Contents

1. Sequences and series

1.1. Sequences

A **sequence** is any function whose domain is $\mathbb{N}$ and whose codomain is $\mathbb{R}$. A sequence is denoted as $\{x_1, x_2, ..., x_n\}$ or $\{x_n\}_{n \in \mathbb{N}}$. In general, the "$n \in \mathbb{N}$" pedix is omitted. Note that, technically, $\{x_1, x_2, ..., x_n\}$ denotes the image of the sequence, and not the function itself, but it is customary to denote sequences as such.

Exercise 1.1.1: Provide some examples of sequences.

Solution:

- The natural numbers, listed from 1 up to a given n, forms a sequence: $\{1, 2, 3, ..., n\} = \{n\}$;
- An infinitely long alternating collection of $+1$ and -1 forms a sequence: $\{-1, 1, -1, ..., (-1)^n\} = \{(-1)^n\}$;
- The integer powers of $1/2$, listed from 1 up to a given $(1/2)^n$, forms a sequence:

$$\left\{1, \frac{1}{2}, \frac{1}{4}, ..., \left(\frac{1}{2}\right)^n\right\} = \left\{\left(\frac{1}{2}\right)^n\right\}$$

□

A sequence $\{x_n\}_{n \in \mathbb{N}}$ is said to be:

- **Increasing** if, for any $n < m$ with $n, m \in \mathbb{N}$, $x_n < x_m$;
- **Decreasing** if, for any $n < m$ with $n, m \in \mathbb{N}$, $x_n > x_m$.
- **Non decreasing** if, for any $n < m$ with $n, m \in \mathbb{N}$, $x_n \leq x_m$;
- **Non increasing** if, for any $n < m$ with $n, m \in \mathbb{N}$, $x_n \geq x_m$;

A sequence that possesses one of the aforementioned properties is said to be **monotone**.

Exercise 1.1.2: Do the sequences in Exercise 1.1.1 possess any of those properties?

Solution:

- By definition, the successor of a natural number is greater than the number itself. Hence, $\{n\}_{n \in \mathbb{N}}$ is an increasing sequence;
- By definition, the successor of the reciprocal a natural number is smaller than the number itself. Hence, $\left\{\left(\frac{1}{2}\right)^n\right\}_{n \in \mathbb{N}}$ is decreasing;
- $\{(-1)^n\}_{n \in \mathbb{N}}$ is not monotone. For example, choosing $n = 0$ and $m = 1$, one has $x_n = (-1)^0 = 1$ and $x_m = (-1)^1 = -1$, hence $x_n < x_m$. However, choosing $n = 1$ and $m = 2$, one has $x_n = (-1)^1 = -1$ and $x_m = (-1)^2 = 1$, hence $x_n > x_m$.

□

A sequence $\{x_n\}_{n \in \mathbb{N}}$ is said to have a **limit** $L \in \mathbb{R}$ if, given any number $\varepsilon > 0$, exists $N \in \mathbb{N}$ dependent on ε, such that for any $n > N$:

$$|x_n - L| < \varepsilon$$

In that case, $\{x_n\}_{n \in \mathbb{N}}$ is said to **converge** to L, or simply to be **convergent**. To denote that $\{x_n\}_{n \in \mathbb{N}}$ has limit L the following notation is used:

$$\lim_{n \to +\infty} x_n = L$$

Proving the convergence of a series is often far from obvious. Lemma 1.1.1 is a useful fact that can aid in such endeavour.

Lemma 1.1.1 (Archimedean principle): For any $a, b \in \mathbb{R}$ with $a > 0$, there exist $n \in \mathbb{N}$ such that $na > b$.

Proof: Stating that $na > b$ is the same as stating that $n > a/b$; since a is assumed to be positive, there are no issues in dividing both sides by a. Since $\mathbb{R}$ is closed under division (the ratio of two real numbers is always a real number), a/b is a real number.

