Problem Seminar Notes

Romanos Diogenes Malikiosis

May 3rd, 2021

Contents

Part I

Analysis

Chapter 1

Sequences and Series

1.1 Sequences

Definition 1.1.1. A sequence is a list of (real) numbers indexed by $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$, i.e.

$$(a_n)_{n=1}^{\infty} = (a_1, a_2, a_3, \ldots).$$

If $n_1 < n_2 < n_3 < \cdots$ are in $\mathbb{N}$, then $(a_{n_k})_{k=1}^{\infty}$ is a *subsequence* of (a_n).

We say that a sequence (a_n) has a limit $a \in \mathbb{R}$ if $\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$ it holds $|a_n - a| < \varepsilon$, and we write

$$\lim a_n = a.$$

Such a sequence is called *convergent*, otherwise it is called *divergent*.

We say that (a_n) *tends* to $+\infty$ (resp. $-\infty$) and we write $\lim a_n = +\infty$ (resp. $\lim a_n = -\infty$), if $\forall M > 0\ \exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$ it holds $a_n > M$ (resp. $a_n < -M$).

The sequence (a_n) is (strictly) increasing if $a_{n+1} \geq a_n$ ($a_{n+1} > a_n$) $\forall n \in \mathbb{N}$ and (strictly) decreasing if $a_{n+1} \leq a_n$ ($a_{n+1} < a_n$) $\forall n \in \mathbb{N}$. (a_n) is called bounded if there is $M > 0$ such that $|a_n| < M$, $\forall n \in \mathbb{N}$.

Example.

1. If $a_n = \frac{1}{n}$ then $(a_n) = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots)$.
2. $(a_{2n}) = (a_2, a_4, a_6, \ldots)$ is a subsequence of $(a_n) = (a_1, a_2, a_3, \ldots)$.
3. $\lim \frac{1}{n} = 0$ and $(\frac{1}{n})$ is strictly decreasing.

Proposition 1.1.2 (Convergence implies boundedness)**.** *A convergent sequence is bounded.*

Theorem 1.1.3 (Monotone Convergence)**.** *Let (a_n) be increasing. Then either $\lim a_n = +\infty$ or $\lim a_n = a \in \mathbb{R}$ (the latter holds if the sequence is bounded). Similarly, if (a_n) is decreasing, then either $\lim a_n = -\infty$ or $\lim a_n = a \in \mathbb{R}$.*

Theorem 1.1.4 (Bolzano-Weierstraß)**.** *Let (a_n) be bounded. Then (a_n) has a convergent subsequence.*

Theorem 1.1.5. *Let $(a_n) \subseteq \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$. If $\lim \frac{a_{n+1}}{a_n} = \ell \in \mathbb{R}^+ \cup \{+\infty\}$ then $\lim \sqrt[n]{a_n} = \ell$.*

We list some basic limits:

$$\lim \alpha^n = \begin{cases} +\infty, & \alpha > 1 \\ 1, & \alpha = 1 \\ 0, & |\alpha| < 1 \\ \text{doesn't exist}, & \alpha \le -1. \end{cases} \tag{1.1.1}$$

$$\lim n^\alpha = \begin{cases} +\infty, & \alpha > 0 \\ 1, & \alpha = 0 \\ 0, & \alpha < 0. \end{cases} \tag{1.1.2}$$

$$\lim \left(1 + \frac{x}{n}\right)^n = e^x, \qquad \forall x \in \mathbb{R}. \tag{1.1.3}$$

Lemma 1.1.6 (Cesàro-Stolz)**.** *Consider two sequences (a_n) and (b_n), and let $\ell \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ be such that*

$$\lim \frac{a_{n+1} - a_n}{b_{n+1} - b_n} = \ell.$$

If either

(i) $\lim a_n = \lim b_n = 0$ *and (b_n) is strictly decreasing for $n \ge n_0$, or*

(ii) $\lim b_n = +\infty$ *and (b_n) is strictly increasing for $n \ge n_0$,*

then

$$\lim \frac{a_n}{b_n} = \ell.$$

Exercises. 1. Find $\lim \sqrt[n]{n}$ (if it exists!).

Solution. Let $a_n = n$. Then $\frac{a_{n+1}}{a_n} = 1 + \frac{1}{n}$ and $\lim(1 + \frac{1}{n}) = 1$, hence $\lim \sqrt[n]{n} = 1$ by Theorem 1.1.5. □

2. Find $\lim \sqrt[n]{n!}$.

Solution. Let $a_n = n!$, hence $\frac{a_{n+1}}{a_n} = n + 1 \xrightarrow[n\to\infty]{} +\infty$, so $\lim \sqrt[n]{n!} = +\infty$ by Theorem 1.1.5. □

3. Find $\lim \frac{n}{\sqrt[n]{n!}}$.

Solution. Let $a_n = \frac{n^n}{n!}$. It holds

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^{n+1}/(n+1)!}{n^n/n!} = \left(\frac{n+1}{n}\right)^n = \left(1 + \frac{1}{n}\right)^n \xrightarrow[n\to\infty]{} e$$

by (1.1.3), therefore the given limit converges also to e by Theorem 1.1.5. □

4. Let $(a_n) \subseteq \mathbb{R}$ satisfy $\lim(2a_{n+1} - a_n) = \ell$. Show $\lim a_n = \ell$.

Proof. If we already knew that (a_n) was convergent, this would have been an easy task; indeed, if $\lim a_n = a$, then the given condition would give

$$\ell = \lim(2a_{n+1} - a_n) = 2a - a = a.$$

However, convergence of (a_n) is not given (and this is the difficult part of the problem)! We'll show that $\lim a_n = \ell$ using the definition of limit. It holds:

$$|a_{n+1}-\ell| = \left|\frac{a_n - \ell}{2} + \frac{2a_{n+1} - a_n - \ell}{2}\right| \leq \frac{|a_n - \ell|}{2} + \frac{|2a_{n+1} - a_n - \ell|}{2}, \quad \forall n \in \mathbb{N}. \tag{1.1.4}$$

For convenience, put $b_n = 2a_{n+1} - a_n$, $\forall n \in \mathbb{N}$, so that $\lim b_n = \ell$ by hypothesis. Let $\varepsilon > 0$ be arbitrary; by definition of limit, there is $n_0 \in \mathbb{N}$ such that $|b_n - \ell| < \frac{\varepsilon}{4}$, $\forall n \geq n_0$. Next, applying (1.1.4) repeatedly, we obtain

$$\begin{aligned}
|a_{n_0+m} - \ell| &< \frac{|a_{n_0+m-1} - \ell|}{2} + \frac{\varepsilon}{4} \\
&< \frac{|a_{n_0+m-2} - \ell|}{4} + \frac{\varepsilon}{8} + \frac{\varepsilon}{4} \\
&\ \vdots \\
&< \frac{|a_{n_0} - \ell|}{2^m} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8} + \cdots + \frac{\varepsilon}{2^{m+1}} \\
&< \frac{|a_{n_0} - \ell|}{2^m} + \frac{\varepsilon}{2}.
\end{aligned}$$

Since $\lim_{m\to\infty} \frac{a_{n_0}-\ell}{2^m} = 0$, there is $m_0 \in \mathbb{N}$ such that $\frac{|a_{n_0}-\ell|}{2^m} < \frac{\varepsilon}{2}$, $\forall m \geq m_0$. Therefore, for $N = m_0 + n_0$ it holds $|a_n - \ell| < \varepsilon$ for all $n \geq N$, which by definition shows that $\lim a_n = \ell$, as desired. □

5. Let $a > 0$ and (x_n) be a sequence of real numbers satisfying the recurrence relation

$$x_{n+1} = a + x_n^2, \quad \forall n \geq 0, \quad x_0 = 0. \tag{1.1.5}$$

Find necessary and sufficient conditions such that $\lim x_n$ exists.

Solution. If this limit exists and equals $\ell \in \mathbb{R}$, then taking limits on both sides of (1.1.5), we obtain the quadratic equation

$$\ell = a + \ell^2.$$

The discriminant Δ equals $1 - 4a$, so a necessary condition for the existence of ℓ is

$$0 < a \leq \frac{1}{4}.$$

We will show that this condition is also sufficient. We assume henceforth that $0 < a \leq \frac{1}{4}$.

First, we show that (x_n) is increasing; this will be done by induction. We observe that

$$x_1 = a > 0 = x_0,$$

so $x_{n+1} \geq x_n \geq 0$ certainly holds for $n = 0$. Assume that it holds for $n = k$; we will show that it also holds for $n = k+1$ (*induction step*). Indeed, we have

$$x_{k+2} - x_{k+1} = (a + x_{k+1}^2) - (a - x_k^2) = x_{k+1}^2 - x_k^2 \geq 0,$$

therefore, $x_{k+2} \geq x_{k+1}$, proving that (x_n) is indeed increasing.

Finally, we will show that (x_n) is bounded. We will use induction once more, to show that

$$x_n \leq \frac{1}{2}, \quad \forall n \in \mathbb{N}.$$

This certainly holds for $n = 0$; assume it holds for $n = k$. Then,

$$x_{k+1} = a + x_k^2 \leq \frac{1}{4} + \left(\frac{1}{2}\right)^2 = \frac{1}{2},$$

completing the solution. □

6. **(IMC '03)** Let (a_n) be a sequence in $\mathbb{R}$ with $a_1 = 1$, $a_{n+1} > \frac{3}{2}a_n$ $\forall n \in \mathbb{N}$.

   (a) Show that the sequence
   $$b_n = \frac{a_n}{(3/2)^{n-1}}$$
   either has a finite limit or tends to $+\infty$.

   (b) Show that for every $\alpha > 1$, there is a sequence (a_n) with the above properties, such that
   $$\lim \frac{a_n}{(3/2)^{n-1}} = \alpha.$$

Proof. (a) We observe that the sequence (b_n) is strictly increasing:

$$b_{n+1} = \frac{a_{n+1}}{(3/2)^n} > \frac{\frac{3}{2}a_n}{(3/2)^n} = b_n, \quad \forall n \in \mathbb{N}.$$

The rest follows from the Bolzano-Weierstraß Theorem 1.1.4.

(b) Define

$$b_n = \alpha - \frac{\alpha - 1}{n}, \quad \forall n \in \mathbb{N},$$

so that (b_n) is strictly increasing and $\lim b_n = \alpha$. Next, define

$$a_n = b_n \left(\frac{3}{2}\right)^{n-1}.$$

The sequence (a_n) obviously satisfies the desired properties. □

7. Let (x_n) be a sequence of real numbers such that

$$\lim(x_{2n} + x_{2n+1}) = 315, \quad \lim(x_{2n} + x_{2n-1}) = 2003.$$

Find

$$\lim \frac{x_{2n}}{x_{2n+1}}.$$

Solution. Let $a_n = x_{2n}$ and $b_n = x_{2n+1}$. By definition,

$$b_n - b_{n-1} = (x_{2n+1} + x_{2n}) - (x_{2n} + x_{2n-1}) \xrightarrow[n\to\infty]{} 315 - 2003 = -1688.$$

Therefore, by definition of the limit, there is $n_0 \in \mathbb{N}$ such that

$$b_n - b_{n-1} < -1000, \quad \forall n \geq n_0$$

(put $\varepsilon = 688$ in the definition of limit). The above clearly shows that

$$\lim b_n = -\infty.$$

Next, observe that

$$\frac{a_{n+1} - a_n}{b_{n+1} - b_n} = \frac{x_{2n+2} - x_{2n}}{x_{2n+3} - x_{2n+1}} = \frac{(x_{2n+2} + x_{2n+1}) - (x_{2n+1} + x_{2n})}{(x_{2n+3} + x_{2n+2}) - (x_{2n+2} + x_{2n+1})}$$
$$\xrightarrow[n\to\infty]{} \frac{2003 - 315}{315 - 2003} = -1.$$

Applying the Cesàro-Stolz Lemma we get

$$\lim \frac{x_{2n}}{x_{2n+1}} = \lim \frac{a_n}{b_n} = -1. \qquad \square$$

8. Let $(a_n) \subseteq \mathbb{R}$ such that

$$\lim_{n\to\infty} a_n \sum_{k=1}^{n} a_k^2 = 1.$$

Show that $\lim_{n\to\infty} \sqrt[3]{3n} \cdot a_n = 1$.

Solution. Let $s_n = \sum_{k=1}^{n} a_k^2$, so that $\lim a_n s_n = 1$, by hypothesis. The sequence (s_n) is obviously increasing, therefore it either tends to $+\infty$ or $\lim s_n = s$ for some $s \in \mathbb{R}$, by Theorem 1.1.3. If the latter holds, then

$$\lim a_n^2 = \lim(s_n - s_{n-1}) = s - s = 0,$$

contradicting the fact that $\lim a_n s_n = 1$. Hence,

$$\lim s_n = +\infty$$

must hold true, which yields $\lim a_n = 0$. Next, we obesrve

$$\sqrt[3]{3n} \cdot a_n = \frac{\sqrt[3]{3n}}{s_n} \cdot a_n s_n,$$

so it suffices to show that

$$\lim \frac{s_n^3}{3n} = 1.$$

We intend to apply the Cesàro-Stolz Lemma; it holds

$$\begin{aligned} s_n^3 - s_{n-1}^3 =& (s_n - s_{n-1})(s_n^2 + s_n s_{n-1} + s_{n-1}^2) \\ =& a_n^2 (s_n^2 + s_n s_{n-1} + s_{n-1}^2) \\ =& (a_n s_n)^2 + (a_n s_n)(a_n s_{n-1}) + (a_n s_{n-1})^2 \xrightarrow[n\to\infty]{} 3 \end{aligned}$$

since

$$a_n s_{n-1} = a_n n s_n - a_n^3 \xrightarrow[n\to\infty]{} 1.$$

A simple application of Cesàro-Stolz Lemma then gives

$$\lim \frac{s_n^3}{3n} = \lim \frac{s_n^3 - s_{n-1}^3}{3n - 3(n-1)} = \frac{3}{3} = 1,$$

as desired. $\square$

9. Show that the sequence $a_n = \sin n$ is divergent.

Proof. Suppose on the contrary that the given sequence is convergent and $a = \lim \sin n$; it must hold $|a| \leq 1$ since $|\sin n| \leq 1$ $\forall n \in \mathbb{N}$. We have

$$\sin(n+1) = \sin n \cos 1 + \cos n \sin 1$$

which solving for $\cos n$ gives

$$\cos n = \frac{\sin(n+1) - \sin n \cos 1}{\sin 1}$$

(we note that $\sin 1 \neq 0$). Letting $n \to \infty$ we get

$$b = \lim \cos n = \frac{a(1-\cos 1)}{\sin(n+1)}.$$

Next,

$$\cos(n+1) = \cos n \cos 1 - \sin n \sin 1 \xrightarrow[n\to\infty]{} b\cos 1 - a \sin 1,$$

therefore $b = b\cos 1 - a\sin 1$ which yields

$$a = -\frac{b(1-\cos 1)}{\sin 1} = -\frac{a(1-\cos 1)^2}{\sin^2 1},$$

hence $a = 0$ (since $\cos 1 \neq 1$) and so $b = 0$ as well. But

$$1 = \lim(\sin^2 n + \cos^2 n) = a^2 + b^2 = 0,$$

a contradiction. Thus, $(\sin n)$ is divergent. □

Remark. The sequence $(\sin n)$ is actually dense in $[0, 1]$, as a consequence of Kronecker's Lemma or Weyl's Criterion.

10. Let $(a_n) \subseteq \mathbb{R}$ satisfy the recurrence relation $a_{n+1} = a_n(1 - a_n)$, $\forall n \in \mathbb{N}$, with $a_1 \in (0, 1)$. Show that $\lim n a_n = 1$.

Proof. Equivalently, we want to show that

$$\lim \frac{1/a_n}{n} = 1.$$

We wish to apply the Cesàro-Stolz Lemma, so we estimate $\frac{1}{a_{n+1}} - \frac{1}{a_n}$:

$$\frac{1}{a_{n+1}} - \frac{1}{a_n} = \frac{a_n - a_{n+1}}{a_n a_{n+1}} = \frac{a_n^2}{a_n a_{n+1}} = \frac{a_n}{a_{n+1}} = \frac{1}{1-a_n},$$

using repeatedly the recurrence relation for (a_n). Next, we will show that $\lim a_n = 0$. First of all, (a_n) is bounded by hypothesis, and

$$a_{n+1} = a_n - a_n^2 < a_n, \quad \forall n \in \mathbb{N},$$

so that (a_n) is a strictly decreasing sequence, therefore (a_n) is convergent by Theorem 1.1.3. Let $a = \lim a_n \in [0, 1]$. Taking limits on the recurrence relation we obtain

$$a = \lim a_{n+1} = \lim a_n(1-a_n) = a(1-a),$$

whence $a = a - a^2$, thus $a = 0$. Applying the Cesàro-Stolz Lemma we finally get

$$\lim \frac{1/a_n}{n} = \lim \frac{\frac{1}{a_{n+1}} - \frac{1}{a_n}}{(n+1) - n} = \lim \frac{1}{1-a_n} = 1,$$

as desired. □

11. Let

$$S_n = \sum_{k=1}^{n} \left(\sqrt{1 + \frac{k}{n^2}} - 1 \right).$$

Show $\lim S_n = \frac{1}{4}$.

Proof. We first estimate the main term in the sum, from above and below:

$$\sqrt{1+\frac{k}{n^2}} - 1 = \frac{\left(\sqrt{1+\frac{k}{n^2}}-1\right)\left(\sqrt{1+\frac{k}{n^2}}+1\right)}{\sqrt{1+\frac{k}{n^2}}+1} = \frac{k/n^2}{\sqrt{1+\frac{k}{n^2}}+1} < \frac{k}{2n^2},$$

and

$$\sqrt{1+\frac{k}{n^2}} - 1 = \frac{k/n^2}{\sqrt{1+\frac{k}{n^2}}+1} \geq \frac{k/n^2}{\sqrt{1+\frac{1}{n^2}}+1},$$

since $1 \leq k \leq n$. Summing both inequalities over k, we obtain

$$\frac{1}{\sqrt{1+\frac{1}{n^2}}+1} \cdot \frac{1}{n^2} \sum_{k=1}^{n} k \leq S_n < \frac{1}{2n^2} \sum_{k=1}^{n} k,$$

which yields

$$\frac{1}{\sqrt{1+\frac{1}{n^2}}+1} \cdot \frac{n+1}{2n} \leq S_n < \frac{n+1}{4n},$$

since it is well-known that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$. Now we have

$$\lim \frac{1}{\sqrt{1+\frac{1}{n^2}}+1} \cdot \frac{n+1}{2n} = \frac{1}{4} = \lim \frac{n+1}{4n},$$

which implies $\lim S_n = \frac{1}{4}$ by the Squeeze Theorem. □

1.2 Series

Consider a sequence (a_n) of real numbers. Under which conditions can we take the sum of all the terms of the sequence,

$$a_1 + a_2 + a_3 + \cdots ?$$

This sum is denoted by $\sum_{n=1}^{\infty} a_n$. We proceed with the formal definition.

Definition 1.2.1. Let (s_n) be the sequence defined by

$$s_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k.$$

This is called the sequence of *partial sums* of the series $\sum_{n=1}^{\infty} a_n$. The series converges when (s_n) converges, in which case we define

$$\sum_{n=1}^{\infty} a_n = \lim s_n.$$

If (s_n) diverges, we say that $\sum_{n=1}^{\infty} a_n$ is divergent.

Theorem 1.2.2. *If the series $\sum_{n=1}^{\infty}$ is convergent, then* $\lim a_n = 0$.

Below are some classical examples of series related to the number $\pi = 3.14159...$

Example. 1.

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1}.$$

2.

$$\frac{\pi^2}{6} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^2}. \tag{1.2.1}$$

The most basic example of a series is the *geometric series*; let $a_n = x^n$, $n \geq 0$. Then

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots = \begin{cases} \frac{1}{1-x}, & |x| < 1 \\ \text{diverges}, & |x| \geq 1. \end{cases} \tag{1.2.2}$$

The following is a simple application of Theorem 1.1.3.

Proposition 1.2.3. *Let (a_n) be a sequence of nonnegative numbers. Then, either $\sum_{n=1}^{\infty} a_n$ converges or tends to $+\infty$.*

Remark. When we have series of nonnegative terms, the convergence of $\sum_{n=1}^{\infty} a_n$ will also be denoted as $\sum_{n=1}^{\infty} a_n < \infty$.

Usually, we ask whether a given series converges or not, rather than finding its sum. To facilitate this task, we compare a series with another one, whose convergence or divergence is already known. This method is mostly applied when the series involved consist of nonnegative terms, so that the sum of the series is well-defined by Proposition 1.2.3.

Theorem 1.2.4 (Comparison Tests)**.** *Suppose that $0 \leq a_n \leq b_n$, $\forall n \in \mathbb{N}$.*

(i) *If $\sum_{n=1}^{\infty} b_n < \infty$, then $\sum_{n=1}^{\infty} a_n < \infty$.*

(ii) *If $\sum_{n=1}^{\infty} a_n = \infty$, then $\sum_{n=1}^{\infty} b_n = \infty$.*

Theorem 1.2.5 (Cauchy Condensation Criterion)**.** *Let (a_n) be a decreasing sequence of nonnegative numbers, converging to 0. Then,*

$$\sum_{n=1}^{\infty} a_n < \infty \Longleftrightarrow \sum_{n=0}^{\infty} 2^n a_{2^n} < \infty.$$

In other words, under the above conditions, the series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=0}^{\infty} 2^n a_{2^n}$ have the same behavior in terms of convergence.

Example. Here we will tackle the convergence of the p-series, $\sum_{n=1}^{\infty} \frac{1}{n^p}$, $p > 0$. The sequence $a_n = \frac{1}{n^p}$ is obviously decreasing and converging to 0, so the behavior of the p-series is identical to that of the "condensed" series

$$\sum_{n=0}^{\infty} 2^n a_{2^n} = \sum_{n=0}^{\infty} \frac{2^n}{(2^n)^p} = \sum_{n=0}^{\infty} \frac{1}{(2^n)^{p-1}} = \sum_{n=0}^{\infty} \left(\frac{1}{2^{p-1}}\right)^n,$$

which is a geometric series, with $x = \frac{1}{2^{p-1}}$. The latter converges if and only if $2^{p-1} > 1$ by (1.2.2), or equivalently, $p > 1$. This, in particular, shows that the harmonic series $\sum_{n=1}^{\infty}$ diverges. Moreover, for $p \leq 0$ the p-series diverges due to Theorem 1.2.2.

The following theorems are criteria for convergence of series of nonnegative numbers.

Theorem 1.2.6 (Root Test). *Let (a_n) be a sequence of nonnegative numbers such that $(\sqrt[n]{a_n})$ converges, and let $\ell = \lim \sqrt[n]{a_n}$.*

(i) *If $\ell < 1$, then $\sum_{n=1}^{\infty} < \infty$.*

(ii) *If $\ell > 1$, then $\sum_{n=1}^{\infty} = \infty$.*

Theorem 1.2.7 (Ratio Test). *Let (a_n) be a sequence of nonnegative numbers such that $\lim\left|\frac{a_{n+1}}{a_n}\right| = \ell \in \mathbb{R}$.*

(i) *If $\ell < 1$, then $\sum_{n=1}^{\infty} < \infty$.*

(ii) *If $\ell > 1$, then $\sum_{n=1}^{\infty} = \infty$.*

Remark. Both of the above tests are inconclusive if $\ell = 1$. This is the case for the p-series.

As mentioned above, it is usually very difficult to compute the sum of a convergent series, even for some simple ones. For example, there is no known formula for the p-series for $p = 3$ (it is a famous Theorem by Apéry in 1978, that the sum is an irrational number). However, the sum can be computed quite easily when the given series can be written as a *telescopic* series. This is achieved if we can write the main term of a series, say a_n, as a difference of two consecutive terms of another sequence. This is akin to finding an antiderivative in order to compute a definite integral.

Proposition 1.2.8. *Let (a_n), (b_n) be two sequences such that $a_n = b_n - b_{n+1}$, and $b = \lim b_n$. Then, the series $\sum_{n=1}^{\infty}$ is convergent and*

$$\sum_{n=1}^{\infty} a_n = b_1 - b.$$

Exercises. 1. Calculate $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$.

Solution. We denote $a_n = \frac{1}{n(n+1)}$. It holds

$$a_n = \frac{(n+1) - n}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

So, for $b_n = \frac{1}{n}$ we have $a_n = b_n - b_{n+1}$ (i.e. we managed to write the given series as a telescopic series), with $b_1 = 1$ and $b = \lim b_n = 0$. Therefore, Proposition 1.2.8 yields

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1. \qquad \square$$

2. Let (a_n) be a sequence of real numbers satisfying the recurrence relation

$$a_{n+1} = \frac{a_n^2}{a_n^2 - a_n + 1}, \quad \forall n \in \mathbb{N}, \tag{1.2.3}$$

and $a_1 = \frac{1}{2}$. Prove that $\sum_{n=1}^{\infty} a_n$ converges and find its sum.

Proof. Our goal is to write the given series as a telescopic one, by manipulating the recurrence relation (1.2.3). It holds

$$\begin{aligned} a_{n+1} = 1 + \frac{a_n - 1}{a_n^2 - a_n + 1} &\iff a_{n+1} - 1 = \frac{a_n - 1}{a_n^2 - a_n + 1} \\ &\iff \frac{1}{a_{n+1} - 1} = a_n + \frac{1}{a_n - 1} \\ &\iff a_n = \frac{1}{1 - a_n} - \frac{1}{1 - a_{n+1}}. \end{aligned}$$

We put $b_n = \frac{1}{1-a_n}$, so that $a_n = b_n - b_{n+1}$, $\forall n \in \mathbb{N}$, and $b_1 = 2$. The steps above make sense, as long as we don't have division by zero, i.e. $a_n \neq 1$, $\forall n \in \mathbb{N}$. We will show that this is indeed the case:

$\underline{a_n > 0,\ \forall n \in \mathbb{N}}$: This can be shown inductively; it holds for $n = 1$ by hypothesis. Assuming it holds for $n = k$, we have

$$a_{k+1} = \frac{a_k^2}{\left(a_k - \frac{1}{2}\right) + \frac{3}{4}} > 0,$$

so it holds for $n = k + 1$ as well, therefore, for all $n \in \mathbb{N}$ by induction.

