know that its sum is $\frac{1}{1 - x}$ for $-1 < x < 1$. And we know that for $|x| \geq 1$, the geometric series does not converge. That is, the series defines a function from $(-1, 1)$ to $\mathbb{R}$ and it is not meaningful for other values of $x$.

**Example 2.1.** Show that the following power series converges for $0 < x < 4$.

$$1 - \frac{1}{2}(x - 2) + \frac{1}{4}(x - 2)^2 + \cdots + \frac{(-1)^n}{2^n}(x - 2)^n + \cdots$$
It is a geometric series with the ratio as \( r = \left( -\frac{1}{2} \right) (x - 2) \). Thus it converges for \( \left| \left( -\frac{1}{2} \right) (x - 2) \right| < 1 \). Simplifying we get the constraint as \( 0 < x < 4 \).

Notice that the power series sums to
\[
\frac{1}{1 - r} = \frac{1}{1 - \left( -\frac{1}{2} \right) (x - 2)} = \frac{2}{x}.
\]

Thus, the power series gives a series expansion of the function \( \frac{2}{x} \) for \( 0 < x < 4 \).

Truncating the series to \( n \) terms give us polynomial approximations of the function \( \frac{2}{x} \).

A fundamental result for the power series is the following. It roughly says that if for some \( c > 0 \), a power series converges with \( x = c \), then it converges for all \( x \) with \( 0 \leq x \leq c \). A similar result holds for the divergence of a power series. For this purpose, we consider power series about \( x = 0 \).

Results on power series about any point \( a \) can be obtained from this particular case in a similar manner.

**Theorem 2.1. (Convergence Theorem for Power Series)** Suppose that the power series \( \sum_{n=0}^{\infty} a_n x^n \) is convergent for \( x = c \) and divergent for \( x = d \) for some \( c > 0 \), \( d > 0 \). Then the power series converges absolutely for all \( x \) with \( \left| x \right| < c \); and it diverges for all \( x \) with \( \left| x \right| > d \).

**Proof:** The power series converges for \( x = c \) means that \( \sum a_n c^n \) converges. Thus \( \lim_{n \to \infty} a_n c^n = 0 \). Then we have an \( M \in \mathbb{N} \) such that for all \( n > M \), \( \left| a_n c^n \right| < 1 \).

Let \( x \in \mathbb{R} \) be such that \( \left| x \right| < c \). Write \( t = \left| \frac{x}{c} \right| \). For each \( n > M \), we have
\[
\left| a_n \right| \left| x \right|^n = \left| a_n x^n \right| = \left| a_n c^n \right| \left| \frac{x}{c} \right|^n < \left| \frac{x}{c} \right|^n = t^n.
\]

As \( 0 \leq t < 1 \), the geometric series \( \sum_{n=M+1}^{\infty} t^n \) converges. By comparison test, for any \( x \) with \( \left| x \right| < c \), the series \( \sum_{n=M+1}^{\infty} \left| a_n x^n \right| \) converges. However, \( \sum_{n=0}^{M} \left| a_n x^n \right| \) is finite. Therefore, the power series \( \sum_{n=0}^{\infty} a_n x^n \) converges absolutely for all \( x \) with \( \left| x \right| < c \).

For the divergence part of the theorem, suppose, on the contrary that the power series converges for some \( \alpha > d \). By the convergence part, the series must converge for \( x = d \), a contradiction. □

Notice that if the power series is about a point \( x = a \), then we take \( t = x - a \) and apply Theorem 2.1. Also, for \( x = 0 \), the power series \( \sum a_n x^n \) always converges.

In view of Theorem 2.1, the following definition makes sense.

Consider the power series \( \sum_{n=0}^{\infty} a_n (x - a)^n \). The real number
\[
R = \text{lub} \{c \geq 0 : \text{the power series converges for all } x \text{ with } |x - a| < c\}
\]
is called the radius of convergence of the power series.

That is, \( R \) is such non-negative number that the power series converges for all \( x \) with \(|x - a| < R\) and it diverges for all \( x \) with \(|x - a| > R\).

If the radius of convergence of the power series \( \sum a_n(x-a)^n \) is \( R \), then the interval of convergence of the power series is

\[
(a - R, a + R) \text{ if it diverges at both } x = a - R \text{ and } x = a + R.
\]

\[
[a - R, a + R] \text{ if it converges at } x = a - R \text{ and diverges at } x = a + R.
\]

\[
(a - R, a + R] \text{ if it diverges at } x = a - R \text{ and converges at } x = a + R.
\]

That is, the interval of convergence of the power series is the open interval \((a - R, a + R)\) along with the point(s) \(a - R\) and/or \(a + R\), wherever it is convergent. Theorem 2.1 guarantees that the power series converges everywhere inside the interval of convergence, it converges absolutely inside the open interval \((a - R, a + R)\), and it diverges everywhere beyond the interval of convergence.

Also, see that when \( R = \infty \), the power series converges for all \( x \in \mathbb{R} \), and when \( R = 0 \), the power series converges only at the point \( x = a \), whence its sum is \( a_0 \).

To determine the interval of convergence, you must find the radius of convergence \( R \), and then test for its convergence separately for the end-points \( x = a - R \) and \( x = a + R \).

### 2.2 Determining radius of convergence

The radius of convergence can be found out by ratio test and/or root test, or any other test.

**Theorem 2.2.** The radius of convergence of the power series \( \sum_{n=0}^{\infty} a_n(x-a)^n \) is given by \( \lim_{n \to \infty} |a_n|^{1/n} \) provided that this limit is either a real number or equal to \( \infty \).

**Proof:** Let \( R \) be the radius of convergence of the power series \( \sum_{n=0}^{\infty} a_n(x-a)^n \). Let \( \lim_{n \to \infty} |a_n|^{1/n} = r \).

We consider three cases and show that

\[
(1) \ r > 0 \Rightarrow R = \frac{1}{r}, \quad (2) \ r = \infty \Rightarrow R = 0, \quad (3) \ r = 0 \Rightarrow R = \infty.
\]

(1) Let \( r > 0 \). By the root test, the series is absolutely convergent whenever

\[
\lim_{n \to \infty} |a_n(x-a)^n|^{1/n} < 1 \quad \text{i.e., } |x-a| \lim_{n \to \infty} |a_n|^{1/n} < 1 \quad \text{i.e., } |x-a| < \frac{1}{r}.
\]

It also follows from the root test that the series is divergent when \(|x-a| > 1/r\). Hence \( R = 1/r \).

(2) Let \( r = \infty \). Then for any \( x \neq a \), \( \lim |a_n(x-a)^n| = \lim |x-a||a_n|^{1/n} = \infty \). By the root test, \( \sum a_n(x-a)^n \) diverges for each \( x \neq a \). Thus, \( R = 0 \).

(3) Let \( r = 0 \). Then for any \( x \in \mathbb{R} \), \( \lim |a_n(x-a)^n|^{1/n} = |x-a| \lim |a_n|^{1/n} = 0 \). By the root test, the series converges for each \( x \in \mathbb{R} \). So, \( R = \infty \).

\[\square\]

Instead of the Root test, if we apply the Ratio test, then we obtain the following theorem.
Theorem 2.3. The radius of convergence of the power series \( \sum_{n=0}^{\infty} a_n (x - a)^n \) is given by \( \lim_{n \to \infty} \left| \frac{a_n}{a_{n+1}} \right| \), provided that this limit is either a real number or equal to \( \infty \).

Example 2.2. For what values of \( x \), do the following power series converge?

(a) \( \sum_{n=0}^{\infty} n! x^n \)  
(b) \( \sum_{n=0}^{\infty} \frac{x^n}{n!} \)  
(c) \( \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \)

(a) \( a_n = n! \). Thus \( \lim |a_n/a_{n+1}| = \lim 1/(n+1) = 0 \). Hence \( R = 0 \). That is, the series is only convergent for \( x = 0 \).

(b) \( a_n = 1/n! \). Thus \( \lim |a_n/a_{n+1}| = \lim (n+1) = \infty \). Hence \( R = \infty \). That is, the series is convergent for all \( x \in \mathbb{R} \).

(c) Here, the power series is not in the form \( \sum b_n x^n \). The series can be thought of as

\[
x(1 - \frac{x^2}{3} + \frac{x^4}{5} + \cdots) = x \sum_{n=0}^{\infty} (-1)^n \frac{t^n}{2n+1} \quad \text{for } t = x^2
\]

Now, for the power series \( \sum_{n=0}^{\infty} (-1)^n \frac{t^n}{2n+1} \), \( a_n = (-1)^n / (2n+1) \).

Thus \( \lim |a_n/a_{n+1}| = \lim \frac{2n+3}{2n+1} = 1 \). Hence \( R = 1 \). That is, for \( |t| = x^2 < 1 \), the series converges and for \( |t| = x^2 > 1 \), the series diverges.

Alternatively, you can use the geometric series. That is, for any \( x \in \mathbb{R} \), consider the series

\[
x(1 - \frac{x^2}{3} + \frac{x^4}{5} + \cdots).
\]

By the ratio test, the series converges if

\[
\lim_{n \to \infty} \left| \frac{u_n}{u_{n+1}} \right| = \lim_{n \to \infty} \frac{2n+3}{2n+1} |x^2| = x^2 < 1.
\]

That is, the power series converges for \(-1 < x < 1 \). Also, by the ratio test, the series diverges for \( |x| > 1 \).

What happens for \( |x| = 1 \)?

For \( x = -1 \), the original power series is an alternating series; it converges due to Liebniz. Similarly, for \( x = 1 \), the alternating series also converges.

Hence the interval of convergence for the original power series (in \( x \)) is \([-1, 1]\).

If \( R \) is the radius of convergence of a power series \( \sum a_n (x - a)^n \), then the series defines a function \( f(x) \) from the open interval \( (a - R, a + R) \) to \( \mathbb{R} \) by

\[
f(x) = a_0 + a_1 (x - a) + a_2 (x - a)^2 + \cdots = \sum_{n=0}^{\infty} a_n (x - a)^n \quad \text{for } x \in (a - R, a + R).
\]

This function can be differentiated and integrated term-by-term and it so happens that the new series obtained by such term-by-term differentiation or integration has the same radius of convergence and they define the derivative and the integral of \( f(x) \). We state it without proof.
Theorem 2.4. Let the power series $\sum_{n=0}^{\infty} a_n(x-a)^n$ have radius of convergence $R > 0$. Then the power series defines a function $f : (a-R, a+R) \to \mathbb{R}$. Further, $f'(x)$ and $\int f(x)dx$ exist as functions from $(a-R, a+R)$ to $\mathbb{R}$ and these are given by

$$f(x) = \sum_{n=0}^{\infty} a_n(x-a)^n, \quad f'(x) = \sum_{n=1}^{\infty} na_n(x-a)^{n-1}, \quad \int f(x)dx = \sum_{n=0}^{\infty} \frac{a_n(x-a)^{n+1}}{n+1} + C,$$

where all the three power series converge for all $x \in (a - R, a + R)$.

Caution: Term by term differentiation may not work for series, which are not power series.

For example, $\sum_{n=1}^{\infty} \frac{\sin(n!x)}{n^2}$ is convergent for all $x$. The series obtained by term-by-term differentiation is $\sum_{n=1}^{\infty} \frac{n! \cos(n!x)}{n^2}$; it diverges for all $x$.

Further, power series about the same point can be multiplied by using a generalization of multiplication of polynomials. We write the multiplication of power series about $x = 0$ for simplicity.

Theorem 2.5. Let the power series $\sum a_n x^n$ and $\sum b_n x^n$ have the same radius of convergence $R > 0$. Then their multiplication has the same radius of convergence $R$. Moreover, the functions they define satisfy the following:

If $f(x) = \sum a_n x^n$, $g(x) = \sum b_n x^n$ then $f(x)g(x) = \sum c_n x^n$ for $a - R < x < a + R$

where $c_n = \sum_{k=0}^{n} a_k b_{n-k} = a_0 b_n + a_1 b_{n-1} + \cdots + a_{n-1} b_1 + a_n b_0$.

Example 2.3. Determine power series expansions of the functions (a) $\frac{2}{(x - 1)^3}$ (b) $\tan^{-1} x$.

(a) For $-1 < x < 1$, $\frac{1}{1 - x} = 1 + x + x^2 + x^3 + \cdots$.

Differentiating term by term, we have

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots$$

Differentiating once more, we get

$$\frac{2}{(1-x)^3} = 2 + 6x + 12x^2 + \cdots = \sum_{n=2}^{\infty} n(n-1)x^{n-2} \quad \text{for} \quad -1 < x < 1.$$

(b) $\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + x^8 - \cdots$ for $|x^2| < 1$.

Integrating term by term we have

$$\tan^{-1} x + C = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots \quad \text{for} \quad -1 < x < 1.$$
Exercises for § 2.1-2.2

1. Find the radius of convergence, the interval of convergence of the series. Also, determine when the series converges conditionally and/or absolutely.

(a) \[ \sum_{n=1}^{\infty} \frac{(3x - 1)n}{n} \]
(b) \[ \sum_{n=1}^{\infty} \frac{x^n}{3^n n \sqrt{n}} \]
(c) \[ \sum_{n=1}^{\infty} \sqrt{n} (2x + 5)^n \]
(d) \[ \sum_{n=1}^{\infty} \frac{(x - \sqrt{2})^{2n+1}}{2n + 2} \]
(e) \[ \sum_{n=0}^{\infty} \left(\frac{x^2 - 1}{2}\right)^n \]. Also find the function it represents.

2. The series \( \tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \frac{62x^9}{2835} + \cdots \) converges for \( -\frac{\pi}{2} < x < \frac{\pi}{2} \).

(a) Find the first five terms of the series for \( \ln |\sec x| \). For what values of \( x \) should this series converge?
(b) Find the first five terms of the series for \( \sec^2 x \). For what values of \( x \) should this series converge?
(c) \( \sec x = 1 + \frac{x^2}{2} + \frac{5x^4}{24} + \frac{61x^6}{720} + \frac{277x^8}{8064} + \cdots \). Check your result in (b) by squaring this series for \( \sec x \).

2.3 Taylor’s formulas

For an elegant power series representation of smooth functions we require Taylor’s formulas. It has two forms: differential form and integral form. The differential form is a generalization of the Mean Value Theorem for differentiable functions, which you already know. We restate it.

**Theorem 2.6. (Taylor’s Formula in Differential Form)** Let \( n \in \mathbb{N} \). Suppose that \( f^{(n)}(x) \) is continuous on \([a, b]\) and is differentiable on \((a, b)\). Then there exists a point \( c \in (a, b) \) such that

\[
 f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.
\]

The polynomial

\[
p(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n
\]

in Taylor’s formula is called as **Taylor’s polynomial of order** \( n \). Notice that the degree of the Taylor’s polynomial may be less than or equal to \( n \). The expression given for \( f(x) \) there is called **Taylor’s formula** for \( f(x) \). Taylor’s polynomial is an approximation to \( f(x) \) with the error

\[
R_n(x) = \frac{f^{(n+1)}(c_{n+1})}{(n+1)!} (x-a)^{n+1}.
\]

How good \( f(x) \) is approximated by \( p(x) \) depends on the smallness of the error \( R_n(x) \).
For example, if we use \( p(x) \) of order 5 for approximating \( \sin x \) at \( x = 0 \), then we get
\[
\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + R_6(x), \quad \text{where} \quad R_6(x) = \frac{\sin \theta}{6!} x^6.
\]
Here, \( \theta \) lies between 0 and \( x \). The absolute error is bounded above by \( |x|^6/6! \). However, if we take the Taylor’s polynomial of order 6, then \( p(x) \) is the same as in the above, but the absolute error is now \( |x|^7/7! \). If \( x \) is near 0, this is smaller than the earlier bound.

Taylor’s theorem can also be written in terms of integrals.

**Theorem 2.7. (Taylor’s Formula in Integral Form)** Let \( f(x) \) be an \( (n + 1) \)-times continuously differentiable function on an open interval \( I \) containing \( a \). Let \( x \in I \). Then
\[
f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x),
\]
where \( R_n(x) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t) \, dt \). An estimate for \( R_n(x) \) is given by
\[
\frac{m x^{n+1}}{(n+1)!} \leq R_n(x) \leq \frac{M x^{n+1}}{(n+1)!}
\]
where \( m \leq f^{(n+1)}(x) \leq M \) for \( x \in I \).

**Proof:** We prove it by induction on \( n \). For \( n = 0 \), we should show that
\[
f(x) = f(a) + R_0(x) = f(a) + \int_a^x f'(t) \, dt.
\]
But this follows from the Fundamental theorem of calculus. Now, suppose that Taylor’s formula holds for \( n = m \). That is, we have
\[
f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(m)}(a)}{m!}(x-a)^m + R_m(x),
\]
where \( R_m(x) = \int_a^x \frac{(x-t)^m}{m!} f^{(m+1)}(t) \, dt \). We evaluate \( R_m(x) \) using integration by parts with the first function as \( f^{(m+1)}(t) \) and the second function as \( (x-t)^m/m! \). Remember that the variable of integration is \( t \) and \( x \) is a fixed number. Then
\[
R_m(x) = \left[ - f^{(m+1)}(t) \frac{(x-t)^{m+1}}{(m+1)!} \right]_a^x + \int_a^x f^{(m+2)}(t) \frac{(x-t)^{m+1}}{(m+1)!} \, dt
\]
\[
= f^{(m+1)}(a) \frac{(x-a)^{m+1}}{(m+1)!} + \int_a^x f^{(m+2)}(t) \frac{(x-t)^{m+1}}{(m+1)!} \, dt
\]
\[
= \frac{f^{(m+1)}(a)}{(m+1)!} (x-a)^{m+1} + R_{m+1}(x).
\]
This completes the proof of Taylor’s formula. The estimate of \( R_n(x) \) follows from
\[
R_n(x) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t) \, dt = (-1)^n \int_a^x \frac{(t-x)^n}{n!} (t) \, dt
\]
\[
= (-1)^n \left[ \frac{(t-x)^{n+1}}{(n+1)!} \right]_a^x = (-1)^n (1-a)^{n+1} \frac{(x-a)^{n+1}}{(n+1)!} = \frac{(x-a)^{n+1}}{(n+1)!}. \quad \square
\]
Notice that if \( f(x) \) is a polynomial of degree \( n \), then Taylor’s polynomial of order \( n \) is equal to the original polynomial.
2.4 Taylor series

As Example 2.2 shows, by clever manipulation of known series and functions they represent we may be able to have a series representation of the function. Which functions can have a power series representation, and how to obtain a power series from such a given function?

Recall: Taylor’s formulas (Theorem 2.6 and Theorem 2.7) say that under suitable hypotheses a function can be written in the following forms:

\[ f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}. \]

\[ f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \int_a^x \frac{(x-t)^n}{n!}f^{(n+1)}(t) \, dt. \]

It is thus clear that whenever one (form) of the remainder term

\[ R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \quad \text{OR} \quad R_n(x) = \int_a^x \frac{(x-t)^n}{n!}f^{(n+1)}(t) \, dt \]

converges to 0 for all \( x \) in an interval around the point \( x = a \), the series on the right hand side would converge and then the function can be written in the form of a series. That is, under the conditions that \( f(x) \) has derivatives of all order, and \( R_n(x) \to 0 \) for all \( x \) in an interval around \( x = a \), the function \( f(x) \) has a power series representation

\[ f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots \]

Such a series is called the Taylor series expansion of the function \( f(x) \). When \( a = 0 \), the Taylor series is called the Maclaurin series.

Conversely, if a function \( f(x) \) has a power series expansion about \( x = a \), then by repeated differentiation and evaluation at \( x = a \) shows that the co-efficients of the power series are precisely of the form \( \frac{f^{(n)}(a)}{n!} \) as in the Taylor series.

**Example 2.4.** Find the Taylor series expansion of the function \( f(x) = 1/x \) at \( x = 2 \). In which interval around \( x = 2 \), the series converges?

We see that

\[ f(x) = x^{-1}, \quad f(2) = \frac{1}{2}; \quad \cdots ; \quad f^{(n)}(x) = (-1)^nn!x^{-(n+1)}, \quad f^{(n)}(2) = (-1)^nn!2^{-(n+1)}. \]

Hence the Taylor series for \( f(x) = 1/x \) is

\[ \frac{1}{2} - \frac{x-2}{2^2} + \frac{(x-2)^2}{2^3} - \cdots + (-1)^n\frac{(x-2)^n}{2^{n+1}} + \cdots \]

A direct calculation can be done looking at the Taylor series so obtained. Here, the series is a geometric series with ratio \( r = -(x-2)/2 \). Hence it converges absolutely whenever

\[ |r| < 1, \quad \text{i.e.,} \quad |x-2| < 2 \quad \text{i.e.,} \quad 0 < x < 4. \]
Does this convergent series converge to the given function? We now require the remainder term in the Taylor expansion. The absolute value of the remainder term in the differential form is (for any \( c, x \) in an interval around \( x = 2 \))

\[
|R_n| = \left| \frac{f^{(n+1)}(c)}{(n+1)!} (x-2)^{n+1} \right| = \left| \frac{(x-2)^{n+1}}{c^{n+2}} \right|
\]

Here, \( c \) lies between \( x \) and 2. Clearly, if \( x \) is near 2, \( |R_n| \to 0 \). Hence the Taylor series represents the function near \( x = 2 \).

**Example 2.5.** Consider the function \( f(x) = e^x \). For its Maclaurin series, we find that

\[
f(0) = 1, \quad f'(0) = 1, \ldots, f^{(n)}(0) = 1, \ldots
\]

Hence its Taylor series is

\[
1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots
\]

By the ratio test, this power series has the radius of convergence

\[
R = \lim_{n \to \infty} \frac{a_n}{a_{n+1}} = \lim_{n \to \infty} \frac{(n+1)!}{n!} = \infty.
\]

Therefore, for every \( x \in \mathbb{R} \) the above series converges. Using the integral form of the remainder,

\[
|R_n(x)| = \left| \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t) \, dt \right| = \left| \int_0^x \frac{(x-t)^n}{n!} e^t \, dt \right| \to 0 \quad \text{as} \quad n \to \infty.
\]

Hence, for each \( x \in \mathbb{R} \),

\[
e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots.
\]

**Example 2.6.** The Taylor series for \( \cos x \) is given by

\[
\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.
\]

The absolute value of the remainder in the differential form is

\[
|R_{2n}(x)| = \frac{|x|^{2n+1}}{(2n+1)!} \to 0 \quad \text{as} \quad n \to \infty
\]

for any \( x \in \mathbb{R} \). Hence the series represents \( \cos x \). That is, for each \( x \in \mathbb{R} \),

\[
\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.
\]

The Taylor polynomials approximating \( \cos x \) are therefore

\[
P_{2n}(x) = \sum_{k=0}^{n} \frac{(-1)^k x^{2k}}{(2k)!}.
\]

The following picture shows how these polynomials approximate \( \cos x \) for \( 0 \leq x \leq 9 \).
Example 2.7. Let \( m \in \mathbb{R} \). Show that, for \(-1 < x < 1\),

\[
(1 + x)^m = 1 + \sum_{n=1}^{\infty} \binom{m}{n} x^n,
\]
where \( \binom{m}{n} = \frac{m(m-1) \cdots (m-n+1)}{n!} \).