Let $a/b = x$; the statement $na > b$ is then equivalent to $n > x$. Suppose that the statement is false, and therefore that there is no n such that $n > x$. Stated otherwise, this would mean that the real number x is greater than any natural number. If that were to be the case, $\mathbb{N}$ would have a supremum, since $\mathbb{N}$ is a non-empty subset of $\mathbb{R}$ and $\mathbb{R}$ is a complete set.

Let $S = \sup(\mathbb{N})$; notice that S might not be equal to x, but is certainly smaller than x. Consider $S - 1$; since $S = \sup(\mathbb{N})$, $S - 1$ is clearly not an upper bound for $\mathbb{N}$, which means that there has to exist $m \in \mathbb{N}$ such that $m > S - 1$. However, this is the same as $m + 1 > S$, which leads to a contradiction since $S = \sup(\mathbb{N})$ and $m + 1 \in \mathbb{N}$ ($\mathbb{N}$ is closed under addition).

Therefore, it must be the case that for any $x \in \mathbb{R}$, there exist $n \in \mathbb{N}$ such that $n > x$. Having chosen $x = a/b$, it then exists $n \in \mathbb{N}$ such that $n > a/b$, which is the same as stating that there exists $n \in \mathbb{N}$ such that $an > b$. $\square$

Exercise 1.1.3: Prove that $\lim\limits_{x \to +\infty} \dfrac{1}{n} = 0$.

Solution: First, notice that:

$$\left| \frac{1}{n} - 0 \right| = \left| \frac{1}{n} \right| = \frac{1}{n}$$

Since 0 is the neutral element of addition for $\mathbb{N}$ and $\mathbb{Q}^+$ is closed under division. Therefore, to prove that $\lim_{n \to +\infty} 1/n = 0$, it is necessary to prove the existence, for any $\varepsilon > 0$, of a $N \in \mathbb{N}$ such that $1/n < \varepsilon$ for any $n > N$.

$1/n < \varepsilon$ is equivalent to $n > 1/\varepsilon$. This means that, by choosing a $N < n$ such that $N > 1/\varepsilon$, one has $1/n < 1/N$, since dividing a number by a bigger number gives a smaller result. Having chosen N such that $N > 1/\varepsilon$, one has:

$$\frac{1}{n} < \frac{1}{N} < \frac{1}{\frac{1}{\varepsilon}} \quad \text{which is} \quad \frac{1}{n} < \frac{1}{N} < \varepsilon$$

The fact that such N exists is guaranteed by Lemma 1.1.1, hence the result is proven. $\square$

Note that the existence of a $L \in \mathbb{R}$ such that $\lim_{n\to+\infty} x_n = L$ is not guaranteed. That is to say, not all sequences are convergent.

Theorem 1.1.1 (Uniqueness of the limit): Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence such that $\lim_{n\to+\infty} x_n = L$ with $L \in \mathbb{R}$. If such an L exists, it is unique.

Theorem 1.1.2: If a sequence $\{x_n\}$ is convergent then it is limited, that is to say it exists an interval $[a, b]$ such that $\{x_n\} \subset [a, b]$.

Theorem 1.1.3: Let $\{x_n\}$ and $\{y_n\}$ be two convergent sequences, converging to x and to y respectively. The following results hold:

- $\{x_n + y_n\}$ converges to $x + y$;
- Given $\alpha \in \mathbb{R}$, $\{\alpha x_n\}$ converges to αx;
- $\{x_n y_n\}$ converges to xy;
- $\{|x_n|\}$ converges to $|x|$;
- If $y \neq 0$ and, for any $n \in \mathbb{N}$, $y_n \neq 0$, the sequence $\left\{\frac{x_n}{y_n}\right\}$ converges to $\frac{x_n}{y_n}$;
- If $x_n < y_n$ for any $n \in \mathbb{N}$, then $x \leq y$;
- If $x_n = k$ for any $n \in \mathbb{N}$, then $\lim_{n\to+\infty} x_n = k$, $k \in \mathbb{R}$;

A sequence $\{x_n\}_{n\in\mathbb{N}}$ is said to have **limit to** $+\infty$, denoted as:

$$\lim_{n\to+\infty} x_n = +\infty$$

If, for any $M > 0$, exists a $N \in \mathbb{N}$ dependent on M such that, for any $n > N$, it is true that $x_n > M$.