$\underline{a_n < 1,\ \forall n \in \mathbb{N}}$: It certainly holds for $n = 1$, by definition. Suppose it holds for $n = k$; then,

$$1 - a_{k+1} = \frac{1 - a_k}{a_k^2 - a_k + 1} = \frac{1 - a_k}{a_k(a_k - 1) + 1} > 1 - a_k > 0,$$

so it holds for $n = k + 1$ and thence for all $n \in \mathbb{N}$ by induction. Moreover, the above inequality shows $a_n > a_{n+1}$ for all $n \in \mathbb{N}$, i.e. (a_n) is strictly decreasing. Since (a_n) is bounded, it is also convergent by Theorem 1.1.3.

Let $\ell = \lim a_n$; since $0 < a_n \leq a_1 = \frac{1}{2}$, $\forall n \in \mathbb{N}$, it holds $0 \leq \ell \leq \frac{1}{2}$. Taking limits on (1.2.3) we obtain

$$\ell = \frac{\ell^2}{\ell^2 - \ell + 1} \iff \ell^3 - \ell^2 + \ell = \ell^2 \iff \ell(\ell - 1)^2 = 0,$$

which yields $\ell = 0$. Therefore,

$$b = \lim b_n = \frac{1}{1 - \lim a_n} = 1,$$

so by Proposition 1.2.8 it holds

$$\sum_{n=1}^{\infty} a_n = b_1 - b = 1.$$

□

3. Calculate

$$\sum_{n=1}^{\infty} \frac{n}{n^4 + n^2 + 1}.$$

Solution. First of all, we observe that the given series converges by Theorem 1.2.4, as

$$0 < \frac{n}{n^4 + n^2 + 1} < \frac{1}{n^3}, \quad \forall n \in \mathbb{N},$$

and $\sum_{n=1}^{\infty} \frac{1}{n^3} < \infty$ (p-series with $p = 3 > 1$). Finding the sum, is more difficult; we will write the given series as a telescopic one.

We will proceed by *Partial Fraction Decomposition*, by observing that

$$n^4 + n^2 + 1 = (n^2 + n + 1)(n^2 - n + 1).$$

We seek constants A and B satisfying

$$\frac{n}{n^4 + n^2 + 1} = \frac{An + B}{n^2 + n + 1} + \frac{Cn + D}{n^2 - n + 1}.$$

The above is equivalent to

$$\begin{aligned} \frac{n}{n^4 + n^2 + 1} &= \frac{(An + B)(n^2 - n + 1) + (Cn + D)(n^2 + n + 1)}{n^4 + n^2 + 1} \\ &= \frac{(A + C)n^3 + (-A + B + C + D)n^2 + (A - B + C + D)n + (B + D)}{n^4 + n^2 + 1}, \end{aligned}$$

which leads us to the following system of linear equalities:

$$A + C = 0 \qquad (1.2.4)$$
$$-A + B + C + D = 0 \qquad (1.2.5)$$
$$A - B + C + D = 1 \qquad (1.2.6)$$
$$B + D = 0. \qquad (1.2.7)$$

Equations (1.2.5) and (1.2.7) yield $-A + C = 0$, which combined with (1.2.4) gives

$$A = C = 0.$$

Thus, the above system reduces to

$$\begin{aligned} B + D &= 0 \\ -B + D &= 1, \end{aligned}$$

whose solution is $B = -\frac{1}{2}$, $D = \frac{1}{2}$, hence

$$\frac{n}{n^4 + n^2 + 1} = \frac{-1/2}{n^2 + n + 1} + \frac{1/2}{n^2 - n + 1} = \frac{1/2}{n^2 - n + 1} - \frac{1/2}{(n + 1)^2 - (n + 1) + 1}.$$

Therefore, the main term of the given series can be written as $b_n - b_{n+1}$, where

$$b_n = \frac{1/2}{n^2 - n + 1}.$$

Since $b_1 = \frac{1}{2}$, and

$$b = \lim b_n = 0,$$

we obtain

$$\sum_{n=1}^{\infty} \frac{n}{n^4 + n^2 + 1} = b_1 - b = \frac{1}{2},$$

by Proposition 1.2.8. □

4. **(SEEMOUS '14)** Let (x_n) be a sequence of real numbers satisfying the recurrence relation

$$x_{n+1} = \frac{x_n + 1 + \sqrt{x_n^2 + 2x_n + 5}}{2}, \quad \forall n \geq 2, \quad x_1 = 2.$$

Prove that $y_n = \sum_{k=1}^{n} \frac{1}{x_k^2 - 1}$ converges and find the limit.

Proof. For every $n \geq 2$ it holds

$$x_{n+1} > \frac{x_n + 1 + \sqrt{x_n^2 + 2x_n + 1}}{2} = \frac{x_n + 1 + |x_n + 1|}{2} \geq x_n + 1,$$

and hence can be easily shown that $x_n \geq n + 1$ holds by induction. Therefore, the given series has positive terms and

$$\sum_{n=1}^{\infty} \frac{1}{x_n^2 - 1} < \sum_{n=1}^{\infty} \frac{1}{n(n+2)} < \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

by virtue of Theorem 1.2.4. We denote the sum of the series by S.

Next, we observe that x_{n+1} is the larger solution of the quadratic equation

$$X^2 - (x_n + 1)X + 1 = 0.$$

Therefore, $x_{n+1}^2 - 1 = (x_n + 1)x_{n+1}$, which yields

$$\begin{aligned} \frac{1}{x_n^2 - 1} &= \frac{1/2}{x_n - 1} - \frac{1/2}{x_n + 1} \\ &= \frac{1/2}{x_n - 1} - \frac{1}{2} \cdot \frac{x_{n+1}}{x_{n+1}^2 - 1} \\ &= \frac{1/2}{x_n - 1} - \frac{1/2}{x_{n+1} - 1} - \frac{1/2}{x_{n+1}^2 - 1}, \end{aligned}$$

or equivalently,

$$\frac{1}{x_n^2 - 1} + \frac{1/2}{x_{n+1}^2 - 1} = \frac{1/2}{x_n - 1} - \frac{1/2}{x_{n+1} - 1}.$$

We sum both sides of the above equation for $n \in \mathbb{N}$; the left hand side gives

$$\sum_{n=1}^{\infty} \left[\frac{1}{x_n^2 - 1} + \frac{1/2}{x_{n+1}^2 - 1} \right] = \sum_{n=1}^{\infty} \frac{1}{x_n^2 - 1} + \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{x_{n+1}^2 - 1} = S + \frac{1}{2}\left(S - \frac{1}{3}\right) = \frac{3}{2}S - \frac{1}{6}.$$

The series of the terms on the right hand side is a telescopic one, hence

$$\sum_{n=1}^{\infty} \left[\frac{1/2}{x_n - 1} - \frac{1/2}{x_{n+1} - 1} \right] = \frac{1/2}{x_1 - 1} - \lim_{n \to \infty} \frac{1/2}{x_{n+1} - 1} = \frac{1}{2}.$$

Combining the above, we get $\frac{3}{2}S - \frac{1}{6} = \frac{1}{2}$, thus $S = \frac{4}{9}$. $\square$

5. Calculate

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{m!n!}{(m+n+2)!}.$$

Solution. We put

$$a_m = m! \sum_{n=0}^{\infty} \frac{n!}{(m+n+2)!}. \tag{1.2.8}$$

The main term in the above series can be rewritten as

$$\begin{aligned}
\frac{n!}{(m+n+2)!} &= \frac{1}{(n+1)\cdots(m+n+2)} \\
&= \frac{1}{m+1} \cdot \frac{(m+n+2)-(n+1)}{(n+1)\cdots(m+n+2)} \\
&= \frac{1}{m+1}\left[\frac{1}{(n+1)\cdots(m+n+1)} - \frac{1}{(n+2)\cdots(m+n+2)}\right] \\
&= \frac{1}{m+1}(b_n - b_{n+1}),
\end{aligned}$$

where

$$b_n = \frac{1}{(n+1)\cdots(m+n+1)}.$$

Therefore, the series in (1.2.8) can be written as a telescopic series:

$$a_m = \frac{m!}{m+1}\sum_{n=0}^{\infty}(b_n - b_{n+1}) = \frac{m!}{m+1}(b_0 - \lim b_n) = \frac{m!}{(m+1)\cdot(m+1)!} = \frac{1}{(m+1)^2}.$$

Thus, by (1.2.1) we obtain

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{m!n!}{(m+n+2)!} = \sum_{m=0}^{\infty} a_m = \sum_{m=0}^{\infty}\frac{1}{(m+1)^2} = \frac{\pi^2}{6}. \qquad \square$$

6. Show that

$$\sum_{n=1}^{9999} \frac{1}{(\sqrt{n}+\sqrt{n+1})(\sqrt[4]{n}+\sqrt[4]{n+1})} = 9.$$

Solution. We rewrite the main term of the sum:

$$\begin{aligned}
\frac{1}{(\sqrt{n}+\sqrt{n+1})(\sqrt[4]{n}+\sqrt[4]{n+1})} &= \frac{\sqrt[4]{(n+1)} - \sqrt[4]{(n)}}{(\sqrt{n}+\sqrt{n+1})(\sqrt{n+1}-\sqrt{n})} \\
&= \frac{\sqrt[4]{(n+1)} - \sqrt[4]{(n)}}{(n+1)-n} \\
&= \sqrt[4]{(n+1)} - \sqrt[4]{(n)}.
\end{aligned}$$

Therefore, the given sum is telescopic, and equals

$$\sum_{n=1}^{9999}\left(\sqrt[4]{(n+1)} - \sqrt[4]{(n)}\right) = \sqrt[4]{10000} - \sqrt[4]{1} = 9. \qquad \square$$

7. **(IMC '19)** Calculate

$$\prod_{n=3}^{\infty} \frac{(n^3+3n)^2}{n^6-64}.$$

Solution. We factorize numerator and denominator as follows:

$$\frac{(n^3+3n)^2}{n^6-64} = \frac{n}{n-2} \cdot \frac{n}{n+2} \cdot \frac{n^2+3}{(n-1)^2+3} \cdot \frac{n^2+3}{(n+1)^2+3}.$$

We separated the fractions as above, because the corresponding infinite products are telescopic. Indeed,

$$\begin{aligned}\prod_{n=3}^{N} \frac{n}{n-2} &= \frac{3 \cdot 4 \cdots (N-1)N}{1 \cdot 2 \cdots (N-3)(N-2)} \\ &= \frac{N(N-1)}{2},\end{aligned}$$

$$\begin{aligned}\prod_{n=3}^{N} \frac{n}{n+2} &= \frac{3 \cdot 4 \cdots (N-1)N}{5 \cdot 6 \cdots (N+1)(N+2)} \\ &= \frac{12}{(N+1)(N+2)},\end{aligned}$$

$$\begin{aligned}\prod_{n=3}^{N} \frac{n^2+3}{(n-1)^2+3} &= \frac{(3^2+3)(4^2+3)\cdots((N-1)^2+3)(N^2+3)}{(2^2+3)(3^2+3)\cdots((N-2)^2+3)((N-1)^2+3)} \\ &= \frac{N^2+3}{7},\end{aligned}$$

$$\begin{aligned}\prod_{n=3}^{N} \frac{n^2+3}{(n+1)^2+3} &= \frac{(3^2+3)(4^2+3)\cdots((N-1)^2+3)(N^2+3)}{(4^2+3)(5^2+3)\cdots((N^2+3)((N+1)^2+3)} \\ &= \frac{12}{(N+1)^2+3},\end{aligned}$$

hence

$$\begin{aligned}\prod_{n=3}^{N} \frac{(n^3+3n)^2}{n^6-64} &= \frac{N(N-1)}{2} \cdot \frac{12}{(N+1)(N+2)} \cdot \frac{N^2+3}{7} \cdot \frac{12}{(N+1)^2+3} \\ &\xrightarrow[N\to\infty]{} \frac{12^2}{2 \cdot 7} = \frac{72}{7}.\end{aligned}$$ □

Chapter 2

Continuity, Derivatives, and Integrals

We recall some basic definitions; a function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* at $x_0 \in \mathbb{R}$ if $\lim_{x \to x_0} f(x) = f(x_0)$ ($\forall \varepsilon > 0\ \exists \delta > 0$ such that $|f(x) - f(x_0)| < \varepsilon$ for every $x \in \mathbb{R}$ satisfying $|x - x_0| < \delta$). $f(x)$ is *differentiable* at x_0 if the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists; when it does, it is denoted by $f'(x_0)$, and is called the *derivative* of f at x_0.

If $F(x)$ is such that $F'(x) = f(x)$, then F is called the *antiderivative* of f. F is unique up to a constant, satisfying this property; F is also called the *indefinite integral* of f, and we write

$$F(x) = \int f(x) \mathrm{d}x + C.$$

If $f : [a, b] \to \mathbb{R}$ is a positive continuous function ($f(x) > 0, \forall x \in [a, b]$), then the *definite integral*

$$\int_a^b f(x) \mathrm{d}x$$

denotes the area between the graph of $y = f(x)$ and the $x - axis$, for $a \leq x \leq b$. If f is a negative function, it is minus this area, and is general, the definite integral denotes the *signed* area between the graph of $y = f(x)$ and the $x - axis$. The Fundamental Theorem of Calculus connects the notion of definite and indefinite integral.

Theorem 2.0.1 (Fundamental Theorem of Calculus)**.** *Let $f : [a, b] \to \mathbb{R}$ be a differentiable function, such that its derivative $f'(x)$ is continuous at every point of $[a, b]$. Then,*

$$\int_a^b f'(x) \mathrm{d}x = f(b) - f(a).$$

2.1 Intermediate and Mean Value Theorems for Derivatives

Theorem 2.1.1 (Continuity)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and let $m \in \mathbb{R}$ be such that $f(a) < m < f(b)$. Then, there is a $x_0 \in (a, b)$ such that $f(x_0) = m$.*

The special case of the above Theorem with $m = 0$, is usualy referred to as *Bolzano's Theorem.*

Theorem 2.1.2 (Rolle's Theorem)**.** *Let f be a differentiable function on $[a, b]$ satisfying $f(a) = f(b)$. Then, there is $x_0 \in (a, b)$, such that $f'(x_0) = 0$.*

Theorem 2.1.3 (Mean Value Theorem for Derivatives)**.** *Let f be a differentiable function on $[a, b]$. There is $x_0 \in (a, b)$ such that*

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}.$$

Theorem 2.1.4 (Darboux's Theorem)**.** *Let f be a differentiable function on $[a, b]$, such that $f'(a) < m < f'(b)$ for some $m \in \mathbb{R}$. Then, there is $x_0 \in (a, b)$, such that $f'(x_0) = m$.*

Remark. If f' were continuous, Theorem 2.1.4 would just be a simple consequence of Theorem 2.1.1. However, there are cases where the derivative f' need not be continuous; in fact, it might not even be Riemann integrable!

Exercises. 1. **(IMC '19)** Suppose $f, g : \mathbb{R} \to \mathbb{R}$. Let f be continuous and g differentiable on $\mathbb{R}$. Assume

$$(f(0) - g'(0))(g'(1) - f(1)) > 0.$$

Show that $\exists c \in (0, 1)$ such that $f(c) = g'(c)$.

Proof. The function F defined by

$$F(x) = \int_0^x f(t)\mathrm{d}t, \quad \forall x \in \mathbb{R},$$

satisfies $F'(x) = f(x)$ by Theorem 2.0.1. By hypothesis, we then have

$$(F'(0) - g'(0))(F'(1) - g(1)) < 0. \tag{2.1.1}$$

Let $h = F' - g' = (F - g)'$; (2.1.1) can be rewritten as $h'(0)h'(1) < 0$, so that 0 is strictly between $h'(0)$ and $h'(1)$ (i.e. it is an *intermediate value*). Then, by Theorem 2.1.4 we obtain the existence of $c \in (0, 1)$ such that $h'(c) = 0$, or equivalently, $F'(c) - g(c) = 0$, which eventually yields $f(c) = g'(c)$, as desired. □

2. Let $f : (-1, 1) \to \mathbb{R}$ be a differentiable function, such that $f(0) = 0$, and

$$|f(x)| \leq |x|^{1/3}, \quad \forall x \in (-1, 1).$$

Show that the series $\sum_{n=1}^{\infty} f\left(\frac{1}{n}\right)$ converges.

Proof. By Theorem 2.1.3, for each $n \in \mathbb{N}$ there is $x_n \in \left(0, \frac{1}{n}\right)$ such that

$$f'(x_n) = \frac{f\left(\frac{1}{n}\right) - f(0)}{\frac{1}{n}},$$

therefore,

$$\left|f\left(\frac{1}{n}\right)\right| = \frac{1}{n}|f'(x_n)| \leq \frac{\sqrt[3]{x_n}}{n} \leq \frac{1}{n^{4/3}},$$

thus,

$$\sum_{n=1}^{\infty} f\left(\frac{1}{n}\right) < \sum_{n=1}^{\infty} \frac{1}{n^{4/3}} < \infty,$$

by Theorem 1.2.4, since the latter series is a p-series with $p = 4/3 > 1$. $\square$

3. **(IMC Selection AUTh '19)** Let $(a_n) \subseteq \mathbb{R}$, such that $0 < a_n < 1$, $\forall n \in \mathbb{N}$, and

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty.$$

Suppose that $f : [0,1] \to \mathbb{R}$ is twice differentiable and $|f''(x)| \leq M$, $\forall x \in [0,1]$. Show that if $\sum_{n=1}^{\infty} f(a_n)$ converges, then $\sum_{n=1}^{\infty} |f(a_n)|$ converges as well.

Proof. First of all, convergence of $\sum_{n=1}^{\infty} a_n^2$ implies that $\lim a_n^2 = 0$ by Theorem 1.2.2, hence $\lim a_n = 0$. Since $\sum_{n=1}^{\infty} f(a_n)$ converges as well, we shall have $\lim f(a_n) = 0$, hence by continuity of f,

$$0 = \lim f(a_n) = f(\lim a_n) = f(0).$$

By Theorem 2.1.3, there is $b_n \in (0, a_n)$ such that

$$f'(b_n) = \frac{f(a_n) - f(0)}{a_n},$$

for every $n \in \mathbb{N}$, or equivalently, $f(a_n) = a_n f'(b_n)$, $\forall n \in \mathbb{N}$. Since $0 < b_n < a_n$, $\forall n \in \mathbb{N}$, and $\lim a_n = 0$, it must also hold $\lim b_n = 0$ by the Squeeze Theorem. By hypothesis, f' is continuous, therefore

$$f'(0) = f'(\lim b_n) = \lim f'(b_n).$$

Suppose that $f'(0) \neq 0$, say $f'(0) > 0$, without loss of generality. Then, there is $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$f'(b_n) > \delta, \quad \forall n \geq n_0.$$

This is impossible, since $\sum_{n=1}^{\infty} f(a_n)$ converges; indeed, for $n \geq n_0$ we obtain

$$f(a_n) = a_n f'(b_n) > \delta a_n,$$

and the series $\sum n = n_0 \infty (\delta a_n)$ diverges by hypothesis, so $\sum_{n=1}^{\infty} f(a_n)$ must diverge if $f'(0) > 0$, by Theorem 1.2.4, contradiction. Hence, $f'(0) = 0$, and for each $n \in \mathbb{N}$ there is $c_n \in (0, b_n)$ such that

$$f''(c_n) = \frac{f'(b_n)}{b_n},$$

by Theorem 2.1.3. Therefore,

$$f(a_n) = a_n f'(b_n) = a_n b_n f''(c_n), \quad \forall n \in \mathbb{N}.$$

But then,

$$|f(a_n)| = a_n b_n |f''(c_n)| \leq M a_n^2,$$

ans since $\sum_{n=1}^{\infty} a_n^2$ converges, the series $\sum_{n=1}^{\infty} |f(a_n)|$ must also converge by Theorem 1.2.4. $\square$

2.2 Mean Value Theorems for Integrals

We recall that a function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if f is discontinuous on a set of measure zero (*Lebesgue's Theorem*); examples of set zero are finite sets and countable sets. A notable example of an uncountable set of measure zero is the Cantor set. We remark that a monotonous function on $[a, b]$ is always integrable, as it has at most countably many discontinuities.

Theorem 2.2.1 (Mean Value Theorem for Integrals). *Let f be a continuous real function on $[a, b]$. Then, there is $x_0 \in [a, b]$ such that*

$$\int_a^b f(x)\mathrm{d}x = f(x_0)(b - a).$$

Also,

$$m(b - a) \leq \int_a^b f(x)\mathrm{d}x \leq M(b - a),$$

where $m = \min\{f(x) : x \in [a, b]\}$, $M = \max f(x) : x \in [a, b]$.

Remark. The fact that f assumes both its minimal as well as its maximal value on $[a, b]$, follows from the theorem in Topology which states that a continuous real function defined on a compact set (i.e. a closed and bounded set, in case of $\mathbb{R}$) always attains its minimum and maximum value. In general, if f is Riemann integrable, it holds

$$(b - a) \inf_{x\in[a,b]} f(x) \leq \int_a^b f(x)\mathrm{d}x \leq (b - a) \sup_{x\in[a,b]} f(x)$$

Corollary 2.2.2. *If f is a Riemann integrable function, satisfying $f(x) \geq 0$, $\forall x \in [a, b]$, then*

$$\int_a^b f(x)\mathrm{d}x \geq 0.$$

If f is continuous, equality holds if and only if $f(x) = 0$, $\forall x \in [a, b]$.

Exercises. 1. Find all continuous functions $f : [0, 1] \to \mathbb{R}$ such that

$$\int_0^1 f(x)\mathrm{d}x = \frac{1}{3} + \int_0^1 f(x^2)^2\mathrm{d}x.$$

Solution. With the substitution $x = u^2$, we obtain $\mathrm{d}x = 2u\mathrm{d}u$ and the integral on the left hand side becomes $\int_0^1 2uf(u^2)\mathrm{d}u$. Observing that $\frac{1}{3} = \int_0^1 x^2\mathrm{d}x$, and reverting u to x, the given equality becomes

$$\begin{aligned}\int_0^1 2xf(x)\mathrm{d}x = \int_0^1 x^2\mathrm{d}x + \int_0^1 f(x^2)^2\mathrm{d}x &\iff \int_0^1 \left[f(x^2)^2 - 2xf(x^2) + x^2\right]\mathrm{d}x \\ &\iff \int_0^1 \left(f(x^2) - x^2\right)\mathrm{d}x = 0 \\ &\iff f(x^2) = x, \forall x \in [0, 1] \text{ (by Corollary 2.2.2)} \\ &\iff f(x) = \sqrt{x}, \forall x \in [0, 1].\end{aligned}$$

□

2. Find all continuous functions such that

$$\int_0^1 f(x)(x - f(x))\mathrm{d}x = \frac{1}{12}.$$

Solution. The integrand is quadratic in terms of $f(x)$; completing the square we get

$$f(x)^2 - xf(x) + \frac{x^2}{4} = \Big(f(x) - \frac{x}{2}\Big)^2,$$

so,

$$\begin{aligned} 0 = \frac{1}{12} + \int_0^1 f(x)(f(x)-x)\mathrm{d}x &= \frac{1}{12} + \int_0^1 \left[\Big(f(x) - \frac{x}{2}\Big)^2 - \frac{x^2}{4}\right]\mathrm{d}x \\ &= \frac{1}{12} + \int_0^1 \Big(f(x) - \frac{x}{2}\Big)^2 \mathrm{d}x - \int_0^1 \frac{x^2}{4}\mathrm{d}x \\ &= \int_0^1 \Big(f(x) - \frac{x}{2}\Big)^2 \mathrm{d}x, \end{aligned}$$

since $\int_0^1 \frac{x^2}{4}\mathrm{d}x = \frac{1}{12}$. The continuity of the integrand in the last integral implies that $f(x) = \frac{x}{2}$, $\forall x \in [0,1]$, by Corollary 2.2.2. □

3. Let $f : [0,1] \to \mathbb{R}$ be a continuous function such that

$$\int_0^1 f(x)\mathrm{d}x = \int_0^1 xf(x)\mathrm{d}x = 1.$$

Prove that

$$\int_0^1 f(x)^2\mathrm{d}x \geq 4.$$

Proof. It holds

$$\int_0^1 (f(x) - (\alpha x + \beta))^2 \mathrm{d}x \geq 0,$$

for every $\alpha, \beta \in \mathbb{R}$, by Theorem 2.2.1. Expanding the square, the above integral equals

$$\begin{aligned} &\int_0^1 f(x)^2\mathrm{d}x - 2\int_0^1 (\alpha x + \beta)f(x)\mathrm{d}x + \int_0^1 (\alpha x+\beta)^2\mathrm{d}x = \\ &\qquad\int_0^1 f(x)^2\mathrm{d}x - 2(\alpha+\beta) + \frac{\alpha^2}{3} + \alpha\beta + \beta^2, \end{aligned}$$

therefore,

$$\int_0^1 f(x)^2\mathrm{d}x - 4 \geq -\left[\frac{\alpha^2}{3} + \alpha\beta + \beta^2 - 2(\alpha+\beta) + 4\right], \quad \forall \alpha, \beta \in \mathbb{R}. \qquad (2.2.1)$$

It suffices to show that the minimum value of the quadratic form

$$\frac{\alpha^2}{3} + \alpha\beta + \beta^2 - 2(\alpha+\beta) + 4$$

is ≤ 0. The above expression is a trinomial with respect to β;

$$\beta^2 + (\alpha - 2)\beta + \left(\frac{\alpha^2}{3} - 2\alpha + 4\right).$$

We recall that the minimum value of $ax^2 + bx + c$ for $a > 0$, is obtained for $x = -\frac{b}{2a}$. Therefore, the minimum value of the above trinomial is obtained for $\beta = \frac{2-\alpha}{2} = 1 - \frac{\alpha}{2}$, which equals

$$1 - \alpha + \frac{\alpha^2}{4} - \frac{\alpha^2}{2} + 2\alpha - 2 + \frac{\alpha^2}{3} - 2\alpha + 4 = \frac{\alpha^2}{12} - \alpha + 3.$$

The last is again, simply a trinomial, whose minimum value is attained for $\alpha = 6$ and equals

$$\frac{36}{12} - 6 + 3 = 0,$$

as desired. One could also argue, that for the specific choice of $\alpha = 6$ and $\beta = -2$, (2.2.1) becomes

$$\int_0^1 f(x)^2 \mathrm{d}x - 4 \geq 0. \qquad \square$$

2.3 Integral Inequalities

We recall some famous inequalities with integrals. In the following, I denotes either a bounded or an undounded interval (could be $[a, b]$, or even $\mathbb{R}$, for example).