To see this, find the derivatives of the given function:

\[
f(x) = (1 + x)^m, \quad f^{(n)}(x) = m(m-1) \cdots (m-n+1)x^{m-n}.
\]

Then the Maclaurin series for \( f(x) \) is the given series. You must show that the series converges for \(-1 < x < 1\) and then the remainder term in the Maclaurin series expansion goes to 0 as \( n \to \infty \) for all such \( x \). The series so obtained is called a binomial series expansion of \( (1 + x)^m \).

Substituting values of \( m \), we get series for different functions. For example, with \( m = 1/2 \), we have

\[
(1 + x)^{1/2} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \cdots \text{ for } -1 < x < 1.
\]

Notice that when \( m \in \mathbb{N} \), the binomial series terminates to give a polynomial and it represents \( (1 + x)^m \) for each \( x \in \mathbb{R} \).

Exercises for §2.2-2.3

1. Find Taylor’s polynomials of orders 0, 1, 2, 3 for the following functions at \( x = a \).
   (a) \( f(x) = \ln x \), \( a = 1 \). \hspace{1cm} (b) \( f(x) = 1/(x + 2) \), \( a = 0 \) \hspace{1cm} (c) \( f(x) = \sin x \), \( a = \pi/4 \).

2. Find Maclaurin’s series for the following functions:
   (a) \( f(x) = (1+x)^{-1} \), \hspace{1cm} (b) \( f(x) = 5 \cos(\pi x) \), \hspace{1cm} (c) \( f(x) = \sinh x = \frac{1}{2}(e^x - e^{-x}) \).

3. Find Taylor series for the following functions at \( x = a \).
   (a) \( f(x) = 3x^5 - x^4 + 2x^3 + x^2 - 2 \), \( a = -1 \) \hspace{1cm} (b) \( f(x) = x/(1-x) \), \( a = 0 \) \hspace{1cm} (c) \( f(x) = 2^x \), \( a = 1 \).
   (d) \( f(x) = \cos(x^{3/2}/\sqrt{2}) \), \( a = 0 \) \hspace{1cm} (e) \( f(x) = \cos^2 x \), \( a = 0 \) \hspace{1cm} (f) \( f(x) = x^2/(1-2x) \), \( a = 0 \).

4. Sum the series \( \frac{\pi}{3} - \frac{\pi^3}{3^3 \cdot 3} + \frac{\pi^5}{3^5 \cdot 5} - \cdots \).

2.5 Fourier series

In the power series for \( \sin x = 1 - x + x^3/3! - \cdots \), the periodicity of \( \sin x \) is not obvious. Also, periodic functions can be expanded in a series involving sines and cosines instead of powers of \( x \). A trigonometric series is of the form

\[
\frac{1}{2} a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)
\]
Since both cosine and sine functions are periodic of period $2\pi$, if the trigonometric series converges to a function $f(x)$, then necessarily $f(x)$ is also periodic of period $2\pi$. Thus,

$$f(0) = f(2\pi) = f(4\pi) = f(6\pi) = \cdots \quad \text{and} \quad f(-\pi) = f(\pi) \quad \text{etc.}$$

Moreover, if $f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$, say, for all $x \in [-\pi, \pi]$, then the coefficients can be determined from $f(x)$. Towards this, multiply $f(t)$ by $\cos mt$ and integrate to obtain:

$$\int_{-\pi}^{\pi} f(t) \cos mt \, dt = \frac{1}{2}a_0 \int_{-\pi}^{\pi} \cos mt \, dt + \sum_{n=1}^{\infty} a_n \int_{-\pi}^{\pi} \cos nt \cos mt \, dt + \sum_{n=1}^{\infty} b_n \int_{-\pi}^{\pi} \sin nt \cos mt \, dt.$$

For $m, n = 0, 1, 2, 3, \ldots$,

$$\int_{-\pi}^{\pi} \cos nt \cos mt = \begin{cases} 0 & \text{if } n \neq m \\ \pi & \text{if } n = m > 0 \\ 2\pi & \text{if } n = m = 0 \end{cases} \quad \text{and} \quad \int_{-\pi}^{\pi} \sin nt \cos mt = 0.$$

Thus, we obtain

$$\int_{-\pi}^{\pi} f(t) \cos mt = \pi a_m, \quad \text{for all } m = 0, 1, 2, 3, \cdots$$

Similarly, by multiplying $f(t)$ by $\sin mt$ and integrating, and using the fact that

$$\int_{-\pi}^{\pi} \sin nt \sin mt = \begin{cases} 0 & \text{if } n \neq m \\ \pi & \text{if } n = m > 0 \\ 0 & \text{if } n = m = 0 \end{cases}$$

we obtain

$$\int_{-\pi}^{\pi} f(t) \sin mt = \pi b_m, \quad \text{for all } m = 1, 2, 3, \cdots$$

Assuming that $f(x)$ has period $2\pi$, we then give the following definition.

Let $f : [-\pi, \pi] \to \mathbb{R}$ be an integrable function extended to $\mathbb{R}$ by periodicity of period $2\pi$, i.e., $f : \mathbb{R} \to \mathbb{R}$ satisfies

$$f(x + 2\pi) = f(x) \quad \text{for all } x \in \mathbb{R}.$$  

Let $a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt \, dt$ for $n = 0, 1, 2, 3, \ldots$, and $b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt \, dt$ for $n = 1, 2, 3, \ldots$.

Then the trigonometric series

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

is called the **Fourier series** of $f(x)$. 

40
A fundamental result, which we state without proof, about Fourier series gives information about its convergence to a function. Recall that function \( f(x) \) is called \textit{piecewise continuous} on an interval iff all points in that interval, where \( f(x) \) is discontinuous, are finite in number; and at such an interior point \( c \), the left and right sided limits \( f(c-) \) and \( f(c+) \) exist.

**Theorem 2.8.** (Convergence of Fourier Series) Let \( f: [-\pi, \pi] \rightarrow \mathbb{R} \) be a function extended to \( \mathbb{R} \) by periodicity of period \( 2\pi \). Suppose \( f(x) \) is piecewise continuous, and \( f(x) \) has both left hand derivative and right hand derivative at each \( x \in (-\pi, \pi) \). Then \( f(x) \) is equal to its Fourier series at all points where \( f(x) \) is continuous; and at a point \( c \), where \( f(x) \) is discontinuous, the Fourier series converges to \( \frac{1}{2}[f(c+) + f(c-)] \).

In particular, if \( f(x) \) and \( f'(x) \) are continuous on \([-\pi, \pi]\) with period \( 2\pi \), then

\[
f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \quad \text{for all} \quad x \in \mathbb{R}.
\]

Further, if \( f(x) \) is an odd function, i.e., \( f(-x) = f(x) \), then for \( n = 0, 1, 2, 3, \ldots \),

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt \, dt = 0
\]

\[
b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt \, dt = \frac{2}{\pi} \int_{0}^{\pi} f(t) \sin nt \, dt.
\]

In this case,

\[
f(x) = \sum_{n=1}^{\infty} b_n \sin nx \quad \text{for all} \quad x \in \mathbb{R}.
\]

Similarly, if \( f(x) \) and \( f'(x) \) are continuous on \([-\pi, \pi]\) with period \( 2\pi \) and if \( f(x) \) is an even function, i.e., \( f(-x) = f(x) \), then

\[
f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx \quad \text{for all} \quad x \in \mathbb{R},
\]

where

\[
a_n = \frac{2}{\pi} \int_{0}^{\pi} f(t) \cos nt \, dt \quad \text{for} \quad n = 0, 1, 2, 3, \ldots
\]

Fourier series can represent functions which cannot be represented by a Taylor series, or a conventional power series; for example, a step function.

**Example 2.8.** Find the Fourier series of the function \( f(x) \) given by the following which is extended to \( \mathbb{R} \) with the periodicity \( 2\pi \):

\[
f(x) = \begin{cases} 
1 & \text{if } 0 \leq x < \pi \\
2 & \text{if } \pi \leq x < 2\pi
\end{cases}
\]
Due to periodic extension, we can rewrite the function \( f(x) \) on \([-\pi, \pi]\) as

\[
 f(x) = \begin{cases} 
 2 & \text{if } -\pi \leq x < 0 \\
 1 & \text{if } 0 \leq x < \pi 
\end{cases}
\]

Then the coefficients of the Fourier series are computed as follows:

\[
 a_0 = \frac{1}{\pi} \int_{-\pi}^{0} f(t) \, dt + \frac{1}{\pi} \int_{0}^{\pi} f(t) \, dt = 3.
\]

\[
 a_n = \frac{1}{\pi} \int_{-\pi}^{0} \cos nt \, dt + \frac{1}{\pi} \int_{0}^{\pi} 2 \cos nt \, dt = 0.
\]

\[
 b_n = \frac{1}{\pi} \int_{-\pi}^{0} \sin nt \, dt + \frac{1}{\pi} \int_{0}^{\pi} 2 \sin nt \, dt = \frac{(-1)^n - 1}{n\pi}.
\]

Notice that \( b_1 = -\frac{2}{\pi}, \ b_2 = 0, \ b_3 = -\frac{2}{3\pi}, \ b_4 = 0, \ldots \). Therefore,

\[
 f(x) = \frac{3}{2} - \frac{2}{\pi} \left( \sin x + \frac{3\sin 3x}{3} + \frac{5\sin 5x}{5} + \cdots \right).
\]

Here, the last expression for \( f(x) \) holds for all \( x \in \mathbb{R} \); however, the function here has been extended to \( \mathbb{R} \) by using its periodicity as \( 2\pi \). If we do not extend but find the Fourier series for the function as given on \([-\pi, \pi]\), then also for all \( x \in [-\pi, \pi] \), the same expression holds.

Once we have a series representation of a function, we should see how the **partial sums** of the series approximate the function. In the above example, let us write

\[
 f_m(x) = \frac{1}{2} a_0 + \sum_{n=1}^{m} (a_n \cos nx + b_n \sin nx).
\]

The approximations \( f_1(x), \ f_3(x), \ f_5(x), \ f_9(x) \) and \( f_{15}(x) \) to \( f(x) \) are shown in the figure below.
Example 2.9. Show that \( x^2 = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n \cos nx}{n^2} \) for all \( x \in [-\pi, \pi] \).

The extension of \( f(x) = x^2 \) to \( \mathbb{R} \) is not the function \( x^2 \). For illustration, in the interval \([\pi, 3\pi]\), its extension looks like \( f(x) = (x - 2\pi)^2 \). Remember that the extension has period \( 2\pi \). Also, notice that \( f(\pi) = f(-\pi) \); thus we have no problem at the point \( \pi \) in extending the function continuously. With this understanding, we go for the Fourier series expansion of \( f(x) = x^2 \) in the interval \([-\pi, \pi]\). We also see that \( f(x) \) is an even function. Its Fourier series is a cosine series. The coefficients of the series are as follows:

\[
a_0 = \frac{2}{\pi} \int_{0}^{\pi} t^2 \, dt = \frac{2}{3} \pi^2.
\]

\[
a_n = \frac{2}{\pi} \int_{0}^{\pi} t^2 \cos nt \, dt = \frac{4}{n^2} (-1)^n.
\]

Therefore,

\[
f(x) = x^2 = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n \cos nx}{n^2} \quad \text{for all } x \in [-\pi, \pi].
\]

In particular, by taking \( x = 0 \) and \( x = \pi \), we have

\[
\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12}, \quad \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.
\]

Due to the periodic extension of \( f(x) \) to \( \mathbb{R} \), we see that

\[
(x - 2\pi)^2 = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n \cos nx}{n^2} \quad \text{for all } x \in [\pi, 3\pi]
\]

which, of course, is simply a verification of the formula for \( x^2 \) for \( x \in [-\pi, \pi] \). It also follows that the same sum is equal to \((x - 4\pi)^2 \) for \( x \in [3\pi, 5\pi] \), etc.

Example 2.10. Show that the Fourier series for \( f(x) = x^2 \) defined on \((0, 2\pi)\) is given by

\[
\frac{4\pi^2}{6} + \sum_{n=1}^{\infty} \left( \frac{4}{\pi^2} \cos nx - \frac{4\pi}{n} \sin nx \right).
\]

Extend \( f(x) \) to \( \mathbb{R} \) by periodicity \( 2\pi \) and by taking \( f(0) = f(2\pi) \). Then

\[
f(-\pi) = f(-\pi+2\pi) = f(\pi) = \pi^2, \quad f(-\pi/2) = f(-\pi/2+2\pi) = f(3\pi/2) = (3\pi/2)^2, \quad f(0) = f(2\pi).
\]

Thus the function \( f(x) \) on \([-\pi, \pi]\) is defined by

\[
f(x) = \begin{cases} (x + 2\pi)^2 & \text{if } -\pi \leq x < 0 \\ x^2 & \text{if } 0 \leq x \leq \pi. \end{cases}
\]

Notice that \( f(x) \) is neither odd nor even. The coefficients of the Fourier series for \( f(x) \) are

\[
a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \, dt = \frac{1}{\pi} \int_{0}^{2\pi} t^2 \, dt = \frac{8\pi^2}{3}.
\]
\[ a_n = \frac{1}{\pi} \int_0^{2\pi} t \cos nt \, dt = \frac{4}{n^2}, \]
\[ b_n = \frac{1}{\pi} \int_0^{2\pi} t \sin nt \, dt = -\frac{4\pi}{n}. \]

Hence the Fourier series for \( f(x) \) is as claimed.

As per the extension of \( f(x) \) to \( \mathbb{R} \), we see that in the interval \((2k\pi, 2(k + 1)\pi)\), the function is defined by \( f(x) = (x - 2k\pi)^2 \). Thus it has discontinuities at the points \( x = 0, \pm 2\pi, \pm 4\pi, \ldots \). At such a point \( x = 2k\pi \), the series converges to the average value of the left and right side limits, i.e., the series when evaluated at \( 2k\pi \) yields the value

\[ \frac{1}{2} \left[ \lim_{x \to 2k\pi^-} f(x) + \lim_{x \to 2k\pi^+} f(x) \right] = \frac{1}{2} \left[ \lim_{x \to 2k\pi^-} (x - 2k\pi)^2 + \lim_{x \to 2k\pi^+} (x - 2(k + 1)\pi)^2 \right] = 2\pi^2. \]

Notice that since \( f(x) \) is extended by periodicity, whether we take the basic interval as \([-\pi, \pi] \) or as \([0, 2\pi] \) does not matter in the calculation of coefficients. We will follow this suggestion elsewhere instead of always redefining \( f(x) \) on \([-\pi, \pi] \). However, the odd or even classification of \( f(x) \) may break down.

**Example 2.11.** Show that for \( 0 < x < 2\pi \), \( \frac{1}{2}(\pi - x) = \sum_{n=1}^{\infty} \frac{\sin nx}{n} \).

Let \( f(x) = x \) for \( 0 < x < 2\pi \). Extend \( f(x) \) to \( \mathbb{R} \) by taking the periodicity as \( 2\pi \) and with the condition that \( f(0) = f(2\pi) \). As in Example 2.9, \( f(x) \) is not an odd function. For illustration, \( f(-\pi/2) = f(3\pi/2) = 3\pi/2 \neq f(\pi/2) = \pi/2 \).

The coefficients of the Fourier series for \( f(x) \) are as follows:

\[ a_0 = \frac{1}{\pi} \int_0^{2\pi} t \, dt = 2\pi, \quad a_n = \frac{1}{\pi} \int_0^{2\pi} t \cos nt \, dt = 0. \]
\[ b_n = \frac{1}{\pi} \int_0^{2\pi} t \sin nt \, dt = \frac{1}{\pi} \left[ \frac{-n \cos nt}{n} \right]_0^{2\pi} + \frac{1}{n\pi} \int_0^{2\pi} \cos nt \, dt = -\frac{2}{n}. \]

By the convergence theorem, \( x = \pi - 2 \sum_{n=1}^{\infty} \frac{\sin nx}{n} \) for \( 0 < x < 2\pi \), which yields the required result.

**Example 2.12.** Find the Fourier series expansion of \( f(x) = \begin{cases} x & \text{if } 0 \leq x \leq \pi/2 \\ \pi - x & \text{if } \pi/2 < x < \pi. \end{cases} \)

Notice that \( f(x) \) has the domain as an interval of length \( \pi \) and not \( 2\pi \). Thus, there are many ways of extending it to \( \mathbb{R} \) by periodicity \( 2\pi \).

1. **Odd Extension:**

First, extend \( f(x) \) to \([-\pi, \pi]\) by requiring that \( f(x) \) is an odd function. This requirement forces \( f(-x) = -f(x) \) for each \( x \in [-\pi, \pi] \). Next, we extend this \( f(x) \) which has been now defined on \([-\pi, \pi] \) to \( \mathbb{R} \) by periodicity \( 2\pi \).
The Fourier series expansion of this extended \( f(x) \) is a sine series, whose coefficients are given by

\[
b_n = \frac{2}{\pi} \int_0^\pi f(t) \sin nt \, dt = \frac{2}{\pi} \int_0^{\pi/2} t \sin nt \, dt + \frac{2}{\pi} \int_{\pi/2}^\pi (\pi - t) \sin nt \, dt = (-1)^{(n-1)/2} \frac{\pi}{4n^2}.
\]

Thus \( f(x) = \frac{\pi}{4} \left( \frac{\sin x}{2} - \sin 3x \frac{1}{3^2} + \sin 5x \frac{1}{5^2} - \cdots \right) \).

In this case, we say that the Fourier series is a **sine series expansion** of \( f(x) \).

2. **Even Extension:**

First, extend \( f(x) \) to \([-\pi, \pi]\) by requiring that \( f(x) \) is an even function. This requirement forces \( f(-x) = f(x) \) for each \( x \in [-\pi, \pi] \). Next, we extend this \( f(x) \) which has been now defined on \([-\pi, \pi]\) to \( \mathbb{R} \) by periodicity \( 2\pi \).

The Fourier series expansion of this extended \( f(x) \) is a cosine series, whose coefficients are

\[
a_n = \frac{2}{\pi} \int_0^\pi f(t) \cos nt \, dt = \frac{2}{\pi} \int_0^{\pi/2} t \cos nt \, dt + \frac{2}{\pi} \int_{\pi/2}^\pi (\pi - t) \cos nt \, dt = -\frac{2}{n^2\pi} \quad \text{for } 4 \not| \, n.
\]

And \( a_0 = \pi/4 \), \( a_{4k} = 0 \). Thus \( f(x) = \frac{\pi}{4} - \frac{2}{\pi} \left( \cos 2x \frac{1}{1^2} + \cos 6x \frac{1}{3^2} + \sin 10x \frac{1}{5^2} + \cdots \right) \).

In this case, we say that the Fourier series is a **cosine series expansion** of \( f(x) \).

3. **Scaling to length** \( 2\pi \):

We define a function \( g : [-\pi, \pi] \rightarrow [0, \pi] \). Then consider the composition \( h = (f \circ g) : [-\pi, \pi] \rightarrow \mathbb{R} \). We find the Fourier series for \( h(y) \) and then resubstitute \( y = g^{-1}(x) \) for obtaining Fourier series for \( f(x) \). Notice that in computing the Fourier series for \( h(y) \), we must extend \( h(y) \) to \( \mathbb{R} \) using periodicity of period \( 2\pi \) and \( h(-\pi) = h(\pi) \).

In this approach, we consider

\[
x = g(y) = \frac{1}{2}(y + \pi), \quad h(y) = f \left( \frac{y + \pi}{2} \right) = \begin{cases} \frac{1}{2}(y + \pi) & \text{if } -\pi \leq y \leq 0 \\ \frac{1}{2}(3\pi - y) & \text{if } 0 \leq y \leq \pi. \end{cases}
\]

The Fourier coefficients are as follows:

\[
a_0 = \frac{1}{\pi} \int_{-\pi}^{0} \frac{t + \pi}{2} \, dt + \frac{1}{\pi} \int_{0}^{\pi} \frac{3\pi - t}{2} \, dt = 2\pi.
\]

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{0} \frac{t + \pi}{2} \cos nt \, dt + \frac{1}{\pi} \int_{0}^{\pi} \frac{3\pi - t}{2} \cos nt \, dt = \pi n^2 (1 - (-1)^n) = \begin{cases} \frac{2}{n\pi^2} & \text{if } n \text{ odd} \\ 0 & \text{if } n \text{ even}. \end{cases}
\]

\[
b_n = \frac{1}{\pi} \int_{-\pi}^{0} \frac{t + \pi}{2} \sin nt \, dt + \frac{1}{\pi} \int_{0}^{\pi} \frac{3\pi - t}{2} \sin nt \, dt = \frac{1}{2n} [2(-1)^n - 1 + 3 - 2(-1)^n] = \frac{1}{n}.
\]

Then the Fourier series for \( h(y) \) is given by

\[
\pi + \sum_{n=1}^{\infty} \left( a_n \cos ny + \frac{1}{n} \sin ny \right).
\]
Using \( y = g^{-1}(x) = 2x - \pi \), we have the Fourier series for \( f(x) \) as

\[
\pi + \sum_{n=1}^{\infty} \left( a_n \cos n(2x - \pi) + \frac{1}{n} \sin n(2x - \pi) \right).
\]

Notice that this is neither a cosine series nor a sine series.

This example suggests three ways of constructing Fourier series for a function \( f(x) \), which might have been defined on any arbitrary interval \([a, b]\).

The first approach says that we define a function \( g(y) : [0, \pi] \rightarrow [a, b] \) and consider the composition \( f \circ g \). Now, \( f \circ g : [0, \pi] \rightarrow \mathbb{R} \). Next, we take an odd extension of \( f \circ g \) with periodicity \( 2\pi \); and call this extended function as \( h \). We then construct the Fourier series for \( h(y) \). Finally, substitute \( y = g^{-1}(x) \). This will give a sine series.

In the second approach, we define \( g(y) \) as in the first approach and take an even extension of \( f \circ g \) with periodicity \( 2\pi \); and call this extended function as \( h \). We then construct the Fourier series for \( h(y) \). Finally, substitute \( y = g^{-1}(x) \). This gives a cosine series.

These two approaches lead to the so-called **half range Fourier series** expansions.

In the third approach, we define a function \( g(y) : [-\pi, \pi] \rightarrow [a, b] \) and consider the composition \( f \circ g \). Now, \( f \circ g : [-\pi, \pi] \rightarrow \mathbb{R} \). Next, we extend \( f \circ g \) with periodicity \( 2\pi \); and call this extended function as \( h \). We then construct the Fourier series for \( h(y) \). Finally, substitute \( y = g^{-1}(x) \). This may give a general Fourier series involving both sine and cosine terms.