A sequence $\{x_n\}_{n\in\mathbb{N}}$ is said to have **limit to** $-\infty$, denoted as:

$$\lim_{n\to+\infty} x_n = -\infty$$

If, for any $M > 0$, exists a $N \in \mathbb{N}$ dependent on M such that, for any $n > N$, it is true that $x_n < -M$.

In general, a sequence having limit to $\pm\infty$ is just called **divergent** (to $+\infty$ or to $-\infty$).

Exercise 1.1.4: Prove that $\lim\limits_{x\to+\infty} -2^n = -\infty$.

Solution: Let M be any strictly positive number, and let $N = [\log_2(M)] + 1$. For any $n > N$, it must be true that $2^n > M$, which in turn is equivalent to $-2^n < -M$. Therefore, $\lim\limits_{x\to+\infty} -2^n = -\infty$ $\square$

If a sequence $\{x_n\}_{n\in\mathbb{N}}$ is neither convergent or divergent, is said to be **indeterminate**. More formally, a sequence is indeterminate if, for any $M > 0$, exists an $N \in \mathbb{N}$ dependent on M such that, for any $n > N$, it is true that $|x_n| > M$. An indeterminate sequence is denoted as:

$$\lim_{n \to +\infty} x_n = \infty$$

Notice the lack of sign on the symbol ∞, meaning that it's neither positive nor negative infinity

Note that the statements "a sequence is limited" and "a sequence has a limit" are not equivalent. Theorem 1.1.2 states that a sequence that has a limit (that is convergent) is limited, but a sequence that is limited might not have a limit.

Theorem 1.1.4: A sequence that is both limited and monotone has a limit.

Note that the opposite of Theorem 1.1.4 is not true. That is, if a sequence has a limit, it might not be monotone.

Theorem 1.1.5 (Squeeze theorem (for sequences)): Let $\{x_n\}$, $\{y_n\}$ and $\{z_n\}$ be three sequences such that, for any $n \in \mathbb{N}$, $x_n \leq y_n \leq z_n$. Suppose that:

$$\lim_{n \to +\infty} x_n = \lim_{n \to +\infty} z_n = L$$

Then, $\lim_{n \to +\infty} y_n = L$.

Theorem 1.1.6: Let $\{x_n\}$ and $\{y_n\}$ be two sequences such that, for any $n \in \mathbb{N}$, $x_n \leq y_n$. If $\lim_{n \to +\infty} y_n = +\infty$, then $\lim_{n \to +\infty} x_n = +\infty$.

Exercise 1.1.5: Prove that $\lim_{n \to +\infty} \left\{ \left(\frac{1}{2}\right)^n \right\} = 0$.

Solution: Consider the sequences $\{0^n\}$ and $\left\{\frac{1}{n}\right\}$. Both sequences are convergent:

$$\lim_{n \to +\infty} 0^n = 0 \qquad \lim_{n \to +\infty} \frac{1}{n} = 0$$

The result on the left is trivial, the result on the right was already proven in Exercise 1.1.3.

Clearly, $0^n \leq (1/2)^n$ for any $n \in \mathbb{N}$. Moreover, the fact that $(1/2)^n \leq 1/n$ for any $n \in \mathbb{N}$ can be proven applying the Principle of Induction:

- With $n = 1$, $(1/2)^1 = 1/2$ and $1/1 = 1$, and clearly $1/2 \leq 1$;
- Assuming $(1/2)^n \leq 1/n$ to be true:

$$\left(\frac{1}{2}\right)^{n+1} \leq \frac{1}{n+1} \Rightarrow \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^n \leq \frac{1}{n} - \frac{1}{n(n+1)} \Rightarrow \left(\frac{1}{2}\right)^n \leq \frac{2}{n} - \frac{2}{n(n+1)} \Rightarrow$$
$$\frac{1}{n} \leq \frac{2}{n} - \frac{2}{n(n+1)} \Rightarrow -\frac{1}{n} \leq -\frac{2}{n(n+1)} \Rightarrow$$
$$\frac{1}{n} \geq \frac{2}{n(n+1)} \Rightarrow 1 \geq \frac{2}{n+1} \Rightarrow n+1 \geq 2 \Rightarrow n \geq 1$$