Theorem 2.3.1 (Cauchy-Schwarz Inequality). *Let f, g be two functions such that f^2, g^2 are integrable on I. Then, fg is also integrable on I, and it holds*

$$\left(\int_I f(x)g(x)\mathrm{d}x\right)^2 \leq \int_I f(x)^2\mathrm{d}x \cdot \int_I g(x)^2\mathrm{d}x.$$

If f and g are continuous, equality is obtained if there are $\lambda, \mu \in \mathbb{R}$, not both zero, such that $\lambda f(x) = \mu g(x)$, $\forall x \in I$ (i.e. f and g are linearly dependent).

The Cauchy-Schwarz inequality holds in any vector space with an inner product (for example, a Hilbert space). For finite dimensional vector spaces it becomes the following inequality:

$$\sqrt{x_1^2 + \cdots + x_n^2}\sqrt{y_1^2 + \cdots + y_n^2} \geq x_1y_1 + \cdots + x_ny_n, \tag{2.3.1}$$

for every $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}$.

Theorem 2.3.2 (Minkowski Inequality). *Let f, g be two functions such that $|f|^p$, $|g|^p$ are integrable on I, with $p \geq 1$. Then, $|f+g|^p$ is also integrable on I, and it holds*

$$\left(\int_I |f(x) + g(x)|^p\mathrm{d}x\right)^{\frac{1}{p}} \leq \left(\int_I |f(x)|^p\mathrm{d}x\right)^{\frac{1}{p}} \left(\int_I |g(x)|^p\mathrm{d}x\right)^{\frac{1}{p}}.$$

Theorem 2.3.3 (Hölder Inequality). *Let f, g be two functions such that $|f|^p$, $|g|^q$ are integrable on I, where $p, q > 1$, and $\frac{1}{p} + \frac{1}{q} = 1$. Then, $|fg|$ is also integrable on I, and it holds*

$$\int_I |f(x)g(x)|\mathrm{d}x \leq \left(\int_I |f(x)|^p\mathrm{d}x\right)^{\frac{1}{p}} \left(\int_I |g(x)|^q\mathrm{d}x\right)^{\frac{1}{q}}.$$

Theorem 2.3.4 (Chebyshev's Inequality)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be two monotonic functions with the same monotonicity (i.e. they are both increasing or both decreasing). Then,*

$$(b-a)\int_a^b f(x)g(x)\mathrm{d}x \geq \int_a^b f(x)\mathrm{d}x \cdot \int_a^b g(x)\mathrm{d}x.$$

If f and g are of opposite monotonicity, then the inequality is reversed.

Exercises. 1. Let $f : [0, 1] \to \mathbb{R}$ be continuous. Show that

$$\left(\int_0^1 f(t)\mathrm{d}t\right)^2 \leq \int_0^1 f(t)^2\mathrm{d}t.$$

Proof. We apply the Cauchy-Schwarz inequality (Theorem 2.3.1) for $g(t) \equiv 1$, and obtain the desired inequality. □

2. Find the maximum value of

$$\frac{\left(\displaystyle\int_0^3 f(x)\mathrm{d}x\right)^3}{\displaystyle\int_0^3 f(x)^3\mathrm{d}x},$$

for $f : [0, 3] \to (0, +\infty)$ continuous.

Solution. Apply Hölder's inequality (Theorem 2.3.3) for $g(x) \equiv 1$, $p = 3$, $q = \frac{3}{2}$, to obtain

$$\int_0^3 f(x)\mathrm{d}x \leq \left(\int_0^3 f(x)^3\mathrm{d}x\right)^{1/3}\left(\int_0^3 \mathrm{d}x\right)^{2/3},$$

or equivalently,

$$\frac{\left(\displaystyle\int_0^3 f(x)\mathrm{d}x\right)^3}{\displaystyle\int_0^3 f(x)^3\mathrm{d}x} \leq 9.$$

We observe that equality is obtained for $f(x) \equiv 1$, so the maximum value is 9. □

3. Let f be an increasing function on $[0, 1]$. Prove that for every $\alpha \in (0, 1)$ it holds

$$\alpha\int_0^1 f(x)\mathrm{d}x \geq \int_0^\alpha f(x)\mathrm{d}x.$$

Proof. Using the substitution $x = \alpha y$ on the second integral, we obtain

$$\int_0^\alpha f(x)\mathrm{d}x = \alpha\int_0^1 f(\alpha y)\mathrm{d}y \leq \alpha\int_0^1 f(y)\mathrm{d}y,$$

as desired, as the inequality $f(\alpha y) \leq f(y)$ holds for all $y \in [0, 1]$ due to the monotonicity of f. □

4. Let $f : [a,b] \to \mathbb{R}$ be a continuously differentiable function with $f(a) = 0$. Show that
$$\int_a^b f(x)^2 \mathrm{d}x \le (b-a)^2 \int_a^b f'(x)^2 \mathrm{d}x.$$

Proof. By the Fundamental Theorem of Calculus (Theorem 2.0.1) we have
$$f(x) = \int_a^x f'(t)\mathrm{d}t,$$
dut to the continuity of f' and the hypothesis $f(a) = 0$. Squaring and applying the Cauchy-Schwarz inequality (Theorem 2.3.1), we get
$$f(x)^2 = \left(\int_a^x f'(t)\mathrm{d}t\right)^2 \le (x-a)\int_a^x f'(t)^2\mathrm{d}t \le (b-a)\int_a^x f'(t)^2\mathrm{d}t.$$
Integrating with respect to x the first and the last expression on $[a,b]$, we obtain the desired inequality (the last expression is independent of x). ☐

<u>*Challenge*</u> Show that the above inequality holds with $\frac{(b-a)^2}{2}$ instead of $(b-a)$.

5. **(SEEMOUS '13)** Find the maximum value of
$$\int_0^1 |f'(x)|^2 |f(x)| \frac{1}{\sqrt{x}} \mathrm{d}x,$$
over all functions $f : [0,1] \to \mathbb{R}$ with continuous derivative, satisfying $f(0) = 0$ and
$$\int_0^1 |f'(x)|^2 \mathrm{d}x \le 1.$$

Solution. Define
$$g(x) = \int_0^x |f'(t)|^2 \mathrm{d}t, \quad x \in [0,1],$$
so that $g'(x) = |f'(x)|^2$, by the Fundamental Theorem of Calculus. Furthermore,
$$|f(x)| \le \int_0^x |f'(t)|\mathrm{d}t \le \left(\int_0^x f'(t)^2 \mathrm{d}t\right)^{1/2} \sqrt{x}, \tag{2.3.2}$$
by applying the Cauchy-Schwarz inequality. The latter gives
$$\frac{|f(x)|}{\sqrt{x}} \le \sqrt{g(x)}.$$

Thus,
$$\begin{aligned}
\int_0^1 |f'(x)|^2 |f(x)| \frac{1}{\sqrt{x}} \mathrm{d}x &\le \int_0^1 g(x)^{1/2} g'(x) \mathrm{d}x \\
&= \int_0^{g(1)} \sqrt{u}\mathrm{d}u = \left[\frac{u^{3/2}}{3/2}\right]_0^{g(1)} \\
&= \frac{2}{3} g(1)^{3/2} = \frac{2}{3}\left(\int_0^1 |f'(t)|^2 \mathrm{d}t\right)^{3/2} \\
&\le \frac{2}{3},
\end{aligned}$$

where we applied the substitution $u = g(x)$. Next, we ask whether equalities can be obtained, and have $\frac{2}{3}$ as the sought after maximum value. We check the inequalities in (2.3.2). The first one becomes an equality precisely when $|f|$ is increasing; the conclusion and hypothesis do not change when we change f to $-f$, so we may assume that f is increasing, hence $f(x) \geq 0$, $\forall x \in [0, 1]$. The second inequality in (2.3.2) becomes an equality when f' and the constant function equal to 1 are linearly dependent on $[0, x]$ for every $x \in [0, 1]$ by Theorem 2.3.1. This shows that f' is a constant, hence $f(x) = cx$, for some $c > 0$. We observe that $\int_0^1 |f'(x)|^2 \mathrm{d}x \leq 1$ implies $c \leq 1$; for $f(x) = x$ we obtain

$$\int_0^1 |f'(x)|^2 |f(x)| \frac{1}{\sqrt{x}} \mathrm{d}x = \frac{2}{3},$$

proving finally that $\frac{2}{3}$ is the desired maximum value. □

6. **(SEEMOUS '11)** Let $n \in \mathbb{N}$ and $f : [0, 1] \to \mathbb{R}$ be an increasing function. Prove that

$$\int_0^1 f(x)\mathrm{d}x \leq (n+1) \int_0^1 x^n f(x)\mathrm{d}x$$

holds. Find all increasing continuous functions for which equality holds.

Proof. Apply the substitution $y = x^{n+1}$ on the second integral, to obtain

$$(n+1) \int_0^1 x^n f(x)\mathrm{d}x = \int_0^1 f(\sqrt[n+1]{y})\mathrm{d}y \geq \int_0^1 f(y)\mathrm{d}y,$$

due to the monotonicity of f, since $f\left(\sqrt[n+1]{y}\right) \geq f(y)$, $\forall y \in [0, 1]$. If f is continuous, then

$$\int_0^1 (f(\sqrt[n+1]{y} - f(y)))\mathrm{d}y \geq 0,$$

with equality if and only if

$$f(\sqrt[n+1]{y}) = f(y), \quad \forall y \in [0, 1],$$

by Corollary 2.2.2. Now, let $\alpha \in (0, 1]$ be arbitrary; for every $m \in \mathbb{N}$ it holds

$$f(\alpha) = f\left(\alpha^{\frac{1}{n+1}}\right) = f\left(\alpha^{\frac{1}{(n+1)^2}}\right) = \cdots = f\left(\alpha^{\frac{1}{(n+1)^m}}\right).$$

By continuity, and the fact that $\lim_{m\to\infty} \alpha^{1/(n+1)^m} = 1$, it holds

$$f(\alpha) = \lim_{m\to\infty} f\left(\alpha^{\frac{1}{(n+1)^m}}\right) = f(1).$$

Since α is arbitrary, we conclude that f is constant on $(0, 1]$; by continuity of f, we must also have $f(0) = f(1)$, thus the only continuous increasing functions satisfying equality are only the constant ones. □

7. **(SEEMOUS '15)** Prove that for every $x \in (0, 1)$ the following inequality holds:

$$\int_0^1 \sqrt{1 + \cos^2 y}\mathrm{d}y > \sqrt{x^2 + \sin^2 x}.$$

Proof. Consider the function $g(x) = x^2 + \sin^2 x$, $x \in [0, 1]$. Its derivative is $g'(x) = 2x + sin2x$, which is positive for $x > 0$ (it follows from $|\sin x| \leq |x|$, with equality if and only if $x = 0$), therefore g is strictly increasing and its maximum value is $g(1) = 1 + \sin^2 1$. It suffices then to show that

$$\int_0^1 \sqrt{1 + \cos^2 y}\mathrm{d}y \geq \sqrt{1 + \sin^2 1}. \tag{2.3.3}$$

The Cauchy-Schwarz inequality for $x_1 = 1$, $x_2 = \cos^2 y$, $y_1 = 1$, $y_2 = \sin^2 1$ (see (2.3.1)) yields

$$\begin{aligned}\int_0^1 \sqrt{1 + \cos^2 y}\sqrt{1 + \sin^2 1}\mathrm{d}y &\geq \int_0^1 (1 + \sin 1 \cos y)\mathrm{d}y \\ &= 1 + \sin 1 \int_0^1 \cos y \mathrm{d}y \\ &= 1 + \sin^2 1.\end{aligned}$$

Dividing by $\sqrt{1 + \sin^2 1}$ we obtain (2.3.3), as desired. □

Second proof. It is known that the length of the graph of $y = f(x)$ for $a \leq x \leq b$, where f is a differentiable function, equals

$$\int_a^b \sqrt{1 + f'(x)^2}\mathrm{d}x. \tag{2.3.4}$$

For this problem, consider the function $f(x) = \sin x$, $x \in [0, 1]$, whose derivative is $f'(x) = \cos x$. The length of the graph $y = f(t)$, $t \in [0, x]$, is greater than the straight line connecting the endpoints $(0, 0)$ and $(x, \sin x)$. Therefore,

$$\int_0^1 \sqrt{1 + \cos^2 t}\mathrm{d}t > \int_0^x \sqrt{1 + \cos^2 t}\mathrm{d}t \geq \sqrt{x^2 + \sin^2 x}.$$ □

8. **(SEEMOUS '16)** Let f be a continuous decreasing function on $\left[0, \frac{\pi}{2}\right]$. Prove

$$\int_{\frac{\pi}{2}-1}^{\frac{\pi}{2}} f(x)\mathrm{d}x \leq \int_0^{\frac{\pi}{2}} f(x) \cos x \mathrm{d}x \leq \int_0^1 f(x)\mathrm{d}x.$$

Proof. The second inequality is obtained via the substitution $y = \sin x$ (so that $\mathrm{d}y = \cos x \mathrm{d}x$):

$$\int_0^{\frac{\pi}{2}} f(x) \cos x \mathrm{d}x \leq \int_0^{\frac{\pi}{2}} f(\sin x) \cos x \mathrm{d}x = \int_0^1 f(y)\mathrm{d}y,$$

since $f(x) \leq f(\sin x)$, $\forall x \geq 0$, due to the fact that f is decreasing.

Next, we apply Chebyshev's Inequality (Theorem 2.3.4) on $f(x)$ and $\cos x$, which are both decreasing on $\left[0, \frac{\pi}{2}\right]$, to obtain

$$\begin{aligned}\int_0^{\frac{\pi}{2}} f(x)\cos x \mathrm{d}x &\geq \frac{2}{\pi}\int_0^{\frac{\pi}{2}} f(x)\mathrm{d}x \int_0^{\frac{\pi}{2}} \cos x \mathrm{d}x \\ &= \frac{2}{\pi}\int_0^{\frac{\pi}{2}-1} f(x)\mathrm{d}x + \frac{2}{\pi}\int_{\frac{\pi}{2}-1}^{\frac{\pi}{2}} f(x)\mathrm{d}x \\ &\geq \frac{2}{\pi}\Big(\frac{\pi}{2}-1\Big) f\Big(\frac{\pi}{2}-1\Big) + \frac{2}{\pi}\int_{\frac{\pi}{2}-1}^{\frac{\pi}{2}} f(x)\mathrm{d}x \\ &\geq \frac{2}{\pi}\Big(\frac{\pi}{2}-1\Big)\int_{\frac{\pi}{2}-1}^{\frac{\pi}{2}-1} f(x)\mathrm{d}x + \frac{2}{\pi}\int_{\frac{\pi}{2}-1}^{\frac{\pi}{2}} f(x)\mathrm{d}x \\ &= \int_{\frac{\pi}{2}-1}^{\frac{\pi}{2}} f(x)\mathrm{d}x,\end{aligned}$$

due to the monotonicity of f and Theorem 2.2.1.

Equality in the first inequality holds precisely all the inequalities above become equalities. Since f is continuous, this holds precisely when f is constant. Equality in the second inequality holds if and only if

$$f(x) = f(\sin x), \quad \forall x \in \Big[0, \frac{\pi}{2}\Big],$$

by continuity of f. Consider $\alpha \in \left[0, \frac{\pi}{2}\right]$; then,

$$f(\alpha) = f(\sin\alpha) = f(\sin\sin\alpha) = \dots .$$

Cosnider the sequence (x_n) with $x_0 = \alpha$ and $x_n = \sin x_{n-1}$. This is a decreasing sequence of nonnegative terms, thus it converges to $\ell \geq 0$, by Theorem 1.1.3, which satisfies $\ell = \sin\ell$, whence $\ell = 0$. Therefore, $f(\alpha) = f(0)$ for every $\alpha \in \left[0, \frac{\pi}{2}\right]$, which eventually shows that the second inequality holds precisely when f is constant. □

9. **(SEEMOUS '18)** Let $f : [0,1] \to (0,1)$ be Riemann integrable. Show that

$$\frac{2\displaystyle\int_0^1 xf(x)^2\mathrm{d}x}{\displaystyle\int_0^1 (f(x)^2+1)\mathrm{d}x} < \frac{\displaystyle\int_0^1 f(x)^2\mathrm{d}x}{\displaystyle\int f(x)\mathrm{d}x}.$$

Proof. For every $x \in [0,1]$ it holds $2f(x) < 1 + f(x)^2$ (since $2a \leq 1 + a^2$ holds for every $a \in \mathbb{R}$, which is equivalent to $0 \leq (a-1)^2$, and equality holds only if $a = 1$), and $xf(x)^2 \leq f(x)^2$, with equality precisely when $x = 1$. Therefore,

$$\frac{2\displaystyle\int_0^1 xf(x)^2\mathrm{d}x}{\displaystyle\int_0^1 (f(x)^2+1)\mathrm{d}x} < \frac{2\displaystyle\int_0^1 f(x)^2\mathrm{d}x}{\displaystyle\int 2f(x)\mathrm{d}x} = \frac{\displaystyle\int_0^1 f(x)^2\mathrm{d}x}{\displaystyle\int f(x)\mathrm{d}x},$$

as desired. □

Chapter 3

Limits of Integrals

So far we have introduced many limiting procedures, such as the limit of a sequence or a function, the derivative and the integral. Could we take for granted that any two of them commute, for example,

$$\lim_{n\to\infty}\int_E f_n(x)\mathrm{d}x = \int_E \lim_{n\to\infty} f_n(x)\mathrm{d}x,$$

or

$$\sum_{n=1}^{\infty}\int_E f_n(x)\mathrm{d}x = \int_E \sum_{n=1}^{\infty} f_n(x)\mathrm{d}x?$$

The answer is in general no; however, we will examine some nice conditions under which the above equations hold.

3.1 Sequences and Series of Functions

Let $E \subseteq \mathbb{R}$ and $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions $f_n : E \to \mathbb{R}$.

Definition 3.1.1 (Pointwise convergence)**.** The sequence $\{f_n\}$ converges *pointwise* to f on E, if for every $x \in E$ the limit $\lim_{n\to\infty} f_n(x)$ exists and equals $f(x)$. We write $f_n \xrightarrow{p} f$.

Revisiting the definition of the limit, pointwise convergence simply means that for every $x \in E$ and $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ we have $|f_n(x) - f(x)| < \varepsilon$. So, this n_0 depends on both x and ε, in principle. When the dependence on x is no longer required, we obtain a stronger form of convergence.

Definition 3.1.2 (Uniform convergence)**.** The sequence $\{f_n\}$ converges *uniformly* to f on E, if for every $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$, such that for every $n \geq n_0$ *and every* $x \in E$ we have $|f_n(x) - f(x)| < \varepsilon$. We write $f_n \xrightarrow{u} f$.

We immediately see that by the wording of the definition, n_0 depends only on ε, so this form of convergence is stronger. An equivalent statement is that

$$\sup_{x\in E}|f_n(x) - f(x)| = \alpha_n \xrightarrow[n\to\infty]{} 0.$$

When the functions f_n of a uniformly convergent sequence share a nice property (e. g. continuity), then uniform convergence carries this property to the limit function as well.

Theorem 3.1.3. *Let* $f_n \xrightarrow{u} f$ *on* $[a, b]$.

1. $f_n \xrightarrow{p} f$ *on* $[a, b]$.
2. *If all* f_n *are continuous on* $[a, b]$*, then so is* f.
3. *If all* f_n *are Riemann integrable, then* f *is also Riemann integrable and*

$$\lim_{n\to\infty} \int_a^b f_n(x)\mathrm{d}x = \int_a^b \lim_{n\to\infty} f_n(x)\mathrm{d}x.$$

The above Theorem tells us that when a sequence converges uniformly then it also converges pointwise. However, a sequence of functions may converge pointwise but not uniformly. We need some extra conditions to make this happen.

Theorem 3.1.4 (Dini)**.** *Suppose that*

(a) $f_n \xrightarrow{p} f$ *on* $K \subseteq \mathbb{R}$.

(b) *All* f_n *are continuous on* K.

(c) f *is continuous on* K.

(d) K *is compact, i. e. closed and bounded (for example,* $K = [a, b]$*).*

(e) *The sequence* $\{f_n(x)\}_{n\in\mathbb{N}}$ *is uniformly monotonous on* K*, i. e. either for all* $x \in K$ *it is increasing or for all* $x \in K$ *it is decreasing.*

Then, $f_n \xrightarrow{u} f$ *on* K.

The notions of pointwise and uniform convergence are naturally defined for series of functions as well; as usual, a series is said to converge pointwise, resp. uniformly, if the corresponding sequence of partial sums converges pointwise, resp. uniformly.

Corollary 3.1.5. *Let* $\{f_n\}_{n\in\mathbb{N}}$ *be a sequence of continuous nonnegative functions on* $[a, b]$*, such that the series* $\sum_{n=1}^{\infty} f_n$ *converges pointwise. If* $f = \sum_{n=1}^{\infty} f_n$ *is also continuous, then*

$$\sum_{n=1}^{\infty} \int_a^b f_n(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x.$$

We conclude this section with the following theorems on uniform convergence.

Theorem 3.1.6 (Weierstraß)**.** *Let* $f : [a, b] \to \mathbb{R}$ *be continuous. Then, there is a sequence of polynomials* $\{p_n\}_{n\in\mathbb{N}}$ *that converges uniformly to* f *on* $[a, b]$.

When a function f accepts a Taylor series on an interval, the Taylor polynomials converge uniformly to f on any compact subset of the interval of convergence.

Theorem 3.1.7. *Let* f *be an infinitely differentiable function, such that its Taylor series converges on an interval* $I \subseteq \mathbb{R}$*. Then, the corresponding Taylor polynomials (i. e. the partial sums of said Taylor series) converge uniformly to* f *on any compact subinterval* $[a, b]$ *of* I.

The above theorem does not hold on I, if it is not compact itself; for example, the function $f(x) = e^x$ has a MacLaurin series (i. e. Taylor series centered at 0)

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

with interval of convergence $I = \mathbb{R}$, but the corresponding MacLaurin polynomials do not converge uniformly to f on $\mathbb{R}$, only pointwise.

3.2 Convergence Theorems

The following theorems hold for the Lebesgue integral. We adjusted them for the Riemann integral as well; for example, the pointwise limit function is *assumed* to be Riemann integrable, whereas the pointwise limit of Lebesgue integrable functions is always Lebesgue integrable.

Theorem 3.2.1 (Monotone Convergence Theorem). *Let $f_n \xrightarrow{p} f$ on $K \subseteq \mathbb{R}$, where f_n and f are nonnegative Riemann integrable functions ($\forall n \in \mathbb{N}$), and such that the sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ is increasing on K. Then*

$$\lim_{n\to\infty} \int_K f_n(x)\mathrm{d}x = \int_K f(x)\mathrm{d}x.$$

The Monotone Convergence Theorem may be applied, even when $\int_K f(x)\mathrm{d}x = +\infty$.

Theorem 3.2.2 (Levi). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of nonnegative integrable functions on K, such that the series $\sum_{n=1}^{\infty} f_n(x)$ converges $\forall x \in K$, and the function $\sum_{n=1}^{\infty} f_n$ is Riemann integrable on K. Then*

$$\sum_{n=1}^{\infty} \int_K f_n(x)\mathrm{d}x = \int_K \sum_{n=1}^{\infty} f_n(x)\mathrm{d}x.$$

Proof. Apply the Monotone Convergence Thoerem 3.2.1 on the sequence of partial sums of $\sum_{n=1}^{\infty} f_n$. □

We remark that Levi's Theorem 3.2.2 is also a corollary of Tonelli's theorem.