In particular, a function \( f : [-\ell, \ell] \rightarrow \mathbb{R} \) which is known to have period \( 2\ell \) can easily be expanded in a Fourier series by considering the new function \( g(x) = f(\ell x/\pi) \). Now, \( g : [-\pi, \pi] \rightarrow \mathbb{R} \) has period \( 2\pi \). We construct a Fourier series of \( g(x) \) and then substitute \( x \) with \( \pi x/\ell \) to obtain a Fourier series of \( f(x) \). This is the reason the third method above is called scaling.

In this case, the Fourier coefficients are given by

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f \left( \frac{\ell}{\pi} t \right) \cos nt \, dt, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f \left( \frac{\ell}{\pi} t \right) \sin nt \, dt
\]

Substituting \( s = \frac{\ell}{\pi} t \), \( dt = \frac{\pi}{\ell} ds \), we have

\[
a_n = \frac{1}{\ell} \int_{-\ell}^{\ell} f(s) \cos ns \, ds, \quad b_n = \frac{1}{\ell} \int_{-\ell}^{\ell} f(s) \sin ns \, ds
\]

And the Fourier series for \( f(x) \) is then, with the original variable \( x \),

\[
\frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi}{\ell} x + b_n \sin \frac{n\pi}{\ell} x \right).
\]

**Remark:** If a function is only defined on an interval \((0, \ell)\), then all the three approaches are applicable. We may extend this function to \((-\ell, \ell)\) by either taking an odd extension or an even extension. Then we may scale \((-\ell, \ell)\) to \((-\pi, \pi)\). Finally extend the function to \(\mathbb{R} \) with periodicity. The Fourier series of this extended function will be a half range expansion. Alternatively, we
may scale \((0, \ell)\) to \((-\pi, \pi)\) and then extend to \(\mathbb{R}\) with periodicity \(2\pi\), and obtain a Fourier series expansion of the resulting function. We may also use the interval \([-\ell, \ell]\) directly in the integrals while evaluating the Fourier coefficients instead of first scaling to \([-\pi, \pi]\) and then constructing the Fourier series.

**Example 2.13.** Construct the half-range Fourier cosine series for \(f(x) = |x|\) on \([0, \ell]\) for some \(\ell > 0\).

Notice that the even extension of the function is \(|x|\) on \([-\ell, \ell]\). Next, it is extended to \(f : \mathbb{R} \to \mathbb{R}\) with period \(2\ell\). It is not \(|x|\) on \(\mathbb{R}\); it is \(|x|\) on \([-\ell, \ell]\). Due to its period as \(2\ell\), it is \(|x - 2\ell|\) on \([\ell, 3\ell]\) etc.

The Fourier coefficients are

\[
\begin{align*}
  b_n &= 0, \quad a_0 = \frac{1}{\ell} \int_{-\ell}^{\ell} |s| \, ds = \frac{2}{\ell} \int_{0}^{\ell} s \, ds = \ell, \\
  a_n &= \frac{2}{\ell} \int_{0}^{\ell} s \cos \left(\frac{n\pi s}{\ell}\right) \, ds = \begin{cases} 0 & \text{if } n \text{ even} \\
 -\frac{4\ell}{n^2\pi^2} & \text{if } n \text{ odd} \end{cases}
\end{align*}
\]

Therefore the Fourier series for \(f(x)\) shows that in \([-\ell, \ell]\),

\[
|x| = \frac{\ell}{2} - \frac{4\ell}{\pi^2} \left[ \frac{\cos(\pi/\ell) x}{1} + \frac{\cos(3\pi/\ell) x}{3^2} + \ldots + \frac{\cos((2n+1)\pi/\ell) x}{(2n+1)^2} + \ldots \right].
\]

As our extension of \(f(x)\) to \(\mathbb{R}\) shows, the above Fourier series represents the function given in the following figure:

![Graph of the function](image)

**A Fun Problem:** Show that the \(n\)th partial sum of the Fourier series for \(f(x)\) can be written as the following integral:

\[
s_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x + t) \sin((2n + 1)t/2) \frac{2 \sin t/2}{2 \sin t/2} \, dt.
\]

We know that \(s_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)\), where

\[
\begin{align*}
  a_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos kt \, dt, \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin kt \, dt
\end{align*}
\]
Substituting these values in the expression for $s_n(x)$, we have

$$s_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \, dt + \frac{1}{\pi} \sum_{k=1}^{n} \left[ \int_{-\pi}^{\pi} f(t) \cos kx \, dt + \int_{-\pi}^{\pi} f(t) \sin kx \, dt \right]$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} \left[ \frac{f(t)}{2} + \sum_{k=1}^{n} \{f(t) \cos kx \cos kt + f(t) \sin kx \sin kt\} \right] \, dt$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[ \frac{1}{2} + \sum_{k=1}^{n} \cos k(t-x) \right] \, dt := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sigma_n(t-x) \, dt.$$

The expression $\sigma_n(z)$ for $z = t - x$ can be re-written as follows:

$$\sigma_n(z) = \frac{1}{2} + \cos z + \cos 2z + \cdots + \cos nz$$

Thus

$$2\sigma_n(z) \cos z = \cos z + 2 \cos z \cos z + 2 \cos z \cos 2z + \cdots + 2 \cos z \cos nz$$

$$= \cos z + [1 + \cos 2z] + [\cos z + \cos 3z] + \cdots + [\cos(n-1)z + \cos(n+1)z]$$

$$= 1 + 2 \cos z + 2 \cos 2z + \cdots + 2 \cos(n-1)z + 2 \cos nz + 2 \cos(n+1)z$$

$$= 2\sigma_n(z) - \cos nz + \cos(n+1)z$$

This gives

$$\sigma_n(z) = \frac{\cos nz - \cos(n+1)z}{2(1 - \cos z)} = \frac{\sin(2n+1)z/2}{2 \sin z/2}$$

Therefore, substituting $\sigma_n(z)$ with $z = t - x$, we have

$$s_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(2n+1)(t-x)/2 \, dt$$

Since the integrand is periodic of period $2\pi$, the value of the integral remains same on any interval of length $2\pi$. Thus

$$s_n(x) = \frac{1}{\pi} \int_{-\pi}^{x+\pi} f(t) \sin(2n+1)(t-x)/2 \, dt$$

Introduce a new variable $y = t - x$, i.e., $t = x + y$. And then write the integral in terms of $t$ instead of $y$ to obtain

$$s_n(x) = \int_{-\pi}^{x+\pi} f(x+y) \sin(2n+1)y/2 \, dy = \int_{-\pi}^{x+\pi} f(x+t) \sin(2n+1)t/2 \, dt$$

This integral is called the Dirichlet Integral. In particular, taking $f(x) = 1$, we see that $a_0 = 2$, $a_k = 0$ and $b_k = 0$ for $k \in \mathbb{N}$; and then we get the identity

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \frac{\sin(2n+1)t/2}{2 \sin t/2} \, dt = 1 \quad \text{for each } n \in \mathbb{N}.$$
Exercises for § 2.5

1. Determine the Fourier series for the following functions with period $2\pi$:
   (a) $f(x) = x^2$, $0 < x < 2\pi$ 
   (b) $f(x) = \begin{cases} -4x & \text{if } -\pi < x < 0 \\ 4x & \text{if } 0 < x < \pi \end{cases}$.

2. Determine the Fourier series of the following functions with period $p$:
   (a) $f(x) = \begin{cases} 0 & \text{if } -2 < x < -1 \\ k & \text{if } -1 < x < 1 \\ 0 & \text{if } 1 < x < 2 \end{cases}$, $p = 4$ 
   (b) $f(x) = \begin{cases} 1 + x & \text{if } -1 < x < 0 \\ 1 - x & \text{if } 0 < x < 1 \end{cases}$, $p = 2$.

3. Find the two half range Fourier series (period is $2\ell$) of the following function by (a) using an odd extension, and (b) using an even extension:
   $$f(x) = \begin{cases} \frac{2k}{\ell}x & \text{if } 0 < x < \frac{\ell}{2} \\ \frac{2k}{\ell}(\ell - x) & \text{if } \frac{\ell}{2} < x < \ell \end{cases}$$

4. Find the Fourier series of the following functions with given period $p$:
   (a) $f(x) = 2x|x|$, $-1 < x < 1$, $p = 2$ 
   (b) $f(x) = \begin{cases} 1 - \frac{1}{2}|x| & \text{if } -2 < x < 2 \\ 0 & \text{if } 2 < x < 6 \end{cases}$, $p = 8$.

5. Find the half range Fourier sine series and Fourier cosine series of the following functions:
   (a) $f(x) = \begin{cases} 0 & \text{if } 0 < x < 2 \\ 1 & \text{if } 2 < x < 4 \end{cases}$ 
   (b) $f(x) = \begin{cases} x & \text{if } 0 < x < \pi/2 \\ \frac{\pi}{2} & \text{if } \pi/2 < x < \pi \end{cases}$.
Part II

Matrices
Chapter 3

Matrix Operations

3.1 Examples of linear equations

Linear equations are everywhere, starting from mental arithmetic problems to advanced defense applications. We start with an example. The system of linear equations

\[
\begin{align*}
    x_1 + x_2 &= 3 \\
    x_1 - x_2 &= 1
\end{align*}
\]

has a unique solution \( x_1 = 2, \ x_2 = 1 \). Substituting these values for the unknowns, we see that the equations are satisfied; but why are there no other solutions? Well, we have not merely guessed this solution; we have solved it! The details are as follows:

Suppose the pair \((x_1, x_2)\) is a solution of the system. Subtracting the first from the second, we get another equation: \( -2x_2 = -2 \). It implies \( x_2 = 1 \). Then from either of the equations, we get \( x_1 = 1 \). To proceed systematically, we would like to replace the original system with the following:

\[
\begin{align*}
    x_1 + x_2 &= 3 \\
    x_2 &= 1
\end{align*}
\]

Substituting \( x_2 = 1 \) in the first equation of the new system, we get \( x_1 = 2 \). In fact, substituting these values of \( x_1 \) and \( x_2 \), we see that the original equation is satisfied.

Convinced? The only solution of the system is \( x_1 = 2, x_2 = 1 \). What about the system

\[
\begin{align*}
    x_1 + x_2 &= 3 \\
    x_1 - x_2 &= 1 \\
    2x_1 - x_2 &= 3
\end{align*}
\]

The first two equations have a unique solution and that satisfies the third. Hence this system also has a unique solution \( x_1 = 2, x_2 = 1 \). So the extra equation does not put any constraint on the solutions that we obtained earlier.

But what about our systematic solution method? We aim at eliminating the first unknown from all
but the first equation. We replace the second equation with the one obtained by second minus the 
fir[l]st. We also replace the third by third minus twice the first. It results in

\[
\begin{align*}
x_1 + x_2 &= 3 \\
-x_2 &= -1 \\
-3x_2 &= 3
\end{align*}
\]

Notice that the second and the third equations *coincide*, hence the conclusion. We give another 
twist. Consider the system

\[
\begin{align*}
x_1 + x_2 &= 3 \\
x_1 - x_2 &= 1 \\
2x_1 + x_2 &= 3
\end{align*}
\]

The first two equations again have the same solution \(x_1 = 2, x_2 = 1\). But this time, the third is 
not satisfied by these values of the unknowns. So, the system has no solution. Also, by using our 
elimination method, we obtain the equations as:

\[
\begin{align*}
x_1 + x_2 &= 3 \\
-x_2 &= -1 \\
-3x_2 &= -3
\end{align*}
\]

The last two equations are not *consistent*. So, the original system has no solution.

Finally, instead of adding another equation, we drop one. Consider the linear equation

\[x_1 + x_2 = 3\]

having only one equation. The old solution \(x_1 = 2, x_2 = 1\) is still a solution of this system. 
But \(x_1 = 1, x_2 = 2\) is also a solution. Moreover, since \(x_1 = 3 - x_2\), by assigning \(x_2\) any real 
number, we get a corresponding value for \(x_1\), which together give a solution. Thus, it has infinitely 
many solutions. Notice that the same conclusion holds if we have more equations, which are some 
multiple of the only given equation. For example,

\[
\begin{align*}
x_1 + x_2 &= 3 \\
2x_1 + 2x_2 &= 6 \\
3x_1 + 3x_2 &= 9
\end{align*}
\]

We see that the number of equations really does not matter, but the number of *independent* equations 
does matter.

*Warning*: the notion of independent equations is not yet clear; nonetheless we have some working 
idea.

It is not also very clear when does a system of equations have a solution, a unique solution, in-
finity many solutions, or even no solutions. And why not a system of equations has more than 
one but finitely many solutions? How do we use our elimination method for obtaining infinite 
number of solutions? To answer these questions, we will introduce matrices. Matrices will help 
us in representing the problem is a compact way and also will lead to a definitive answer. We will 
also study the eigenvalue problem for matrices which come up often in applications.
3.2 Basic matrix operations

A matrix is a rectangular array of symbols. For us these symbols are real numbers or, in general, complex numbers. The individual numbers in the array are called the entries of the matrix. The number of rows and the number of columns in any matrix are necessarily positive integers. A matrix with $m$ rows and $n$ columns is called an $m \times n$ matrix and it may be written as

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix},$$

or as $A = [a_{ij}]$ for short with $a_{ij} \in \mathbb{F}$ for $i = 1, \ldots, m, j = 1, \ldots, n$. The number $a_{ij}$ which occurs at the entry in $i$th row and $j$th column is referred to as the $(ij)$th entry (sometimes as $(i, j)$-th entry) of the matrix $[a_{ij}]$.

As usual, $\mathbb{R}$ denotes the set of all real numbers and $\mathbb{C}$ denotes the set of all complex numbers. We will write $\mathbb{F}$ for either $\mathbb{R}$ or $\mathbb{C}$. The numbers in $\mathbb{F}$ will also be referred to as scalars. Thus each entry of a matrix is a scalar.

Any matrix with $m$ rows and $n$ columns will be referred as an $m \times n$ matrix. The set of all $m \times n$ matrices with entries from $\mathbb{F}$ will be denoted by $\mathbb{F}^{m \times n}$.

A row vector of size $n$ is a matrix in $\mathbb{F}^{1 \times n}$. Similarly, a column vector of size $m$ is a matrix in $\mathbb{F}^{m \times 1}$. Sometimes we will write $\mathbb{F}^{1 \times n}$ as $\mathbb{F}^n$. The vectors in $\mathbb{F}^n$ will be written as

$$(a_1, \ldots, a_n) \text{ or as } [a_1, \ldots, a_n] \text{ or as } [a_1 \cdots a_n].$$

We will sometimes write a column vector as $[b_1 \cdots b_n]^t$, for saving vertical space.

Any matrix in $\mathbb{F}^{m \times n}$ is said to have its size as $m \times n$. If $m = n$, the rectangular array becomes a square array with $m$ rows and $m$ columns; and the matrix is called a square matrix of order $m$.

Naturally, two matrices of the same size are considered equal when their corresponding entries coincide, i.e., if $A = [a_{ij}]$ and $B = [b_{ij}]$ are in $\mathbb{F}^{m \times n}$, then

$$A = B \iff a_{ij} = b_{ij}$$

for each $i \in \{1, \ldots, m\}$ and for each $j \in \{1, \ldots, n\}$. Thus matrices of different sizes are unequal.

The zero matrix is a matrix each entry of which is 0. We write 0 for all zero matrices of all sizes. The size is to be understood from the context.

Let $A = [a_{ij}] \in \mathbb{F}^{n \times n}$ be a square matrix of order $n$. The entries $a_{ij}$ are called as the diagonal entries of $A$. The diagonal of $A$ consists of all diagonal entries; the first entry on the diagonal is $a_{11}$, and the last diagonal entry is $a_{nn}$. The entries of $A$, which are not on the diagonal, are called as off diagonal entries of $A$; they are $a_{ij}$ for $i \neq j$. The diagonal of the following matrix is shown in red:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 0 \end{bmatrix}.$$
Here, 1 is the first diagonal entry, 3 is the second diagonal entry and 5 is the third and the last diagonal entry.

If all off-diagonal entries of $A$ are 0, then $A$ is said to be a **diagonal matrix**. Only a square matrix can be a diagonal matrix. There is a way to generalize this notion to any matrix, but we do not require it. Notice that the diagonal entries in a diagonal matrix need not all be nonzero. For example, the zero matrix of order $n$ is also a diagonal matrix. The following is a diagonal matrix. We follow the convention of not showing the 0 entries in a matrix.

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 3 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

We also write a diagonal matrix with diagonal entries $d_1, \ldots, d_n$ as $\text{diag}(d_1, \ldots, d_n)$. Thus the above diagonal matrix is also written as

\[\text{diag}(1, 3, 0)\].

The **identity matrix** is a square matrix of which each diagonal entry is 1 and each off-diagonal entry is 0.

\[I = \text{diag}(1, \ldots, 1)\].

When identity matrices of different orders are used in a context, we will use the notation $I_m$ for the identity matrix of order $m$.

We write $e_i$ for a column vector whose $i$th component is 1 and all other components 0. When we consider $e_i$ as a column vector in $\mathbb{F}^{n \times 1}$, the $j$th component of $e_i$ is $\delta_{ij}$. Here,

\[\delta_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{if } i \neq j
\end{cases}\]

is the Kroneker’s delta. Notice that the identity matrix $I = [\delta_{ij}]$.

There are then $n$ distinct column vectors $e_1, \ldots, e_n$. The list of column vectors $e_1, \ldots, e_n$ is called the **standard basis** for $\mathbb{F}^{n \times 1}$. Accordingly, the $e_i$s are referred to as the **standard basis vectors**. These are the columns of the identity matrix of order $n$, in that order; that is, $e_i$ is the $i$th column of $I$. The transposes of these $e_i$s are the rows of $I$. That is, the $i$th row of $I$ is $e_i^t$. Thus

\[I = \begin{bmatrix}
e_1 & \cdots & e_n\end{bmatrix} = \begin{bmatrix}
e_1^t \\
\vdots \\
e_n^t\end{bmatrix}.
\]

A **scalar matrix** is a matrix of which each diagonal entry is a scalar, the same scalar, and each off-diagonal entry is 0. Each scalar matrix is a diagonal matrix with same scalar on the diagonal. The following is a scalar matrix:

\[
\begin{bmatrix}
3 & 0 \\
0 & 3 \\
0 & 0
\end{bmatrix}
\]
It is also written as diag(3, 3, 3). If \( A, B \in \mathbb{F}^{m \times m} \) and \( A \) is a scalar matrix, then \( AB = BA \). Conversely, if \( A \in \mathbb{F}^{m \times m} \) is such that \( AB = BA \) for all \( B \in \mathbb{F}^{m \times m} \), then \( A \) must be a scalar matrix. This fact is not obvious, and its proof will require much more than discussed until now.

A matrix \( A \in \mathbb{F}^{m \times n} \) is said to be **upper triangular** iff all entries above the diagonal are zero. That is, \( A = [a_{ij}] \) is upper triangular when \( a_{ij} = 0 \) for \( i > j \). In writing such a matrix, we simply do not show the zero entries below the diagonal. Similarly, a matrix is called **lower triangular** iff all its entries above the diagonal are zero. Both upper triangular and lower triangular matrices are referred to as **triangular** matrices. A diagonal matrix is both upper triangular and lower triangular.

The following are examples of lower triangular matrix \( L \) and upper triangular matrix \( U \), both of order 3.

\[
L = \begin{bmatrix}
1 & 2 & 3 \\
2 & 3 & 4 \\
3 & 4 & 5
\end{bmatrix}, \quad U = \begin{bmatrix}
1 & 2 & 3 \\
3 & 4 & 5 \\
3 & 4 & 5
\end{bmatrix}.
\]

**Sum** of two matrices of the same size is a matrix whose entries are obtained by adding the corresponding entries in the given two matrices. That is, if \( A = [a_{ij}] \) and \( B = [b_{ij}] \) are in \( \mathbb{F}^{m \times n} \), then

\[
A + B = [a_{ij} + b_{ij}] \in \mathbb{F}^{m \times n}.
\]

For example,

\[
\begin{bmatrix}
1 & 2 & 3 \\
2 & 3 & 1 \\
3 & 1 & 2
\end{bmatrix} + \begin{bmatrix}
3 & 1 & 2 \\
2 & 1 & 3 \\
1 & 3 & 4
\end{bmatrix} = \begin{bmatrix}
4 & 3 & 5 \\
4 & 4 & 4 \\
4 & 8 & 0
\end{bmatrix}.
\]

We informally say that matrices are added entry-wise. Matrices of different sizes can never be added.

It then follows that \( A + B = B + A \).

Similarly, matrices can be **multiplied by a scalar** entry-wise. If \( A = [a_{ij}] \in \mathbb{F}^{m \times n} \), and \( \alpha \in \mathbb{F} \), then

\[
\alpha A = [\alpha a_{ij}] \in \mathbb{F}^{m \times n}.
\]

Therefore, a scalar matrix with \( \alpha \) on the diagonal is written as \( \alpha I \). Notice that

\[
A + 0 = 0 + A = A
\]

for all matrices \( A \in \mathbb{F}^{m \times n} \), with an implicit understanding that \( 0 \in \mathbb{F}^{m \times n} \). For \( A = [a_{ij}] \), the matrix \(-A \in \mathbb{F}^{m \times n}\) is taken as one whose \((ij)\)th entry is \(-a_{ij} \). Thus

\[
-A = (-1)A \quad \text{and} \quad A + (-A) = -A + A = 0.
\]

We also abbreviate \( A + (-B) \) to \( A - B \), as usual.