It is then possible to apply Theorem 1.1.5 to prove that $\lim_{n \to +\infty} \left\{ \left(\frac{1}{2} \right)^n \right\} = 0$. □

Let $\{x_n\}_{n \in \mathbb{N}}$ be an increasing sequence of natural numbers. $\{k_n\}_{n \in \mathbb{N}}$. The sequence $\left\{x_{k_n}\right\}_{n \in \mathbb{N}}$ is then said to be a **subsequence** of the sequence $\{x_n\}_{n \in \mathbb{N}}$.

Theorem 1.1.7: Let $\{x_n\}_{n \in \mathbb{N}}$ be a sequence. If $\{x_n\}_{n \in \mathbb{N}}$ converges to L, then any subsequence of $\{x_n\}_{n \in \mathbb{N}}$ also converges to L.

Theorem 1.1.8: Every limited sequence has (at least) a converging subsequence.

1.2. Series

Let $\{a_n\}_{n \in \mathbb{N}}$ be a sequence. The following sum having an infinite amount of elements is called a **series**:

$$\sum_{n=i}^{\infty} a_n$$

The a_n elements are called the **terms** of the series. In most cases, i is equal to either 0 or 1.

A summation of this kind, even though having infinite terms, may still give a finite value. In particular, a series is said to be **convergent** if the sequence:

$$\{s_n\} = \left\{ \sum_{k=i}^{n} a_k \right\} = \left\{ \sum_{k=i}^{1} a_k, \sum_{k=i}^{2} a_k, \dots, \sum_{k=i}^{n} a_k \right\} = \{a_i, a_i + a_{i+1}, \dots, a_i + a_{i+1} + \dots + a_{i+n}\}$$

called the **partial sums sequence**, is itself convergent. If L is the limit of $\{s_n\}$, the series $\sum_{n=i}^{\infty} a_n$ is equal to L, and L is called the **sum** of the series:

$$\lim_{n \to +\infty} s_n = L = \sum_{n=i}^{\infty} a_n$$

Similarly, a series is said to be **divergent** (to $+\infty$ or to $-\infty$) if its partial sums sequence is divergent (to $+\infty$ or to $-\infty$, respectively). Finally, a series is said to be **indeterminate** if its partial sums sequence is indeterminate.

Exercise 1.2.1: Let r and a be two real numbers. Study the behaviour of the series

$$\sum_{n=0}^{\infty} ar^n$$

with respect to how r and a vary. Any series defined as such is called a **geometric series of common ratio** r.

Solution: If either a or r are equal to 0, the series is clearly equal to 0. This is why such cases are not taken into account.

The n-th partial sum of the series can be rewritten as:

$$s_n = \sum_{i=0}^{n} ar^n = ar^0 + ar^1 + ar^2 + ar^3 + \dots + ar^n = a\left(\frac{1 - r^{n+1}}{1 - r}\right)$$

As long as $r \neq 1$. Therefore:

$$\sum_{n=0}^{\infty} ar^n = \lim_{n\to+\infty} s_n = \lim_{n\to+\infty} a\frac{1 - r^{n+1}}{1 - r} = \frac{a}{1 - r} - \left(\frac{a}{1 - r}\right)\lim_{n\to+\infty} r^{n+1} =$$

$$= \frac{a}{1 - r}\left(1 - \lim_{n\to+\infty} r^{n+1}\right)$$

There are three possibilities:

- If $r \geq 1$, it is easy to see that $\{r^{n+1}\}$ diverges to $+\infty$, and therefore $\sum_{n=0}^{\infty} ar^n$ diverges to $+\infty$ if a is positive or to $-\infty$ if a is negative. This is particularly evident when $r = 1$:

$$s_n = \sum_{i=0}^{n} a1^n = a1^0 + a1^1 + a1^2 + \dots + a1^n = a + a + a + \dots + a = a(n + 1)$$