Theorem 3.2.3 (Dominated Convergence Theorem). *Let $f_n \xrightarrow{p}$ on $K \subseteq \mathbb{R}$, and let $|f_n(x)| \leq g(x)$, $\forall x \in K$, $\forall n \in \mathbb{N}$, where f, g, and f_n are Riemann integrable $\forall n \in \mathbb{N}$. Then,*

$$\lim_{n\to\infty} \int_K f_n(x)\mathrm{d}x = \int_K \lim_{n\to\infty} f_n(x)\mathrm{d}x.$$

Corollary 3.2.4 (Bounded Convergence Theorem). *Let $f_n \xrightarrow{p} f$ on $[a, b]$, and let $|f_n(x)| \leq M$, $\forall x \in K$, $\forall n \in \mathbb{N}$, where f and f_n are Riemann integrable $\forall n \in \mathbb{N}$. Then,*

$$\lim_{n\to\infty} \int_K f_n(x)\mathrm{d}x = \int_K \lim_{n\to\infty} f_n(x)\mathrm{d}x.$$

Proof. Apply the Dominated Convergence Theorem 3.2.3 for the Riemann integrable function $g(x) = M$, $\forall x \in [a, b]$. □

3.3 Exercises

1. **(SEEMOUS 2009/1)**

   (a) For $k \in \mathbb{N}$, calculate

$$\lim_{n\to\infty} \frac{(2n+1)!}{(n!)^2} \int_0^1 (x(1-x))^n x^k \mathrm{d}x.$$

(b) Let $f : [0,1] \to \mathbb{R}$ be continuous. Calculate

$$\lim_{n\to\infty} \frac{(2n+1)!}{(n!)^2} \int_0^1 (x(1-x))^n f(x)\mathrm{d}x.$$

Proof. **(a)** Define

$$I_{m,n} = \int_0^1 x^m (1-x)^n \mathrm{d}x,$$

so that the given integral is $I_{n+k,n}$. Applying integrations by parts, we obtain

$$I_{m,n} = \left[\frac{x^{m+1}(1-x)^n}{m+1}\right]_0^1 + \frac{n}{m+1}\int_0^1 x^{m+1}(1-x)^{n-1}\mathrm{d}x = \frac{n}{m+1} I_{m+1,n-1},$$

as long as $n > 0$. Using the above equality repeatedly, we get

$$\begin{aligned} I_{m,n} &= \frac{n}{m+1} I_{m+1,n-1} = \frac{n(n-1)}{(m+1)(m+2)} I_{m+2,n-2} = \cdots \\ &= \frac{n!}{(m+1)\cdots(m+n)} I_{m+n,0} = \frac{n!}{(m+1)\cdots(m+n)}\left[-\frac{(1-x)^{m+n+1}}{m+n+1}\right]_0^1 \\ &= \frac{m!n!}{(m+n)!}\cdot\frac{1}{m+n+1} = \frac{m!n!}{(m+n+1)!}. \end{aligned}$$

Therefore,

$$\begin{aligned} \lim_{n\to\infty} \frac{(2n+1)!}{(n!)^2} \int_0^1 (x(1-x))^n x^k \mathrm{d}x &= \lim_{n\to\infty} \frac{(2n+1)!}{(n!)^2} \frac{(n+k)!n!}{(2n+k+1)!} \\ &= \lim_{n\to\infty} \prod_{j=0}^{k-1} \frac{n+1+j}{2n+2+j} \\ &= \prod_{j=0}^{k-1} \lim_{n\to\infty} \frac{n+1+j}{2n+2+j} \\ &= \left(\frac{1}{2}\right)^k. \end{aligned}$$

(b) Let

$$L_n(f) = \frac{(2n+1)!}{(n!)^2} \int_0^1 (x(1-x))^n f(x)\mathrm{d}x.$$

Some useful properties of the functional L_n are the following:

$$L_n(f) \le L_n(|f|), \quad \text{if } f(x) \le g(x), \forall x \in [0,1], \text{ then } L_n(f) \le L_n(g).$$

Denote by $\mathcal{A}$ the class of continuous functions f such that the given limit equals $f(\frac{1}{2})$. By **(a)**, the functions $f(x) = x^k$ belong to $\mathcal{A}$. Moreover, by linearity of the integral, $\mathcal{A}$ is a *real vector space*, i. e. if $f, g \in \mathcal{A}$ and $\lambda \in \mathbb{R}$, then also $f+g, \lambda f \in \mathcal{A}$. Indeed,

$$\lim_{n\to\infty} L_n(f+g) = \lim_{n\to\infty} L_n(f) + \lim_{n\to\infty} L_n(g) = f\left(\frac{1}{2}\right) + g\left(\frac{1}{2}\right) = (f+g)\left(\frac{1}{2}\right),$$

and

$$\lim_{n\to\infty} L_n(\lambda f) = \lambda \lim_{n\to\infty} L_n(f) = \lambda f\left(\frac{1}{2}\right) = (\lambda f)\left(\frac{1}{2}\right).$$

Therefore, all polynomials belong to $\mathcal{A}$. Next, we will show that $\mathcal{A}$ is closed under uniform limits, i. e. if $\{f_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ and $f_n \xrightarrow{u} f$, then also $f \in \mathcal{A}$; f will definitely be continuous by Theorem 3.1.3. To this end, let $\varepsilon > 0$ be arbitrary. By definition, there is some (fixed!) $m \in \mathbb{N}$ such that

$$|f_m(x) - f(x)| < \frac{\varepsilon}{3}, \quad \forall x \in [0,1].$$

Also, let $n_0 \in \mathbb{N}$ such that

$$\left|L_n(f_m) - f_m\left(\frac{1}{2}\right)\right| < \frac{\varepsilon}{3}, \quad \forall n \geq n_0.$$

Therefore, for all $n \geq n_0$ it holds

$$\begin{aligned}|L_n(f) - f\left(\frac{1}{2}\right)| &\leq |L_n(f) - L_n(f_m)| + \left|L_n(f_m) - f_m\left(\frac{1}{2}\right)\right| + \left|f_m\left(\frac{1}{2}\right) - f\left(\frac{1}{2}\right)\right| \\ &< L_n(|f - f_m|) + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &\leq L_n\left(\frac{\varepsilon}{3}\chi_{[0,1]}\right) + \frac{2\varepsilon}{3} \\ &= \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} = \varepsilon.\end{aligned}$$

By definition of the limit, this clearly shows that $\lim_{n\to\infty} L_n(f) = f(\frac{1}{2})$, whence $f \in \mathcal{A}$, as desired. We conclude by observing that every continuous function f is the uniform limit of a sequence of polynomials on $[0,1]$, due to Theorem 3.1.6 by Weierstraß, thus $\mathcal{A}$ contains *all* continuous functions, completing the proof. □

2. **(SEEMOUS 2010/1)** Let $f_0 : [0,1] \to \mathbb{R}$ be a continuous function, and for $n \in \mathbb{N}$ define recursively $f_n(x) = \int_0^x f_{n-1}(t)\mathrm{d}t$.

   (a) Prove that the series $\sum_{n=1}^{\infty} f_n(x)$ converges $\forall x \in [0,1]$.

   (b) Find an explicit formula for this series, with respect to f_0.

Proof. **(a)** Since f_0 is continuous on $[0,1]$, it also bounded. Suppose that $|f_0(x)| \leq M$, $\forall x \in [0,1]$. We will show inductively that

$$|f_n(x)| \leq M\frac{x^n}{n!}, \quad \forall n \in \mathbb{N}, \forall x \in [0,1].$$

It obviously holds for $n = 0$, so suppose it holds for $n-1$, when $n > 0$. We have

$$|f_n(x)| \leq \int_0^x |f_{n-1}(t)|\mathrm{d}t \leq M\int_0^x \frac{t^{n-1}}{(n-1)!}\mathrm{d}t = M\left[\frac{t^n}{n!}\right]_0^x = M\frac{x^n}{n!},$$

as desired. We will show now that $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely $\forall x \in [0,1]$:

$$\sum_{n=1}^{\infty}|f_n(x)| \leq M\sum_{n=1}^{\infty}\frac{x^n}{n!} = M(e^x - 1) < \infty.$$

(b) Define $F(x) = \sum_{n=1}^{\infty} f_n(x)$ for $x \in [0,1]$. We will show that this series converges uniformly on $[0,1]$. It suffices to show that

$$\sup_{x\in[0,1]} |F(x) - \sum_{k=1}^{n} f_k(x)| = \alpha_n \xrightarrow[n\to\infty]{} 0.$$

It holds

$$0 \le \alpha_n = \sup_{x\in[0,1]} |\sum_{k=n+1}^{\infty} f_n(x)| \le M \sup_{x\in[0,1]} \sum_{k=n+1}^{\infty} \frac{x^k}{k!} = M \sum_{k=n+1}^{\infty} \frac{1}{k!},$$

and the latter converges to 0 since the series $\sum_{k=0}^{\infty} \frac{1}{k!}$ converges. F is then continuous as the uniform limit of a series of continuous functions by Theorem 3.1.3, so again by Theorem 3.1.3 we obtain

$$F(x) = \sum_{n=0}^{\infty} \int_0^x f_n(t)\mathrm{d}t = \int_0^x \sum_{n=0}^{\infty} f_n(t)\mathrm{d}t = \int_0^x (F(t) + f_0(t))\mathrm{d}t.$$

By the Fundamental Theorem of Calculus 2.0.1 the above is equivalent to the differential equation

$$F'(x) = F(x) + f_0(x) \Leftrightarrow e^{-x} f_0(x) = e^{-x}(F'(x) - F(x)) = (e^{-x}F(x))',$$

and the latter is equivalent to

$$F(x) = e^x \int_0^x e^{-t} f_0(t)\mathrm{d}t,$$

since $F(0) = 0$. □

3. **(SEEMOUS 2019/1)** A sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq [0,1]$ is called "Devin" if for every continuous $f : [0,1] \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i) = \int_0^1 f(x)\mathrm{d}x. \tag{3.3.1}$$

Prove that $\{x_n\}$ is Devin if and only if for every nonnegative integer k it holds

$$\lim_{n\to\infty} \frac{1}{n} x_i^k = \frac{1}{k+1}. \tag{3.3.2}$$

Proof. One direction is obvious; if $\{x_n\}$ is Devin, then (3.3.2) holds as the functions $f(x) = x^k$ are continuous. Conversely, assume that (3.3.2) holds for every nonnegative integer k; we'll show that (3.3.1) also holds, for every continuous function f. Define

$$L_n(f) = \frac{1}{n} \sum_{i=1}^{n} f(x_i),$$

and let $\mathcal{A}$ be the class of continuous functions that are Devin, i. e.

$$\lim_{n\to\infty} L_n(f) = \int_0^1 f(x)\mathrm{d}x \Leftrightarrow f \in \mathcal{A}.$$

By assumption, $\mathcal{A}$ contains all powers of x, i. e. the functions x^k. Hence, it also contains all polynomials, as $\mathcal{A}$ is a *real vector space*; if $f, g \in \mathcal{A}$ and $\lambda \in \mathbb{R}$, then $\lambda f, f+g \in \mathcal{A}$. Indeed,

$$\begin{aligned}\lim_{n\to\infty} L_n(f+g) &= \lim_{n\to\infty}(L_n(f)+L_n(g))\\ &= \lim_{n\to\infty} L_n(f) + \lim_{n\to\infty} L_n(g)\\ &= \int_0^1 f(x)\mathrm{d}x + \int_0^1 g(x)\mathrm{d}x\\ &= \int_0^1 (f(x)+g(x))\mathrm{d}x,\end{aligned}$$

hence $f, g \in \mathcal{A} \Rightarrow f+g \in \mathcal{A}$, and similarly,

$$\begin{aligned}\lim_{n\to\infty} L_n(\lambda f) &= \lim_{n\to\infty} \lambda L_n(f) = \lambda \lim_{n\to\infty}\\ &= \lambda \int_0^1 f(x)\mathrm{d}x = \int_0^1 (\lambda f)(x) = dix,\end{aligned}$$

which shows $f \in \mathcal{A} \Rightarrow \lambda f \in \mathcal{A}$. Next, consider an arbitrary continuous f and a sequence of polynomials $\{p_m\}_{m\in\mathbb{N}}$ that converges uniformly to f on $[0,1]$ (such a sequence exists by Theorem 3.1.6 of Weierstraß). Let $\varepsilon > 0$ be arbitrary, so by definition there is $m_0 \in \mathbb{N}$ such that

$$|p_m(x) - f(x)| < \frac{\varepsilon}{3}, \quad \forall x \in [0,1].$$

Fix such an m; since $p_m \in \mathcal{A}$ there is $n_0 \in \mathbb{N}$ such that

$$\left| L_n(p_m) - \int_0^1 p_m(x)\mathrm{d}x \right| < \frac{\varepsilon}{3}, \quad \forall n \geq n_0.$$

Therefore, for $n \geq n_0$ we have

$$\begin{aligned}\left| L_n(f) - \int_0^1 f(x)\mathrm{d}x \right| &\leq |L_n(f) - L_n(p_m)| + \left| L_n(p_m) - \int_0^1 p_m(x)\mathrm{d}x \right| + \left| \int_0^1 p_m(x)\mathrm{d}x - \int_0^1 f(x)\mathrm{d}x \right|\\ &< L_n(|f-p_m|) + \frac{\varepsilon}{3} + \int_0^1 |p_m(x) - f(x)|\mathrm{d}x\\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,\end{aligned}$$

which shows by definition that

$$\lim_{n\to\infty} L_n(f) = \int_0^1 f(x)\mathrm{d}x,$$

so that $f \in \mathcal{A}$, and thus $\mathcal{A}$ contains all continuous functions, as desired. □

4. **(SEEMOUS 2012/4)**

   (a) Compute

$$\lim_{n\to\infty} n \int_0^1 \left(\frac{1-x}{1+x}\right)^n \mathrm{d}x.$$

(b) For $k \in \mathbb{N}$, compute

$$\lim_{n\to\infty} n^{k+1} \int_0^1 \left(\frac{1-x}{1+x}\right)^n x^k \mathrm{d}x.$$

Solution. **(a)** Applying the substitution $y = \frac{1-x}{1+x}$ we obtain

$$\begin{aligned}
\int_0^1 n\left(\frac{1-x}{1+x}\right)^n \mathrm{d}x &= 2n \int_0^1 \frac{y^n}{(1+y)^2}\mathrm{d}y \\
&= \left[\frac{2ny^{n+1}}{(n+1)(1+y)^2}\right]_0^1 + \frac{4n}{n+1}\int_0^1 \frac{y^{n+1}}{(1+y)^3}\mathrm{d}y \\
&= \frac{2n}{4(n+1)} + \frac{4n}{n+1}\int_0^1 \frac{y^{n+1}}{(1+y)^3}\mathrm{d}y,
\end{aligned}$$

and the latter tends to $\frac{1}{2}$ as $n \to \infty$, since

$$0 \le \int_0^1 \frac{y^{n+1}}{(1+y)^3}\mathrm{d}y \le \int_0^1 y^{n+1}\mathrm{d}y = \left[\frac{y^{n+2}}{n+2}\right]_0^1 = \frac{1}{n+2} \xrightarrow[n\to\infty]{} 0.$$

(b) Applying the same substitution as before, we obtain

$$n^{k+1}\int_0^1 \left(\frac{1-x}{1+x}\right)^n x^k \mathrm{d}x = 2n^{k+1}\int_0^1 y^n \frac{(1-y)^k}{(1+y)^{k+2}}\mathrm{d}y.$$

We put

$$f(y) = \frac{(1-y)^k}{(1+y)^{k+2}}, \quad y \in [0,1].$$

Applying integration by parts we obtain

$$\begin{aligned}
\int_0^1 y^n f(y)\mathrm{d}y &= \left[\frac{y^{n+1}}{n+1}f(y)\right]_0^1 - \frac{1}{n+1}\int_0^1 y^{n+1}f'(y)\mathrm{d}y \\
&= -\frac{1}{n+1}\int_0^1 y^{n+1}f'(y)\mathrm{d}y.
\end{aligned}$$

The derivatives of higher order of f satisfy

$$f^{(j)}(y) = \sum_{i=0}^{j} \binom{j}{i} [(1-y)^k]^{(i)}[(1+y)^{-(k+2)}]^{(j-i)}.$$

So, if $j \le k-1$ we get $f^{(j)}(1) = 0$, whereas if $j = k$ the only nonzero term in the above sum is obtained for $i = j = k$, hence

$$f^{(k)}(1) = \frac{(-1)^k k!}{2^{k+2}}.$$

Applying repeatedly integration by parts, we obtain inductively

$$\begin{aligned}
\int_0^1 y^n f(y)\mathrm{d}y &= -\frac{1}{n+1}\int_0^1 y^{n+1} f'(y)\mathrm{d}y \\
&= -\frac{1}{n+1}\left[\frac{y^{n+2}}{n+2} f'(y)\right] + \frac{1}{(n+1)(n+2)}\int_0^1 y^{n+2} f''(y)\mathrm{d}y \\
&= \frac{1}{(n+1)(n+2)}\int_0^1 y^{n+2} f''(y)\mathrm{d}y \\
&= \cdots \\
&= \frac{(-1)^k}{(n+1)\cdots(n+k)}\int_0^1 y^{n+k} f^{(k)}(y)\mathrm{d}y \\
&= \frac{(-1)^k}{(n+1)\cdots(n+k+1)}\left[\left[y^{n+k+1} f^{(k)}(y)\right]_0^1 - \int_0^1 y^{n+k+1} f^{(k+1)}(y)\mathrm{d}y\right] \\
&= \frac{(-1)^k f^{(k)}(1)}{(n+1)\cdots(n+k+1)} + \frac{(-1)^{k+1}}{(n+1)\cdots(n+k+1)}\int_0^1 y^{n+k+1} f^{(k+1)}(y)\mathrm{d}y.
\end{aligned}$$

Our goal is to compute $\lim_{n\to\infty} 2n^{k+1}\int_0^1 y^n f(y)\mathrm{d}y$; we first have

$$\lim_{n\to\infty}\frac{2(-1)^k n^{k+1} f^{(k)}(1)}{(n+1)\cdots(n+k+1)} = \frac{k!}{2^{k+2}}\prod_{j=1}^{k+1}\lim_{n\to\infty}\frac{n}{n+j} = \frac{k!}{2^{k+2}},$$

and then

$$\lim_{n\to\infty}\frac{2(-1)^{k+1} n^{k+1}}{(n+1)\cdots(n+k+1)}\int_0^1 y^{n+k+1} f^{(k+1)}(y)\mathrm{d}y = \lim_{n\to\infty}\int_0^1 y^{n+k+1} f^{(k+1)}(y)\mathrm{d}y = 0.$$

Indeed, $f^{(k+1)}$ is continuous, and as such satisfies $|f^{(k+1)}(y)| \le M$, $\forall y \in [0,1]$, for some $M > 0$. Hence,

$$\begin{aligned}
\left|\int_0^1 y^{n+k+1} f^{(k+1)}(y)\mathrm{d}y\right| &\le \int_0^1 y^{n+k+1}|f^{(k+1)}(y)|\mathrm{d}y \\
&\le M\int_0^1 y^{n+k+1}\mathrm{d}y = M\left[\frac{y^{n+k+2}}{n+k+2}\right]_0^1 \\
&= \frac{M}{n+k+2} \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

Thus, the desired limit equals $\frac{k!}{2^{k+1}}$. □

5. **(SEEMOUS 2014/4)**

(a) Show that

$$\lim_{n\to\infty} n\int_0^n \frac{\arctan\frac{x}{n}}{x(x^2+1)}\mathrm{d}x = \frac{\pi}{2}.$$

(b) Calculate

$$\lim_{n\to\infty} n\left(n\int_0^n \frac{\arctan\frac{x}{n}}{x(x^2+1)}\mathrm{d}x - \frac{\pi}{2}\right).$$

Proof. **(a)** Define

$$f_n(x) = \begin{cases} 1, & x = 0 \\ \frac{n \arctan \frac{x}{n}}{x(x^2+1)}, & 0 < x \leq n \\ 0, & x > n \end{cases}$$

so that the given integral equals $\int_0^\infty f_n(x)\mathrm{d}x$. We remark that each f_n is continuous at 0 since $\lim_{y\to 0} \frac{\arctan y}{y} = 1$. Moreover, from $0 \leq \arctan y \leq y$, $\forall y \geq 0$, we get

$$0 \leq f_n(x) \leq \frac{1}{1+x^2}, \quad \forall x \geq 0, \forall n \in \mathbb{N},$$

and

$$\lim_{n\to\infty} f_n(x) = \frac{1}{1+x^2} \lim_{n\to\infty} \frac{\arctan \frac{x}{n}}{\frac{x}{n}} = \frac{1}{1+x^2} = f(x), \quad \forall x \geq 0.$$

So, $f_n \xrightarrow{p} f$, and the sequence $\{f_n\}$ is dominated by the Riemann integrable function f on $[0, +\infty)$. Therefore, by the Dominated Convergence Theorem 3.2.3 we obtain

$$\begin{aligned} \lim_{n\to\infty} \int_0^\infty f_n(x)\mathrm{d}x &= \int_0^\infty \frac{1}{1+x^2}\mathrm{d}x = \lim_{N\to\infty} \int_0^N \frac{1}{1+x^2}\mathrm{d}x \\ &= \lim_{N\to\infty} [\arctan x]_0^N = \lim_{N\to\infty} \arctan N = \frac{\pi}{2}. \end{aligned}$$

(b) The given limit, if it exists, equals

$$-\lim_{n\to\infty} \int_0^\infty n(f(x) - f_n(x))\mathrm{d}x.$$

We have

$$\int_0^\infty n(f(x) - f_n(x))\mathrm{d}x = n \int_0^n \frac{\frac{x}{n} - \arctan \frac{x}{n}}{\frac{x}{n}(1+x^2)}\mathrm{d}x + n(\pi/2 - \arctan n).$$

We calculate first

$$\lim_{n\to\infty} n\Big(\frac{\pi}{2} - \arctan n\Big) = \lim_{n\to\infty} \frac{\frac{\pi}{2} - \arctan n}{\frac{1}{n}}.$$

We will calculate the above limit as $n \to \infty$ and n is a real number, so applying L'Hôpital's rule we get

$$\lim_{n\to\infty} \frac{-\frac{1}{n^2+1}}{-\frac{1}{n^2}} = 1.$$

Next, we will calculate

$$\lim_{n\to\infty} n \int_0^n \frac{\frac{x}{n} - \arctan \frac{x}{n}}{\frac{x}{n}(1+x^2)}\mathrm{d}x.$$

Applying the substitution $y = \frac{x}{n}$, the above expression equals

$$\int_0^1 \frac{1 - \frac{\arctan y}{y}}{y^2 + \frac{1}{n^2}}\mathrm{d}y.$$

Let $g_n(y)$ denote the integrand above; it holds $g_n(y) \geq 0$, $\forall n \in \mathbb{N}$, $\forall y \in [0,1]$, as $0 \leq \arctan y \leq y$ for $y \in [0,1]$. The sequence $\{g_n\}_{n\in\mathbb{N}}$ converges *increasingly* to the continuous[1] function

$$g(y) = \begin{cases} \frac{1}{3}, & y = 0 \\ \frac{1-\frac{\arctan y}{y}}{y^2}, & y > 0 \end{cases}$$

so, by the Monotone Convergence Theorem 3.2.1 we obtain

$$\lim_{n\to\infty} \int_0^1 g_n(x)\mathrm{d}x = \int_0^1 g(x)\mathrm{d}x.$$

We recall the power series of $\arctan y$:

$$\arctan y = \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n+1}}{2n+1} = y - \frac{y^3}{3} + \frac{y^5}{5} - \cdots, \quad y \in (-1,1].$$

Therefore, the power series expansion of g on $[0,1]$ is

$$g(y) = \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n}}{2n+3}.$$

By Theorems 3.1.3 and 3.1.7 we obtain

$$\begin{aligned}
\int_0^1 g(y)\mathrm{d}y &= \sum_{n=0}^{\infty} \int_0^1 \frac{(-1)^n y^{2n}}{2n+3}\mathrm{d}y = \sum_{n=0}^{\infty} \left[\frac{(-1)^n y^{2n+1}}{(2n+1)(2n+3)}\right]_0^1 \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)(2n+3)} = \frac{1}{1\cdot 3} - \frac{1}{3\cdot 5} + \frac{1}{5\cdot 7} - \cdots \\
&= \frac{1}{2}\left[\left(1 - \frac{1}{3}\right) - \left(\frac{1}{3} - \frac{1}{5}\right) + \left(\frac{1}{5} - \frac{1}{7}\right) - \cdots\right] \\
&= \frac{1}{2}\sum_{n=0}^{\infty}(-1)^n\left(\frac{1}{2n+1} - \frac{1}{2n+3}\right) \\
&= -\frac{1}{2} + \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = -\frac{1}{2} + \arctan 1 = \frac{\pi}{4} - \frac{1}{2}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\lim_{n\to\infty} n\left(n\int_0^n \frac{\arctan\frac{x}{n}}{x(x^2+1)}\mathrm{d}x - \frac{\pi}{2}\right) &= -\lim_{n\to\infty}\int_0^\infty n(f(x) - f_n(x))\mathrm{d}x \\
&= -\int_0^1 g(y)\mathrm{d}y - 1 \\
&= -\frac{\pi}{4} - \frac{1}{2}.
\end{aligned}$$

□

6. **(SEEMOUS 2017/4)**

   (a) Calculate $\displaystyle\int_0^1 (1-t)^n e^t \mathrm{d}t$, for $n \in \mathbb{N}$.

[1] The limit as $y \longrightarrow 0+$ can easily be found using L'Hôpital's rule.

(b) Let k be a fixed nonnegative integer. Define the sequence $\{x_n\}_{n\geq k}$ with

$$x_n = \sum_{i=k}^{n} \binom{i}{k}\left(e - 1 - \frac{1}{1!} - \cdots - \frac{1}{i!}\right).$$

Prove that x_n converges and find its limit.