For example,

\[
3 \begin{bmatrix}
1 & 2 & 3 \\
2 & 3 & 1 \\
3 & 1 & 2
\end{bmatrix} - \begin{bmatrix}
3 & 1 & 2 \\
2 & 1 & 3 \\
1 & 3 & 4
\end{bmatrix} = \begin{bmatrix}
0 & 5 & 7 \\
4 & 8 & 0
\end{bmatrix}.
\]

The addition and scalar multiplication as defined above satisfy the following properties:

Let \( A, B, C \in \mathbb{F}^{m \times n} \). Let \( \alpha, \beta \in \mathbb{F} \).
1. \( A + B = B + A. \)
2. \( (A + B) + C = A + (B + C). \)
3. \( A + 0 = 0 + A = A. \)
4. \( A + (-A) = (-A) + A = 0. \)
5. \( \alpha(\beta A) = (\alpha \beta)A. \)
6. \( \alpha(A + B) = \alpha A + \alpha B. \)
7. \( (\alpha + \beta)A = \alpha A + \beta A. \)
8. \( 1A = A. \)

Notice that whatever we discuss here for matrices apply to row vectors and column vectors, in particular. But remember that a row vector cannot be added to a column vector unless both are of size \( 1 \times 1, \) when both become numbers in \( \mathbb{F}. \)

Another operation that we have on matrices is **multiplication of matrices**, which is a bit involved. Let \( A = [a_{ik}] \in \mathbb{F}^{m \times n} \) and \( B = [b_{kj}] \in \mathbb{F}^{n \times r}. \) Then their **product** \( AB \) is a matrix \([c_{ij}] \in \mathbb{F}^{m \times r}, \) where the entries are

\[
c_{ij} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.
\]

Notice that the matrix product \( AB \) is defined only when the number of columns in \( A \) is equal to the number of rows in \( B. \)

A particular case might be helpful. Suppose \( A \) is a row vector in \( \mathbb{F}^{1 \times n} \) and \( B \) is a column vector in \( \mathbb{F}^{n \times 1}. \) Then their product \( AB \in \mathbb{F}^{1 \times 1}; \) it is a matrix of size \( 1 \times 1. \) Often we will identify such matrices with numbers. The product now looks like:

\[
\begin{bmatrix}
a_1 & \cdots & a_n
\end{bmatrix}
\begin{bmatrix}
b_1 \\
\vdots \\
b_n
\end{bmatrix}
= [a_1b_1 + \cdots + a_nb_n]
\]

This is helpful in visualizing the general case, which looks like

\[
\begin{bmatrix}
a_{11} & a_{1k} & a_{1n} \\
a_{11} & \cdots & a_{ik} & \cdots & a_{1n} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{m1} & a_{mk} & a_{mn}
\end{bmatrix}
\begin{bmatrix}
b_{11} & b_{1j} & b_{1r} \\
\vdots & \vdots & \vdots \\
b_{\ell1} & b_{\ell j} & b_{\ell r} \\
\vdots & \vdots & \vdots \\
b_{n1} & b_{nj} & b_{nr}
\end{bmatrix}
= \begin{bmatrix}
c_{11} & c_{1j} & c_{1r} \\
c_{11} & \cdots & \cdots \\
c_{i1} & c_{ij} & c_{ir} \\
\vdots & \vdots & \vdots \\
c_{m1} & c_{mj} & c_{mr}
\end{bmatrix}
\]

The \( i \)th row of \( A \) multiplied with the \( j \)th column of \( B \) gives the \((ij)\)th entry in \( AB. \) Thus to get \( AB, \) you have to multiply all \( m \) rows of \( A \) with all \( r \) columns of \( B. \) Besides writing a linear system in compact form, we will see later why matrix multiplication is defined this way.
For example, 
\[
\begin{bmatrix}
3 & 5 & -1 \\
4 & 0 & 2 \\
-6 & -3 & 2
\end{bmatrix}
\begin{bmatrix}
2 & -2 & 3 & 1 \\
5 & 0 & 7 & 8 \\
9 & -4 & 1 & 1
\end{bmatrix}
= 
\begin{bmatrix}
22 & -2 & 43 & 42 \\
26 & -16 & 14 & 6 \\
-9 & 4 & -37 & -28
\end{bmatrix}
\]

If \( u \in \mathbb{F}^{1 \times n} \) and \( v \in \mathbb{F}^{n \times 1} \), then \( uv \in \mathbb{F}^{1 \times 1} \); but \( vu \in \mathbb{F}^{n \times n} \).

\[
\begin{bmatrix}
3 & 6 & 1 \\
4
\end{bmatrix}
\begin{bmatrix}
1 & 2 \\
2 & 3
\end{bmatrix}
= 
\begin{bmatrix}
19
\end{bmatrix},
\begin{bmatrix}
1 & 2 \\
2 & 3
\end{bmatrix}
\begin{bmatrix}
3 & 6 & 1 \\
4
\end{bmatrix}
= 
\begin{bmatrix}
3 & 6 & 1 \\
6 & 12 & 2 \\
12 & 24 & 4
\end{bmatrix}
\]

It shows clearly that matrix multiplication is not commutative. Commutativity can break down due to various reasons. First of all when \( AB \) is defined, \( BA \) may not be defined. Secondly, even when both \( AB \) and \( BA \) are defined, they may not be of the same size; and thirdly, even when they are of the same size, they need not be equal. For example,

\[
\begin{bmatrix}
1 & 2 \\
2 & 3
\end{bmatrix}
\begin{bmatrix}
0 & 1 \\
2 & 3
\end{bmatrix}
= 
\begin{bmatrix}
4 & 7 \\
6 & 11
\end{bmatrix}
\quad \text{but}\quad
\begin{bmatrix}
0 & 1 \\
2 & 3
\end{bmatrix}
\begin{bmatrix}
1 & 2 \\
2 & 3
\end{bmatrix}
= 
\begin{bmatrix}
2 & 3 \\
8 & 13
\end{bmatrix}
\]

It does not mean that \( AB \) is never equal to \( BA \). There can be some particular matrices \( A \) and \( B \) both in \( \mathbb{F}^{m \times n} \) such that \( AB = BA \).

If \( A \in \mathbb{F}^{m \times n} \), then \( AI_n = A \) and \( I_mA = A \).

Unlike numbers, product of two nonzero matrices can be a zero matrix. For example,

\[
\begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
0 & 0 \\
0 & 1
\end{bmatrix}
= 
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
\]

It is easy to verify the following properties of matrix multiplication:

1. If \( A \in \mathbb{F}^{m \times n} \), \( B \in \mathbb{F}^{n \times r} \) and \( C \in \mathbb{F}^{r \times p} \), then \( (AB)C = A(BC) \).
2. If \( A, B \in \mathbb{F}^{m \times n} \) and \( C \in \mathbb{F}^{n \times r} \), then \( (A + B)C = AB + AC \).
3. If \( A \in \mathbb{F}^{m \times n} \) and \( B, C \in \mathbb{F}^{n \times r} \), then \( A(B + C) = AB + AC \).
4. If \( \alpha \in \mathbb{F} \), \( A \in \mathbb{F}^{m \times n} \) and \( B \in \mathbb{F}^{n \times r} \), then \( \alpha(AB) = (\alpha A)B = A(\alpha B) \).

You can see matrix multiplication in a block form. Suppose \( A \in \mathbb{F}^{m \times n} \). Write its \( i \)th row as \( A_{i*} \). Also, write its \( k \)th column as \( A_{*k} \). Then we can write \( A \) as a row of columns and also as a column of rows in the following manner:

\[
A = [a_{ik}] = [A_{1*} \cdots A_{n*}] =
\begin{bmatrix}
A_{1*} \\
\vdots \\
A_{m*}
\end{bmatrix}.
\]

Write \( B \in \mathbb{F}^{n \times r} \) similarly as

\[
B = [b_{kj}] = [B_{*1} \cdots B_{*r}] =
\begin{bmatrix}
B_{1*} \\
\vdots \\
B_{n*}
\end{bmatrix}.
\]
Then their product $AB$ can now be written as
\[
AB = [AB_{*1} \cdots AB_{*r}] = \begin{bmatrix} A_{1*}B \\ \vdots \\ A_{m*}B \end{bmatrix}.
\]
When writing this way, we ignore the extra brackets $[ \text{ and } ]$.

**Powers** of square matrices can be defined inductively by taking
\[
A^0 = I \quad \text{and} \quad A^n = AA^{n-1} \quad \text{for } n \in \mathbb{N}.
\]
A square matrix $A$ of order $m$ is called **invertible** iff there exists a matrix $B$ of order $m$ such that
\[
AB = I = BA.
\]
Such a matrix $B$ is called an **inverse** of $A$. If $C$ is another inverse of $A$, then
\[
C = CI = C(AB) = (CA)B = IB = B.
\]
Therefore, an inverse of a matrix is unique and is denoted by $A^{-1}$. We talk of invertibility of square matrices only; and all square matrices are not invertible. For example, $I$ is invertible but $0$ is not.

If both $A, B \in \mathbb{F}^{n \times n}$ are invertible, then $(AB)^{-1} = B^{-1}A^{-1}$.

Reason?
\[
B^{-1}A^{-1}AB = B^{-1}IB = I = AIA^{-1} = ABB^{-1}A^{-1}.
\]

Invertible matrices play a crucial role in solving linear systems uniquely. We will come back to the issue later.

**Exercises for § 3.2**

1. Compute $AB, CA, \ DC, \ DCAB, \ A^2, \ D^2$ and $A^3B^2$, where
   \[
   A = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}, \ B = \begin{bmatrix} 4 & -1 \\ 2 & -1 \end{bmatrix}, \ C = \begin{bmatrix} -1 & 2 \\ 2 & -1 \\ 1 & 3 \end{bmatrix}, \ D = \begin{bmatrix} 3 & 2 & 1 \\ 4 & -6 & 0 \\ 1 & -2 & -2 \end{bmatrix}.
   \]

2. Let $E_{ij}$ be the $n \times n$ matrix whose $ij$th entry is 1 and all other entries are 0. Show that each $A = [a_{ij}] \in \mathbb{C}^{n \times n}$ can be expressed as $A = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} E_{ij}$. Also show that $E_{ij}E_{km} = 0$ if $j \neq k$, and $E_{ij}E_{jm} = E_{im}$.

3. Let $A \in \mathbb{C}^{m \times n}, \ B \in \mathbb{C}^{n \times p}$. Let $B_1, \ldots, B_p$ be the columns of $B$. Show that $AB_1, \ldots, AB_p$ are the columns of $AB$.

4. Let $A \in \mathbb{C}^{m \times n}, \ B \in \mathbb{C}^{n \times p}$. Let $A_1, \ldots, A_m$ be the rows of $A$. Show that $A_1B, \ldots, A_mB$ are the rows of $AB$.

5. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$. Show that $A^n = \begin{bmatrix} 1 & n & n(n-1) \\ 0 & 1 & 2n \\ 0 & 0 & 1 \end{bmatrix}$ for $n \in \mathbb{N}$.

6. Construct two $3 \times 3$ matrices $A$ and $B$ such that $AB = 0$ but $BA \neq 0$.
3.3 Transpose and adjoint

We consider another operation on matrices. Given a matrix \( A \in \mathbb{F}^{m \times n} \), its transpose is a matrix in \( \mathbb{F}^{n \times m} \), which is denoted by \( A^t \), and is defined by

the \((ij)\)th entry of \( A^t \) is the \((ji)\)th entry of \( A \).

That is, the \(i\)th column of \( A^t \) is the column vector \([a_{i1}, \cdots, a_{in}]^t\). The rows of \( A \) are the columns of \( A^t \) and the columns of \( A \) become the rows of \( A^t \). In particular, if \( u = [a_1 \cdots a_m] \) is a row vector, then its transpose is

\[
\begin{bmatrix}
  a_1 \\
  \vdots \\
  a_m
\end{bmatrix}
\]

which is a column vector, as mentioned earlier. Similarly, the transpose of a column vector is a row vector. If you write \( A \) as a row of column vectors, then you can express \( A^t \) as a column of row vectors, as in the following:

\[
A = \begin{bmatrix}
  A_{*1} & \cdots & A_{*n}
\end{bmatrix} \Rightarrow A^t = \begin{bmatrix}
  A_{*1}^t \\
  \vdots \\
  A_{*n}^t
\end{bmatrix}.
\]

For example,

\[
A = \begin{bmatrix}
  1 & 2 & 3 \\
  2 & 3 & 1
\end{bmatrix} \Rightarrow A^t = \begin{bmatrix}
  1 & 2 \\
  2 & 3 \\
  3 & 1
\end{bmatrix}.
\]

It then follows that transpose of the transpose is the original matrix. The following are some of the properties of this operation of transpose.

1. \((A^t)^t = A\).
2. \((A + B)^t = A^t + B^t\).
3. \((\alpha A)^t = \alpha A^t\).
4. \((AB)^t = B^t A^t\).
5. If \( A \) is invertible, then \( A^t \) is invertible, and \((A^t)^{-1} = (A^{-1})^t\).

In the above properties, we assume that the operations are allowed, that is, in (2), \( A \) and \( B \) must be of the same size. Similarly, in (4), the number of columns in \( A \) must be equal to the number of rows in \( B \); and in (5), \( A \) must be a square matrix.
It is easy to see all the above properties, except perhaps the fourth one. For this, let $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times r}$. Now, the $(ji)$th entry in $(AB)^t$ is the $(ij)$th entry in $AB$; and it is given by

$$a_{i1}b_{j1} + \cdots + a_{in}b_{jn}.$$ 

On the other side, the $(ji)$th entry in $B^tA^t$ is obtained by multiplying the $j$th row of $B^t$ with the $i$th column of $A^t$. This is same as multiplying the entries in the $j$th column of $B$ with the corresponding entries in the $i$th row of $A$, and then taking the sum. Thus it is

$$b_{j1}a_{i1} + \cdots + b_{jn}a_{in}.$$ 

This is the same as computed earlier.

The fifth one follows from the fourth one and the fact that $(AB)^{-1} = B^{-1}A^{-1}$.

Observe that transpose of a lower triangular matrix is an upper triangular matrix, and vice versa.

Close to the operations of transpose of a matrix is the adjoint. Let $A = [a_{ij}] \in \mathbb{F}^{m \times n}$. The adjoint of $A$ is denoted as $A^*$, and is defined by

the $(ij)$th entry of $A^* = \text{the complex conjugate of } (ji)$th entry of $A$.

We write $\overline{\alpha}$ for the complex conjugate of a scalar $\alpha$. That is, $\overline{\alpha + i\beta} = \alpha - i\beta$. Thus, if $a_{ij} \in \mathbb{R}$, then $\overline{a}_{ij} = a_{ij}$. Thus, when $A$ has only real entries, $A^* = A^t$. Also, the $i$th column of $A^t$ is the column vector $(\overline{a}_{i1}, \cdots, \overline{a}_{in})^t$. For example,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix} \Rightarrow A^* = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{bmatrix}.$$ 

$$A = \begin{bmatrix} 1+i & 2 & 3 \\ 2 & 3 & 1-i \end{bmatrix} \Rightarrow A^* = \begin{bmatrix} 1 - i & 2 \\ 2 & 3 \\ 3 & 1 + i \end{bmatrix}.$$ 

Similar to the transpose, the adjoint satisfies the following properties:

1. $(A^*)^* = A$.
2. $(A + B)^* = A^* + B^*$.
3. $(\alpha A)^* = \overline{\alpha}A^*$.
4. $(AB)^* = B^*A^*$.
5. If $A$ is invertible, then $A^*$ is invertible, and $(A^*)^{-1} = (A^{-1})^*$.

Here also, in (2), the matrices $A$ and $B$ must be of the same size, and in (4), the number of columns in $A$ must be equal to the number of rows in $B$. The adjoint of $A$ is also called the conjugate transpose of $A$. Notice that if $A \in \mathbb{R}^{m \times n}$, then $A^* = A^t$.

Occasionally, we will use $\overline{A}$ for the matrix obtained from $A$ by taking complex conjugate of each
entry. That is, the \((ij)\)th entry of \(\overline{A}\) is the complex conjugate of the \((ij)\)th entry of \(A\). Hence \(A^* = (\overline{A})^t\).

Further, the familiar dot product in \(\mathbb{R}^{1\times 3}\) can be generalized to \(\mathbb{F}^{1\times n}\) or to \(\mathbb{F}^{n\times 1}\). For vectors \(u, v \in \mathbb{F}^{1\times n}\), we define their **inner product** as

\[
\langle u, v \rangle = uv^*.
\]

For example, \(u = [1 \ 2 \ 3]\), \(v = [2 \ 1 \ 3]\) \(\Rightarrow \langle u, v \rangle = 1 \times 2 + 2 \times 1 + 3 \times 3 = 13\).

Similarly, for \(x, y \in \mathbb{F}^{n\times 1}\), we define their inner product as

\[
\langle x, y \rangle = y^* x.
\]

In case, \(\mathbb{F} = \mathbb{R}\), the \(x^*\) becomes \(x^t\). The inner product satisfies the following properties:

Let \(x, y, z \in \mathbb{F}^{n\times 1}\) (or in \(\mathbb{F}^{1\times n}\)); \(\alpha, \beta \in \mathbb{F}\).

1. \(\langle x, x \rangle \geq 0\).
2. \(\langle x, x \rangle = 0\) iff \(x = 0\).
3. \(\langle x, y \rangle = \overline{\langle y, x \rangle}\).
4. \(\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle\).
5. \(\langle z, x + y \rangle = \langle z, x \rangle + \langle z, y \rangle\).
6. \(\langle \alpha x, y \rangle = \alpha \langle x, y \rangle\).
7. \(\langle x, \beta y \rangle = \overline{\beta} \langle x, y \rangle\).

The inner product gives rise to the length of a vector as in the familiar case of \(\mathbb{R}^{1\times 3}\). We now call the generalized version of length as the **norm**. If \(u\) is in \(\mathbb{F}^{1\times n}\) or in \(\mathbb{F}^{n\times 1}\), we define its **norm**, denoted by \(\|u\|\) as the nonnegative square root of \(\langle u, u \rangle\). That is,

\[
\|u\| = \sqrt{\langle u, u \rangle}.
\]

The norm satisfies the following properties:

Let \(x, y \in \mathbb{F}^{1\times n}\) (or in \(\mathbb{F}^{n\times 1}\)); \(\alpha \in \mathbb{F}\).

1. \(\|x\| \geq 0\).
2. \(\|x\| = 0\) iff \(x = 0\).
3. \(\|\alpha x\| = |\alpha| \|x\|\).
4. \(|\langle x, y \rangle| \leq \|x\| \|y\|\). (*Cauchy-Schwartz inequality*)
5. \(\|x + y\| \leq \|x\| + \|y\|\). (*Triangle inequality*)
Using these properties, the acute angle between any two nonzero vectors can be defined. Let \( x, y \in \mathbb{F}^{1 \times n} \) (or in \( \mathbb{F}^{n \times 1} \)). The angle \( \theta \) between \( x \) and \( y \), denoted by \( \theta(x, y) \) is defined by

\[
\cos \theta(x, y) = \frac{|\langle x, y \rangle|}{\|x\| \|y\|}.
\]

In particular, when \( \theta(x, y) = \pi/2 \), we say that the vectors \( x \) and \( y \) are **orthogonal**, and we write this as \( x \perp y \). That is,

\[
x \perp y \iff \langle x, y \rangle = 0.
\]

Notice that this definition allows \( x \) and \( y \) to be zero vectors. Also, the zero vector is orthogonal to every vector.

It follows that if \( x \perp y \), then \( \|x\|^2 + \|y\|^2 = \|x + y\|^2 \). This is referred to as **Pythagoras law**. The converse of Pythagoras law holds when \( \mathbb{F} = \mathbb{R} \). For \( \mathbb{F} = \mathbb{C} \), it does not hold, in general.

**Exercises for § 3.3**

1. Determine \( A^t, \bar{A}, A^*, A^*A \) and \( AA^* \), where
   
   (a) \[
   A = \begin{bmatrix}
   -1 & 2 & 3 & 1 \\
   2 & -1 & 0 & 3 \\
   0 & -1 & -3 & 1
   \end{bmatrix}
   \]
   
   (b) \[
   A = \begin{bmatrix}
   1 & -2 + i & 3 - i \\
   i & -1 - i & 2i \\
   1 + 3i & -i & -3 \\
   -2 & 0 & -i
   \end{bmatrix}
   \]

2. Let \( u \) and \( v \) be the first and second rows of the matrix \( A \) in Exercise 1(a), and let \( x \) and \( y \) be the first and second columns of \( A \) in Exercise 1(b), respectively. Compute the inner products \( \langle u, v \rangle \) and \( \langle x, y \rangle \).

3. In \( \mathbb{C} \), consider the inner product \( \langle x, y \rangle = x\overline{y} \). Let \( x = 1 \) and \( y = i \) be two vectors in \( \mathbb{C} \). Show that \( \|x\|^2 + \|y\|^2 = \|x + y\|^2 \) but \( \langle x, y \rangle \neq 0 \).

4. In \( \mathbb{F}^{n \times 1} \), show that the parallelogram law holds. That is, for all \( x, y \in \mathbb{F}^{n \times 1} \), we have \( \|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \).

5. In \( \mathbb{F}^{1 \times n} \), derive the triangle inequality from Cauchy-Schwartz inequality.

6. Let \( A \in \mathbb{C}^{m \times n} \). Suppose \( AA^* = I_m \). Does it follow that \( A^*A = I_n \)?

**3.4 Elementary row operations**

Recall that while solving linear equations in two or three variables, you try to eliminate a variable from all but one equation by adding an equation to the other, or even adding a constant times one equation to another. We do similar operations on the rows of a matrix. These are achieved by multiplying a given matrix with some special matrices, called elementary matrices.

Let \( e_1, \ldots, e_m \in \mathbb{F}^{m \times 1} \) be the standard basis vectors. Let \( 1 \leq i, j \leq m \). The product \( e_i e_j^t \) is an \( m \times m \) matrix whose \( (i, j) \)th entry is 1 and all other entries are 0. We write such a matrix as \( E_{ij} \).
For instance, when \( m = 3 \), we have
\[
e_2e_3^t = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = E_{23}.
\]

An elementary matrix of order \( m \) is one of the following three types:

1. \( E[i, j] := I - E_{ii} - E_{jj} + E_{ij} + E_{ji} \) with \( i \neq j \).
2. \( E_\alpha[i] := I - E_{ii} + \alpha E_{ii} \), where \( \alpha \) is a nonzero scalar.
3. \( E_\alpha[i, j] := I + \alpha E_{ij} \), where \( \alpha \) is a nonzero scalar and \( i \neq j \).

Here, \( I \) is the identity matrix of order \( m \). Similarly, the order of the elementary matrices will be understood from the context; we will not show that in our symbolism.

**Example 3.1.** The following are instances of elementary matrices of order 3.
\[
E[1, 2] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_-1[2] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2[3, 1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}.
\]

We observe that for a matrix \( A \in \mathbb{F}^{m \times n} \), the following are true:

1. \( E[i, j] A \) is the matrix obtained from \( A \) by exchanging its \( i \)th and \( j \)th rows.
2. \( E_\alpha[i] A \) is the matrix obtained from \( A \) by replacing its \( i \)th row with \( \alpha \) times the \( i \)th row.
3. \( E_\alpha[i, j] A \) is the matrix obtained from \( A \) by replacing its \( i \)th row with the \( i \)th row plus \( \alpha \) times the \( j \)th row.

We call these operations of pre-multiplying a matrix with an elementary matrix as elementary row operations. Thus there are three kinds of elementary row operations as listed above. Sometimes, we will refer to them as of Type-1, 2, or 3, respectively. Also, in computations, we will write
\[
A \xrightarrow{E} B
\]

to mean that the matrix \( B \) has been obtained by an elementary row operation \( E \), that is, \( B = EA \).