- If $-1 < r < 1$, (almost) as shown in Exercise 1.1.3, $\lim_{n\to+\infty} r^{n+1} = 0$. Therefore, $s_n = \dfrac{a}{1 - r}$.
- If $r \leq -1$, it is easy to see that $\{r^{n+1}\}$ is indeterminate, and therefore $\sum_{n=0}^{\infty} ar^n$ is in turn indeterminate. This is particularly evident when $r = -1$:

$$s_n = \sum_{i=0}^{n} a(-1)^n = a(-1)^0 + a(-1)^1 + a(-1)^2 + a(-1)^3 + \dots = \cancel{a} - \cancel{a} + \cancel{a} - \cancel{a} + \dots$$

  This means that s_n is either equal to a if n is even and equal to 0 if n is odd. Therefore, the limit $\lim_{n\to+\infty} s_n$ does not exist.

□

Exercise 1.2.2: Prove that the series

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

converges to 1. This series is called the **Mengoli series**.

Solution: Notice that, for any $k \in \mathbb{N}$ with $k > 0$:

$$\frac{1}{k(k + 1)} = \frac{1}{k} - \frac{1}{k + 1}$$

It is then possible to expand the series as:

$$
\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n} - \frac{1}{n+1} = \\
&= \frac{1}{1} - \frac{1}{1+1} + \frac{1}{2} - \frac{1}{2+1} + \frac{1}{3} - \frac{1}{3+1} + \frac{1}{4} - \frac{1}{4+1} + \dots = \\
&= 1 - \cancel{\frac{1}{2}} + \cancel{\frac{1}{2}} - \cancel{\frac{1}{3}} + \cancel{\frac{1}{3}} - \cancel{\frac{1}{4}} + \cancel{\frac{1}{4}} - \frac{1}{5} + \dots = \\
&= 1 - \lim_{n \to +\infty} \frac{1}{n+1} = 1 - 0 = 1
\end{aligned}
$$

□

Any series that can be rewritten as $\sum_{n=1}^{+\infty} b_n - b_{n+1}$ with $b \in \mathbb{R}$ is called a **telescopic series**. The series in Exercise 1.2.2 is an example of a telescopic series.

Exercise 1.2.3: Prove that the series

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges to $+\infty$. This series is called the **harmonic series**.

Solution: Let $\{s_n\}$ be the sequence of partial sums of the harmonic series. Notice that, for any $m \in \mathbb{N}$:

$$s_{2^m} = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \dots + \left(\frac{1}{2^{m-1}+1} + \dots + \frac{1}{2^m}\right) > 1 + \frac{1}{2} + \frac{1}{2} + \dots + \frac{1}{2} = 1 + \frac{m}{2}$$

Since the first brackets contain 2 terms, the second brackets 4 terms, ecc... until the last brackets, that contain 2^{m-1} terms.

Therefore, $s_{2^m} > 1 + \frac{m}{2}$, from which stems $\lim_{n \to +\infty} s_{2^m} = +\infty$. Also notice that $\{s_{2^m}\}$ is a subsequence of $\{s_n\}$.

The sequence $\{s_n\}$ is (strictly) increasing and monotone, therefore it is either convergent or divergent to $+\infty$. If it were convergent, by Theorem 1.1.7 any of its subsequences would also be convergent, but since $\{s_{2^m}\}$ is not, it cannot be convergent. Then, $\{s_n\}$ ought to diverge to $+\infty$, and therefore $\sum_{n=1}^{\infty} \frac{1}{n}$ also diverges to $+\infty$. □

Theorem 1.2.1 (Cauchy necessary condition for the convergence of series): If the series $\sum_{n=i}^{\infty} a_n$ is convergent, then $\lim_{n \to +\infty} a_n = 0$.

Note that the opposite of Theorem 1.2.1 is not true, since it could be that $\lim_{n \to +\infty} a_n = 0$ and yet the series $\sum_{n=i}^{\infty} a_n$ might not converge. The harmonic series, as shown in Exercise 1.2.3, is one such example.