Proof. **(a)** Denote the given integral by I_n; we will find first a recursive formula for I_n. It holds

$$I_0 = \int_0^1 e^t \mathrm{d}t = e - 1.$$

For $n > 0$, we apply integration by parts to obtain

$$I_n = \int_0^1 (1-t)^n (e^t)' \mathrm{d}t = [(1-t)^n e^t]_0^1 + n\int_0^1 (1-t)^{n-1} e^t \mathrm{d}t = -1 + nI_{n-1}.$$

The sequence $\{\alpha_n\}_{n\geq 0}$ with $\alpha_n = \frac{1}{n!} I_n$ satisfies

$$\alpha_n = -\frac{1}{n!} + \alpha_{n-1},$$

or equivalently, $\alpha_n - \alpha_{n-1} = -\frac{1}{n!}$, hence

$$\alpha_n = \alpha_0 + \sum_{j=1}^{n}(\alpha_j - \alpha_{j-1}) = (e-1) - \sum_{j=1}^{n} \frac{1}{j!},$$

and thus

$$I_n = n!e - n!\left(1 + \frac{1}{1!} + \cdots + \frac{1}{n!}\right).$$

(b) Using **(a)** we may write

$$\begin{aligned}
x_n &= \sum_{i=k}^{n} \binom{i}{k}\left(e - 1 - \frac{1}{1!} - \cdots - \frac{1}{i!}\right) \\
&= \sum_{i=k}^{n} \binom{i}{k} \frac{1}{i!} I_i = \sum_{i=k}^{n} \int_0^1 \frac{1}{k!(i-k)!}(1-t)^i e^t \mathrm{d}t \\
&= \int_0^1 e^t \sum_{i=k}^{n} \frac{(1-t)^i}{k!(i-k!)} \mathrm{d}t \\
&= \frac{1}{k!}\int_0^1 e^t (1-t)^k \sum_{j=0}^{n-k} \frac{(1-t)^j}{j!} \mathrm{d}t.
\end{aligned}$$

We recall that the Taylor series of e^{1-t},

$$e^{1-t} = 1 + \frac{1-t}{1!} + \frac{(1-t)^2}{2!} + \cdots$$

converges pointwise for all $t \in \mathbb{R}$, hence uniformly on $[0,1]$ by Theorem 3.1.7. Therefore, the sequence of functions

$$e^t (1-t)^k \sum_{j=0}^{n-k} \frac{(1-t)^j}{j!} \xrightarrow[n\to\infty]{u} e^t(1-t)^k e^{1-t} = e(1-t)^k,$$

and by Theorem 3.1.3 we obtain

$$\lim_{n\to\infty} x_n = \frac{e}{k!}\int_0^1 (1-t)^k \mathrm{d}t = \frac{e}{k!}\left[-\frac{(1-t)^{k+1}}{k+1}\right]_0^1 = \frac{e}{(k+1)!}.$$

□

7. **(SEEMOUS 2018/4)**

(a) Let f be a polynomial. Show that

$$\int_0^\infty e^{-x} f(x)\mathrm{d}x = f(0) + f'(0) + f''(0) + \cdots$$

(b) Let $f : \mathbb{R} \to \mathbb{R}$ be an infinitely differentiable function, whose Taylor series at 0 has radius of convergence $R = +\infty$. If $\sum_{n=0}^\infty f^{(n)}(0)$ converges absolutely, then the integral $\int_0^\infty e^{-x} f(x)\mathrm{d}x$ converges and

$$\sum_{n=0}^\infty f^{(n)}(0) = \int_0^\infty e^{-x} f(x)\mathrm{d}x.$$

Proof. **(a)** We first compute the integrals

$$I_n = \int_0^\infty e^{-x} x^n \mathrm{d}x,$$

by finding a recursive formula for I_n, *assuming they converge.* It holds

$$I_0 = \lim_{N\to\infty} \int_0^N e^{-x}\mathrm{d}x = \lim_{N\to\infty} [-e^{-x}]_0^N = \lim_{N\to\infty} (1 - e^{-N}) = 1.$$

We proceed inductively for $n > 0$; assuming that I_{n-1} converges, we apply integration by parts to obtain

$$\begin{aligned} I_n &= \int_0^\infty e^{-x} x^n \mathrm{d}x = \lim_{N\to\infty}\int_0^N (-e^{-x})' x^n \mathrm{d}x = \\ &= \lim_{N\to\infty}\left([-e^{-x}x^n]_0^N + n\int_0^N e^{-x}x^{n-1}\mathrm{d}x\right) \\ &= \lim_{N\to\infty}(-e^{-N}N^n) + n\lim_{N\to\infty}\int_0^N e^{-x}x^{n-1}\mathrm{d}x \\ &= nI_{n-1}. \end{aligned}$$

Applying repeatedly this recursive formula we get

$$I_n = nI_{n-1} = n(n-1)I_{n-2} = \cdots = n!I_0 = n!.$$

Now, suppose that $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, hence

$$\begin{aligned} \int_0^\infty e^{-x} f(x)\mathrm{d}x &= \int_0^\infty e^{-x}(a_n x^n + \cdots + a_1 x + a_0)\mathrm{d}x \\ &= a_n\int_0^\infty e^{-x}x^n\mathrm{d}x + \cdots + a_1\int_0^\infty e^{-x}x\mathrm{d}x + a_0\int_0^\infty e^{-x}\mathrm{d}x \\ &= a_n I_n + \cdots + a_1 I_1 + a_0 I_0 \\ &= a_n n! + \cdots + a_1 + a_0 \\ &= f^{(n)}(0) + \cdots + f'(0) + f(0), \end{aligned}$$

as desired.

(b) Let $f_n(x) = e^{-x}\sum_{k=0}^{n}\frac{x^k}{x!}f^{(k)}(0)$, so that

$$\int_0^\infty f_n(x)\mathrm{d}x = \sum_{k=0}^{n} f^{(k)}(0),$$

by **(a)**. We want to show that

$$\int_0^\infty f_n(x)\mathrm{d}x \xrightarrow[n\to\infty]{} \int_0^\infty e^{-x}f(x)\mathrm{d}x.$$

Since the Taylor series of f has radius of convergence $R = +\infty$, we have that $f_n(x) \xrightarrow{p} e^{-x}f(x)$. In order to show the convergence of integrals, we need a stronger form of convergence (uniform convergence does not necessarily hold here, and besides, the interval of integration is unbounded). For every $n \in \mathbb{N}$ and $x \geq 0$ we have

$$|f_n(x)| \leq e^{-x}\sum_{k=0}^{n}\frac{x^k}{x!}|f^{(k)}(0)| =: g_n(x),$$

and

$$g_n(x) \leq e^{-x}\sum_{k=0}^{\infty}\frac{x^k}{x!}f^{(k)}(0) =: g(x).$$

The latter is well defined for all $x \geq 0$, since $\sum_{k=0}^{\infty}|f^{(k)}(0)| < \infty$. Since $\{g_n(x)\}_{n\geq 0}$ is increasing by definition $\forall x \geq 0$, we get by the Monotone Convergence Theorem 3.2.1

$$\int_0^\infty g(x)\mathrm{d}x = \lim_{n\to\infty}\int_0^\infty g_n(x)\mathrm{d}x = \lim_{n\to\infty}\sum_{k=0}^{n}|f^{(k)}(0)| = \sum_{k=0}^{\infty}|f^{(k)}(0)| < \infty,$$

therefore g is integrable on $[0, +\infty)$. Since $|f_n(x)| \leq g(x)\ \forall n \geq 0,\ \forall x \geq 0$[2], we get by the Dominated Convergence Theorem 3.2.3,

$$\int_0^\infty e^{-x}f(x)\mathrm{d}x = \lim_{n\to\infty}\int_0^\infty f_n(x)\mathrm{d}x = \lim_{n\to\infty}\sum_{k=0}^{n} f^{(k)}(0) = \sum_{k=0}^{\infty} f^{(k)}(0). \qquad \square$$

8. **(SEEMOUS 2019/4)**

   (a) For $n \in \mathbb{N}$, calculate the integral

$$\int_0^1 x^{n-1}\ln x\mathrm{d}x.$$

   (b) Calculate the sum

$$\sum_{n=0}^{\infty}(-1)^n\left(\frac{1}{(n+1)^2} - \frac{1}{(n+2)^2} + \frac{1}{(n+3)^2} - \cdots\right).$$

[2]This means that the pointwise convergence $f_n(x) \xrightarrow{p} e^{-x}f(x)$ is dominated by the integrable function g; that is why it was crucial to show that $\int_0^\infty g(x)\mathrm{d}x$ converges.

Proof. **(a)** The given integral is improper; for $t > 0$ we have

$$\begin{aligned}\int_t^1 x^{n-1}\ln x\mathrm{d}x &= \left[\frac{x^n}{n}\ln x\right]_t^1 - \frac{1}{n}\int_t^1 x^{n-1}\mathrm{d}x\\ &= -\frac{t^n\ln t}{n} - \frac{1}{n}\int_t^1 x^{n-1}\mathrm{d}x.\end{aligned}$$

We will examine the behavior of the above expression as $t \to 0^+$; first,

$$\lim_{t\to 0+}\frac{-t^n\ln t}{n} = \lim_{t\to 0+}\frac{\ln t^{-1}}{nt^{-n}} = \lim_{y\to\infty}\frac{\ln y}{ny^n} = 0,$$

by putting $y = \frac{1}{t}$. Therefore,

$$\lim_{t\to 0+}\int_t^1 x^{n-1}\ln x\mathrm{d}x = -\frac{1}{n}\lim_{t\to 0+}\int_t^1 x^{n-1}\mathrm{d}x = -\frac{1}{n}\int_0^1 x^{n-1}\mathrm{d}x = -\frac{1}{n^2}.$$

(b) First of all, we note that each term of the series is the sum of a series itself, which converges absolutely (it is the tail of a p-series, with $p = 2$). The nth partial sum may be rewritten as follows:

$$\begin{aligned}\sum_{m=0}^{n-1}(-1)^m\left(\frac{1}{(m+1)^2} - \frac{1}{(m+2)^2} + \cdots\right) &= \sum_{m=0}^{n-1}\left[\sum_{k=m+1}^{n}\frac{(-1)^{k-1}}{k^2} + \sum_{k=n+1}^{\infty}\frac{(-1)^{k-1}}{k^2}\right]\\ &= \sum_{k=1}^{n}\sum_{m=0}^{k-1}\frac{(-1)^{k-1}}{k^2} + n\sum_{k=n+1}^{\infty}\frac{(-1)^{k-1}}{k^2}\\ &= \sum_{k=1}^{n}\frac{(-1)^{k-1}}{k} + n\sum_{k=n+1}^{\infty}\frac{(-1)^{k-1}}{k^2}.\end{aligned}$$

We will estimate the second term first; we have

$$\begin{aligned}\left|\sum_{k=n+1}^{\infty}\frac{(-1)^{k-1}}{k^2}\right| &= \frac{1}{(n+1)^2} - \frac{1}{(n+2)^2} + \frac{1}{(n+3)^2} - \cdots\\ &= \frac{1}{(n+1)^2} - \sum_{k=0}^{\infty}\left[\frac{1}{(n+2k)^2} - \frac{1}{(n+2k+1)^2}\right]\\ &\le \frac{1}{(n+1)^2},\end{aligned}$$

hence

$$\lim_{n\to\infty} n\sum_{k=n+1}^{\infty}\frac{(-1)^{k-1}}{k^2} = 0.$$

Therefore, the sum of the given series equals

$$\lim_{n\to\infty}\sum_{k=1}^{n}\frac{(-1)^{k-1}}{k} = \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}.$$

We recall the power series of $\ln(x-1)$:

$$\ln(x+1) = \sum_{k=1}^{\infty}\frac{(-1)^{k-1}x^k}{k}, \quad x \in (-1, 1].$$

Plugging in $x = 1$, we obtain

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} = \ln 2,$$

so, the desired sum equals $\ln 2$. $\square$

Part II

Complex Numbers

Chapter 1

Construction and basic definitions

Here we will introduce the notion of the *complex number*, which is required to understand polynomials in terms of their roots, and this is important especially for understanding the behavior and properties of matrices with respect to their eigenvalues.

The set (or more appropriately, the *field*) of complex numbers is just an extension of the set of real numbers, $\mathbb{R}$, which nevertheless satisfies some very nice properties. We recall in passsing, that $\mathbb{R}$ itself was axiomatically obtained through successive extensions of other number systems; each extension satisfies a certain need. For example, we begin with the set of natural numbers,

$$\mathbb{N} = \{1, 2, 3, \ldots\},$$

where addition and multiplication are well-defined. It is certainly necessary to define subtraction for all pairs of natural numbers; this leads us to the first extension to the set of integers,

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

Next, we wish to define division for all pairs of integers; this leads us to the set of rational numbers,

$$\mathbb{Q} = \Big\{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\Big\}.$$

The set $\mathbb{Q}$ is simply the set of all fractions whose numerators and denominators are integers (denominator is always nonzero); addition and multiplication are defined as is already known. The next extension is more subtle; if one wants to approximate numbers, i.e. take limits, then $\mathbb{Q}$ is not sufficient, it has "holes". This simply means that there are convergent sequences $(q_n) \subseteq \mathbb{Q}$, whose limit lies outside $\mathbb{Q}$ (this means that $\mathbb{Q}$ is not *complete*). If we wanted to include all possible limits of convergent sequences in $\mathbb{Q}$, we would have to extend $\mathbb{Q}$ to a *complete* set, and thus obtain $\mathbb{R}$, which is usually depicted as a straight line, stretching indefinitely on both sides.

$\mathbb{R}$ is closed under addition, subtraction, multiplication and (nonzero) division; along with some other properties (i.e. commutativity of addition and multiplication), this means that $\mathbb{R}$ is a *field*. This means that whenever $a, b \in \mathbb{R}$, the numbers $a + b, a - b, ab, a/b$ (the latter as long as $b \neq 0$) all belong to $\mathbb{R}$. In other words, $\mathbb{R}$ contains the roots of all polynomials of first degree with real coefficients (i.e. the product of a and $b \neq 0$ is simply the solution to the equation $\frac{x}{b} - a = 0$). However, when we search for roots of second degree polynomials, this is no longer true; for example, the polynomial $x^2 + 1$ always takes positive values in $\mathbb{R}$. If we had to extend $\mathbb{R}$ so as to include the roots of this polynomial, we would obtain a further extension, the set of *complex numbers* denoted by $\mathbb{C}$, and this is the last step in these successive extensions.

$\mathbb{C}$ is the smallest field containg both $\mathbb{R}$ as well as the root of $x^2 + 1$, the so-called *imaginary* root, denoted by i (the other root of this polynomial will then be $-i$). Topologically, $\mathbb{C}$ is depicted as a plane; as we shall say in the next part, $\mathbb{C}$ is isomorphic to $\mathbb{R}^2$ as *real vector space.* A succinct description of its elements is the following:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}.$$

So, every complex number is written in the form $z = a + bi$ for *unique* real numbers a and b, respectively called the *real* and *imaginary* part of z, denoted by

$$\operatorname{Re} z = a, \quad \operatorname{Im} z = b.$$

We will see now the basic arithmetic operations on $\mathbb{C}$; let $z = a + bi$ and $w = c + di$.
<u>*Addition*</u> It holds

$$z + w = (a + c) + (b + d)i,$$

or equivalently,

$$\operatorname{Re}(z + w) = \operatorname{Re} z + \operatorname{Re} w, \quad \operatorname{Im}(z + w) = \operatorname{Im} z + \operatorname{Im} w.$$

From the above, we can easily deduce the formulae for subtraction, namely

$$z - w = (a - c) + (b - d)i,$$

or equivalently,

$$\operatorname{Re}(z - w) = \operatorname{Re} z - \operatorname{Re} w, \quad \operatorname{Im}(z - w) = \operatorname{Im} z - \operatorname{Im} w.$$

<u>*Multiplication*</u> Here, it holds

$$zw = ac + (ad + bc)i + bdi^2 = (ac - bd) + (ad + bc)i,$$

since $i^2 = -1$. In other words,

$$\operatorname{Re}(zw) = \operatorname{Re} z \operatorname{Re} w - \operatorname{Im} z \operatorname{Im} w, \quad \operatorname{Im}(zw) = \operatorname{Re} z \operatorname{Im} w + \operatorname{Im} z \operatorname{Re} w.$$

For the division between two complex numbers, the following notion is very helpful.

Definition 1.0.1. If $z = a + bi$, $a, b \in \mathbb{R}$, then the *complex conjugate* of z equals $a - bi$ and is denoted by $\bar{z}$. The distance from z to 0 on the complex plane is called the *modulus* of z and is denoted by $|z|$; it equals $\sqrt{a^2 + b^2}$.

<u>**Properties**</u>

(i) It holds $z + \bar{z} = 2 \operatorname{Re} z$ and $z - \bar{z} = 2i \operatorname{Im} z$.

(ii) For all $z \in \mathbb{C}$ it holds $z\bar{z} = |z|^2$. Indeed,

$$z\bar{z} = (a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2 = |z|^2.$$

Now we can proceed to describe division:
Division For $z \in \mathbb{C} \setminus \{0\}$ it holds

$$\frac{1}{z} = \frac{\bar{z}}{|z|^2},$$

so that

$$\operatorname{Re}\left(\frac{1}{z}\right) = \frac{\operatorname{Re} z}{|z|^2}, \quad \operatorname{Im}\left(\frac{1}{z}\right) = -\frac{\operatorname{Im} z}{|z|^2}.$$

Therefore, if $w \in \mathbb{C} \setminus \{0\}$ it holds

$$\frac{z}{w} = \frac{z\bar{w}}{|w|^2}.$$

Chapter 2

Polynomials

2.1 Quadratic polynomials

The construction of $\mathbb{C}$ is arisen by adjoining a root of the quadratic polynomial x^2+1, which as we know does not exist in $\mathbb{R}$. We will now see that every quadratic polynomial has precisely two complex roots (this includes the case where there is only one *double* root; as we will see, there is a good reason for calling it double).

So, let

$$P(x) = ax^2 + bx + c, \quad a, b, c \in \mathbb{R}, a \neq 0,$$

and denote by

$$\Delta = b^2 - 4ac$$

the *discriminant* of P. We distinguish three cases:

1. If $\Delta > 0$, then P has two *distinct* real roots, namely

$$\frac{-b \pm \sqrt{\Delta}}{2a}.$$

2. If $\Delta = 0$, then P has a unique *double* real root, namely

$$-\frac{b}{2a}.$$

   The reason for this term is the factorization of $P(x)$ into linear terms:

$$P(x) = a\left(x + \frac{b}{2a}\right)^2.$$

3. If $\Delta < 0$, we obtain

$$P(x) = a\left(x + \frac{b}{2a}\right)^2 - \frac{\Delta}{4a^2}$$

   when we complete the square, thus P has two distinct complex conjugate roots, namely

$$z = -\frac{b}{2a} + i\frac{\sqrt{|\Delta|}}{2a}, \quad \bar{z} = -\frac{b}{2a} - i\frac{\sqrt{|\Delta|}}{2a},$$

   and we factorize P as

$$P(x) = a(x - z)(x - \bar{z}).$$

We tackled the case of quadratic polynomials with real coefficients. It is not hard to see that every quadratic polynomial with complex coefficients has also two complex roots (though not necessarily conjugate) by completing the square. However, we skip this part since it requires taking square roots in $\mathbb{C}$; this becomes a lot easier when we introduce the *polar form* of a complex number.

2.2 The Fundamental Theorem of Algebra

Theorem 2.2.1 (Fundamental Theorem of Algebra)**.** *A polynomial*

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \quad a_j \in \mathbb{C}, 0 \le j \le n, a_n \neq 0$$

has exactly *n complex roots, counting multiplicities. If the roots are $r_1, \dots r_n$, then we can factorize $P(x)$ into linear terms as*

$$P(x) = a_n(x - r_1)\cdots(x - r_n).$$

The multiplicity of a root r in P, is the number of times that the factor $(x - r)$ appears in the above product.

This Theorem tells us that we need to look no further, if we wish to solve polynomial equations with complex coefficients, therefore, the sequence of extensions of number systems is now complete. The above property shows precisely that $\mathbb{C}$ is an *algebraically closed field.*

Now that we know that a polynomial with complex coefficients of degree n has exactly n roots, we may calculate their sum with respect to the coefficients; this is a result due to Vieta.

Theorem 2.2.2 (Vieta's formula)**.** *Suppose*

$$P(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{C}[x], a_n \neq 0,$$

has roots $\rho_1, \dots, \rho_n$. Then

$$\rho_1 + \rho_2 + \cdots + \rho_n = -\frac{a_{n-1}}{a_n}$$

and

$$\rho_1 \rho_2 \cdots \rho_n = (-1)^n \frac{a_0}{a_n}.$$

Proof. By Theorem 2.2.1, $P(x)$ is factorized into linear terms as

$$a_n(x - \rho_1)\cdots(x - \rho_n).$$

Expanding the above product, we observe that the coefficient of x^{n-1} equals

$$-a_n(\rho_1 + \cdots + \rho_n),$$

which must equal a_{n-1}, and the constant term equals

$$(-1)^n a_n \rho_1 \cdots \rho_n,$$

which proves the second part. □

Remark. Vieta's formulae actually involve all coefficients of $P(x)$. It is known in particular, that the value of the jth elementary symmetric polynomial on the roots of $P(x)$ equals $(-1)^j \frac{a_{n-j}}{a_n}$.

Chapter 3

Polar form of a complex number

If $z = a + bi$, we may apply conversion to polar coordinates for $(a, b) \in \mathbb{R}^2$ to (ρ, θ), to obtain the polar form of z:

$$z = \rho(\cos\theta + i\sin\theta),$$

where $\rho = |z|$ and $\theta = \arg z$, the *argument* of z. The *principal* argument of z, denoted by $\operatorname{Arg} z$, satisfies

$$\operatorname{Arg} z \in (-\pi, \pi].$$

We will try to rewrite the polar form of z into yet another convenient form; the power series of e^x, $\sin x$, $\cos x$ can be extended to $\mathbb{C}$ as follows:

$$\begin{aligned} e^x &= 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \\ \cos z &= 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots \\ \sin z &= z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots . \end{aligned}$$

Replacing z by iz in the latter two, we obtain

$$e^{iz} = \cos z + i\sin z. \tag{3.0.1}$$

For $z = \pi$ we obtain the famous *Euler identity*:

$$e^{i\pi} + 1 = 0. \tag{3.0.2}$$

Thus, the polar form of z can also be written as $z = \rho e^{i\theta}$. We can now easily describe certain operations with the polar form:
<u>*Multiplication*</u> Let $z = \rho e^{i\theta}$, $w = re^{i\varphi}$. Then,

$$zw = (\rho r)e^{i(\theta+\varphi)},$$

or equivalently,

$$|zw| = |z| \cdot |w|, \quad \arg(zw) = \arg z + \arg w.$$

When $\rho = r = 1$, we obtain the following well-known trigonometric identities:

$$\begin{aligned} \cos(\theta + \varphi) &= \operatorname{Re}\left(e^{i\theta+\varphi}\right) \\ &= \operatorname{Re} e^{i\theta} \operatorname{Re} e^{i\varphi} - \operatorname{Im} e^{i\theta} \operatorname{Im} e^{i\varphi} \\ &= \cos\theta\cos\varphi - \sin\theta\sin\varphi, \end{aligned}$$

and

$$\begin{aligned}\sin(\theta+\varphi) &= \operatorname{Im}\left(e^{i\theta+\varphi}\right) \\ &= \operatorname{Im} e^{i\theta} \operatorname{Re} e^{i\varphi} + \operatorname{Re} e^{i\theta} \operatorname{Im} e^{i\varphi} \\ &= \sin\theta\cos\varphi + \cos\theta\sin\varphi.\end{aligned}$$

Conjugation If $z = \rho e^{i\theta}$, then

$$\bar{z} = \overline{\rho e^{i\theta}} = \rho e^{-i\theta}.$$

In particular, if $|z| = 1$, then

$$\bar{z} = \frac{1}{z}.$$

Division If $z = \rho e^{i\theta}$, $w = re^{i\varphi} \neq 0$, then

$$\frac{z}{w} = \frac{\rho}{r} e^{i(\theta-\varphi)}.$$

Chapter 4

Roots of unity

According to the Fundamental Theorem of Algebra, the polynomial $x^n - 1$ has exactly n complex roots, which are called the nth *roots of unity.* For example, there is only one first root of unity, namely 1; the second roots of unity are ± 1, and the fourth roots of unity are $\pm 1, \pm i$, since

$$(\pm i)^4 = (i^2)^2 = (-1)^2 = 1.$$

The description of the nth roots of unity is below:

Theorem 4.0.1. *The nth roots of unity are*

$$e^{\frac{2k\pi i}{n}} = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}, \quad k = 0, 1, \dots, n-1.$$

Proof. Let z be a root of the polynomial $x^n - 1$; it satisfies $z^n = 1$, which in polar form is equivalent to

$$1 = (\rho e^{i\theta})^n = \rho^n e^{in\theta},$$

where $z = \rho e^{i\theta}$, $\rho > 0$, $\theta \in [0, 2\pi)$. Taking moduli on both sides of the above equality we obtain $\rho^n = 1$ which yields $\rho = 1$ (the only positive real root of $x^n - 1$ is 1). Therefore, $e^{in\theta} = 1$, or equivalently $n\theta = 2k\pi$ for some $k \in \mathbb{Z}$. Since $\theta \in [0, 2\pi)$, we must have

$$\theta = \frac{2k\pi}{n}, \quad k = 0, 1, \dots, n-1,$$

completing the proof. $\square$

All roots of unity lie on the unit circle

$$S^1 = \{z \in \mathbb{C} : |z| = 1\}.$$

In particular, the nth roots of unity are the vertices of a regular n-gon, inscribed in S^1.

A root of unity is called a *primitive* nth root of unity if it is not a mth root of unity for any $m < n$.

Theorem 4.0.2. *If $\omega \neq 1$ is an nth root of unity, then*

$$1 + \omega + \omega^2 + \cdots + \omega^{n-1} = 0.$$

Proof. It holds

$$1 + \omega + \cdots + \omega^{n-1} = \frac{1 - \omega^n}{1 - \omega} = 0. \qquad \square$$

Chapter 5

Exercises

1. Let $\omega_0, \omega_1, \ldots, \omega_{n-1}$ be the nth roots of unity. Show that

   (a) $\omega_0\omega_1 \cdots \omega_{n-1} = (-1)^{n-1}$.