**Example 3.2.** See the following applications of elementary row operations:
\[
\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \xrightarrow{E_{-3}[3, 1]} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{E_{-2}[2, 1]} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\]

Often we will apply elementary row operations in a sequence. In this way, the above operations could be shown in one step as \( E_{-3}[3, 1], E_{-2}[2, 1] \). However, remember that the result of application of this sequence of elementary row operations on a matrix \( A \) is \( E_{-2}[2, 1] E_{-3}[3, 1] A \); the products are in reverse order.
Exercises for § 3.4


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

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

2. Argue in general terms why the following are true:

   (a) $E[i, j]A$ is the matrix obtained from $A$ by exchanging its $i$th and $j$th rows.

   (b) $E_\alpha[i]A$ is the matrix obtained from $A$ by replacing its $i$th row with $\alpha$ times the $i$th row.

   (c) $E_\alpha[i, j]A$ is the matrix obtained from $A$ by replacing its $i$th row with the $i$th row plus $\alpha$ times the $j$th row.

3.5 Row reduced echelon form

Elementary operations can be used to reduce a matrix to a nice form, bringing in many zero entries. Recall that this corresponds to eliminating a variable from an equation of a linear system. The first, from left, nonzero entry in a nonzero row of a matrix is called a pivot. We denote a pivot in a row by putting a box around it. A column where a pivot occurs is called a pivotal column.

A matrix $A \in \mathbb{F}^{m \times n}$ is said to be in row reduced echelon form (RREF) iff the following conditions are satisfied:

(1) Each pivot is equal to 1.

(2) The column index of the pivot in the $(i + 1)$th row is greater than the column index of the pivot in the $i$th row, for all $i$ from 1 to $m - 1$.

(3) In a pivotal column, all entries other than the pivot are zero.

(4) All zero rows are at the bottom.

Example 3.3. The matrix

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

is in row reduced echelon form whereas the matrices

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

are not in row reduced echelon form.
Any matrix can be brought to a row reduced echelon form by using elementary row operations. We give an algorithm to achieve this.

**Reduction to Row Reduced Echelon Form**

1. Set the work region $R$ as the whole matrix $A$.

2. If all entries in $R$ are 0, then stop.

3. If there are nonzero entries in $R$, then find the leftmost nonzero column. Mark it as the pivotal column.

4. Find the topmost nonzero entry in the pivotal column. Box it; it is a pivot.

5. If the pivot is not on the top row of $R$, then exchange the row of $A$ which contains the top row of $R$ with the row where the pivot is.

6. If the pivot, say, $\alpha$ is not equal to 1, then replace the top row of $R$ in $A$ by $1/\alpha$ times that row.

7. Make all entries, except the pivot, in the pivotal column as zero by replacing each row above and below the top row of $R$ using elementary row operations in $A$ with that row and the top row of $R$.

8. Find the sub-matrix to the right and below the pivot. If no such sub-matrix exists, then stop. Else, reset the work region $R$ to this sub-matrix, and go to 2.

We will refer to the output of the above reduction algorithm as the *row reduced echelon form* (the RREF) of a given matrix.

**Example 3.4.**

\[
A = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 3 & 5 & 7 & 1 \\ 1 & 5 & 4 & 5 \\ 2 & 8 & 7 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 4 & 2 & 5 \\ 0 & 6 & 3 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & \frac{1}{2} \\ 0 & 4 & 2 & 5 \\ 0 & 6 & 3 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & \frac{1}{2} \\ 0 & 4 & 2 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B
\]

Here, $R1 = E_{-3}[2, 1], E_{-1}[3, 1], E_{-2}[4, 1]$; $R2 = E_{-1}[2, 1], E_{-4}[3, 2], E_{-6}[4, 2]$; and $R3 = E_{1/2}[1, 3], E_{-1/2}[2, 3], E_{-6}[4, 3]$. The matrix $B$ is the RREF of $A$. Notice that

\[
\]

The products are in reverse order.
Exercises for § 3.5

1. Compute row reduced echelon forms of the following matrices:

(a) \[
\begin{bmatrix}
0 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 0
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
2 & -1 & -1 & 0 \\
1 & 0 & 1 & -4 \\
0 & 1 & -1 & -4
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
1 & 2 & 1 & -1 \\
0 & 2 & 3 & 3 \\
1 & -1 & -3 & -4 \\
1 & 1 & 5 & -2
\end{bmatrix}
\]

2. Suppose that a matrix is already in RREF with r number of pivots. Show that the pivotal columns are the basis vectors $e_1, \ldots, e_r$ in that order, from left to right.

3. Suppose that a matrix is already in RREF with the r pivotal columns as $C_1, \ldots, C_r$. Is it true that each non-pivotal column can be expressed as $\alpha_1 C_1 + \cdots + \alpha_r C_r$ for some suitable scalars $\alpha_1, \ldots, \alpha_r$?

4. Argue why our algorithm for reducing a matrix to its RREF gives a unique output.

3.6 Determinant

There are two important quantities associated with a square matrix. One is the trace and the other is the determinant.

The sum of all diagonal entries of a square matrix is called the trace of the matrix. That is, if $A = [a_{ij}] \in \mathbb{F}^{n \times m}$, then

$$\text{tr}(A) = a_{11} + \cdots + a_{nn} = \sum_{k=1}^{n} a_{kk}.$$ 

In addition to $\text{tr}(I_m) = m$, $\text{tr}(0) = 0$, the trace satisfies the following properties:

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

1. $\text{tr}(\alpha A) = \alpha \text{tr}(A)$ for each $\alpha \in \mathbb{F}$.

2. $\text{tr}(A^t) = \text{tr}(A)$ and $\text{tr}(A^*) = \overline{\text{tr}(A)}$.

3. $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$ and $\text{tr}(AB) = \text{tr}(BA)$.

4. $\text{tr}(A^* A) = 0$ iff $\text{tr}(AA^*) = 0$ iff $A = 0$.

Observe that $\text{tr}(A^* A) = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 = \text{tr}(AA^*)$. Form this (4) follows.

The second quantity, called the determinant of a square matrix $A = [a_{ij}] \in \mathbb{F}^{n \times n}$, written as $\det(A)$, is defined inductively as follows:

If $n = 1$, then $\det(A) = a_{11}$.

If $n > 1$, then

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

where the matrix $A_{1j} \in \mathbb{F}^{(n-1) \times (n-1)}$ is obtained from $A$ by deleting the first row and the $j$th column of $A$.  

66
When \( A = [a_{ij}] \) is written showing all its entries, we also write \( \det(A) \) by replacing the two big closing brackets \( [ \) and \( ] \) by two vertical bars \( | \) and \( | \). For a \( 2 \times 2 \) matrix, its determinant is seen as follows:

\[
\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (-1)^{1+1}a_{11}\det[a_{22}] + (-1)^{1+2}a_{12}\det[a_{21}] = a_{11}a_{22} - a_{12}a_{21}.
\]

Similarly, for a \( 3 \times 3 \) matrix, we need to compute three \( 2 \times 2 \) determinants. For example,

\[
\det \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{vmatrix} = (-1)^{1+1} \times 1 \times \begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} + (-1)^{1+2} \times 2 \times \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} + (-1)^{1+3} \times 3 \times \begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}
\]

\[
= 1 \times \begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} - 2 \times \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} + 3 \times \begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}
\]

\[
= (3 \times 2 - 1 \times 1) - 2 \times (2 \times 2 - 1 \times 3) + 3 \times (2 \times 1 - 3 \times 3)
\]

\[
= 5 - 2 \times 1 + 3 \times (-7) = -18.
\]

For a lower triangular matrix, we see that

\[
\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & \cdots & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{33} & \cdots & \cdots & \cdots \\ a_{23} & \cdots & \cdots & \ddots & \ddots \\ a_{33} & \cdots & \cdots & \ddots & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots \\ a_{nn} \end{vmatrix} = \cdots = a_{11}a_{22}\cdots a_{nn}.
\]

In general, the determinant of any triangular matrix (upper or lower), is the product of its diagonal entries. In particular, the determinant of a diagonal matrix is also the product of its diagonal entries. Thus, if \( I \) is the identity matrix of order \( n \), then \( \det(I) = 1 \) and \( \det(-I) = (-1)^n \).

Our definition of determinant expands the determinant in the first row. In fact, the same result may be obtained by expanding it in any other row, or even any other column. Along with this, some more properties of the determinant are listed in the following.

Let \( A \in \mathbb{F}^{n \times n} \). The sub-matrix of \( A \) obtained by deleting the \( i \)th row and the \( j \)th column is called the \((ij)\)th \textbf{minor} of \( A \), and is denoted by \( A_{ij} \). The \((ij)\)th \textbf{co-factor} of \( A \) is \((-1)^{i+j}\det(A_{ij})\); it is denoted by \( C_{ij}(A) \). Sometimes, when the matrix \( A \) is fixed in a context, we write \( C_{ij}(A) \) as \( C_{ij} \).

The \textbf{adjugate} of \( A \) is the \( n \times n \) matrix obtained by taking transpose of the matrix whose \((ij)\)th entry is \( C_{ij}(A) \); it is denoted by \( \text{adj}(A) \). That is, \( \text{adj}(A) \in \mathbb{F}^{n \times n} \) is the matrix whose \((ij)\)th co-factor is \( C_{ji}(A) \). Also, we write \( A_i(x) \) for the matrix obtained from \( A \) by replacing its \( i \)th row by a row vector \( x \) of appropriate size.

Let \( A \in \mathbb{F}^{n \times n} \). Let \( i, j, k \in \{1, \ldots, n\} \). Let \( E[i, j] \), \( E_\alpha[i] \) and \( E_\alpha[i, j] \) be the elementary matrices of order \( n \) with \( 1 \leq i \neq j \leq n \) and \( \alpha \neq 0 \), a scalar. Then the following statements are true.
1. \( \det(E[i,j] A) = -\det(A) \).

2. \( \det(E_\alpha[i] A) = \alpha \det(A) \).

3. \( \det(E_\alpha[i,j] A) = \det(A) \).

4. If some row of \( A \) is the zero vector, then \( \det(A) = 0 \).

5. If one row of \( A \) is a scalar multiple of another row, then \( \det(A) = 0 \).

6. For any \( i \in \{1, \ldots, n\} \), \( \det(A_i(x + y)) = \det(A_i(x)) + \det(A_i(y)) \).

7. \( \det(A^t) = \det(A) \).

8. If \( A \) is a triangular matrix, then \( \det(A) \) is equal to the product of the diagonal entries of \( A \).

9. \( \det(AB) = \det(A) \det(B) \) for any matrix \( B \in \mathbb{F}^{n \times n} \).

10. \( \det(A^t) = \det(A) \).

11. \( A \text{adj}(A) = \text{adj}(A)A = \det(A)I \).

12. \( A \) is invertible iff \( \det(A) \neq 0 \).

Elementary column operations are operations similar to row operations, but with columns instead of rows. Notice that since \( \det(A^t) = \det(A) \), the facts concerning elementary row operations also hold true if elementary column operations are used. Using elementary operations, the computational complexity for evaluating a determinant can be reduced drastically. The trick is to bring a matrix to a triangular form by using elementary row operations, so that the determinant of the triangular matrix can be computed easily.

**Example 3.5.**

\[
\begin{bmatrix}
1 & 0 & 0 & 1 \\
-1 & 1 & 0 & 1 \\
-1 & -1 & 1 & 1 \\
-1 & -1 & -1 & 1
\end{bmatrix} \xrightarrow{R_1} \begin{bmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 2 \\
0 & -1 & 1 & 2 \\
0 & -1 & -1 & 2
\end{bmatrix} \xrightarrow{R_2} \begin{bmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 2 \\
0 & 0 & 1 & 4 \\
0 & 0 & 1 & 4
\end{bmatrix} \xrightarrow{R_3} \begin{bmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 2 \\
0 & 0 & 1 & 4 \\
0 & 0 & 0 & 8
\end{bmatrix} = 8.
\]

Here, \( R_1 = E_1[2, 1]; E_1[3, 1]; E_1[4, 1] \), \( R_2 = E_1[3, 2]; E_1[4, 2] \), and \( R_3 = E_1[4, 3] \).

Finally, the upper triangular matrix has the required determinant.

**Example 3.6.** See that the following is true, for verifying Property (6) as mentioned above:

\[
\begin{bmatrix}
3 & 1 & 2 & 4 \\
-1 & 1 & 0 & 1 \\
-1 & -1 & 1 & 1 \\
-1 & -1 & -1 & 1
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 1 \\
-1 & 1 & 0 & 1 \\
-1 & -1 & 1 & 1 \\
-1 & -1 & -1 & 1
\end{bmatrix} + \begin{bmatrix}
2 & 1 & 2 & 3 \\
-1 & 1 & 0 & 1 \\
-1 & -1 & 1 & 1 \\
-1 & -1 & -1 & 1
\end{bmatrix}.
\]
Exercises for § 3.6

1. Construct an $n \times n$ nonzero matrix, where no row is a scalar multiple of another row but its determinant is 0.

2. Let $A \in \mathbb{C}^{n \times n}$. Show that if $\text{tr}(A^* A) = 0$, then $A = 0$.

3. Let $a_1, \ldots, a_n \in \mathbb{C}$. Let $A$ be the $n \times n$ matrix whose first row has all entries as 1 and whose $k$th row has entries $a_1^{k-1}, \ldots, a_n^{k-1}$ in that order. Show that $\det(A) = \prod_{i<j} (a_i - a_j)$.

4. Let $A$ be an $n \times n$ matrix with integer entries. Prove that if $\det(A) = \pm 1$, then $A^{-1}$ has only integer entries.

5. Determine $\text{adj}(A)$ and $A^{-1}$ using the adjugate, where $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$.

3.7 Computing inverse of a matrix

The adjugate property of the determinant provides a way to compute the inverse of a matrix, provided it is invertible. However, it is very inefficient. We may use elementary row operations to compute the inverse. Our computation of the inverse bases on the following fact.

**Theorem 3.1.** A square matrix is invertible iff it is a product of elementary matrices.

**Proof:** Each elementary matrix is invertible since $E[i,j]$ is its own inverse, $E_{1/\alpha}[i]$ is the inverse of $E_\alpha[i]$, and $E_{-\alpha}[i,j]$ is the inverse of $E_\alpha[i,j]$. Therefore, any product of elementary matrices is invertible.

Conversely, suppose that $A$ is an invertible matrix. Let $EA^{-1}$ be the RREF of $A^{-1}$. If $EA^{-1}$ has a zero row, then $EA^{-1}A$ also has a zero row. That is, $E$ has a zero row. But $E$ is a product of elementary matrices, which is invertible; it does not have a zero row. Therefore, $EA^{-1}$ does not have a zero row. Then each row in the square matrix $EA^{-1}$ has a pivot. But the only square matrix in RREF having a pivot at each row is the identity matrix. Therefore, $EA^{-1} = I$. That is, $A = E$, a product of elementary matrices. \[\square\]

The computation of inverse will be easier if we write the matrix $A$ and the identity matrix $I$ side by side and apply the elementary operations on both of them simultaneously. For this purpose, we introduce the notion of an augmented matrix.

If $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{m \times k}$, then the matrix $[A|B] \in \mathbb{F}^{m \times (n+k)}$ obtained from $A$ and $B$ by writing first all the columns of $A$ and then the columns of $B$, in that order, is called an augmented matrix. The vertical bar shows the separation of columns of $A$ and of $B$, though, conceptually unnecessary. For computing the inverse of a matrix, start with the augmented matrix $[A|I]$. Apply elementary row operations for reducing $A$ to its row reduced echelon form, while simultaneously applying the
same operations on the entries of $I$. This means we pre-multiply the matrix $[A|I]$ with a product $B$ of elementary matrices. In block form, our result is the augmented matrix $[BA|BI]$. If $BA = I$, then $BI = A^{-1}$. That is, the part that contained $I$ originally will give the matrix $A^{-1}$ after the elementary row operations have been applied. If after row reduction, it turns out that $B \neq I$, then $A$ is not invertible; this information is a bonus.

**Example 3.7.** For illustration, consider the following square matrices:

$$
A = \begin{bmatrix}
1 & -1 & 2 & 0 \\
-1 & 0 & 0 & 2 \\
2 & 1 & -1 & -2 \\
1 & -2 & 4 & 2 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
1 & -1 & 2 & 0 \\
-1 & 0 & 0 & 2 \\
2 & 1 & -1 & -2 \\
0 & -2 & 0 & 2 \\
\end{bmatrix}.
$$

We want to find the inverses of the matrices, if at all they are invertible.

Augment $A$ with an identity matrix to get

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

Use elementary row operations. Since $a_{11} = 1$, we leave row(1) untouched. To zero-out the other entries in the first column, we use the sequence of elementary row operations $E_1[2,1]$, $E_{-2}[3,1]$, $E_{-1}[4,1]$ to obtain

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

The pivot is $-1$ in $(2,2)$ position. Use $E_{-1}[2]$ to make the pivot 1.

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

Use $E_1[1,2]$, $E_{-3}[3,2]$, $E_1[4,2]$ to zero-out all non-pivot entries in the pivotal column to 0:

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

Since a zero row has appeared in the $A$ portion of the augmented matrix, we conclude that $A$ is not invertible. You see that the second portion of the augmented matrix has no meaning now. However,
it records the elementary row operations which were carried out in the reduction process. Verify that this matrix is equal to


and that the first portion is equal to this matrix times \( A \).

For \( B \), we proceed similarly. The augmented matrix \([B|I]\) with the first pivot looks like:

\[
\begin{bmatrix}
1 & -1 & 2 & 0 \\
-1 & 0 & 0 & 2 \\
2 & 1 & -1 & -2 \\
0 & -2 & 0 & 2
\end{bmatrix}
\]

The sequence of elementary row operations \( E_1[2, 1]; E_{-2}[3, 1] \) yields

\[
\begin{bmatrix}
1 & -1 & 2 & 0 \\
0 & -1 & 2 & 2 \\
0 & 3 & -5 & -2 \\
0 & -2 & 0 & 2
\end{bmatrix}
\]

Next, the pivot is \(-1\) in \((2, 2)\) position. Use \( E_{-1}[2] \) to get the pivot as \(1\).

\[
\begin{bmatrix}
1 & -1 & 2 & 0 \\
0 & 1 & 0 & 6 \\
0 & -2 & 0 & 2 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

Next pivot is \(1\) in \((3, 3)\) position. Now, \( E_2[2, 3]; E_3[4, 3] \) produces

\[
\begin{bmatrix}
1 & 0 & 0 & -2 \\
0 & 1 & 0 & 6 \\
0 & 0 & 1 & 4 \\
0 & 0 & 0 & 14
\end{bmatrix}
\]

Next pivot is \(14\) in \((4, 4)\) position. Use \([4; 1/14] \) to get the pivot as \(1\):

\[
\begin{bmatrix}
1 & 0 & 0 & -2 \\
0 & 1 & 0 & 6 \\
0 & 0 & 1 & 4 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

Use \( E_2[1, 4]; E_{-6}[2, 4]; E_{-4}[3, 4] \) to zero-out the entries in the pivotal column:
Thus \( B^{-1} = \frac{1}{7} \begin{bmatrix} 2 & 3 & 4 & 1 \\ 1 & 5 & 2 & -3 \\ 3 & 1 & -1 & -2 \\ 1 & 5 & 2 & \frac{1}{2} \end{bmatrix} \). Verify that \( B^{-1}B = BB^{-1} = I \).

Observe that if a matrix is not invertible, then our algorithm for reduction to RREF produces a pivot in the \( I \) portion of the augmented matrix.

**Exercises for § 3.7**

1. Compute the inverses of the following matrices, if possible:
   
   (a) \( \begin{bmatrix} 2 & 1 & 2 \\ 1 & 3 & 1 \\ -1 & 1 & 2 \end{bmatrix} \)  
   (b) \( \begin{bmatrix} 1 & 4 & -6 \\ -1 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix} \)  
   (c) \( \begin{bmatrix} 3 & 1 & 1 & 2 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 2 & -1 \\ -2 & 1 & -1 & 3 \end{bmatrix} \)

2. Let \( A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -b & -c \end{bmatrix} \), where \( b, c \in \mathbb{C} \). Show that \( A^{-1} = bI + cA \).

3. Show that if a matrix \( A \) is upper triangular and invertible, then so is \( A^{-1} \).

4. Show that if a matrix \( A \) is lower triangular and invertible, then so is \( A^{-1} \).

5. Show that every \( n \times n \) matrix can be written as a sum of two invertible matrices.

6. Show that every \( n \times n \) invertible matrix can be written as a sum of two non-invertible matrices.
Chapter 4

Rank and Linear Equations

4.1 Matrices as linear maps

Let $A \in \mathbb{F}^{m \times n}$. We may view the matrix $A$ as a function from $\mathbb{F}^{n \times 1}$ to $\mathbb{F}^{m \times 1}$. It goes as follows. Let $x \in \mathbb{F}^{n \times 1}$. Then define the matrix $A$ as a function $A : \mathbb{F}^{n \times 1} \to \mathbb{F}^{m \times 1}$ by

$$A(x) = Ax.$$ 

That is, the value of the function $A$ at any vector $x \in \mathbb{F}^{n \times 1}$ is the vector $Ax$ in $\mathbb{F}^{m \times 1}$.

Since the matrix product $Ax$ is well defined, such a function is meaningful. We see that due to the properties of matrix product, the following are true:

1. $A(u + v) = A(u) + A(v)$ for all $u, v \in \mathbb{F}^{n \times 1}$.
2. $A(\alpha v) = \alpha A(v)$ for all $v \in \mathbb{F}^{n \times 1}$ and for all $\alpha \in \mathbb{F}$.

In this manner a matrix is considered as a linear map. A linear map is any function from $\mathbb{F}^{n \times 1}$ to $\mathbb{F}^{m \times 1}$ satisfying the above two properties. To see the connection between the matrix as a rectangular array and as a function, consider the values of the matrix $A$ at the standard basis vectors $e_1, \ldots, e_n$ in $\mathbb{F}^{n \times 1}$. Recall that $e_j$ is a column vector in $\mathbb{F}^{n \times 1}$ where the $j$th entry is 1 and all other entries are 0. Let $A = [a_{ij}] \in \mathbb{F}^{m \times n}$. We see that

$$Ae_j = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} = j\text{th column of } A.$$ 

A matrix $A \in \mathbb{F}^{m \times n}$ is viewed as the linear map $A : \mathbb{F}^{n \times 1} \to \mathbb{F}^{m \times 1}$, where $A(e_j)$ is the $j$th column of $A$, and $A(v) = Av$ for each $v \in \mathbb{F}^{n \times 1}$.

The range of the matrix $A$ (of the linear map $A$) is the set $R(A) = \{Ax : x \in \mathbb{F}^{n \times 1}\}$. Now, each vector $x = [\alpha_1, \ldots, \alpha_n]^t \in \mathbb{F}^{n \times 1}$ can be written as

$$x = \alpha_1 e_1 + \cdots + \alpha_n e_n.$$
If \( y \in R(T) \), then there exists an \( x \in F^{n \times 1} \) such that \( y = Ax \). Then such a \( y \) can be written as
\[
y = Ax = \alpha_1 A e_1 + \cdots + \alpha_n A e_n.
\]
Conversely we see that each vector \( \alpha_1 A e_1 + \cdots + \alpha_n A e_n \) is in \( R(A) \). Since \( A e_j \) is the \( j \)th column of \( A \), we find that
\[
R(A) = \{ \alpha_1 A_1 + \cdots + \alpha_n A_n : a_1, \ldots, a_n \in F \},
\]
where \( A_1, \ldots, A_n \) are the \( n \) columns of \( A \).