Given a series, there is no formula that can be applied to determine whether the limit of its partial sums is a finite value, and whether it exists at all. A different approach to determine the convergence or non convergence of a series is to apply what is called a **convergence test**.

Theorem 1.2.2:
- If the series $\sum a_n$ and $\sum b_n$ are convergent, the series $\sum(a_n + b_n)$ is also convergent, and it converges to $\sum a_n + \sum b_n$;
- If the series $\sum a_n$ is convergent then, given $\lambda \in \mathbb{R}$, the series $\sum \lambda a_n$ is also convergent, and it converges to $\lambda \sum a_n$;
- If there exists $N \in \mathbb{N}$ such that, for any $n \geq N$, $a_n = b_n$, then the series $\sum a_n$ and $\sum b_n$ have the same behaviour.

Exercise 1.2.4: Study the behaviour of the series $\displaystyle\sum_{n=0}^{+\infty} \frac{17}{6^n}$.

Solution:

$$\sum_{n=0}^{+\infty} \frac{17}{6^n} = \sum_{n=0}^{+\infty} 17\frac{1}{6^n} = \sum_{n=0}^{+\infty} 17\left(\frac{1}{6}\right)^n = 17\sum_{n=0}^{+\infty}\left(\frac{1}{6}\right)^n = 17\frac{1}{1-\frac{1}{6}} = 17 \cdot \frac{6}{5} = \frac{102}{5}$$

The application of Theorem 1.2.2 is justified because $\sum_{n=0}^{+\infty} (1/6)^n$ is a geometric series of common ratio $1/6$, that was proven to be convergent in Exercise 1.2.1. □

Lemma 1.2.1: If the terms of a series are all strictly positive, such series is either convergent or divergent to $+\infty$.

Theorem 1.2.3 (Comparison test): Let $\sum_{n=i}^{+\infty} a_n$ and $\sum_{n=i}^{+\infty} b_n$ be two series. Suppose that, for any $n \in \mathbb{N}$, $0 \leq a_n \leq b_n$. Then:

1. If $\sum_{n=i}^{+\infty} b_n$ is convergent, $\sum_{n=i}^{+\infty} a_n$ is also convergent;
2. If $\sum_{n=i}^{+\infty} a_n$ is divergent, $\sum_{n=i}^{+\infty} b_n$ is also divergent.

Exercise 1.2.5: Prove that the series $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2}$ and $\displaystyle\sum_{n=1}^{+\infty} \frac{n+3}{n^3+25}$ are convergent.

Solution:
- The ratio $\frac{1}{n^2}$ is always strictly positive for any $n \in \mathbb{N}$ with $n > 0$. Also, for any strictly positive $n \in \mathbb{N}$, $\frac{1}{n^2} \leq \frac{2}{n^2+n}$. But $\frac{2}{n^2+n} = 2\left(\frac{1}{n(n+1)}\right)$, and the series $\sum_{n=1}^{+\infty} \frac{1}{n(n+1)}$ is known to converge, as shown in Exercise 1.2.2. Applying Theorem 1.2.2, the series $\sum_{n=1}^{+\infty} \frac{2}{n^2+n}$ converges and in turn, since $0 \leq \frac{1}{n^2} \leq \frac{2}{n^2+n}$ for any strictly positive integer n, applying Theorem 1.2.3 gives that $\sum_{n=1}^{+\infty} \frac{1}{n^2}$ converges.
- The ratio $\frac{n+3}{n^3+25}$ is always strictly positive for any strictly positive interger n. It also less than $\frac{4}{n^2}$ for any strictly positive integer n. In the previous point it was shown that $\sum_{n=1}^{+\infty} \frac{1}{n^2}$ converges, and therefore applying Theorem 1.2.2 it can be shown that $\sum_{n=1}^{+\infty} \frac{4}{n^2}$ also converges. There-

fore, since $0 \leq \frac{n+3}{n^3+25} \leq \frac{4}{n^2}$ for any $n \in \mathbb{N}$ strictly positive, applying Theorem 1.2.3 gives that $\sum_{n=1}^{+\infty} \frac{n+3}{n^3+25}$ converges.