   (b)
$$\sum_{j=0}^{n-1} \omega_j^k = \begin{cases} 0, & 1 \leq k \leq n-1 \\ n, & k = n. \end{cases}$$

   Proof. (a) is a simple application of Theorem 2.2.2 for the polynomial $P(x) = x^n - 1$. The sum in (b) is just a sum of consecutive terms of a geometric progression, as
$$\omega_j = e^{\frac{2j\pi i}{n}} = \left(e^{\frac{2\pi i}{n}}\right)^j = \omega_i^j,$$
therefore,
$$\sum_{j=0}^{n-1} \omega_j^k = \sum_{j=0}^{n-1} \omega_1^{kj} = \sum_{j=0}^{n-1} \omega_k^j.$$
So, if $k = n$ the above sum equals n (we put $\omega_0 = \omega_n$), while if $1 \leq k \leq n-1$ it holds
$$\sum_{j=0}^{n-1} \omega_k^j = \frac{1-\omega_k^n}{1-\omega_k} = 0. \qquad \square$$

2. If $\theta \notin \{2k\pi : k \in \mathbb{Z}\}$, show
$$1 + \cos\theta + \cos 2\theta + \cdots + \cos n\theta = \frac{1}{2} + \frac{\sin\left(n+\frac{1}{2}\right)\theta}{2\sin\frac{\theta}{2}}.$$

   Proof. The condition $\theta \notin \{2k\pi : k \in \mathbb{Z}\}$ implies that $\sin\frac{\theta}{2} \neq 0$, so the right hand side makes sense. Moreover, $e^{i\theta} \neq 1$, hence the left hand side equals
$$\begin{aligned}
\operatorname{Re}\left(1 + e^{i\theta} + e^{2i\theta} + \cdots + e^{ni\theta}\right) &= \operatorname{Re}\left(\frac{e^{i(n+1)\theta} - 1}{e^{i\theta} - 1}\right) \\
&= \operatorname{Re}\left[\frac{(e^{i(n+1)\theta} - 1)(e^{-i\theta} - 1)}{2 - 2\operatorname{Re}(e^{i\theta} - 1)}\right] \\
&= \frac{\operatorname{Re}(e^{in\theta} - e^{i(n+1)\theta} - e^{-i\theta} + 1)}{2 - 2\cos\theta} \\
&= \frac{1 - \cos\theta + \cos n\theta - \cos(n+1)\theta}{2 - 2\cos\theta}.
\end{aligned}$$

Using the formula

$$\cos\varphi - \cos\psi = 2\sin\frac{\varphi+\psi}{2}\sin\frac{\psi-\varphi}{2},$$

we obtain

$$\cos n\theta - \cos(n+1)\theta = 2\sin\left(n+\frac{1}{2}\right)\theta\sin\frac{\theta}{2}$$

and

$$1-\cos\theta = \cos 0 - \cos\theta = 2\sin^2\frac{\theta}{2}.$$

Subsituting these in the equality above we get

$$\frac{1-\cos\theta+\cos n\theta-\cos(n+1)\theta}{2-2\cos\theta} = \frac{1}{2}+\frac{2\sin\left(n+\frac{1}{2}\right)\theta\sin\frac{\theta}{2}}{2\sin^2\frac{\theta}{2}} = \frac{1}{2}+\frac{\sin\left(n+\frac{1}{2}\right)\theta}{2\sin\frac{\theta}{2}}.$$

□

3. ω is a primitive nth root of unity if and only if $\omega = e^{\frac{2d\pi i}{n}}$, where $\gcd(d,n)=1$.

Proof. By Theorem 4.0.1, ω is an nth root of unity if and only if $\omega = e^{\frac{2d\pi i}{n}}$. Under which additional condition is it a primitive root? Let $om^m = 1$, so that $e^{\frac{2md\pi i}{n}} = 1$, which implies that $n \mid dm$. If $\gcd(d,n)=1$, this is only possible if $n \mid m$, which further implies that $m \geq n$, so that ω is a primitive nth root of unity. On the other hand, if $\gcd(d,n)=\delta>1$, then $m = n/\delta < n$ satisfies $n \mid dm$, so that ω is not a primitive nth root of unity, as desired. □

4. Prove that

$$1-\binom{n}{2}+\binom{n}{4}-\binom{n}{6}+\cdots = 2^{\frac{n}{2}}\cos\frac{n\pi}{4}, \qquad n\in\mathbb{N}.$$

Proof. Due to the binomial theorem we have

$$\begin{aligned}(1+i)^n &= 1+i\binom{n}{1}-\binom{n}{2}-i\binom{n}{3}+\binom{n}{4}+\cdots\\(1-i)^n &= 1-i\binom{n}{1}-\binom{n}{2}+i\binom{n}{3}+\binom{n}{4}+\cdots.\end{aligned}$$

Adding the corresponding sides, we get

$$\begin{aligned}1-\binom{n}{2}+\binom{n}{4}-\binom{n}{6}+\cdots &= \frac{1}{2}[(1+i)^n+(1-i)^n]\\&= \mathrm{Re}\left[\left(\sqrt{2}e^{\frac{2\pi i}{8}}\right)^n\right]\\&= \mathrm{Re}\left(2^{\frac{n}{2}}e^{\frac{2n\pi i}{8}}\right)\\&= 2^{\frac{n}{2}}\cos\frac{n\pi}{4}.\end{aligned}$$

□

5. Compute

$$\binom{n}{1}\cos x+\binom{n}{2}\cos 2x+\cdots+\binom{n}{n}\cos nx.$$

Solution. It holds

$$\begin{aligned}1+\binom{n}{1}\cos x+\binom{n}{2}\cos 2x+\cdots+\binom{n}{n}\cos nx &= \operatorname{Re}\left(1+\binom{n}{1}e^{ix}+\cdots+\binom{n}{n}e^{inx}\right)\\ &= \operatorname{Re}\left[(1+e^{ix})^n\right].\end{aligned}$$

We will try to find the polar form of $1+e^{ix}$; it holds

$$1+e^{ix}=1+\cos x+i\sin x=2\cos^2\frac{x}{2}+i\cdot 2\sin\frac{x}{2}\cos\frac{x}{2}=2\cos\frac{x}{2}e^{i\frac{x}{2}},$$

therefore,

$$\begin{aligned}\binom{n}{1}\cos x+\binom{n}{2}\cos 2x+\cdots+\binom{n}{n}\cos nx &= \operatorname{Re}\left[\left(2\cos\frac{x}{2}e^{i\frac{x}{2}}\right)^n\right]-1\\ &= \operatorname{Re}\left[2^n\cos^n\frac{x}{2}e^{\frac{inx}{2}}\right]-1\\ &= 2^n\cos^n\frac{x}{2}\cos\frac{nx}{2}-1. \qquad \square\end{aligned}$$

Part III

Linear Algebra

Chapter 1

Basic Notions

1.1 Vector Spaces

A set V is called a *vector space* over $\mathbb{R}$ or $\mathbb{C}$ (or more generally, over a field $\mathbb{F}$) if it satisfies the following properties:

(i) V is endowed with an operation ("addition" of vectors) satisfying

1. $v + w = w + v$, $\forall v, w \in V$.
2. $(v + w) + u = v + (w + u)$, $\forall v, w, u \in V$.
3. $\exists 0 \in V$, such that $u + 0 = 0 + u = u$, $\forall u \in V$.
4. $\forall v \in V$, $\exists -v \in V$ such that $v + (-v) = 0$.

(ii) There is a scalar multiplication, $\lambda \cdot v \in V$, where $\lambda \in \mathbb{R}$ (or $\mathbb{C}$), such that

5. $\lambda(u + v) = \lambda u + \lambda v$, $\forall u, v \in V$, $\forall \lambda$.
6. $(\lambda + \mu)v = \lambda u + \mu u$, $\forall u \in V$, $\forall \lambda, \mu$.
7. $1 \cdot u = u$, $\forall u \in V$.
8. $(\lambda\mu)u = \lambda(\mu u)$, $\forall u \in V$, $\forall \lambda, \mu$.

The elements λ, μ, are called *scalars*.

The basic example of a real vector space is $\mathbb{R}^n$ with componentwise addition:

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_1, \dots, x_n \in \mathbb{R} \right\},$$

with

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{bmatrix}.$$

Also, $\mathbb{C}$ itself is a vector space over $\mathbb{R}$.

Definition 1.1.1 (Linear (in)dependence)**.** Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). The elements $v_1, \dots, v_m \in V$ are called *linearly dependent* if there are scalars $\lambda_1, \dots, \lambda_m$, not all zero, such that

$$\lambda_1 v_1 + \cdots + \lambda_m v_m = 0,$$

i. e. a linear combination of $v_1, \dots, v_m$ equals zero. Otherwise, they are called *linearly independent*.

Example. The vectors

$$\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix} \in \mathbb{R}^2$$

are linearly dependent, as

$$\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}2\\1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix},$$

but

$$\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}$$

are linearly independent.

Remark. It is important to emphasize that linear (in)dependence depends very much on the field of scalars as well; for example, consider $\mathbb{C}$ as a vector space over $\mathbb{R}$ and over $\mathbb{C}$. The elements 1, i, are linearly independent over $\mathbb{R}$; if $a, b \in \mathbb{R}$ such that $a+bi = 0$, then $a = \operatorname{Re} 0 = 0 = \operatorname{Im} 0 = b$. However, they are linearly dependent over $\mathbb{C}$, as

$$i \cdot 1 + (-1) \cdot i = 0.$$

1.2 Dimension and Basis

Let V be a vector space. The maximum number of linearly independent elements of V is the *dimension* of V; such a set is called a *basis*. We emphasize that the dimensions depends on the field of scalars, and is denoted by $\dim_{\mathbb{R}} V$ or $\dim_{\mathbb{C}} V$, depending on whether we consider V as a real or complex vector space. When the field of scalars is unambiguous, we simply write $\dim V$.

Example.

1. $\mathbb{R}^n$ has dimension n; the standard basis is

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$$

2. It holds $\dim_{\mathbb{R}} \mathbb{C} = 2$; a basis is 1, i. On the other hand, $\dim_{\mathbb{C}} \mathbb{C} = 1$. Generally, $\dim_{\mathbb{R}} \mathbb{C}^n = 2 \dim_{\mathbb{C}} \mathbb{C}^n = 2n$.

3. Let $\mathbb{R}_n[x]$ denote the set of polynomials with real coefficients of degree $\leq n$. Then, $\dim_{\mathbb{R}} \mathbb{R}_n[x] = n + 1$. A basis is

$$1, x, x^2, \dots, x^n.$$

4. Let $C[a, b]$ denote the set of continuous real functions $f : [a, b] \to \mathbb{R}$, with the usual addition and scalar multiplication:

$$(f + g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x).$$

   Then, $C[a, b]$ is a real vector space of infinite dimension.

5. $\mathbb{R}$ is a vector space over $\mathbb{Q}$ of infinite dimension.

1.3 Inner Product and Orthogonality

Between two vectors $x, y \in \mathbb{R}^n$ with

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

we define their *inner product* as follows:

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j y_j.$$

If $x, y \in \mathbb{C}^n$, then the inner product is defined as

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j \bar{y}_j.$$

We say that x is *orthogonal* to y if $\langle x, y \rangle = 0$. The inner product satisfies the following properties for every $x, y, z \in V$ and scalars λ, μ (V is either $\mathbb{R}^n$ or $\mathbb{C}^n$):

1. $\langle x, x \rangle \geq 0$, with equality if and only if $x = 0$.
2. $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
3. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$.
4. $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$.

1.4 Exercises

1. Suppose that $x_1, \ldots, x_m \in \mathbb{R}^n \setminus \{0\}$ are pairwise orthogonal. Show that they are linearly independent.

   Proof. Let $\lambda_1 x_1 + \cdots + \lambda_m x_m = 0$, for some scalars $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$. We will show that this forces all λ_j to be zero. Indeed, for every j we have

   $$0 = \langle \lambda_1 x_1 + \cdots + \lambda_m x_m, x_j \rangle = \sum_{k=1}^{m} \lambda_k \langle x_k, x_j \rangle = \lambda_j \langle x_j, x_j \rangle,$$

   whence $\lambda_j = 0$, as desired. □

2. **(IMC '18)** Given $k \in \mathbb{N}$, find the smallest $n \in \mathbb{N}$ such that there are nonzero vectors

   $$u_1, \ldots, u_k \in \mathbb{R}^n$$

   with $\langle u_i, u_j \rangle = 0$ whenever $|i - j| > 1$.

Solution. For $k = 1, 2$, the restriction $|i - j| > 1$ is never valid, therefore we can choose $n = 1$ (the smallest natural number). Let's see what happens when $k = 3$; the only restriction we have is that $u_1 \perp u_3$, therefore u_1 and u_3 are linearly independent, so that $n \geq 2$. We can actually have $n = 2$, and choose $u_1 = \mathbf{e}_1$, $u_3 = \mathbf{e}_2$, and $u_2 \in \mathbb{R}^2 \setminus \{0\}$ be arbitrary

In order to get a better idea, let's also try $k = 4$. The restrictions we have this time are

$$u_1 \perp u_3, u_1 \perp u_4, u_2 \perp u_4.$$

We can again choose $n = 2$, and $u_1 = u_2 = \mathbf{e}_1$, $u_3 = u_4 = \mathbf{e}_2$.

It seems that for given k, n is roughly half; it seems also evident that we can choose two consecutive vectors be equal to a standard basis vector. In particular, we'll show that

$$n = \lceil \frac{k}{2} \rceil = \begin{cases} \frac{k}{2}, & k \text{ even} \\ \frac{k+1}{2}, & k \text{ odd.} \end{cases}$$

This is indeed so; if k is even, we observe that $u_1, u_3, \ldots, u_{k-1}$ are pairwise orthogonal, hence linearly independent, so that $n \geq \frac{k}{2}$. In fact, we may choose $n = \frac{k}{2}$ and

$$u_1 = u_2 = \mathbf{e}_1, u_3 = u_4 = \mathbf{e}_2, \ldots, u_{2\ell-1} = u_{2\ell} = \mathbf{e}_\ell, \ldots, u_{k-1} = u_k = \mathbf{e}_n.$$

On the other hand, if k is odd, then $u_1, u_3, \ldots, u_k$ are pairwise orthogonal and $n \geq \frac{k+1}{2}$. In fact, we may choose $n = \frac{k+1}{2}$ and

$$u_1 = u_2 = \mathbf{e}_1, u_3 = u_4 = \mathbf{e}_2, \ldots, u_{2\ell-1} = u_{2\ell} = \mathbf{e}_\ell, \ldots, u_{k-2} = u_{k-1} = \mathbf{e}_{n-1}, u_k = \mathbf{e}_n.$$

It is clear that the above selections satisfy the given restriction, proving the desired result. $\square$

Chapter 2

Matrices

2.1 Definitions

A matrix is a rectangular array of elements in a field of scalars ($\mathbb{R}$ or $\mathbb{C}$), having m rows and n columns:

$$A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

If $m = n$ the matrix is called *square*. Some special square matrices are the *identity* and *zero* matrix, which are respectively

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \quad O_n = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

So, for the elements of the identity matrix it holds

$$a_{ij} = \begin{cases} 1, & i = j \\ 0, & i \ne j, \end{cases}$$

while for those of the zero matrix we simply have $a_{ij} = 0$ for all i, j.

The set of all matrices having m rows and n columns shall be denoted henceforth by $\mathbb{R}^{m\times n}$ or $\mathbb{C}^{m\times n}$, depending on the field of scalars we consider. Given a matrix $A = (a_{ij}) \in \mathbb{C}^{m\times n}$, we define the *conjugate* of A to be $\bar{A} = (\bar{a}_{ij})$. The *transpose* of A is the matrix

$$A^\top = (a_{ji})_{\substack{1\le j\le n\\ 1\le i\le m}} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}.$$

The *conjugate transpose* is denoted by $A^\star$, which satisfies

$$A^\star = (\bar{A})^\top = \overline{A^\top}.$$

A matrix is called *symmetric* if $A = A^\top$ and *Hermitian* if $A = A^\star$. Transposition and conjugate trasnposition are *involutions*, that is, they are operations that if applied twice, one ends up where they started, i.e.

$$(A^\top)^\top = A, \quad (A^\star)^\star = A.$$

The sum of the diagonal elements of a square matrix $A \in \mathbb{C}^{n\times n}$ is called the *trace* of A:

$$\operatorname{Tr} A = a_{11} + a_{22} + \cdots + a_{nn}.$$

Lemma 2.1.1. *Let $A \in \mathbb{R}^n$. Then $\operatorname{Tr} AA^\top = 0$ if and only if $A = O_n$.*

Proof. Let $A = (a_{ij})$, so that $A^\top = (b_{ij})$ with $b_{ij} = a_{ji}$. The (i,i) entry of $AA^\top$ is

$$\sum_{j=1}^n a_{ij} b_{ji} = \sum_{j=1}^n a_{ij} a_{ij} = \sum_{j=1}^n a_{ij}^2.$$

Summing over i, we obtain the trace of $AA^\top$, hence

$$\operatorname{Tr} AA^\top = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2,$$

that is, the trace of $AA^\top$ is the sum of the squares of *all* entries of A. Therefore, this trace is zero if and only if $A = O_n$. ☐

2.2 Multiplication

The product $A \cdot B$ of two matrices A, B is defined only if the number of columns of A equals the number of rows of B. So, suppose that $A = (a_{ij}) \in \mathbb{C}^{m\times n}$ and $B = (b_{jk}) \in \mathbb{C}^{n\times p}$. Then, their product equals

$$A \cdot B = C = (c_{ik}) = \left(\sum_{j=1}^n a_{ij} b_{jk} \right)_{\substack{1\le i\le m\\ 1\le k\le p}} \in \mathbb{C}^{m\times p}.$$

If A and B are real matrices, we observe that the (i,k) entry of the matrix C is simply the inner product of the ith row of A with the kth column of B, considered as vectors in $\mathbb{R}^n$.

From the definition we immediately observe that the product $B \cdot A$ is not defined, unless $m = p$; even in that case, we have $A \cdot B \in \mathbb{C}^{m\times m}$ and $B \cdot A \in \mathbb{C}^{n\times n}$. But even when both A and B are square matrices of the same size, it is not guaranteed that $A \cdot B = B \cdot A$. Matrix multiplication in $\mathbb{C}^{n\times n}$ is not commutative; however, there are square matrices that commute with every other matrix of the same size. If $A \in \mathbb{C}^{n\times n}$, we have

$$AI_n = I_n A = A, \quad AO_n = O_n A = O_n.$$

If for $A \in \mathbb{C}^{n\times n}$ there is $B \in \mathbb{C}^{n\times n}$ such that $AB = I_n$, then A is called *invertible*, and the matrix B is the *inverse* of A, denoted by A^{-1}. It holds

$$AA^{-1} = A^{-1}A = I_n.$$

For two square matrices A and B, the commutator is defined as

$$[A, B] = AB - BA,$$

which, roughly speaking, shows the failure of commutation between A and B. A and B commute if and only if $[A, B] = O$. Despite the noncommutativity of multiplication, the following result shows us that the trace is *blind* to this phenomenon, even though the products AB and BA might have different size.

Theorem 2.2.1. *Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times m}$. Then,*

$$\operatorname{Tr} AB = \operatorname{Tr} BA$$

Proof. Suppose that $A = (a_{ij})$, $B = (b_{jk})$. The diagonal elements of AB have the form

$$\sum_{j=1}^{n} a_{ij}b_{ji}, \quad 1 \le i \le m$$

and those of BA have the form

$$\sum_{i=1}^{m} b_{ji}a_{ij}, \quad 1 \le j \le n.$$

Summing over i and j respectively, we obtain the corresponding traces, therefore

$$\operatorname{Tr} AB = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} = \operatorname{Tr} BA. \qquad \square$$

2.3 Exercises

1. Let $M \in \mathbb{C}^{n\times n}$. Prove that there are Hermitian A and B, such that $M = A + iB$.

   Solution. Assuming that M had already that form, we will try to determine A and B in terms of M. We would have $\bar{M} = \bar{A} - i\bar{B}$, and taking transposes on both sides we obtain

   $$M^\star = A^\star - iB^\star = A - iB,$$

   the latter equality holding if A and B are Hermitian. If that is the case, then

   $$A = \frac{1}{2}[(A + iB) + (A - iB)] = \frac{1}{2}[M + M^\star]$$

   and

   $$B = \frac{1}{2i}[(A + iB) - (A - iB)] = \frac{1}{2i}[M - M^\star].$$

   Obviously, the above formulae for A and B determine Hermitian matrices, as

   $$(\frac{1}{2}[M + M^\star])^\star = \frac{1}{2}[M^\star + (M^\star)^\star] = \frac{1}{2}[M^\star + M]$$

   and

   $$(\frac{1}{2i}[M - M^\star])^\star = -\frac{1}{2i}[M^\star - (M^\star)^\star] = -\frac{1}{2i}[M^\star - M],$$

   completing the proof. $\square$

2. Are there $A, B \in \mathbb{C}^{n\times n}$ such that $AB - BA = I_n$?

Answer. No. Taking traces on both sides we obtain

$$\operatorname{Tr} I_n = n \neq 0 = \operatorname{Tr} AB - \operatorname{Tr} BA = \operatorname{Tr}(AB - BA),$$

showing that we can never have $AB - BA = I_n$. □

Remark. The above shows that the trace of any commutator is always zero, i. e. $\operatorname{Tr}[A, B] = 0$.

3. Let $A, B \in \mathbb{C}^{n\times n}$ be two non-commuting matrices, for which there are $p, q, r \in \mathbb{R}$, all nonzero, such that

$$pAB + qBA = I_n \tag{2.3.1}$$
$$A^2 = rB^2. \tag{2.3.2}$$

Prove that $p = q$.

Proof. Multiplying both sides of (2.3.1) by A on the left and applying (2.3.2) we obtain

$$A = pA^2B + qABA = prB^3 + qABA,$$

and when we multiply on the right we get

$$A = pABA + qBA^2 = pABA + qrB^3.$$

Hence,

$$O = prB^3 + qABA - (pABA + qrB^3) = (p - q)[rB^3 - ABA].$$

If $p \neq q$, then

$$(p - q)A = pA - qA = p(prB^3 + qABA) - q(pABA + qrB^3) = (p^2 - q^2)rB^3,$$

and dividing by $p - q$ we obtain

$$A = (p + q)rB^3,$$

showing that A and B commute, contradiction. Thus, $p = q$. □

4. Compute the nth power of

$$J_m(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix} \in \mathbb{C}^{m\times m}.$$

Solution. We write $J_m(\lambda) = \lambda I_m + N_m$, where

$$N_m = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Since I_m and N_m commute, we will have

$$J_m(\lambda)^n = (\lambda I_m + N_m) = \lambda^n I_m + \binom{n}{1}\lambda^{n-1}N_m + \binom{n}{2}\lambda^{n-2}N_m^2 + \cdots$$

So, it suffices to find the powers of N_m. If $N_m = (n_{ij})$, then

$$n_{ij} = \begin{cases} 1, & j = i+1 \\ 0, & j \neq i+1. \end{cases}$$

We calculate first N_m^2:

$$N_m^2 = \left(\sum_{j=1}^{m} n_{ik}n_{kj} \right)_{1 \le i,j \le m}.$$

The (i,j) entry of the above matrix can only be nonzero when $j = i+2$; in that case, we have $n_{i,i+2} = 1$, as the sum above has only zero terms, except for the term with $k = i+1$, which equals 1. We may use induction to show that $N_m^k = \left(n_{ij}^{(k)}\right)$, where

$$n_{ij}^{(k)} = \begin{cases} 1, & j = i+k \\ 0, & j \neq i+k, \end{cases},$$

when $k < m$, and $N_m^k = O_m$, when $k \ge m$. The latter holds since

$$N_m^m = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix} = O_m.$$

We conclude by writing

$$J_m(\lambda)^n = \begin{bmatrix} \lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \dots & \binom{n}{m-1}\lambda^{n-m+1} & \binom{n}{m}\lambda^{n-m} \\ 0 & \lambda^n & \binom{n}{1}\lambda^{n-1} & \dots & \binom{n}{m-2}\lambda^{n-m+2} & \binom{n}{m-1}\lambda^{n-m+1} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda^n & \binom{n}{1}\lambda^{n-1} \\ 0 & 0 & 0 & \dots & 0 & \lambda^n \end{bmatrix},$$

where the symbol $\binom{n}{m}$ equals zero when $m > n$. □

5. Let $A, B \in \mathbb{R}^{n \times n}$ satisfy

$$\mathrm{Tr}(AA^\top + BB^\top) = \mathrm{Tr}(AB + A^\top B^\top).$$

Show that $A = B^\top$.