**Exercises for § 4.1**

1. Let \( A \in C^{m \times n} \). Define \( T : C^{1 \times m} \to C^{1 \times n} \) by \( T(x) = xA \) for \( x \in C^{1 \times m} \). Show that \( T \) is a linear map. Identify \( T(e_j) \), where \( e_j \) is a vector in \( C^{1 \times m} \) whose \( j \)th component is 1 and all other components are 0.

2. Define \( T : R^{3 \times 1} \to R^{2 \times 1} \) by \( T([a, b, c]^t) = [c, b + a]^t \). Show that \( T \) is a linear map. Find a matrix \( A \in R^{2 \times 3} \) such that \( T([a, b, c]^t) = A[a, b, c]^t \).

### 4.2 Linear independence

We give a name to an expression of the type we have seen in the last section.

If \( v_1, \ldots, v_m \in F^{1 \times n} \), (or in \( F^{n \times 1} \)) then the vector
\[
\alpha_1 v_1 + \cdots + \alpha_m v_m
\]

is called a linear combination of \( v_1, \ldots, v_m \), where \( \alpha_1, \ldots, \alpha_m \in F \) are some scalars.

For example, in \( F^{1 \times 2} \), one linear combination of \( v_1 = [1, 1] \) and \( v_2 = [1, -1] \) is as follows:
\[
2[1, 1] + 1[1, -1].
\]
This linear combination evaluates to \([3, 1]\). Thus \([3, 1]\) is a linear combination of \( v_1, v_2 \).

Is \([4, -2]\) a linear combination of \( v_1 \) and \( v_2 \)? Yes, since
\[
[4, -1] = 1[1, 1] + 3[1, -1].
\]
In fact, every vector in \( F^{1 \times 2} \) is a linear combination of \( v_1 \) and \( v_2 \). Reason:
\[
[a, b] = \frac{a+b}{2} [1, 1] + \frac{a-b}{2} [1, -1].
\]

However, every vector in \( F^{1 \times 2} \) is not a linear combination of \([1, 1]\) and \([2, 2]\). Reason? Any linear combination of these two vectors is a multiple of \([1, 1]\). Then \([1, 0]\) is not a linear combination of these two vectors.

The vectors \( v_1, \ldots, v_m \) in \( F^{n \times 1} \) are called linearly dependent iff at least one of them is a linear combination of others. The vectors are called linearly independent iff none of them is a linear combination of others.
For example, \([1, 1], [1, -1], [4, -1]\) are linearly dependent vectors whereas \([1, 1], [1, -1]\) are linearly independent vectors.

Notice that if \(\alpha_1 = \cdots = \alpha_m = 0\), then obviously, the linear combination \(\alpha_1 v_1 + \cdots + \alpha_m v_m\) evaluates to 0. That is, the zero vector can always be written as a trivial linear combination.

Suppose the vectors \(v_1, \ldots, v_m\) are linearly dependent. Then one of them, say, \(v_i\) is a linear combination of others. That is,

\[
v_i = \alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} + \alpha_{i+1} v_{i+1} + \cdots + \alpha_m v_m.
\]

Then

\[
\alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} + (-1)v_i + \alpha_{i+1} v_{i+1} + \cdots + \alpha_m v_m = 0.
\]

Here, we see that a linear combination becomes zero, where at least one of the coefficients, that is, the \(i\)th one is nonzero.

Conversely, suppose that we have scalars \(\beta_1, \ldots, \beta_m\) not all zero such that

\[
\beta_1 v_1 + \cdots + \beta_m v_m = 0.
\]

Suppose that the \(k\)th scalar \(\beta_k\) is nonzero. Then

\[
v_k = \frac{-1}{\beta_k} \left( \beta_1 v_1 + \cdots + \beta_{k-1} v_{k-1} + \beta_{k+1} v_{k+1} + \cdots + \beta_m v_m \right).
\]

That is, the vectors \(v_1, \ldots, v_m\) are linearly dependent.

Thus we have proved the following:

\(v_1, \ldots, v_m\) are linearly dependent iff \(\alpha_1 v_1 + \cdots + \alpha_m v_m = 0\) for scalars \(\alpha_1, \ldots, \alpha_m\) not all zero.

The same may be written in terms of linear independence.

**Theorem 4.1.** The vectors \(v_1, \ldots, v_m \in \mathbb{F}^{1 \times n}\) are linearly independent iff for all \(\alpha_1, \ldots, \alpha_m \in \mathbb{F}\),

\[
\alpha_1 v_1 + \cdots + \alpha_m v_m = 0 \text{ implies that } \alpha_1 = \cdots = \alpha_m = 0.
\]

The same is true when \(v_1, \ldots, v_n \in \mathbb{F}^{n \times 1}\).

Theorem 4.1 provides a way to determine whether a finite number of vectors are linearly independent or not. You start with a linear combination of the given vectors; and equate it to 0. Then you must be able to derive that each coefficient in that linear combination is 0. If this is the case, then the given vectors are linearly independent. If it is not possible, then from its proof you must be able to find a way of expressing one of the vectors as a linear combination of the others, showing that the vectors are linearly dependent.

**Example 4.1.** Are the vectors \([1, 1, 1], [2, 1, 1], [3, 1, 0]\) linearly independent?

We start with an arbitrary linear combination and equate it to the zero vector. Solve the resulting linear equations to determine whether all the coefficients are necessarily 0 or not.

So, let

\[
a[1, 1, 1] + b[2, 1, 1] + c[3, 1, 0] = [0, 0, 0].
\]
Comparing the components, we have
\[ a + 2b + 3c = 0, \quad a + b + c = 0, \quad a + b = 0. \]
The last two equations imply that \( c = 0 \). Substituting in the first, we see that
\[ a + 2b = 0. \]
This and the equation \( a + b = 0 \) give \( b = 0 \). Then it follows that \( a = 0 \).
We conclude that the given vectors are linearly independent.

**Example 4.2.** Are the vectors \([1, 1, 1], [2, 1, 1], [3, 2, 2] \) linearly independent?

Clearly, the third one is the sum of the first two. So, the given vectors are linearly dependent. To illustrate our method, we start with an arbitrary linear combination and equate it to the zero vector. Solve the resulting linear equations to determine whether all the coefficients are necessarily 0 or not.

So, as earlier, let
\[ a[1, 1, 1] + b[2, 1, 1] + c[3, 2, 2] = [0, 0, 0]. \]
Comparing the components, we have
\[ a + 2b + 3c = 0, \quad a + b + 2c = 0, \quad a + b + 2c = 0. \]
The last equation is redundant. Subtracting the second from the first, we have
\[ b + c = 0. \]
We may choose \( b = 1, \ c = -1 \) to satisfy this equation. Then from the second equation, we have \( a = 1 \). Then our starting equation says that
\[ 1[1, 1, 1] + 1[2, 1, 1] + (-1)[3, 2, 2] = [0, 0, 0]. \]
That is, the third vector is the sum of the first two.

Be careful with the direction of implication here. Your work-out must be in the form
\[ \alpha_1 v_1 + \cdots + \alpha_m v_m = 0 \Rightarrow \cdots \Rightarrow \alpha_1 = \cdots = \alpha_m = 0. \]
And that would prove linear independence.

To see how linear independence is helpful, consider the following system of linear equations:
\[
\begin{align*}
x_1 + 2x_2 - 3x_3 &= 2 \\
2x_1 - x_2 + 2x_3 &= 3 \\
4x_1 + 3x_2 - 4x_3 &= 7
\end{align*}
\]
Here, we find that the third equation is redundant, since 2 times the first plus the second gives the third. That is, the third one linearly depends on the first two. (You can of course choose any other
Now, take the row vectors of coefficients of the unknowns as in the following:

\[ v_1 = [1, 2, -3, 2], \quad v_2 = [2, -1, 2, 3], \quad v_3 = [4, 3, -4, 7]. \]

We see that \( v_3 = 2v_1 + v_2 \), as it should be. We see that the vectors \( v_1, v_2, v_3 \) are linearly dependent. But the vectors \( v_1, v_2 \) are linearly independent. Thus, solving the given system of linear equations is the same thing as solving the system with only first two equations. For solving linear systems, it is of primary importance to find out which equations linearly depend on others. Once determined, such equations can be thrown away, and the rest can be solved.

**Exercises for § 4.2**

1. Determine whether the given vectors are linearly independent, in each case:
   
   (a) \([1, 2, 6], [-1, 3, 4], [-1, -4, 2]\) in \(\mathbb{R}^{1\times 3}\).
   
   (b) \([1, 0, 2, 1], [1, 3, 2, 1], [4, 1, 2, 2]\) in \(\mathbb{C}^{1\times 4}\).

2. Suppose that \(u, v, w\) are linearly independent in \(\mathbb{C}^{1\times 5}\). Are the following lists of vectors linearly independent?
   
   (a) \(u, v + \alpha w, w\), where \(\alpha\) is a nonzero complex number.
   
   (b) \(u + v, v + w, w + u\).
   
   (c) \(u - v, v - w, w - u\).

### 4.3 Gram-Schmidt orthogonalization

It is easy to see that if the nonzero vectors \(v_1, \ldots, v_n\) are orthogonal, then they are also linearly independent. How?

Suppose \(v_1, \ldots, v_n\) are nonzero orthogonal vectors. Assume that

\[ \alpha_1 v_1 + \cdots + \alpha_n v_n = 0. \]

Take inner product with \(v_j\) for \(j \in \{1, \ldots, n\}\). For \(i \neq j\), \(\langle v_i, v_j \rangle = 0\). So,

\[ \alpha_j \langle v_j, v_j \rangle = 0. \]

But \(v_j \neq 0\) implies that \(\langle v_j, v_j \rangle \neq 0\). Therefore, \(\alpha_j = 0\). That is,

\[ \alpha_1 = \cdots = \alpha_n = 0. \]

Therefore, the vectors \(v_1, \ldots, v_n\) are linearly independent.

Conversely, given \(n\) linearly independent vectors \(v_1, \ldots, v_n\) (necessarily all nonzero), we can orthogonalize them. If \(v_1, \ldots, v_k\) are linearly independent but \(v_1, \ldots, v_k, v_{k+1}\) are linearly dependent, then we will see that our orthogonalization process will yield the \((k+1)\)th vector as the zero vector.

We now discuss this method, called *Gram-Schmidt orthogonalization*. 
Given two linearly independent vectors \( u_1, u_2 \) on the plane how do we construct two orthogonal vectors?

Keep \( v_1 = u_1 \). Take out the projection of \( u_2 \) on \( u_1 \) to get \( v_2 \). Now, \( v_2 \perp v_1 \).

What is the projection of \( u_2 \) on \( u_1 \)?

Its length is \( \langle u_2, u_1 \rangle \). Its direction is that of \( u_1 \). Thus taking \( v_1 = u_1 \) and \( v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 \) does the job. You can now verify that \( \langle v_2, v_1 \rangle = 0 \).

We may continue this process of taking away projections in \( \mathbb{R}^{n \times 1} \), or in \( \mathbb{R}^{1 \times n} \). It results in the following process.

**Theorem 4.2.** (Gram-Schmidt orthogonalization) Let \( u_1, u_2, \ldots, u_n \) be linearly independent vectors. Define

\[
\begin{align*}
v_1 &= u_1 \\
v_2 &= u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 \\
&\quad \vdots \\
v_{n+1} &= u_{n+1} - \frac{\langle u_{n+1}, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \cdots - \frac{\langle u_{n+1}, v_n \rangle}{\langle v_n, v_n \rangle} v_n
\end{align*}
\]

Then \( v_1, v_2, \ldots, v_n \) are orthogonal and the set of linear combinations of \( v_1, v_2, \ldots, v_n \) is equal to the set of linear combinations of \( u_1, u_2, \ldots, u_n \).

**Proof:** \( \langle v_2, v_1 \rangle = \langle u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1, v_1 \rangle = \langle u_2, v_1 \rangle - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} \langle v_1, v_1 \rangle = 0 \). Use Induction. \( \square \)

Observe that if \( u_1, \ldots, u_k \) are linearly independent but \( u_1, \ldots, u_{k+1} \) are linearly dependent, then Gram-Schmidt process will compute nonzero orthogonal vectors \( v_1, \ldots, v_k \) and it will give \( v_{k+1} \) as the zero vector.

**Example 4.3.** The vectors \( u_1 = [1, 0, 0], u_2 = [1, 1, 0], u_3 = [1, 1, 1] \) are linearly independent in \( \mathbb{R}^{1 \times 3} \). Apply Gram-Schmidt Orthogonalization.

\[
\begin{align*}
v_1 &= [1, 0, 0]. \\
v_2 &= u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = [1, 1, 0] - 1 [1, 0, 0] = [0, 1, 0]. \\
v_3 &= u_3 - \frac{\langle u_3, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle u_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 = [1, 1, 1] - [1, 0, 0] - [0, 1, 0] = [0, 0, 1].
\end{align*}
\]

The vectors \( [1, 0, 0], [0, 1, 0], [0, 0, 1] \) are orthogonal.

**Example 4.4.** Apply Gram-Schmidt orthogonalization process on the vectors \( u_1 = [1, 1, 0, 1], u_2 = [0, 1, 1, -1] \) and \( u_3 = [1, 3, 2, -1] \).

\[
\begin{align*}
v_1 &= [1, 1, 0, 1]. \\
v_2 &= u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = [0, 1, 1, -1] - 0 [1, 1, 0, 1] = [0, 1, 1, -1].
\end{align*}
\]
\[ v_3 = u_3 - \frac{\langle u_3, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle u_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 \]
\[ = [1, 3, 2, -1] - [1, 1, 0, 1] - 2[0, 1, 1, -1] = [0, 0, 0, 0]. \]

Discarding \( v_3 \), which is the zero vector, we have only two linearly independent vectors out of \( u_1, u_2, u_3 \). They are \( u_1 \) and \( u_2 \); and \( u_3 \) is a linear combination of these two. In fact, the process also revealed that \( u_3 = u_1 - 2u_2 \).

An orthogonal set can be made orthonormal by dividing each vector by its norm. Also you can modify Gram-Schmidt orthogonalization process to directly output orthonormal vectors.

**Exercises for § 4.3**

1. Find \( u \in \mathbb{R}^{1 \times 3} \) so that \( [1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}], [1/\sqrt{2}, 0, -1/\sqrt{2}] \), \( u \) are orthonormal. Form a matrix with the vectors as rows, in that order. Verify that the columns of the matrix are also orthonormal.

2. Using Gram-Schmidt process, orthonormalize the vectors \( [1, 1, 1], [1, 0, 1], [0, 1, 2] \).

3. Show that the cross product \( u \times v \) of two linearly independent vectors \( u, v \) in \( \mathbb{R}^{1 \times 3} \) is orthogonal to both \( u \) and \( v \). How to obtain this third vector as \( u \times v \) by Gram-Schmidt process?

### 4.4 Determining linear independence

Gram-Schmidt process has the main goal of orthogonalizing a list of vectors. We could use it to determine linear independence. The elementary row operations provide a more efficient way. Given \( m \) row vectors \( v_1, \ldots, v_m \in \mathbb{F}^{1 \times n} \), we form a matrix \( A \) with its \( i \)th row as \( v_i \). Then using elementary row operations, we bring it to its RREF.

Observe that exchanging \( v_i \) with \( v_j \) in the list of vectors does not change linear independence of the vectors. Multiplying a nonzero scalar with \( v_i \) does not affect linear independence. Also, replacing \( v_i \) with \( v_i + \alpha v_j \) does not alter linear independence.

To see the last one, suppose \( v_1, \ldots, v_m \) are linearly independent. Let \( w_i = v_i + \alpha v_j, i \neq j \). To show the linear independence of \( v_1, \ldots, v_{i-1}, w_i, v_{i+1}, \ldots, v_n \), suppose that

\[ \beta_1 v_1 + \cdots + \beta_{i-1} v_{i-1} + \beta_i w_i + \beta_{i+1} v_{i+1} + \cdots + \beta_m v_m = 0. \]

Then

\[ \beta_1 v_1 + \cdots + \beta_{i-1} v_{i-1} + \beta_i (v_i + \alpha v_j) + \beta_{i+1} v_{i+1} + \cdots + \beta_m v_m = 0. \]

Simplifying, we have

\[ \beta_1 v_1 + \cdots + \beta_i v_i + \cdots + (\beta_j + \alpha \beta_i) v_j + \cdots + \beta_m v_m = 0. \]

Using linear independence of \( v_1, \ldots, v_m \), we obtain

\[ \beta_1 = \cdots = \beta_i = \cdots = \beta_j + \alpha \beta_i = \cdots = \beta_m = 0. \]
This gives $\beta_j = -\alpha \beta_i = 0$ and all other $\beta$s are zero. Thus $v_1, \ldots, w_i, \ldots, v_m$ are linearly independent. Similarly, the converse also holds.

Thus, we take these vectors as the rows of a matrix and apply our reduction to RREF algorithm. From the RREF, we know that all rows where a pivot occurs are linearly independent. If you want to determine exactly which vectors among these are linearly independent, you must keep track of the row exchanges. A summary of the discussion in terms of a matrix is as follows.

**Theorem 4.3.** Let $A \in \mathbb{F}^{m \times n}$. The rows of $A$ are linearly independent iff the rows of the RREF of $A$ are linearly independent.

**Example 4.5.** To determine whether the vectors $[1, 1, 0, 1], [0, 1, 1, -1]$ and $[1, 3, 2, -1]$ are linearly independent or not, we proceed as follows.

$$
\begin{pmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 1 & -1 \\
1 & 3 & 2 & -1
\end{pmatrix}
\xrightarrow{E_{-1}[3,1]}
\begin{pmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 1 & -1 \\
0 & 2 & 2 & -2
\end{pmatrix}
\xrightarrow{R_1}
\begin{pmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 1 & -1 \\
0 & 0 & 0 & -4
\end{pmatrix}
\xrightarrow{R_2}
\begin{pmatrix}
1 & 0 & -1 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
$$

Here, $R_1 = E_{-1}[1, 2], E_{-2}[3, 2]$ and $R_2 = E_{-1/4}[3], E_{-2}[1, 3], E_1[2, 3]$.

The last matrix is in RREF in which each row has a pivot. Thus all the rows in the RREF are independent. Therefore, the original vectors are linearly independent.

**Example 4.6.** Are the vectors $[1, 1, 0, 1]^t, [0, 1, 1, -1]^t$ and $[2, -1, -3, 5]^t$ linearly independent?

The vectors are in $\mathbb{F}^{4 \times 1}$. These are linearly independent iff their transposes are. Forming a matrix with the transpose of the given vectors as rows, and reducing it to its RREF, we see that

$$
\begin{pmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 1 & -1 \\
2 & -1 & -3 & 5
\end{pmatrix}
\xrightarrow{E_{-2}[3,1]}
\begin{pmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 1 & -1 \\
0 & -3 & -3 & 3
\end{pmatrix}
\xrightarrow{R_1}
\begin{pmatrix}
1 & 0 & -1 & -2 \\
0 & 1 & 1 & -1 \\
0 & 0 & 0 & 0
\end{pmatrix}
$$

Here, $R_1 = E_{-1}[2, 1], E_3[3, 2]$. Since a zero row has appeared, the original vectors are linearly dependent. Also, notice that no row exchanges were carried out in the reduction process. Therefore, the third vector is a linear combination of the first two vectors; and the first two vectors are linearly independent.

**Exercises for § 4.4**

1. Using elementary row operations determine whether the given vectors are linearly dependent or independent in each of the following cases.

   (a) $[1, 0, -1, 2, -3], [-2, 1, 2, 4, -1], [3, 0, -1, 1, 1], [-2, 1, 1, -1, -2]$.

   (b) $[1, 0, -1, 2, -3], [-2, 1, 2, 4, -1], [3, 0, -1, 1, 1], [-2, -1, 0, -7, 3]$.

   (c) $[1, i, -1, 1 - i], [i, -1, -i, 1 + i], [2, 0, 1, i], [1 + i, 1 - i, -1, -i]$.

2. Suppose $A \in \mathbb{F}^{n \times n}$ is an invertible matrix, and $v_1, \ldots, v_m \in \mathbb{F}^{n \times 1}$. Prove that $v_1, \ldots, v_m$ are linearly independent iff $Av_1, \ldots, Av_n$ are linearly independent.
4.5 Rank

Consider an \( m \times n \) matrix \( A \). It may very well happen that all its rows are linearly independent. It may also happen that only some \( r \) of the \( m \) rows are linearly independent and other \( m - r \) rows are linear combinations of those \( r \) rows. This number \( r \), which is the maximum number of linearly independent rows in \( A \), is called the row rank of \( A \). The row rank is equal to 0 when all entries of the matrix are 0. If all the rows of an \( m \times n \) matrix are linearly independent, its row rank is \( m \).

Similarly, the maximum number of linearly independent columns of a matrix is called its column rank. We take the column rank of a zero matrix as 0; and if all columns are linearly independent, the column rank is the number of columns. Recall that for a matrix \( A \), its range \( R(A) \) is the set of all linear combinations of its columns. Therefore, the column rank of \( A \) is \( k \) means that there are exactly \( k \) columns of \( A \) such that each vector in \( R(A) \) is a linear combination of these \( k \) columns.

**Example 4.7.** Find the row rank and the column rank of the matrix

\[
A = \begin{bmatrix}
1 & 1 & 1 & 2 & 1 \\
1 & 2 & 1 & 1 & 1 \\
3 & 5 & 3 & 4 & 3 \\
-1 & 0 & -1 & -3 & -1
\end{bmatrix}.
\]

Here, we see that \( \text{row}(3) = \text{row}(1) + 2\text{row}(2) \), \( \text{row}(4) = \text{row}(2) - 2\text{row}(1) \). But \( \text{row}(2) \) is not a scalar multiple of \( \text{row}(1) \), that is, \( \text{row}(1), \text{row}(2) \) are linearly independent. Therefore, the row rank of \( A \) is 2.

For the linear independence of columns, we see that

\[ \text{col}(3) = \text{col}(5) = \text{col}(1), \text{col}(4) = 3\text{col}(1) - \text{col}(2). \]

And columns one and two are linearly independent. So, the column rank of \( A \) is also 2.

It can be proved that an elementary row operation neither alters the row rank nor the column rank of a matrix. Therefore, the row rank of a matrix is the same as the row rank of its RREF. Similarly, the column rank of a matrix is same as the column rank of its RREF. However, the row rank and the column rank of a matrix in RREF is equal to the number of pivots. Therefore, the row rank and the column rank of a matrix are equal.

We thus define the rank of a matrix \( A \) as its row rank, which is also equal to its column rank, and denote it by \( \text{rank}(A) \). Since the row rank of \( A^t \) is the column rank of \( A \), it follows that \( \text{rank}(A^t) = \text{rank}(A) \). The number \( n - \text{rank}(A) \) is called the nullity of \( A \). We will connect the nullity to the number of linearly independent solutions of a linear homogeneous system \( Ax = 0 \) later.

For instance, look at Examples 4.5-4.6. There, the rank of

\[
\begin{bmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 1 & -1 \\
1 & 3 & 2 & -1
\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}
1 & 1 & 0 & 1 \\
0 & 1 & 1 & -1 \\
2 & -1 & -3 & 5
\end{bmatrix}
\]

81
are, respectively, 3 and 2. Their nullity are 0 and 1, respectively.

It thus follows that if \( A \) is an \( m \times n \) matrix, then \( \text{rank}(A) \) must be less than or equal to \( \min\{m, n\} \).

Also, if we take more than \( n \) number of vectors in \( \mathbb{F}^{1 \times n} \) (or in \( \mathbb{F}^{n \times 1} \)), then they are bound to be linearly dependent. Hint: look at the RREF of the matrix whose rows are the given vectors.

Further, we see that a square matrix is invertible iff its rank is equal to its order.

**Exercises for § 4.5**

1. Reduce the following matrix to its RREF and determine its rank:

\[
\begin{bmatrix}
  1 & 1 & 1 & 2 & 1 \\
  1 & 2 & 1 & 1 & 1 \\
  3 & 5 & 3 & 4 & 3 \\
 -1 & 0 & -1 & -3 & -1
\end{bmatrix}
\]

2. Let \( T : \mathbb{R}^{3 \times 1} \to \mathbb{R}^{3 \times 1} \) be defined by \( T([a, b, c]^t) = [a + b, 2a - b - c, a + b + c]^t \). Let \( A \in \mathbb{R}^{3 \times 3} \) be the matrix such that \( T(x) = Ax \) for \( x \in \mathbb{R}^{3 \times 1} \). Find \( \text{rank}(A) \).

3. If \( E \in \mathbb{F}^{m \times m} \) is an elementary matrix and \( A \in \mathbb{F}^{m \times n} \), then show that the row rank of \( EA \) is equal to the row rank of \( A \).

4. If \( B \in \mathbb{F}^{m \times m} \) is an invertible matrix and \( A \in \mathbb{F}^{m \times n} \), then show that the column rank of \( BA \) is equal to the column rank of \( A \).

5. From previous two exercises, conclude that an elementary row operation neither alters the row rank nor the column rank of a matrix.

6. Let \( A \in \mathbb{C}^{m \times k} \), \( B \in \mathbb{R}^{m \times k} \), \( C \in \mathbb{C}^{k \times m} \). Show that \( \text{rank}(A^t) = \text{rank}(A) \).

7. Show that if \( v_1, \ldots, v_n \in \mathbb{F}^{1 \times m} \) and \( n > m \), then these vectors are linearly dependent.

8. Prove that if \( A \in \mathbb{F}^{n \times n} \), then \( \text{rank}(A) = n \) iff \( \det(A) \neq 0 \).

### 4.6 Solvability of linear equations

We can now use our knowledge about matrices to settle some issues regarding solvability of linear systems. A **linear system** with \( m \) equations in \( n \) unknowns looks like:

\[
\begin{align*}
  a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
  a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
  & \vdots \\
  a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{align*}
\]
Solving such a linear system amounts to determining the unknowns $x_1, \ldots, x_n$ with known scalars $a_{ij}$ and $b_i$. Using the abbreviation $x = [x_1, \ldots, x_n]^t$, $b = [b_1, \ldots, b_m]^t$ and $A = [a_{ij}]$, the system can be written in the compact form:

$$Ax = b.$$ 

Here, $A \in \mathbb{F}^{m \times n}$, $x \in \mathbb{F}^{n \times 1}$ and $b \in \mathbb{F}^{m \times 1}$. We also say that the matrix $A$ is the **system matrix** of the linear system $Ax = b$. Observe that the matrix $A$ is a linear transformation from $\mathbb{F}^{n \times 1}$ to $\mathbb{F}^{m \times 1}$, where $m$ is the **number of equations** and $n$ is the **number of unknowns** in the system.

There is a slight deviation from our accepted symbolism. In case of linear systems, we write $b$ as a column vector and $x_i$ are unknown scalars.

Let $A \in \mathbb{F}^{m \times n}$ and $b \in \mathbb{F}^{m \times 1}$. A **solution** of the system $Ax = b$ is any vector $y \in \mathbb{F}^{n \times 1}$ such that $Ay = b$. In such a case, if $y = [a_1, \ldots, a_n]^t$, then $a_i$ is called as the **value of the unknown** $x_i$ in the solution $y$. In this language a solution of the system is also written informally as

$$x_1 = a_1, \ldots, x_n = a_n.$$ 

The system $Ax = b$ has a solution iff $b \in R(A)$; and it has a unique solution iff $b \in R(A)$ and $A$ is a one-one map. Corresponding to the linear system $Ax = b$ is the **homogeneous system**

$$Ax = 0.$$ 

The homogeneous system always has a solution since $y := 0$ is a solution. It has infinitely many solutions when it has a nonzero solution. For, if $y$ is a solution of $Ax = 0$, then so is $\alpha x$ for any scalar $\alpha$.

To study the non-homogeneous system, we use the **augmented matrix** $[A|b] \in \mathbb{F}^{m \times (n+1)}$ which has its first $n$ columns as those of $A$ in the same order, and the $(n+1)$th column is $b$. For example,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \quad \Rightarrow \quad [A|b] = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 5 \end{bmatrix}.$$ 

**Theorem 4.4.** Let $A \in \mathbb{F}^{m \times n}$ and $b \in \mathbb{F}^{m \times 1}$. Then the following statements are true.

1. $Ax = b$ has a solution iff rank($[A|b]$) = rank($A$).

2. If $u$ is a particular solution of $Ax = b$, then each solution of $Ax = b$ is given by $u + y$, where $y$ is a solution of the homogeneous system $Ax = 0$.

3. If $[A'|b']$ is obtained from $[A|b]$ by a finite sequence of elementary row operations, then each solution of $Ax = b$ is a solution of $A'x = b'$, and vice versa.

4. If $r = \text{rank}([A|b]) = \text{rank}(A) < n$, then there are $n - r$ unknowns which can take arbitrary values and other $r$ unknowns be determined from the values of these $n - r$ unknowns.

5. $Ax = b$ has a unique solution iff rank($[A|b]$) = rank($A$) = $n$.

6. If $m = n$, then $Ax = b$ has a unique solution iff $\det(A) \neq 0$. 

83
Proof: (1) \(Ax = b\) has a solution iff \(b \in R(A)\) iff \(b\) is a linear combination of columns of \(A\) iff \(\text{rank}([A|b]) = \text{rank}(A)\).

(2) Let \(u\) be a particular solution of \(Ax = b\). Then \(Au = b\). Now, \(y\) is a solution of \(Ax = b\) iff \(Ay = Au\) iff \(A(y - u) = 0\) iff \(y - u\) is a solution of \(Ax = 0\).

(3) If \([A'|b']\) has been obtained from \([A|b]\) by a finite sequence of elementary row operations, then \(A' = EA\) and \(b' = Eb\), where \(E\) is the product of corresponding elementary matrices. Thus \(E\) is invertible. Now, \(A'x = b'\) iff \(EAx = Eb\) iff \(Ax = E^{-1}Eb = b\).

(4) Due to (2), consider solving the corresponding homogeneous system. Let \(\text{rank}(A) = r < n\). Due to (3), assume that \(A\) is in RREF. There are \(r\) number of pivots in \(A\) and \(m - r\) number of zero rows. Omit all the zero rows. It does not affect the solutions. The \(n - r\) unknowns which do not correspond to pivots can take arbitrary values, and the unknowns corresponding to pivots can be expressed in terms of these \(n - r\) unknowns.

(5) It follows from (1) and (4).

(6) Notice that for a matrix \(A \in \mathbb{F}^{n \times n}\), it is invertible iff \(\text{rank}(A) = n\) iff \(\det(A) \neq 0\). Then the statement follows from (5). \(\blacksquare\)

A system of linear equations \(Ax = b\) is said to be consistent iff \(\text{rank}([A|b]) = \text{rank}(A)\).

Theorem 4.4(1) says that only consistent systems have solutions. Conversely, if a system has a solution, then the system must be consistent. The statement in Theorem 4.4(5) is sometimes informally stated as follows:

A consistent system has \(n - \text{rank}(A)\) number of linearly independent solutions.

Exercises for § 4.6

1. Show that a linear system \(Ax = b\) is solvable iff \(b\) is a linear combination of columns of \(A\).

2. Let \(A \in \mathbb{C}^{n \times n}\). Show that \(A\) is invertible iff for each \(B \in \mathbb{C}^{n \times n}\), \(AB = 0\) implies that \(B = 0\).

3. Consider the linear system \(Ax = b\), where \(A \in \mathbb{F}^{m \times n}\) and \(\text{rank}(A) = r\). Write explicit conditions on \(m, n, r\) so that the system has
   (a) no solution     (b) unique solution     (c) infinite number of solutions

4.7 Gauss-Jordan elimination

Gauss-Jordan elimination is an application of converting the augmented matrix to its row reduced echelon form for solving linear systems.

To determine whether a system of linear equations is consistent or not, we convert the augmented matrix \([A|b]\) to its RREF. In the RREF, if an entry in the \(b\) portion has become a pivot, then the system is inconsistent; otherwise, the system is consistent.
Example 4.8. Is the following system of linear equations consistent?

\[
\begin{align*}
5x_1 + 2x_2 - 3x_3 + x_4 &= 7 \\
x_1 - 3x_2 + 2x_3 - 2x_4 &= 11 \\
3x_1 + 8x_2 - 7x_3 + 5x_4 &= 8 \\
\end{align*}
\]

We take the augmented matrix and reduce it to its row reduced echelon form by elementary row operations.

\[
\begin{bmatrix}
5 & 2 & -3 & 1 & 7 \\
1 & -3 & 2 & -2 & 11 \\
3 & 8 & -7 & 5 & -15 \\
\end{bmatrix}
\xrightarrow{R_1}
\begin{bmatrix}
1 & 2/5 & -3/5 & 1/5 & 7/5 \\
0 & -17/5 & 13/5 & -11/5 & 48/5 \\
0 & 34/5 & -26/5 & 22/5 & -19/5 \\
\end{bmatrix}
\xrightarrow{R_2}
\begin{bmatrix}
1 & 0 & -5/17 & -1/17 & 43/17 \\
0 & 1 & -13/17 & 11/17 & -48/17 \\
0 & 0 & 0 & 0 & 77/5 \\
\end{bmatrix}
\]

Here, \( R_1 = E_{1/5}[1], E_{-1}[2, 1], E_{-3}[3, 1] \) and \( R_2 = E_{-5/17}[2], E_{-2/5}[1, 2], E_{-34/5}[3, 2] \). Since an entry in the \( b \) portion has become a pivot, the system is inconsistent. In fact, you can verify that the third row in \( A \) is simply first row minus twice the second row, whereas the third entry in \( b \) is not the first entry minus twice the second entry. Therefore, the system is inconsistent.

Example 4.9. We change the last equation in the previous example to make it consistent. We consider the new system

\[
\begin{align*}
5x_1 + 2x_2 - 3x_3 + x_4 &= 7 \\
x_1 - 3x_2 + 2x_3 - 2x_4 &= 11 \\
3x_1 + 8x_2 - 7x_3 + 5x_4 &= -15 \\
\end{align*}
\]

The reduction to echelon form will change the changed entry as follows:

\[
\begin{bmatrix}
5 & 2 & -3 & 1 & 7 \\
1 & -3 & 2 & -2 & 11 \\
3 & 8 & -7 & 5 & -15 \\
\end{bmatrix}
\xrightarrow{R_1}
\begin{bmatrix}
1 & 2/5 & -3/5 & 1/5 & 7/5 \\
0 & -17/5 & 13/5 & -11/5 & 48/5 \\
0 & 34/5 & -26/5 & 22/5 & -19/5 \\
\end{bmatrix}
\xrightarrow{R_2}
\begin{bmatrix}
1 & 0 & -5/17 & -1/17 & 43/17 \\
0 & 1 & -13/17 & 11/17 & -48/17 \\
0 & 0 & 0 & 0 & 77/5 \\
\end{bmatrix}
\]

with \( R_1 = E_{1/5}[1], E_{-1}[2, 1], E_{-3}[3, 1] \) and \( R_2 = E_{-5/17}[2], E_{-2/5}[1, 2], E_{-34/5}[3, 2] \) as the row operations. This expresses the fact that the third equation is redundant. Now, solving the new system in row reduced echelon form is easier. Writing as linear equations, we have

\[
\begin{align*}
1x_1 - \frac{5}{17}x_3 - \frac{1}{17}x_4 &= \frac{43}{17} \\
1x_2 - \frac{13}{17}x_3 + \frac{11}{17}x_4 &= -\frac{48}{17} \\
\end{align*}
\]
The unknowns corresponding to the pivots are called the **basic variables** and the other unknowns are called the **free variable**. The number of basic variables is equal to the number of pivots, which is the rank of the system matrix. By assigning the free variables \( x_i \) to any arbitrary values, say, \( \alpha_i \), the basic variables can be evaluated in terms of \( \alpha_i \).

In the above reduced system, the basic variables are \( x_1 \) and \( x_2 \); and the unknowns \( x_3, x_4 \) are free variables. We assign \( x_3 \) to \( \alpha \) and \( x_4 \) to \( \beta \). Then we have

\[
\begin{align*}
x_1 &= \frac{43}{17} + \frac{5}{17} \alpha + \frac{1}{17} \beta, \\
x_2 &= -\frac{48}{17} + \frac{13}{17} \alpha - \frac{11}{17} \beta.
\end{align*}
\]

Therefore, any vector \( y \in \mathbb{F}^{4 \times 1} \) in the form

\[
y := \begin{bmatrix} \frac{43}{17} + \frac{5}{17} \alpha + \frac{1}{17} \beta \\ -\frac{48}{17} + \frac{13}{17} \alpha - \frac{11}{17} \beta \\ \alpha \\ \beta \end{bmatrix} \quad \text{for } \alpha, \beta \in \mathbb{F}
\]

is a solution of the linear system. Observe that

\[
y = \begin{bmatrix} 43/17 \\ -48/17 \\ 0 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 5/17 \\ 13/17 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 1/17 \\ -11/17 \\ 0 \\ 1 \end{bmatrix}.
\]

Here, the first vector is a particular solution of the original system. The two vectors

\[
\begin{bmatrix} 5/17 \\ 13/17 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1/17 \\ -11/17 \\ 0 \\ 1 \end{bmatrix}
\]

are linearly independent solutions of the corresponding homogeneous system. There should be exactly two such linearly independent solutions of the homogeneous system, because the nullity of the system matrix is the number of unknowns minus its rank, which is \( 4 - 2 = 2 \).

There are variations of Gauss-Jordan elimination. Instead of reducing the augmented matrix to its row reduced echelon form, if we reduce it to another intermediary form, called the **row echelon form**, then we obtain the method of **Gaussian elimination**. In the row echelon form, we do not require the entries above a pivot to be 0; also the pivots need not be equal to 1. In that case, we will require back-substitution in solving a linear system. To illustrate this process, we redo Example 4.9 starting with the augmented matrix, as follows:

\[
\begin{bmatrix} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & -15 \end{bmatrix} \xrightarrow{R_1} \begin{bmatrix} 5 & 2 & -3 & 1 & 7 \\ 0 & -17/5 & 13/5 & -11/5 & 48/5 \\ 0 & 34/5 & -26/5 & 22/5 & -96/5 \end{bmatrix}
\]

\[
\xrightarrow{E_2[3,2]} \begin{bmatrix} 5 & 2 & -3 & 1 & 7 \\ 0 & -17/5 & 13/5 & -11/5 & 48/5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\]
Here, \( R_1 = E_{-1/5}[2, 1], E_{-3/5}[3, 1] \). The augmented matrix is now in row echelon form. It is a consistent system, since no entry in the \( b \) portion is a pivot. The pivots say that \( x_1, x_2 \) are basic variables and \( x_3, x_4 \) are free variables. We assign \( x_3 \) to \( \alpha \) and \( x_4 \) to \( \beta \). Writing in equations form, we have

\[
x_1 = 7 - 2x_2 + 3\alpha - \beta, \quad x_2 = -\frac{5}{17}(\frac{48}{5} - \frac{13}{5}\alpha + \frac{11}{5}\beta).
\]

First we determine \( x_2 \) and then back-substitute. We obtain

\[
x_1 = \frac{43}{17} + \frac{5}{17}\alpha + \frac{1}{17}, \quad x_2 = -\frac{48}{17} + \frac{13}{17}\alpha - \frac{11}{17}\beta, \quad x_3 = \alpha, \quad x_4 = \beta.
\]

As you see we end up with the same set of solutions as in Gauss-Jordan elimination.

**Exercises for § 4.7**

1. Using Gauss-Jordan elimination, and also by Gaussian elimination, solve the following linear systems:

   (a) \( 3w + 2x + 2y - z = 2, \quad 2x + 3y + 4z = -2, \quad y - 6z = 6 \).

   (b) \( w + 4x + y + 3z = 1, \quad 2x + y + 3z = 0, \quad w + 3x + y + 2z = 1, \quad 2x + y + 6z = 0 \).

   (c) \( w - x + y - z = 1, \quad w + x - y - z = 1, \quad w - x - y + z = 2, \quad 4w - 2x - 2y = 1 \).

2. Show that the following linear system \( x + y + kz = 1, \quad x - y - z = 2, \quad 2x + y - 2z = 3 \) has no solution for \( k = 1 \), and has a unique solution for each \( k \neq 1 \).
Chapter 5

Matrix Eigenvalue Problem

5.1 Eigenvalues and eigenvectors

Let \( A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \). Here, \( A : \mathbb{R}^{2 \times 1} \rightarrow \mathbb{R}^{2 \times 1} \). It transforms straight lines to straight lines or points.

Get me a straight line which is transformed to itself.

\[
A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.
\]

Thus, the line \( \{(x, x) : x \in \mathbb{R}\} \) never moves. So also the line \( \{(x, -x) : x \in \mathbb{R}\} \).

Observe: \( A \begin{bmatrix} x & x \\ x & x \end{bmatrix} = 1 \begin{bmatrix} x & x \\ x & x \end{bmatrix} \) and \( A \begin{bmatrix} x & -x \\ -x & -x \end{bmatrix} = (-1) \begin{bmatrix} x & -x \\ -x & -x \end{bmatrix} \).

Let \( A \in \mathbb{F}^{n \times n} \). A scalar \( \lambda \in \mathbb{F} \) is called an eigenvalue of \( A \) iff there exists a non-zero vector \( v \in \mathbb{F}^{n \times 1} \) such that \( Av = \lambda v \). Such a vector \( v \) is called an eigenvector of \( A \) for (or, associated with, or, corresponding to) the eigenvalue \( \lambda \).

\textbf{Example 5.1.} Consider the matrix \( A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \). It has an eigenvector \([0, 0, 1]^t\) associated with the eigenvalue 1. Is \([0, 0, c]^t\) also an eigenvector associated with the same eigenvalue 1?

In fact, corresponding to an eigenvalue, there are infinitely many eigenvectors.

5.2 Characteristic polynomial

Notice that \( Av = \lambda v \) iff \((A - \lambda I)v = 0\). Therefore, a nonzero vector \( v \) is an eigenvector for the eigenvalue \( \lambda \) iff \( v \) is a nonzero solution of the homogeneous system \((A - \lambda I)x = 0\). Further, the linear system \((A - \lambda I)x = 0\) has a nonzero solution iff \( \text{rank}(A - \lambda I) < n \), where \( A \) is an \( m \times n \) matrix. And this happens iff \( \det(A - \lambda I) = 0 \). Therefore, we have the following result.

\textbf{Theorem 5.1.} Let \( A \in \mathbb{F}^{n \times n} \). A scalar \( \lambda \in \mathbb{F} \) is an eigenvalue of \( A \) iff \( \det(A - \lambda I) = 0 \).
The polynomial $\det(A - tI)$ is called the **characteristic polynomial** of the matrix $A$. Each eigenvalue of $A$ is a zero of the characteristic polynomial of $A$. Even if $A$ is a matrix with real entries, some of the zeros of its characteristic polynomial may turn out to be complex numbers. Considering $A$ as a linear transformation from $\mathbb{R}^{n \times 1}$ to $\mathbb{R}^{m \times 1}$, the scalars are now only real numbers. Thus each zero of the characteristic polynomial may not be an eigenvalue; only the real zeros are. We say that each zero of the characteristic polynomial is a **complex eigenvalue** of $A$.

**Convention:** We regard $A$ as a matrix with complex entries. We look at $A$ as a linear transformation $A : \mathbb{C}^{n \times 1} \to \mathbb{C}^{m \times 1}$. Then each complex eigenvalue, that is, a zero of the characteristic polynomial of $A$, is considered as an eigenvalue of $A$.

Since the characteristic polynomial of a matrix $A$ of order $n$ is a polynomial of degree $n$ in $t$, it has exactly $n$, not necessarily distinct, zeros. And these are the eigenvalues (complex eigenvalues) of $A$. Notice that, here, we are using the fundamental theorem of algebra which says that each polynomial of degree $n$ with complex coefficients can be factored into exactly $n$ linear factors.

**Caution:** When $\lambda$ is a complex eigenvalue of $A \in \mathbb{F}^{n \times n}$, a corresponding eigenvector $x$ is, in general, a vector in $\mathbb{C}^{n \times 1}$.

**Example 5.2.** Find the eigenvalues and corresponding eigenvectors of the matrix

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

The characteristic polynomial is

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

The eigenvalues of $A$ are its zeros, that is, 1, 1, 1.

To get an eigenvector, we solve $A(a, b, c)^t = (a, b, c)^t$ or that

$$a = a, \ a + b = b, \ a + b + c = c.$$ 

It gives $b = c = 0$ and $a \in \mathbb{F}$ can be arbitrary. Since an eigenvector is nonzero, all the eigenvectors are given by $(a, 0, 0)^t$, for $a \neq 0$.

**Example 5.3.** For $A \in \mathbb{R}^{2 \times 2}$, given by

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

the characteristic polynomial is $t^2 + 1 = 0$. It has no real zeros. Then $A$ has no eigenvalue. However, $i$ and $-i$ are its complex eigenvalues. That is, the same matrix $A \in \mathbb{C}^{2 \times 2}$ has eigenvalues as $i$ and $-i$. The corresponding eigenvectors are obtained by solving

$$A(a, b)^t = i(a, b)^t \text{ and } A(a, b) = -i(a, b)^t.$$
For $\lambda = i$, we have $b = ia, -a = ib$. Thus, $(a, ia)^t$ is an eigenvector for $a \neq 0$.