Proof. The given condition is equivalent to

$$\begin{aligned}0 &= \mathrm{Tr}(AA^\top - AB - A^\top B^\top + BB^\top)\\ &= \mathrm{Tr}(A(A^\top - B) - (A^\top - B)B^\top)\\ &= \mathrm{Tr}(A(A^\top - B)) - \mathrm{Tr}((A^\top - B)B^\top)\\ &= \mathrm{Tr}(A(A^\top - B) - \mathrm{Tr}(B^\top(A^\top - B))\\ &= \mathrm{Tr}(A(A^\top - B) - B^\top(A^\top - B))\\ &= \mathrm{Tr}((A - B^\top)(A^\top - B))\\ &= \mathrm{Tr}((A - B^\top)(A - B^\top)^\top),\end{aligned}$$

which due to Lemma 2.1.1 is equivalent to $A - B^\top = O_n$, or $A = B^\top$, as desired. ☐

6. **(SEEMOUS '19)** Let $m, n \in \mathbb{N}$. Prove that for any $A_1, \ldots, A_m \in \mathbb{R}^{n\times n}$, there are $\varepsilon_1, \ldots, \varepsilon_m \in \{\pm 1\}$, such that

$$\mathrm{Tr}\left((\varepsilon_1 A_1 + \cdots + \varepsilon_m A_m)^2\right) \geq \mathrm{Tr}(A_1^2) + \cdots + \mathrm{Tr}(A_m^2).$$

Proof. We will apply induction on m. For $m = 1$ it is trivial, as the given inequality becomes $\mathrm{Tr}\, A_1^2 \geq \mathrm{Tr}\, A_1^2$. Assume that the conclusion holds for any selection of $m = k$ matrices, namely for any $A_1, \ldots, A_k \in \mathbb{R}^{n\times n}$, there are $\varepsilon_1, \ldots, \varepsilon_k \in \{\pm 1\}$, such that

$$\mathrm{Tr}\left((\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)^2\right) \geq \mathrm{Tr}(A_1^2) + \cdots + \mathrm{Tr}(A_k^2).$$

We will try to prove it for $m = k+1$, so suppose that in the above selection of matrices we have one more, namely $A_{k+1} \in \mathbb{R}^{n\times n}$. For $\varepsilon_{k+1} = \pm 1$ we have:

$$\begin{aligned}&\mathrm{Tr}\left((\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k + \varepsilon_{k+1} A_{k+1})^2\right)\\ &= \mathrm{Tr}((\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)^2 + A_{k+1}^2 + \varepsilon_{k+1} A_{k+1}(\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)\\ &+ \varepsilon_{k+1}(\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k) A_{k+1})\\ &= \mathrm{Tr}\left((\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)^2\right) + \mathrm{Tr}\, A_{k+1}^2 + 2\varepsilon_{k+1}\, \mathrm{Tr}\left(A_{k+1}(\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)\right)\\ &\geq \sum_{j=1}^{k+1} \mathrm{Tr}\, A_j^2 + 2\varepsilon_{k+1}\, \mathrm{Tr}\left(A_{k+1}(\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)\right),\end{aligned}$$

using the inductive hypothesis and Theorem 2.2.1. So, if

$$\mathrm{Tr}\left(A_{k+1}(\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)\right) \geq 0$$

we select $\varepsilon_{k+1} = 1$, while if

$$\mathrm{Tr}\left(A_{k+1}(\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k)\right) < 0$$

we select $\varepsilon_{k+1} = -1$. Therefore, there is always a selection of sign for ε_{k+1}, such that

$$\mathrm{Tr}\left((\varepsilon_1 A_1 + \cdots + \varepsilon_k A_k + \varepsilon_{k+1} A_{k+1})^2\right) \geq \sum_{j=1}^{k+1} \mathrm{Tr}\, A_j^2,$$

proving the conslusion for $m = k+1$ as well. By induction, it should then hold for every $m \in \mathbb{N}$, as desired. ☐

Chapter 3

Linear Maps

3.1 Matrices as linear maps

Let V and W be two finite dimensional vector spaces, and let $T : V \to W$ be a map satisfying

$$T(av + bu) = aT(v) + bT(w), \quad \forall v, w \in V,$$

for all scalars a, b. Such a map is called a *linear map*. The image of the map

$$T(V) = \{w \in W : \exists v, w = T(v)\}$$

is called the *range* of T, while the set of vectors mapping to zero in W (i.e. $T^{-1}(\{0\})$), is called the *kernel* of T, and is denoted by $\ker T$.

If $T(V) = W$, then T is called *surjective* (i.e. onto), whereas if $\ker T = \{0\}$, then T is called injective (i.e. $1-1$). If T satisfies both properties, it is called *bijective*.

In the usual case, we may have $V = \mathbb{R}^n$ or $\mathbb{C}^n$, and $W = \mathbb{R}^m$ or $\mathbb{C}^m$, respectively. Suppose that

$$\mathbf{e}_1, \ldots, \mathbf{e}_n$$

is the standard basis of V while

$$\mathbf{f}_1, \ldots, \mathbf{f}_m$$

is the standard basis of W. It is convenient to express each element of V or W as a linear combination of the standard basis in each space (the coefficients appearing are the coordinates with respect to each basis). We observe that the values of T on $\mathbf{e}_1 \ldots, \mathbf{e}_n$ completely determine the values of T on V, as

$$T(\lambda_1 \mathbf{e}_1 + \cdots + \lambda_n \mathbf{e}_n) = \lambda_1 T(\mathbf{e}_1) + \cdots + \lambda_n T(\mathbf{e}_n),$$

by linearity, for all scalars $\lambda_1, \ldots, \lambda_n$. The value of $T(\mathbf{e}_j)$ is a linear combination of $\mathbf{f}_1, \ldots, \mathbf{f}_m$ as follows

$$T(\mathbf{e}_j) = a_{1j}\mathbf{f}_1 + \cdots + a_{mj}\mathbf{f}_m.$$

The matrix $A = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$ is the *matrix representation* of T with respect to the given bases.

How does A change, if we consider different bases? Let's consider the special case where $m = n$ and A is the matrix representation of T with respect to the same basis on $V = W$, the standard one. First, we note that $T(v) = Av$, where the right-hand side is simply a matrix product between A and $v \in \mathbb{R}^n$, expressed as a $n \times 1$ matrix;

the result is again a $n \times 1$ matrix. Now, suppose that we want to express T with respect to the basis

$$v_1, \dots, v_n.$$

Let S be the $n \times n$ matrix whose jth column equals v_j; S is invertible. Then, the matrix representation of T with respect to $v_1, \dots, v_n$ is $S^{-1}AS$. The matrices A and $S^{-1}AS$ are called *similar*, and represent the same linear map with respect to different bases. For this reason, these two matrices share a lot of features in common as we shall see, for example, eigenvalues, trace, etc.

3.2 Eigenvalues and eigenvectors

Let $T : V \to V$ be a linear map. A vector $v \in V$ is called an *eigenvector* if there is some scalar λ such that $T(v) = \lambda v$. The scalar λ is called *eigenvalue* of T. Of course, we can talk about eigenvalues and eigenvectors of a matrix A, viewed as a linear map. From the definition of an eigenvalue, it is evident that the set of eigenvalues remains invariant when we change the basis of V. Therefore, A shares the set of eigenvalues with all similar matrices SAS^{-1}, where $S : V \to V$ is invertible.

The set of eigenvalues, denoted by $\lambda(A)$, is more accurately a *multiset*, as we count each eigenvalue with a multiplicity. For example, consider the matrix

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

The matrix has three eigenvectors, namely the vectors of the standard orthonormal basis, $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. The first two correspond to the eigenvalue 1, while the third to the eigenvalue 2. Actually, every linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$ is an eigenvector corresponding to 1, which implies that the set of eigenvectors corresponding to 1 is a subspace of $\mathbb{R}^3$ of dimension 2 (the subspace of eigenvectors corresponding to an eigenvalue λ is called *eigenspace*, and is usually denoted by E_λ). That is why, we write $\lambda(A) = \{1, 1, 2\}$, i. e. 1 is included twice to emphasize the fact that it has multiplicity 2. Here, we use the notion of the *geometric* multiplicity, which is precisely the dimension of the *eigenspace* E_1; however, the more appropriate term is the *algebraic* multiplicity, to be introduced below (in this example, the two notions coincide). We also note that the set of eigenvalues $\lambda(A)$, counting algebraic multiplicities is also called *spectrum* of A, or the set of *characteristic roots* of A.

If A had a basis of eigenvectors $v_1, \dots, v_n$ corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$, then the matrix representation with respect to the basis $v_1, \dots, v_n$ would be the diagonal matrix

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

We observe that the trace of this matrix is precisely the sum of its eigenvalues, which also equals $\operatorname{Tr} A$, since A is similar to the above matrix. This holds generally, not necessarily for matrices A that accept a basis of eigenvectors.

Theorem 3.2.1. *Let $A \in \mathbb{C}^{n\times n}$, with eigenvalues (counting multiplicities) $\lambda_1, \dots, \lambda_n$. Then*

$$\operatorname{Tr} A = \lambda_1 + \dots + \lambda_n.$$

Chapter 4

Determinants

4.1 Basics, definitions

The notion of determinant is defined for square matrices only. For a 2×2 matrix

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

its determinant equals $ad - bc$, and is denoted by $\det A$ or $|A|$. In 3×3 matrices we have

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31}.$$

Given $A = (a_{ij})_{1\leq i,j\leq n} \in \mathbb{C}^{n\times n}$ we denote by A_{ij} the determinant of the matrix which is obtained from A by removing the ith row and jth column (also called the (i,j)-minor). Then, $\det A$ is defined inductively as

$$|A| = a_{11}A_{11} - a_{12}A_{12} + \cdots + (-1)^{1+n}a_{1n}A_{1n}.$$

The *adjugate* matrix of A is the transpose of the $n \times n$ matrix formed by these minors with the appropriate sign:

$$\operatorname{adj} A = ((-1)^{i+j}A_{ij})^{\top}_{1\leq i,j\leq n}.$$

A is invertible if and only if $\det A \neq 0$, in which case it holds

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A.$$

It also holds

$$\det A = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}$$

where S_n is the set of permutations of $\{1, 2, \ldots, n\}$ and $\operatorname{sign}(\sigma) = \pm 1$ is the sign of the permutation σ (we will not define it here rigorously). The above sum has $n!$ terms, and each term is the product of elements of A on a so-called *generalized diagonal*.

The determinant has also a nice geometric interpretation. Suppose that the columns of A are the vectors $v_1, \ldots, v_n$. Then, $|\det A|$ equals the volume of the parallelotope spanned by the vectors $v_1, \ldots, v_n$. Therefore, $\det A = 0$ if and only if the columns are linearly independent, which in turn is equivalent to A being invertible (or $\operatorname{rank} A = n$).

The calculation of a determinant seems difficult in the general case. However, there are some very simple cases which are very helpful to compute determinants of arbitrary matrices, namely upper triangular matrices:

$$\begin{vmatrix} a_1 & * & \dots & * \\ 0 & a_2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{vmatrix} = a_1 a_2 \cdots a_n.$$

A determinant can be expressed in terms of the eigenvalues of a matrix in the following simple way.

Theorem 4.1.1. *Let $A \in \mathbb{C}^{n\times n}$ having eigenvalues $\lambda_1, \dots, \lambda_n$. Then*

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n.$$

4.2 Systems of linear equations

We will restrict to systems of n linear equations in n unknowns. Let $A \in \mathbb{C}^{n\times n}$ be the coefficient matrix, $\mathbf{b} \in \mathbb{C}^{n\times 1}$ the constant matrix, and

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

the variable vector. We consider the system

$$A\mathbf{x} = \mathbf{b}.$$

If $\mathbf{b} = 0$, the system is called homogeneous and always accepts the zero solution $\mathbf{x} = 0$. If $\det A \neq 0$, this solution is unique, otherwise it accepts infinitely many solutions. In the general (inhomogeneous) case, if $\det A \neq 0$, the system has the unique solution

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

If $\det A = 0$, it has either no solutions or infinitely many solutions.

4.3 Properties

We list a handful of properties for the determinant; below, $A, B \in \mathbb{C}^{n\times n}$ and $\lambda \in \mathbb{C}$.

1. $\det(\lambda A) = \lambda^n \det A$.
2. $\det(\bar{A}) = \overline{\det A}$, $\det A^\top = \det A$, $\det A^* = \overline{\det A}$.
3. $\det AB = \det A \det B$.
4. We define the *direct sum* of the square matrices A and B as follows (here, they may have different size):
$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$
   Then, $\det A \oplus B = \det A \det B$.

5. Exchanging two rows (or columns) changes the sign of the determinant. To be more precise, let $(v_1, \ldots, v_n)$ be the $n \times n$ matrix whose jth column is v_j. Then,

$$\det(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -\det(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n),$$

where in the latter matrix we have exchanged the columns v_i and v_j.

6. $$\det(v_1 + v_1', \ldots, v_n) = \det(v_1, \ldots, v_n) + \det(v_1', \ldots, v_n).$$

Exercises. 1. Let $A, B \in \mathbb{R}^{n \times n}$ be invertible and n odd. Show that

$$AB + BA \neq O_n.$$

Proof. Assume otherwise, hence $AB = -BA$. Taking determinants we obtain

$$0 \neq \det A \det B = \det AB = \det(-BA) = (-1)^n \det B \det A = -\det A \det B,$$

which is a contradiction. □

2. Let $A, B \in \mathbb{R}^{n \times n}$ be two commuting matrices (i.e. $AB = BA$). Show that $\det(A^2 + B^2) \geq 0$.

Proof. Since A and B commute we have

$$(A + iB)(A - iB) = A^2 - A(iB) + iBA - i^2B^2 = A^2 + B^2,$$

therefore

$$\det(A^2+B^2) = \det(A+iB)\det(A-iB) = \det(A+iB)\overline{\det(A + iB)} = |\det(A+iB)|^2 \geq 0,$$

due to $\overline{A - iB} = A + iB$, since A and B are real matrices. This concludes the proof. □

3. Let n be odd, $A, B \in \mathbb{R}^{n \times n}$, such that $A^2 + B^2 = O_n$. Show that $AB - BA$ is not invertible.

Proof. It holds

$$(A + iB)(A - iB) = A^2 + B^2 + i(BA - AB) = i(BA - AB),$$

by hypothesis. Taking determinants we obtain

$$|\det(A + iB)|^2 = (-i)^n \det(AB - BA).$$

If $AB - BA$ were invertible then

$$i^n = (-1)^n \frac{|\det(A + iB)|^2}{\det(AB - BA)} \in \mathbb{R},$$

a contradiction, since n is odd (whence $i^n = \pm i \notin \mathbb{R}$). □

4.4 The characteristic and minimum polynomials

Let A be a $n \times n$ matrix. The *characteristic polynomial* of A is defined as

$$\chi_A(\lambda) = \det(\lambda I_n - A),$$

and is a polynomial of degree n. The coefficients of λ^{n-1} and the constant one are known, while the others are more complicated:

$$\chi_A(\lambda) = \lambda^n - \operatorname{Tr} A \lambda^{n-1} + \cdots + (-1)^n \det A.$$

If A is a real matrix, then χ_A has real coefficients. The most important aspect of χ_A is tha fact that its roots are precisely the eigenvalues of A; indeed, λ is an eigenvalue if and only if there is nonzero v such that $Av = \lambda v$, or equivalently $(\lambda I_n - A)v = 0$. The latter can be viewed as a $n \times n$ homogeneous linear system, having a nonzero solution. This holds if and only if $\chi_A(\lambda) = \det(\lambda I_n - A) = 0$. The multiplicity of a root λ appearing in χ_A, is called the *algebraic* multiplicity of the eigenvalue λ.

A polynomial expression of a matrix A has the form

$$p(A) = a_n A^n + a_{n-1} A^{n_1} + \cdots + a_1 A + a_0 I,$$

where

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{C}[x]$$

is a polynomial of degree n ($a_n \neq 0$) with complex coefficients.

Theorem 4.4.1. *[Cayley-Hamilton] Let $A \in \mathbb{C}^{n\times n}$ and χ_A be its characteristic polynomial. Then, $\chi_A(A) = O_n$.*

The polynomial expression of A related to the characteristic polynomial is thus the zero matrix. All polynomials with this property, are multiples of another polynomial attached to A, the *minimum* polynomial.

Definition 4.4.2. Let $A \in \mathbb{C}^{n\times n}$. The monic polynomial $\varphi(x) \in \mathbb{C}[x]$ of smallest degree that satisfies $\varphi(A) = O_n$ is called the minimum polynomial of A, denoted by m_A.

We list some basic and important properties of the minimum, characteristic polynomials, and the eigenvalues. The proofs of the statements below are left as an exercise to the reader.

Properties

(1) m_A is unique.

(2) If $f(x) \in \mathbb{C}[x]$ with $f(A) = O$, then m_A divides f, i. e. there is $g(x) \in \mathbb{C}[x]$ such that $f(x) = m_A(x) g(x)$.

(3) Suppose that

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k}.$$

Then,

$$m_A(x) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_k)^{m_k},$$

where

$$0 < m_j \leq n_j, \quad 1 \leq j \leq k.$$

In other words, the minimum polynomial has exactly the same roots with the characteristic polynomial, albeit with a smaller multiplicity.

(4) Suppose that
$$A = \begin{bmatrix} B & E \\ O & D \end{bmatrix}.$$
Then $\chi_A(\lambda) = \chi_B(\lambda)\chi_D(\lambda)$ and thence $\lambda(A) = \lambda(B) \cup \lambda(D)$ (counting multiplicities).

(5) If A is upper triangular, then the diagonal elements of A are precisely its eigenvalues, appearing with the same multiplicity.

(6) A is invertible if and only if zero is not an eigenvalue of A.

(7) If $\lambda(A) = \{\lambda_1, \ldots, \lambda_n\}$ and $P(x) \in \mathbb{C}[x]$, then $\lambda(P(A)) = \{P(\lambda_1), \ldots, P(\lambda_n)\}$. In particular,
$$\lambda(aI_n + bA) = a + b\lambda(A) = \{a + b\lambda_1, \ldots, a + b\lambda_n\},$$
and
$$\lambda(A^p) = \lambda(A)^p = \{\lambda_1^p, \ldots, \lambda_n^p\},$$
for $p \in \mathbb{N}$.

Exercises. 1. Show that there does not exist a matrix $A \in \mathbb{R}^{2021\times 2021}$ satisfying $A^2 - A + I = O$.

Proof. The minimum polynomial $m_A(\lambda)$ divides $\lambda^2 - \lambda + 1$, by Property **(2)**. Since A is real, m_A cannot equal either $\lambda - \rho_1$ or $\lambda - \rho_2$, where
$$\rho_{1,2} = \frac{1 \pm i\sqrt{3}}{2} \notin \mathbb{R}.$$
Therefore,
$$m_A(\lambda) = \lambda^2 - \lambda + 1.$$
By Property **(3)**, we must have
$$\chi_A(\lambda) = (\lambda - \rho_1)^n(\lambda - \rho_2)^k,$$
where $n + k = 2021$. But $\chi_A(\lambda) \in \mathbb{R}[\lambda]$, therefore its roots must appear in complex conjugate pairs (it has no real roots), which shows that $n = k$. However, this is a contradiction, as $2n = 2021$ has no integer solution. □

Second proof. A more elementary proof is the following: subtract $-A$ on both sides and take determinants, to obtain
$$-\det A = (-1)^{2021} \det A = \det(-A) = \det[(A - I)^2] = (\det(A - I))^2 \geq 0,$$
whence $\det A < 0$, since A is invertible, as it holds $A(I - A) = I$. On the other hand, if we add $3A$ on both sides of the given equation, we obtain
$$3^{2021} \det A = \det(3A) = \det[(A + I)^2] = (\det(A + I))^2 \geq 0,$$
which yields $\det A > 0$, a contradiction. □

2. Let $N \in \mathbb{R}^{n\times n}$ be such that $N^n = O \neq N^{n-1}$, for $n > 1$. Show that there does not exist a matrix X such that $X^2 = N$.

Proof. By hypothesis, we deduce that $m_N(\lambda) = \lambda^n = \chi_N(\lambda)$. If such a matrix X existed, it would also be nilpotent, as $X^{2n} = N^n = O$. Therefore, the only eigenvalue of X is zero with multiplicity n, which shows that $\chi_X(\lambda) = \lambda^n$. By the Cayley-Hamilton Theorem, we obtain $X^n = O$. If n is even, then

$$O = X^n = (X^2)^{\frac{n}{2}} = N^{\frac{n}{2}},$$

which is a conrtadiction since $\frac{n}{2} \leq n-1$. If $n \geq 3$ is odd, then

$$O = X^{n+1} = N^{\frac{n+1}{2}},$$

again a contradiction, since $\frac{n+1}{2} \leq n-1$. This shows that such a matrix X cannot exist. □

Chapter 5

Diagonalization and Jordan Canonical Form

Let $A \in \mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$. The goal of this section is to change the basis of the vector space, so that the linear operator $T(v) = Av$ has the simplest possible form.

5.1 Diagonalization

Definition 5.1.1. The matrix A is called *diagonalizable* if there is a basis of eigenvectors, i. e. $v_1, \ldots, v_n$ form a basis and

$$Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2, \ldots, Av_n = \lambda_n v_n,$$

for $\lambda_1, \ldots \lambda_n \in \mathbb{C}$ (the eigenvalues of A). We then have

$$A = V \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} V^{-1}$$

where

$$V = \begin{bmatrix} | & | & \ldots & | \\ v_1 & v_2 & \ldots & v_n \\ | & | & \ldots & | \end{bmatrix},$$

the matrix whose jth column is v_j.

Remark. If a matrix is diagonalizable, it is much easier to compute the powers of A (i. e. the *dynamics* of A), and this is the principal reason for diagonalizing a matrix. This is accomplished by simply computing the corresponding powers of the eigenvalues:

$$A^k = V \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix} V^{-1}.$$

Some basic initial properties are listed below.

Theorem 5.1.2.

(1) *If the eigenvalues of A are all distinct (i. e. the characteristic polynomial of A possesses only simple roots), then A is diagonalizable.*

(2) *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then A is diagonalizable over $\mathbb{R}$ and all eigenvalues are real.*

5.2 Jordan Canonical Form

Let $N_m \in \mathbb{C}^{m\times m}$ be defined as

$$N_m = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}$$

which was introduced in exercise 4. In that exercise we had defined the matrix $J_m(\lambda) = \lambda I_m + N_m$ and computed its powers; this matrix is called the *Jordan block* associated with the eigenvalue λ. We recall that N_m is a *nilpotent* matrix (a power of such a matrix equals the zero matrix), satisfying

$$N_m^m = O_m \neq N_m^{m-1}.$$

Not every square matrix A is diagonalizable; the best we can achieve when we change the basis in this case, is to have some 1's just above the diagonal (where the eigenvalues of the given matrix lie) and zeroes everywhere else. We write thus the matrix in a different basis as a diagonal concatenation of Jordan blocks associated to all eigenvalues of A, i. e. a similar matrix to A is a *direct sum* of Jordan blocks.

Let $A \in \mathbb{C}^{k\times k}$, $B \in \mathbb{C}^{m\times m}$. The direct sum of A and B is denoted by $A \oplus B \in \mathbb{C}^{(k+m)\times(k+m)}$ and is defined as

$$A \oplus B = \left[\begin{array}{c|c} A & O_{k\times m} \\ \hline O_{m\times k} & B \end{array}\right]$$

where $O_{k\times m}$ is the zero matrix of size $k \times m$. Considered as an operator on $\mathbb{C}^{(k+m)\times(k+m)}$, it sends the vector $(u, v) \in \mathbb{C}^{k+m}$ to (Au, Bv), where $u \in \mathbb{C}^k$, $v \in \mathbb{C}^m$.

Theorem 5.2.1. *[Jordan Canonical Form] Every $A \in \mathbb{C}^{n\times n}$ has a unique* Jordan Canonical Form *(JCF for short), that is, a matrix $D \in \mathbb{C}^{n\times n}$ such that $A = VDV^{-1}$ for $V \in \mathbb{C}^{n\times n}$ invertible, and*

$$D = \bigoplus_{j=1}^{k} B_j,$$

where the $B_1, \ldots, B_k$ are Jordan blocks associated with the eigenvalues of A. The columns of V are called generalized eigenvectors *of A. D is unique up to permutations of the Jordan blocks B_j.*

Corollary 5.2.2. *All possible JCFs for $n = 2$ are*

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

For $n = 3$ they are

$$\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}, \begin{bmatrix} \lambda_1 & 1 & \\ & \lambda_1 & \\ & & \lambda_2 \end{bmatrix}, \begin{bmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{bmatrix}.$$

The minimum polynomial $m_A(x)$ of A, gives us information about the largest Jordan block associated to an eigenvalue of A.

Theorem 5.2.3. *Let $A \in \mathbb{C}^{n\times n}$ and let*

$$m_A(x) = (x-\lambda_1)^{n_1}\cdots(x-\lambda_m)^{n_m}.$$

Then the JCF of A has at least one Jordan block of size n_j associated with λ_j, for $1 \le j \le m$.

Corollary 5.2.4. *The minimum polynomial m_A has only simple roots if and only if A is diagonalizable.*

If $A \in \mathbb{R}^{n\times n}$, the JCF of A might not belong to $\mathbb{R}^{n\times n}$; this happens if A has eigenvalues outside of $\mathbb{R}$. However, there is some version of the JCF, if we wish to restrict solely to real matrices. For this purpose, we need to define what a *real Jordan block* is. If the eigenvalue is real, we define $J_m(\lambda) = \lambda I_m + N_m$, as before. However, if $\lambda = a + bi$ is not real (i. e. $b \neq 0$), then $\bar{\lambda} = a - bi$ is also an eigenvalue of A, and a *real* Jordan block associated with *both* λ and $\bar{\lambda}$ is the following $2m \times 2m$ matrix

$$J_m(\lambda) = \left[\begin{array}{c|c|c|c|c|c} C & I_2 & O & \dots & O & O \\ \hline O & C & I_2 & \dots & O & O \\ \hline O & O & C & \dots & O & O \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline O & O & O & \dots & C & I_2 \\ \hline O & O & O & \dots & O & C \end{array}\right],$$

where

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad O = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Theorem 5.2.5. *[Real Jordan Canonical Form] Let $A \in \mathbb{R}^{n\times n}$. Then there is an invertible $V \in \mathbb{R}^{n\times n}$ such that $A = VRV^{-1}$, where R is the direct sum of real Jordan blocks associated with the eigenvalues of A. For a given eigenvalue λ, the largest Jordan block associated with λ appearing in the direct sum for R is $J_m(\lambda)$, where m is the multiplicity of λ in the minimum polynomial $m_A(\lambda)$.*

5.3 The Spectral Theorem and Simultaneous Diagonalization/Triangularization

A version of the spectral Theorem has already been given in Theorem 5.1.2**(2)**. This holds in greater generality.