For the eigenvalue $-i$, the eigenvectors are $(a, -ia)$ for $a \neq 0$. Following our convention, we regard $A$ as a matrix with complex entries; and it has eigenvalues $i$ and $-i$. Our convention allows us to take the second perspective.

Two matrices $A, B \in \mathbb{F}^{n \times n}$ are called **similar** iff there exists an invertible matrix $P \in \mathbb{F}^{n \times n}$ such that $P^{-1}AP = B$. The following theorem lists some important facts about eigenvalues.

**Theorem 5.2.** Let $A \in \mathbb{F}^{n \times n}$. Then the following are true.

1. $A$ and $A^t$ have the same eigenvalues.
   
   **Reason:** $\det(A^t - tI) = \det((A - tI)^t) = \det(A - tI)$.

2. Similar matrices have the same eigenvalues.
   
   **For:** $\det(P^{-1}AP - tI) = \det(P^{-1}(A - tI)P) = \det(P^{-1})\det(A - tI)\det(P) = \det(A - tI)$.

3. If $A$ is a diagonal or an upper triangular or a lower triangular matrix, then its diagonal elements are precisely its eigenvalues.
   
   **Reason:** In all these cases, $\det(A - tI) = (a_{11} - t) \cdots (a_{nn} - t)$.

4. $\det(A)$ equals the product and $\text{tr}(A)$ equals the sum of all eigenvalues of $A$.
   
   **Proof:** Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A$, not necessarily distinct. Now,
   
   $\det(A - tI) = (\lambda_1 - t) \cdots (\lambda_n - t)$.
   
   Put $t = 0$. It gives $\det(A) = \lambda_1 \cdots \lambda_n$.
   
   Expand $\det(A - tI)$ and equate the coefficients of $t^{n-1}$ to get
   
   Coeff of $t^{n-1}$ in $\det(A - tI) = \text{Coeff of } t^{n-1} \text{ in } (a_{11} - t) \cdot A_{11}
   
   = \cdots = \text{Coeff of } t^{n-1} \text{ in } (a_{11} - t) \cdot (a_{22} - t) \cdots (a_{nn} - t) = (-1)^{n-1}\text{tr}(A)$.
   
   But Coeff of $t^{n-1}$ in $\det(A - tI) = (-1)^{n-1} \cdot (\lambda_1 + \cdots + \lambda_n)$.

5. (Cayley-Hamilton) Any square matrix satisfies its characteristic polynomial.
   
   This can be proved by using the adjugate property of a determinant.

Cayley-Hamilton theorem helps us in computing powers of matrices and also the inverse of a matrix if at all it is invertible. For instance, suppose that a matrix $A$ has the characteristic polynomial

$$ a_0 + a_1 t + \cdots + a_n t^n. $$

By Cayley-Hamilton theorem, we have

$$ a_0 I + a_1 A + \cdots + a_n A^n = 0. $$

Then $A^n = - (a_0 I + a_1 A + \cdots + a_{n-1} A^{n-1})$. Then $A^n, A^{n+1}, \ldots$ can be reduced to computing $A, A^2, \ldots, A^{n-1}$.

For computing the inverse, suppose that $A$ is invertible. Then $\det(A) \neq 0$; thus $\lambda = 0$ does not
satisfy \( \det(A - \lambda I) = 0 \). That is, \( \lambda = 0 \) is not an eigenvalue of \( A \). It implies that \( (t - \lambda) \) is not a factor of the characteristic polynomial of \( A \). Therefore, the constant term \( a_0 \) in the characteristic polynomial of \( A \) is nonzero. Then we can rewrite the above equation as

\[
a_0I + A(a_1I + \cdots + a_nA^{n-1}) = 0.
\]

Multiplying \( A^{-1} \) and simplifying, we obtain

\[
A^{-1} = -\frac{1}{a_0}(a_1I + a_2A + \cdots + a_nA^{n-1}).
\]

This way, \( A^{-1} \) can also be computed.

**Exercises for § 5.1-5.2**

1. Find eigenvalues and corresponding eigenvectors of the matrix

\[
\begin{bmatrix}
3 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 0 & -2 \\
0 & 0 & 2 & 0 \\
\end{bmatrix}
\]

2. Let \( A, B, P \in \mathbb{C}^{n \times n} \) be such that \( B = P^{-1}AP \). Let \( \lambda \) be an eigenvalue of \( A \). Show that a vector \( v \) is an eigenvector of \( B \) corresponding to the eigenvalue \( \lambda \) iff \( Pv \) is an eigenvector of \( A \) corresponding to the same eigenvalue \( \lambda \).

3. An \( n \times n \) matrix \( A \) is said to be **idempotent** if \( A^2 = A \). Show that the only possible eigenvalues of an idempotent matrix are 0 or 1.

4. An \( n \times n \) matrix \( A \) is said to be **nilpotent** if \( A^k = 0 \) for some natural number \( k \). Show that 0 is the only eigenvalue of a nilpotent matrix.

5. Show that if rank of an \( n \times n \) matrix is 1, then its trace is one of its eigenvalues. What are its other eigenvalues?

6. Find all eigenvalues and their corresponding eigenvectors of the matrix \( A \in \mathbb{C}^{n \times n} \), where the \( j \)th row has all entries as \( j \).

### 5.3 Special types of matrices

A square matrix \( A \) is called **self-adjoint**, and also **hermitian**, iff \( A^* = A \). A hermitian matrix with real entries satisfies \( A^t = A \); and accordingly, such a matrix is called a **real symmetric** matrix. In general, \( A \) is called a **symmetric** matrix iff \( A^t = A \). And \( A \) is called **skew hermitian** iff \( A^* = -A \); also, a matrix is called **skew symmetric** iff \( A^t = -A \). In the following, \( B \) is symmetric, \( C \) is skew-symmetric, \( D \) is hermitian, and \( E \) is skew-hermitian. \( B \) is also hermitian and \( C \) is also skew-hermitian.

\[
B = \begin{bmatrix}
1 & 2 & 3 \\
2 & 3 & 4 \\
3 & 4 & 5 \\
\end{bmatrix}, \quad C = \begin{bmatrix}
0 & 2 & -3 \\
-2 & 0 & 4 \\
3 & -4 & 0 \\
\end{bmatrix}, \quad D = \begin{bmatrix}
i & 2i & 3 \\
2i & 3 & 4 \\
3 & 4 & 5 \\
\end{bmatrix}, \quad E = \begin{bmatrix}
0 & 2 + i & 3 \\
2 - i & i & 4i \\
3 & -4i & 0 \\
\end{bmatrix}
\]
Notice that a skew-symmetric matrix must have a zero diagonal, and the diagonal entries of a skew-hermitian matrix must be 0 or purely imaginary. Reason:

\[ a_{ii} = -\overline{a_{ii}} \Rightarrow 2\text{Re}(a_{ii}) = 0. \]

Let \( A \) be a square matrix. Since \( A + A^t \) is symmetric and \( A - A^t \) is skew symmetric, every square matrix can be written as a sum of a symmetric matrix and a skew symmetric matrix:

\[ A = \frac{1}{2}(A + A^t) + \frac{1}{2}(A - A^t). \]

Similar rewriting is possible with hermitian and skew hermitian matrices:

\[ A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*). \]

A square matrix \( A \) is called **unitary** iff \( A^*A = I = AA^* \). In addition, if \( A \) is real, then it is called an orthogonal matrix. That is, an **orthogonal matrix** is a matrix with real entries satisfying \( A^tA = I = AA^t \). Notice that a square matrix is unitary iff it is invertible and its inverse is equal to its adjoint. Similarly, a real matrix is orthogonal iff it is invertible and its inverse is its transpose. In the following, \( B \) is a unitary matrix of order 2, and \( C \) is an orthogonal matrix (also unitary) of order 3:

\[
B = \frac{1}{2} \begin{bmatrix} 1 + i & 1 - i \\ 1 - i & 1 + i \end{bmatrix}, \quad C = \frac{1}{3} \begin{bmatrix} 2 & 1 & 2 \\ -2 & 2 & 1 \\ 1 & 2 & -2 \end{bmatrix}.
\]

The following are examples of orthogonal \( 2 \times 2 \) matrices. \( O_1 \) is said to be a **rotation by an angle \( \theta \)** and \( O_2 \) is called a **reflection by an angle \( \theta/2 \)** along the \( x \)-axis. Can you say why are they so called?

\[
O_1 := \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad O_2 := \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & -\cos \theta \end{bmatrix}.
\]

Unitary or orthogonal matrices preserve inner product and also the norm. Reason: Suppose \( A \) is a unitary matrix. \( \langle Ax, Ay \rangle = \langle x, A^*Ay \rangle = \langle x, y \rangle \). Taking \( x = y \) we have \( ||Ax||^2 = ||x||^2 \).

The columns of such a matrix are orthonormal, and so are its rows. Reason: Since \( A^tA = I \), the \( i \)th row of \( A^* \) multiplied with the \( j \)th column of \( A \) gives \( \delta_{ij} \). However, this product is simply the inner product of the \( j \)th column of \( A \) with the \( i \)th column of \( A \). Therefore, columns of \( A \) are orthonormal. Similarly, \( AA^* = I \) implies that rows of \( A \) are orthonormal.

In general, if \( A \in \mathbb{F}^{m \times n} \), then \( A^*A = I \) is equivalent to asserting that the columns of \( A \) are orthonormal; and \( AA^* = I \) is equivalent to the rows of \( A \) are orthonormal.

Let \( A \in \mathbb{F}^{n \times n} \). Let \( \lambda \) be any complex eigenvalue of \( A \) with an eigenvector \( v \in \mathbb{C}^{n \times 1} \). Now, \( Av = \lambda v \). Pre-multiplying with \( v^* \), we have \( v^*Av = \lambda v^*v \in \mathbb{C} \). Using this we see the following:

1. **If \( A \) is hermitian or real symmetric, then \( \lambda \in \mathbb{R} \).**
   Reason: If \( A \) is hermitian, then \( A = A^* \). Now,

   \[
   (v^*Av)^* = v^*A^*v = v^*Av \quad \text{and} \quad (v^*v)^* = v^*v.
   \]

   So, both \( v^*Av \) and \( v^*v \) are real. Therefore, in \( v^*Av = \lambda v^*v \), \( \lambda \) is also real.
2. If $A$ is skew-hermitian or skew-symmetric, then $\lambda$ is purely imaginary or zero.

   Reason: When $A$ is skew-hermitian, $(v^*Av)^* = -v^*Av$. Then $v^*Av = \lambda v^*v$ implies that
   $$(\lambda v^*v)^* = -\lambda(v^*v).$$

   Since $v \neq 0$, $v^*v \neq 0$. Therefore, $\lambda^* = \bar{\lambda} = -\lambda$. That is, $2\text{Re}(\lambda) = 0$. This shows that $\lambda$ is purely imaginary or zero.

3. If $A$ is unitary or orthogonal, then $|\lambda| = 1$.

   Reason: Suppose $A^*A = I$. Now, $Av = \lambda v$ implies $v^*A^* = (\lambda v)^* = \bar{\lambda}v^*$. Then
   $$v^*v = v^*Iv = v^*A^*Av = \bar{\lambda}\lambda v^*v = |\lambda|^2v^*v.$$

   Since $v^*v \neq 0$, $|\lambda| = 1$.

4. If $A$ is unitary then $|\det(A)| = 1$.

   Reason: The determinant is the product of eigenvalues.

5. If $A$ is orthogonal, then $\det(A) = \pm 1$.

   Reason: The determinant is the product of eigenvalues.

Not only each eigenvalue of a real symmetric matrix is real, but also a corresponding real eigenvector can be chosen. To see this, let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Let $\lambda \in \mathbb{R}$ be an eigenvalue of $A$. If $v = x + iy \in \mathbb{C}^{n \times 1}$ is a corresponding eigenvector with $x, y \in \mathbb{R}^{n \times 1}$, then

$$A(x + iy) = \lambda(x + iy).$$

Comparing the real and imaginary parts, we have

$$Ax = \lambda x, \quad Ay = \lambda y.$$

Since $x + iy \neq 0$, at least one of $x$ or $y$ is nonzero. Choose one nonzero vector out of $x$ and $y$. That is a real eigenvector corresponding to the eigenvalue $\lambda$ of $A$.

Exercises for § 5.3

1. Construct an orthogonal $2 \times 2$ matrix whose determinant is 1.

2. Construct an orthogonal $2 \times 2$ matrix whose determinant is $-1$.

3. Construct a $3 \times 3$ Hermitian matrix with no zero entry whose eigenvalues are 1, 2 and 3.

4. Construct a $2 \times 2$ skew-Hermitian matrix whose eigenvalues are purely imaginary.

5. Show that if a matrix $A$ is real symmetric and invertible, then so is $A^{-1}$.

6. Show that if a matrix $A$ is Hermitian and invertible, then so is $A^{-1}$.
5.4 Diagonalization

Since diagonal matrices are easy to tackle, we ask whether it is possible to transform a matrix to a diagonal matrix by similarity. That is, given a square matrix \( A \), whether it is possible to have an invertible matrix \( P \) such that \( P^{-1}AP \) is a diagonal matrix. If such a matrix \( P \) exists, then \( A \) is called diagonalizable. To diagonalize a matrix \( A \) means that we determine an invertible matrix \( P \) and a diagonal matrix \( D \) such that \( P^{-1}AP = D \).

Recall that similar matrices have the same eigenvalues; and a diagonal matrix has all its eigenvalues on the diagonal. Thus whenever \( P^{-1}AP = D \) is a diagonal matrix, we have all the eigenvalues of \( A \) appearing as the entries on the diagonal of \( D \). Moreover, \( P^{-1}AP = D \) implies that \( AP = PD \). Suppose the columns of \( P \) are the vectors \( v_1, \ldots, v_n \). Then this equation says that

\[
A[v_1 \cdots v_n] = [v_1 \cdots v_n] \text{diag}(\lambda_1, \ldots, \lambda_n),
\]

where \( \lambda_1, \ldots, \lambda_n \) are the eigenvalues of \( A \). Multiplying the matrices, we have

\[
Av_j = \lambda_j v_j \quad \text{for each } j = 1, \ldots, n.
\]

That is, the columns of \( P \) are exactly the eigenvectors corresponding to the \( n \) eigenvalues of \( A \). Since it is required that \( P \) is invertible, these vectors \( v_1, \ldots, v_n \) must be linearly independent. If a matrix does not have \( n \) number of linearly independent eigenvectors, then it is not diagonalizable.

**Example 5.4.** Consider the matrix

\[
\begin{bmatrix}
1 & 1 \\
0 & 1
\end{bmatrix}
\]

Since it is upper triangular, its eigenvalues are the diagonal entries. That is, 1 is the only eigenvalue of \( A \) occurring twice. To find the eigenvectors, we solve

\[
(A - I) \begin{bmatrix} a \\ b \end{bmatrix} = 0.
\]

The equation can be rewritten as

\[
a + b = a, \ b = b.
\]

Solving the equations, we have \( a = 0 \) and \( b \) arbitrary. Thus there is only one linearly independent eigenvector, namely, \( (0, b)^t \) for a nonzero scalar \( b \). Therefore, \( A \) is not diagonalizable.

We quote a result which guarantees that under some suitable conditions on the matrix \( A \), there must exist \( n \) number of linearly independent eigenvectors of \( A \); and then \( A \) would be diagonalizable.

**Theorem 5.3.** Let \( A \in \mathbb{F}^{n \times n} \). Then the following are true:

1. If there are \( n \) distinct eigenvalues of \( A \), then \( A \) is diagonalizable.
2. If \( A \) is hermitian, then \( A \) is unitarily diagonalizable.
3. If \( A \) is real symmetric, then \( A \) is orthogonally diagonalizable.
In Theorem 5.3(2), the phrase *unitarily diagonalizable* means that there exists a unitary matrix $P$ such that $P^{-1}AP = D$ is a diagonal matrix. Once $P$ is determined, we may use $P^*AP = D$, since $P$ is unitary. Similarly, in Theorem 5.3(3), this $P$ is supposed to be an orthogonal matrix. Again, if such a $P$ has been found out, then $P^tAP$ will be the diagonal matrix.

Assuming that the given $n \times n$ matrix $A$ has $n$ linearly independent eigenvectors, we have the following procedure for diagonalization:

Determine the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A$ and corresponding eigenvectors of $A$.

Construct linearly independent eigenvectors $v_1, \ldots, v_n$ for the eigenvalues $\lambda_1, \ldots, \lambda_n$.

Take $P = [v_1 \ldots v_n]$.

Then $P^{-1}AP = D = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

**Example 5.5.** Let $A = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$. It is real symmetric. It has eigenvalues $-1, 2, 2$. To find the associated eigenvectors, we must solve the linear systems of the form $Ax = \lambda x$.

For the eigenvalue $-1$, the system $Ax = -x$ gives

$$x_1 - x_2 - x_3 = -x_1, \\
-x_1 + x_2 - x_3 = -x_2, \\
-x_1 - x_2 + x_3 = -x_3.$$ 

It yields $x_1 = x_2 = x_3$. One eigenvector is $(1, 1, 1)^t$.

For the eigenvalue $2$, we have the equations as

$$x_1 - x_2 - x_3 = 2x_1, \\
-x_1 + x_2 - x_3 = 2x_2, \\
-x_1 - x_2 + x_3 = 2x_3.$$ 

It gives $x_1 + x_2 + x_3 = 0$. We can have two linearly independent eigenvectors such as $(-1, 1, 0)^t$ and $(-1, -1, 2)^t$.

The three eigenvectors are orthogonal to each other. To orthonormalize, we divide each by its norm. We end up at the following orthonormal eigenvectors:

$$\begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \quad \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}.$$

They are linearly independent due to orthonormality. Taking

$$P = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix},$$

we see that $P^{-1} = P^t$ and

$$P^{-1}AP = P^tAP = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$
If your choice of linearly independent eigenvectors are not orthonormal, then you can use Gram-Schmidt orthogonalization, and then orthonormalize each by dividing it with its norm. However, \( P \) is chosen to be unitary or orthogonal according as \( A \) is hermitian or real symmetric.

**Exercises for § 5.4**

1. Show that the matrix
   \[
   \begin{bmatrix}
   2 & -1 & 0 \\
   -1 & 2 & 0 \\
   2 & 2 & 3 \\
   \end{bmatrix}
   \]
   is diagonalizable with a matrix in \( \mathbb{R}^{3 \times 3} \).

2. Determine whether \( A \) is diagonalizable by a matrix with complex entries:
   (a) \[
   \begin{bmatrix}
   2 & 3 \\
   6 & -1 \\
   \end{bmatrix}
   \]
   (b) \[
   \begin{bmatrix}
   1 & -10 & 0 \\
   -1 & 3 & 1 \\
   -1 & 0 & 4 \\
   \end{bmatrix}
   \]
   (c) \[
   \begin{bmatrix}
   2 & 1 & 0 & 0 \\
   0 & 2 & 0 & 0 \\
   0 & 0 & 2 & 0 \\
   0 & 0 & 0 & 5 \\
   \end{bmatrix}
   \]

3. Diagonalize \( A = \begin{bmatrix}
   7 & -5 & 15 \\
   6 & -4 & 15 \\
   0 & 0 & 1 \\
   \end{bmatrix} \). Then compute \( A^6 \).
Bibliography


Index

max(A), 6
min(A), 6

absolutely convergent, 28
absolute value, 6
adjoint of a matrix, 60
adjugate, 67
angle between vectors, 62
Archimedian property, 5

basic variables, 86
binomial series, 39

Cayley-Hamilton, 90
center power series, 30
characteristic polynomial, 89
coefficients power series, 30
cofactor, 67
column rank, 81
column vector, 53
comparison test, 15, 20
completeness property, 5
complex conjugate, 60
complex eigenvalue, 89
conditionally convergent, 28
conjugate transpose, 60
consistent system, 84
constant sequence, 10
convergence theorem power series, 31
convergent series, 11
converges improper integral, 19
converges integral, 18
converges sequence, 8, 9
cosine series expansion, 45
dense, 6

Determinant, 66
diagonalizable, 94
diagonal entries, 53
diagonal matrix, 54
diagonal of a matrix, 53
Dirichlet integral, 48
diverges improper integral, 19
diverges integral, 18
diverges to $-\infty$, 9, 11
diverges to $\infty$, 9, 11
diverges to $\pm\infty$, 19
eigenvalue, 88
eigenvector, 88
elementary matrix, 63
elementary row operation, 63
equal matrices, 53
error in Taylor’s formula, 35
even extension, 45

Fourier series, 40
free variables, 86

Gaussian elimination, 86
geometric series, 12
glb, 5
Gram-Schmidt orthogonalization, 77
greatest integer function, 6

half-range Fourier series, 46
harmonic series, 12
Homogeneous system, 83

identity matrix, 54
improper integral, 18
inner product, 61
integral test, 24
interval of convergence, 32
Leibniz theorem, 27
limit comparison series, 16
limit comparison test, 20
linearly dependent, 74
linearly independent, 74
linear combination, 74
linear map, 73
Linear system, 82
lub, 5
Maclaurin series, 37
Matrix, 53
  augmented, 69
  entry, 53
  hermitian, 91
  inverse, 58
  invertible, 58
  lower triangular, 55
  multiplication, 56
  multiplication by scalar, 55
  order, 53
  orthogonal, 92
  real symmetric, 91
  size, 53
  skew hermitian, 91
  skew symmetric, 91
  sum, 55
  symmetric, 91
  trace, 66
  unitary, 92
minor, 67
neighborhood, 6
norm, 61
nullity, 81
odd extension, 44
off diagonal entries, 53
orthogonal vectors, 62
partial sum, 11
partial sum of Fourier series, 42
piecewise continuous, 41
pivot, 64
pivotal column, 64
powers of matrices, 58
power series, 30
Pythagoras, 62
radius of convergence, 32
range, 73
rank, 81
ratio comparison test, 16
ratio test, 25
re-indexing series, 15
Reduction
  row reduced echelon form, 65
root test, 26
row rank, 81
Row reduced echelon form, 64
row vector, 53
sandwich theorem, 10
scalars, 53
scalar matrix, 54
scaling extension, 45
sequence, 8
similar matrices, 90
sine series expansion, 45
solution of linear system, 83
standard basis, 54
standard basis vectors, 54
system matrix, 83
Taylor series, 37
Taylor’s formula, 35
Taylor’s formula differential, 35
Taylor’s formula integral, 36
Taylor’s polynomial, 35
terms of sequence, 8
to diagonalize, 94
transpose of a matrix, 59
triangular matrix, 55
trigonometric series, 39
upper triangular matrix, 55
value of unknown, 83
zero matrix, 53