Theorem 5.3.1. *[Spectral Theorem] Let $A \in \mathbb{C}^{n\times n}$ be a* normal *matrix, that is $AA^* = A^*A$. Then, there is an orthonormal basis of eigenvectors of A, say $v_1, \dots, v_n \in \mathbb{C}^n$. Moreover, $A = UDU^{-1}$, where $U \in \mathbb{C}^{n\times n}$ is the unitary matrix whose jth column is v_j, and D is a diagonal matrix, whose diagonal elements are the eigenvalues of A.*

When a set of matrices pairwise commute to each other, we can find a *common* basis of eigenvectors, under certain conditions.

Theorem 5.3.2. *[Simultaneous Diagonalization] Let $A_1, \dots, A_m \in \mathbb{C}^{n\times n}$ be pairwise commutative matrices, i. e. $A_iA_j = A_jA_i$ for every $1 \le i, j \le m$.*

(1) If all A_j are normal, then there is an orthonormal basis of common eigenvectors. Equivalently, there is a unitary matrix U, such that each UA_jU^* is diagonal.

(2) If all A_j are diagonalizable, then there is a basis of common eigenvectors (not necessarily orthogonal). Equivalently, there is an invertible $V \in \mathbb{C}^{n\times n}$, such that each VA_jV^{-1} is diagonal.

If we impose no restriction on the matrices A_j besides commutativity, the best thing we have is the simultaneous triangularization.

Theorem 5.3.3. *[Schur's Theorem] Let $A_1, \dots, A_m \in \mathbb{C}^{n\times n}$ be pairwise commutative matrices. Then, there is a unitary matrix U such that each UA_jU^* is upper triangular.*

Remark. Under the conditions of the above Theorem, it is not always true that we can bring all A_j simultaneously to their Jordan Canonical Form.

5.4 Special matrices

There are some determinants that appear frequently, both in exercises as well as in theory. Perhaps the most famous is the *Vandermonde* determinant, given by

$$V(x_1, \dots, x_n) = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \dots & x_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \dots & x_n^{n-1} \end{vmatrix} = \prod_{1\le i<j\le n} (x_j - x_i).$$

A circulant matrix is a matrix whose every row is obtained through cyclic permutations of the top row. For example,

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_{n-1} & x_n \\ x_n & x_1 & x_2 & \dots & x_{n-2} & x_{n-1} \\ x_{n-1} & x_n & x_1 & \dots & x_{n-3} & x_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_3 & x_4 & x_5 & \dots & x_1 & x_2 \\ x_2 & x_3 & x_4 & \dots & x_n & x_1 \end{bmatrix}.$$

An important special case of a Vandermonde matrix is the *Fourier* matrix, also known as the matrix of DFT (discrete Fourier transform). If $\omega = e^{\frac{2\pi i}{n}}$, then

$$F_n = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{n-2} & \dots & \omega \end{bmatrix} = \left(\omega^{(i-1)(j-1)}\right)_{1\le i<j\le n}.$$

It holds

$$F_n^2 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}$$

and $F_n^4 = I_n$. We note that the matrix $U_n = \frac{1}{\sqrt{n}} F_n$ is unitary, as

$$\left(\frac{1}{\sqrt{n}} F_n \right)^{-1} = \frac{1}{\sqrt{n}} F_n^*,$$

which implies $F_n^{-1} = \frac{1}{n} F_n^*$. Moreover, the columns of U_n form an orthonormal basis of eigenvectors of <u>*any*</u> circulant matrix, therefore by the spectral theorem $U_n^* X U_n$ is diagonal, and the eigenvalues of X are the values of the DFT of $(x_1, \dots, x_n)$, namely

$$x_1 + x_2\omega^k + \dots + x_n\omega^{(n-1)k}, 0 \le k \le n-1.$$

Thus,

$$\det X = \prod_{k=0}^{n-1} (x_1 + x_2\omega^k + \dots + x_n\omega^{(n-1)k}). \tag{5.4.1}$$

5.5 Exercises

1. Find the JCF of

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}.$$

Solution. For each matrix A from the above, we compute m_A, χ_A. The JCF of A will be denoted by D_A.

$\boxed{A = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]}$ It holds

$$\chi_A(\lambda) = \det(\lambda I_2 - A) = \begin{vmatrix} \lambda - 1 & 0 \\ -1 & \lambda - 1 \end{vmatrix} = (\lambda - 1)^2.$$

Since $\lambda I_2 - A \neq O$, we must have $m_A(\lambda) = \chi_A(\lambda) = (\lambda - 1)^2$. Therefore, A has only one double eigenvalue, $\lambda = 1$. Since it is also a double root of m_A, D_A must have a Jordan block of size 2 associated with 1 due to Theorem 5.2.3. Since A is already a 2×2 matrix, we shall have

$$D_A = J_2(1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

$\boxed{A = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right]}$ It holds

$$\chi_A(\lambda) = \det(\lambda I_2 - A) = \begin{vmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 2 \end{vmatrix} = (\lambda - 1)(\lambda - 2).$$

It is evident the characteristic polynomial of A has two simple roots, $\lambda_1 = 1$ and $\lambda_2 = 2$, therefore $m_A = \chi_a$ and A is diagonalizable, hence

$$D_A = J_1(1) \oplus J_1(2) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

$\boxed{A = \left[\begin{smallmatrix} 2 & -1 \\ 4 & -2 \end{smallmatrix}\right]}$ Here, we have

$$\chi_A(\lambda) = \det(\lambda I_2 - A) = \begin{vmatrix} \lambda - 2 & 1 \\ -4 & \lambda + 2 \end{vmatrix} = \lambda^2.$$

Since $A \neq O$, we must also have $m_A(\lambda) = \lambda^2$, therefore by Theorem 5.2.3, D_A must consist of a single Jordan block of size 2 associated to the unique eigenvalue of A, namely

$$D_A = J_2(0) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$ □

2. Find the JCF of

$$A = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

Proof. The characteristic polynomial is

$$\chi_A(\lambda) = \det(\lambda I_3 - A) = \begin{vmatrix} \lambda - \lambda_1 & -1 & 0 \\ 0 & \lambda - \lambda_2 & -1 \\ 0 & 0 & \lambda - \lambda_3 \end{vmatrix} = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3),$$

hence the eigenvalues are precisely the diagonal elements of A, namely λ_1, λ_2, λ_3. If they are all distinct, then $m_A = \chi_A$, and by Theorem 5.2.3 A is diagonalizable, therefore

$$D_A = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}.$$

Next, suppose that $\lambda_1 = \lambda_2 \neq \lambda_3$, so that $\chi_A(\lambda) = (\lambda - \lambda_1)^2(\lambda - \lambda_3)$. The minimum polynomial is either $(\lambda - \lambda_1)(\lambda - \lambda_3)$ or χ_A itself. We first compute $(A - \lambda_1 I_3)(A - \lambda_3 I_3)$:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_3 - \lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1 - \lambda_3 & 1 & 0 \\ 0 & \lambda_1 - \lambda_3 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \lambda_1 - \lambda_3 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq O,$$

hence $m_A(\lambda)$ does not equal $(\lambda - \lambda_1)(\lambda - \lambda_3)$. Thus, $m_A(\lambda) = \chi_A(\lambda)$, and by Theorem 5.2.3 we obtain

$$D_A = J_2(\lambda_1) \oplus J_1(\lambda_3) = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

The last case is if all eigenvalues are equal, so that $\chi_A(\lambda) = (\lambda - \lambda_1)^3$. The minimum polynomial is $(\lambda - \lambda_1)^m$, for some $1 \leq m \leq 3$. We have

$$A - \lambda_1 I_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = N_3 \neq O,$$

and $(A - \lambda_1 I_3)^2 = N_3^2 \neq O$, therefore we must have $m_a = \chi_A$. By Theorem 5.2.3 once again, we obtain

$$D_A = A. \qquad \square$$

3. If $A \in \mathbb{C}^{n \times n}$ is nilpotent and diagonalizable, then it must be the zero matrix.

Proof. Suppose that $m \in \mathbb{N}$ is the smallest with $A^m = O$, hence $m_A(\lambda) = \lambda^m$. However, since A is diagonalizable, m_A must only have simple roots, which yields $m = 1$, thus $A = O$. $\square$

4. Find the JCF of the matrices A and B, where

$$m_A(\lambda) = (\lambda - 1)^3(\lambda - 2), \chi_A(\lambda) = (\lambda - 1)^3(\lambda - 2)^3,$$

and

$$m_A(\lambda) = (\lambda - 1)^3(\lambda - 2), \chi_A(\lambda) = (\lambda - 1)^4(\lambda - 2)^3.$$

Solution. Since $\deg \chi_A = 6$, A must be a 6×6 matrix. The only eigenvalues of A are 1 and 2, both with multiplicity 3. The size of the largest Jordan block associated with $\lambda = 1$ is 3, while the size of the largest Jordan block associated with $\lambda = 2$ is 1, by Theorem 5.2.3. The sum of the sizes of the blocks in each eigenvalue equals the multiplicity of said eigenvalue; this shows that we must have one Jordan block of size 3 associated with $\lambda = 1$ and three Jordan blocks of size 1 associated with $\lambda = 2$. Hence,

$$D_A = J_3(1) \oplus J_1(2) \oplus J_1(2) \oplus J_1(2) = \begin{bmatrix} 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & 1 & & & \\ & & & 2 & & \\ & & & & 2 & \\ & & & & & 2 \end{bmatrix}.$$

We work in a similar fashion for B. Since $\deg \chi_B = 7$, B must be a 7×7 matrix. Here, 1 is an eigenvalue of multiplicity 4, while 2 has multiplicity 3. As before, the size of the largest Jordan block associated with $\lambda = 1$ is 3, while the size of the largest Jordan block associated with $\lambda = 2$ is 1, again by Theorem 5.2.3. Therefore, we must have a Jordan block of size 3 associated with 1. Since the multiplicity of 1 is 4, we must also have another one with size 1. For the other eigenvalue, we can only have Jordan blocks of minimal size, therefore

$$D_B = J_3(1) \oplus J_1(1) \oplus J_1(2) \oplus J_1(2) \oplus J_1(2) = \begin{bmatrix} 1 & 1 & & & & & \\ & 1 & 1 & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & & & & 2 & & \\ & & & & & 2 & \\ & & & & & & 2 \end{bmatrix}. \qquad \square$$

5. **(IMC Selection AUTh)** Let $A \in \mathbb{R}^{n\times n}$ be a matrix with no real eigenvalues. Show that

   (a) n is even.

   (b) there is $B \in \mathbb{R}^{n\times n}$ such that $AB = BA$ and $B^2 = -I$.

Proof. **(a)** By hypothesis, the roots of the characteristic polynomial χ_A appear in complex conjugate pairs, hence its degree must be even. But $n = \deg \chi_A$, proving the desired fact.

(b) Let R be the real Jordan Canonical Form of A, and suppose $A = VRV^{-1}$ for some invertible $V \in \mathbb{R}^{n\times n}$. The desired B should satisfy

$$(VRV^{-1})B = B(VRV^{-1}),$$

or equivalently,

$$R(V^{-1}BV) = (V^{-1}BV)R,$$

which is obtained from the previous equation by multiplying by V^{-1} on the left and by V on the right. We observe that $B^2 = -I$ if and only if $(v^{-1}BV)^2 = -I$. Replacing $V^{-1}BV$ by B, it suffices to find B such that $BR = RB$ and $B^2 = -I$. Next, suppose that

$$R = \bigoplus_{j=1}^{k} J_{m_j}(\lambda_j),$$

which holds by Theorem 5.2.5; here, $\lambda_1, \bar{\lambda}_1, \ldots, \lambda_k, \bar{\lambda}_k$ are all the eigenvalues of A. Some may appear multiple times, due to the fact that there might be multiple Jordan blocks associated with one eigenvalue. We observe that if B had the same block decomposition, i.e.

$$B = \bigoplus_{j=1}^{k} B_j,$$

where $B_j \in \mathbb{R}^{2m_j\times 2m_j}$, we would have

$$RB = \bigoplus_{j=1}^{k} J_{m_j}(\lambda_j)B_j, \quad BR = \bigoplus_{j=1}^{k} B_j J_{m_j}(\lambda_j)$$

and

$$B^2 = \bigoplus_{j=1}^{k} B_j^2,$$

therefore $B^2 = -I$ if and only if $B_j^2 = -I_{2m_j}$ for all $1 \le j \le k$. Thus, we reduce to the case where R is a *single* Jordan block, say $J_m(\lambda)$, for $\lambda = a + bi$. Now, take

$$B = \bigoplus_{\ell=1}^{m} J = \begin{bmatrix} J & O & \ldots & O \\ O & J & \ldots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \ldots & J \end{bmatrix},$$

where

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

It is easy to check that $CJ = JC$ and $J^2 = -I_2$, where

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

Therefore,

$$B^2 = \bigoplus_{\ell=1}^{m} J^2 = -I,$$

and

$$\begin{aligned} RB &= \begin{bmatrix} C & I_2 & O & \dots & O & O \\ O & C & I_2 & \dots & O & O \\ O & O & C & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & C & I_2 \\ O & O & O & \dots & O & C \end{bmatrix} \begin{bmatrix} J & O & O & \dots & O & O \\ O & J & O & \dots & O & O \\ O & O & J & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & J & O \\ O & O & O & \dots & O & J \end{bmatrix} \\ &= \begin{bmatrix} CJ & J & O & \dots & O & O \\ O & CJ & J & \dots & O & O \\ O & O & CJ & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & CJ & J \\ O & O & O & \dots & O & CJ \end{bmatrix} \\ &= \begin{bmatrix} JC & J & O & \dots & O & O \\ O & JC & J & \dots & O & O \\ O & O & JC & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & JC & J \\ O & O & O & \dots & O & JC \end{bmatrix} \\ &= BR, \end{aligned}$$

as desired. □

6. A matrix $A \in \mathbb{C}^{n\times n}$ is called an *involution* if $A^2 = I$. Find the maximum number of distinct pairwise commuting involutions in $\mathbb{C}^{n\times n}$.

Solution. Let $A_1, \dots, A_m \in \mathbb{C}^{n\times n}$ be pairwise commuting involutions. We observe that the minimum polynomial of each A_j divides $\lambda^2 - 1$, therefore all such minimum polynomials have only simple roots. By Corollary 5.2.4, each A_j is diagonalizable. By Theorem 5.3.2**(2)**, there is an invertible matrix V such that each $D_j = VA_jV^{-1}$ is diagonal; the matrices D_j are also involutions themselves, and they are pairwise commutative. The diagonal entries must be ± 1. The problem then amounts to finding the maximum number of distinct matrices of the form

$$\begin{bmatrix} \pm 1 & 0 & \dots & 0 \\ 0 & \pm 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pm 1 \end{bmatrix}.$$

The proof is combinatorial; we have two choices for each diagonal entry, among n diagonal entries. Therefore, the maximum number of such matrices is 2^n. □

7. Let $A \in \mathbb{R}^{n\times n}$ be such that $A + A^\top = O$. Prove that

$$\det(I_n + \lambda A^2) \geq 0,$$

for every $\lambda \in \mathbb{R}$.

Proof. It holds $A = -A^\top$; such a matrix is called *skew-symmetric*. A is real, so $A^* = A^\top$; hence, A is also normal, as $AA^\top = -A^2 = A^\top A$. By Theorem 5.3.2(**1**), there is a unitary matrix $U \in \mathbb{C}^{n\times n}$ such that $A = UDU^*$, where D is diagonal, and its diagonal elements are precisely the eigenvalues of A, counting multiplicities. It holds $A^\top = A^* = (UDU^*)^* = UD^*U^*$, therefore $D^* = -D$ (such a matrix is called *skew-Hermitian*). Therefore, for every eigenvalue λ of A, it holds $\bar{\lambda} = -\lambda$. This shows that $\operatorname{Re}\lambda = 0$, hence the real eigenvalues are zero, while the rest appear are imaginary and appear in pairs, λ, $-\lambda$. Let $\lambda_1 = i\mu_1, -\lambda_1 = -i\mu_1, \ldots, \lambda_m = i\mu_m, -\lambda_m = -i\mu_m$ be the nonzero eigenvalues of A, where $\mu_1, \ldots, \mu_m \in \mathbb{R}$. Then,

$$\begin{aligned}
\det(I_n + \lambda A^2) &= \det(I_n + \lambda UD^2U^*) = \det[U(I_n + \lambda D^2)U^*] \\
&= \det U \det(I_n + \lambda D^2) \det U^* = \det(I_n + \lambda D^2) \\
&= \begin{vmatrix} 1 &&&&&&&& \\ & 1 &&&&&&& \\ && \ddots &&&&&& \\ &&& 1 &&&&& \\ &&&& 1-\lambda^2\mu_1^2 &&&& \\ &&&&& 1-\lambda^2\mu_1^2 &&& \\ &&&&&& \ddots && \\ &&&&&&& 1-\lambda^2\mu_m^2 & \\ &&&&&&&& 1-\lambda^2\mu_m^2 \end{vmatrix} \\
&= \prod_{j=1}^{m} (1 - \lambda^2\mu_j^2)^2 \geq 0,
\end{aligned}$$

as desired. □

8. Let n be odd and $A \in \mathbb{R}^{n\times n}$, such that either $A^2 = O_n$ or $A^2 = I_n$. Show that

$$\det(A + I_n) \geq \det(A - I_n). \tag{5.5.1}$$

Proof. $\boxed{A^2 = O_n}$ The minimum polynomial of A is either λ or λ^2. Hence, the only eigenvalue of A is zero with multiplicity n, and by Theorem 5.2.3 the Jordan blocks of the JCF of A are either $\left[\begin{smallmatrix} 0 \end{smallmatrix}\right]$ or $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. Let $V \in \mathbb{R}^{n\times n}$ be an invertible matrix, such that $A = VJV^{-1}$, where J is the JCF of A. It holds

$$\det(A + I_n) = \det(VJV^{-1} + I_n) = \det[V(J + I_n)V^{-1}] = \det(J + I_n) = 1,$$

since $J + I_n$ is an upper triangular matrix whose diagonal elements are all 1. Similarly,

$$\det(A - I_n) = \det(J - I_n) = (-1)^n = -1,$$

hence (5.5.1) holds.

$\boxed{A^2 = I_n}$ The minimal polynomial of A is either $\lambda - 1$, $\lambda + 1$, or $\lambda^2 - 1$. In any case, A is diagonalizable by Corollary 5.2.4. If $A = \pm I_n$, the result follows easily due to the fact that n is odd, whence $\det(-I_n) = -1$. If $A \neq \pm I_n$, then the minimal polynomial of A is $\lambda^2 - 1$, hence both ± 1 are eigenvalues of A. But then both sides of (5.5.1) are zero. □

9. Let $A, B \in \mathbb{R}^{n\times n}$ be such that $AB = BA$. Suppose also that $\det(A + B) = 0$. Show that

$$\det(A^k + B^k) \geq 0, \quad \forall k \in \mathbb{N}.$$

Proof. Since A and B commute, if k is odd we have

$$A^k + B^k = (A + B)(A^{k-1} - A^{k-2}B + \cdots - AB^{k-2} + B^{k-1}),$$

therefore,

$$\det(A^k + B^k) = \det(A + B)\det(A^{k-1} - A^{k-2}B + \cdots - AB^{k-2} + B^{k-1}) = 0.$$

If $k = 2m$ is even, we have the factorization

$$A^{2m} + B^{2m} = (A^m + iB^m)(A^m - iB^m).$$

Since A and B are real matrices, we obtain

$$\begin{aligned} \det(A^{2m} + B^{2m}) &= \det(A^m + iB^m)\det(A^m - iB^m) \\ &= \det(A^m + iB^m)\overline{\det(A^m + iB^m)} \\ &= |\det(A^m + iB^m)|^2 \geq 0, \end{aligned}$$

completing the proof. □

10. Let $A \in \mathbb{C}^{n\times n}$. Show that there is $B \in \mathbb{C}^{n\times n}$ such that $ABA = A$.

Proof. Let U be a unitary matrix such that $J = U^*AU$ is the JCF of A. If such a matrix B exists, then U^*BU satisfies the same property with respect to J; indeed,

$$J(U^*BU)J = U^*AU \cdot U^*BU \cdot U^*AU = U^*(ABA)U = U^*AU = J.$$

So, without loss of generality, we may assume that A itself is a direct sum of Jordan blocks. Let

$$A = \bigoplus_{j=1}^{k} J_{m_j}(\lambda_j).$$

If $B_j \in \mathbb{C}^{m_j\times m_j}$ satisfies the same property with respect to $J_{m_j}(\lambda_j)$, i. e.

$$J_{m_j}(\lambda_j)B_j J_{m_j}(\lambda_j) = J_{m_j}(\lambda_j),$$

then the matrix $B = \bigoplus_{j=1}^{k} B_j$ satisfies

$$ABA = \bigoplus_{j=1}^{k} J_{m_j}(\lambda_j)B_j J_{m_j}(\lambda_j) = \bigoplus_{j=1}^{k} J_{m_j}(\lambda_j) = A.$$

Thus, we may assume without loss of generality that A is a *single* Jordan block, i. e. $A = J_n(\lambda) = \lambda I_n + N_n$. If $\lambda \neq 0$, A is invertible so we may take $B = A^{-1}$ and obviously have $ABA = AA^{-1}A = A$. If $\lambda = 0$, so that $A = N_n$, we may take $B = N_n^\top$; it is not hard to show that

$$\begin{aligned} N_n N_n^\top &= \begin{bmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \end{aligned}$$

and then it is easily shown that $N_n N_n^\top N_n = N_n$, as desired. $\square$

11. **(IMC '19)** A 4-digit number YEAR is called very good if the system

$$\begin{aligned} Yx + Ey + Az + Rw &= Y \\ Rx + Yy + Ez + Aw &= E \\ Ax + Ry + Yz + Ew &= A \\ Ex + Ay + Rz + Yw &= R \end{aligned}$$

has at least two solutions in x, y, z, w. Find all good YEARs in the 21st century (2001-2100).

Solution. Let

$$C = \begin{bmatrix} Y & E & A & R \\ R & Y & E & A \\ A & R & Y & E \\ E & A & R & Y \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}, \quad D = \begin{bmatrix} Y \\ E \\ A \\ R \end{bmatrix},$$

so that the given system of equations may be rewritten as $CX = D$. As is already known, such a system has exactly one solution if and only if $\det C \neq 0$. So, we require first that $\det C = 0$. Since C is circulant, it holds by (5.4.1)

$$\det C = (Y + E + A + R)(Y + Ei - A - Ri)(Y - E + A - R)(Y - Ei + A + Ri).$$

Since $Y + E + A + R > 0$, some of the other three eigenvalues must be zero. We distinguish the following cases:

$\boxed{Y - E + A - R = 0}$ This cannot occur if YEAR=2100, therefore $Y = 2$, $E = 0$, hence $R = A + 2$. Let $J = U_4^* C U_4$ be the diagonal form of C, so that

$$\begin{aligned} J &= \begin{bmatrix} Y + E + A + R & & & \\ & Y - Ei - A + Ri & & \\ & & Y - E + A - R & \\ & & & Y + Ei - A - Ri \end{bmatrix} \\ &= \begin{bmatrix} 2A + 4 & & & \\ & 2 - A + i(2 + A) & & \\ & & 0 & \\ & & & 2 - A - i(2 + A) \end{bmatrix} \end{aligned}$$

The system $CX = D$ is equivalent to $U_4 J U_4^* X = D$ or $J U_4^* X = U_4^* D$, which in turn is equivalent to

$$J(F_4^* X) = F_4^* D = \begin{bmatrix} Y+E+A+R \\ Y+Ei-A-Ri \\ Y-E+A-R \\ Y-Ei-A+Ri \end{bmatrix} = \begin{bmatrix} 2A+4 \\ 2-A-i(2+A) \\ 0 \\ 2-A+i(2+A) \end{bmatrix}.$$

We put

$$X' = F_4^* X = \begin{bmatrix} x' \\ y' \\ z' \\ w' \end{bmatrix},$$

so the original system has infinitely many solutions in x, y, z, w if and only if the system $JX' = F_4^* D$ has infinitely many solutions in x', y', z', w'. The new system can be rewritten as

$$\begin{aligned} (2A+4)x' &= 2A+4 \\ [2-A+i(2+A)]y' &= 2-A-i(2+A) \\ 0z' &= 0 \\ [2-A-i(2+A)]w' &= 2-A+i(2+A), \end{aligned}$$

which obviously has unique solutions for x', y', w' but infinitely many for z'.

$\boxed{Y+Ei-A-Ri=0}$ Then, the eigenvalue $Y-Ei-A+Ri$ must also be zero. We must have $Y = A$ and $E = R$. This is only possible for the year 2020. Proceeding as above, the matrices we obtain this time are

$$J = \begin{bmatrix} 4 & & & \\ & 0 & & \\ & & 4 & \\ & & & 0 \end{bmatrix}$$

and

$$F_4^* D = \begin{bmatrix} 4 \\ 0 \\ 4 \\ 0 \end{bmatrix},$$

which obviously has the solution

$$F_4^* X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

hence it has infinitely many solutions, since $\det J = 0$.

We conclude that the good years are 2002, 2013, 2020, 2024, 2035, 2046, 2057, 2068, 2079. $\square$

Bibliography

[1] T. Andreescu, *Essential Linear Algebra with Applications: A Problem Solving Approach*, Springer 2014.

[2] R. Gelca, T. Andreescu, *Putnam and Beyond*. Second edition, Springer 2017.

[3] T.-L. Radulescu, V. Radulescu, T. Andreescu, *Problems in Real Analysis: Advanced Calcuclus on the Real Axis*. Springer-Verlag New York 2009.