□

Theorem 1.2.4 (Limit test): Let $\sum_{n=i}^{+\infty} a_n$ and $\sum_{n=i}^{+\infty} b_n$ be two series. Suppose that $a_n \geq 0$ and $b_n > 0$ for any $n \in \mathbb{N}$, and also suppose that:

$$\lim_{n \to +\infty} \frac{a_n}{b_n} = L$$

- If $L \neq 0$ and $L \neq +\infty$, the two series have the same behaviour;
- If $L = 0$ and $\sum_{n=i}^{+\infty} b_n$ is convergent, then $\sum_{n=i}^{+\infty} a_n$ is also convergent;
- If $L = +\infty$ and $\sum_{n=i}^{+\infty} b_n$ is divergent, then $\sum_{n=i}^{+\infty} a_n$ is also divergent.

Exercise 1.2.6: Prove that the series $\displaystyle\sum_{n=1}^{+\infty} \frac{n+7}{n^3-8n}$ is convergent.

Solution: Consider the series $\sum_{n=1}^{+\infty} \frac{1}{n^2}$. Since both this series and $\sum_{n=1}^{+\infty} \frac{n+7}{n^3-8n}$ have only positive terms, it is possible to apply Theorem 1.2.4:

$$\lim_{n \to +\infty} \frac{\frac{n+7}{n^3-8n}}{\frac{1}{n^2}} = \lim_{n \to +\infty} \frac{n^{\cancel{2}}(n+7)}{\cancel{n}(n^2-8)} = \lim_{n \to +\infty} \frac{n^2+7n}{n^2-8} = \lim_{n \to +\infty} \frac{\cancel{n^2}\left(1+\frac{7}{n}\right)}{\cancel{n^2}\left(1-\frac{8}{n^2}\right)} = \frac{1+0}{1-0} = 1$$

Since $1 \neq 0$ and $1 \neq +\infty$ and since $\sum_{n=1}^{+\infty} \frac{1}{n^2}$ is known to be convergent (see Exercise 1.2.5), $\sum_{n=1}^{+\infty} \frac{n+7}{n^3-8n}$ is also convergent. □

Theorem 1.2.5 (Cauchy condensation test): Let $\{a_n\}$ be a decreasing sequence of positive numbers. The series $\sum_{n=1}^{+\infty} a_n$ is convergent if and only if the series $\sum_{n=1}^{+\infty} 2^n a_{2^n}$ is convergent.

A series is said to be **absolutely convergent** if the sum of the absolute values of the summands is finite. More precisely, a series $\sum_{n=0}^{+\infty} a_n$ is said to be absolutely convergent if $\sum_{n=0}^{+\infty} |a_n|$ is convergent.

Theorem 1.2.6 (Absolute convergence test): If a series is absolutely convergent, it is also convergent.

The opposite of Theorem 1.2.6 is not necessarely true, since a series can be convergent but not absolutely convergent.

Theorem 1.2.7 (Ratio test): Let $\sum_{n=1}^{+\infty} a_n$ be a series. Suppose that:

$$\lim_{n\to+\infty} \frac{|a_{n+1}|}{|a_n|} = L,\ a_n \neq 0$$

Then:

- If $L < 1$, the series is (absolutely) convergent;
- If $L > 1$, the series is divergent;
- If $L = 1$, the test is inconclusive.

Exercise 1.2.7: Prove that $\displaystyle\sum_{n=1}^{+\infty} \frac{n}{5^n} = \frac{1}{5}$

Solution: Since $\dfrac{n}{5^n}$ is always strictly positive, Theorem 1.2.7 can be applied:

$$\lim_{n\to+\infty} \frac{\left|\frac{n+1}{5^{n+1}}\right|}{\left|\frac{n}{5^n}\right|} = \lim_{n\to+\infty} \frac{n+1}{5\cdot\cancel{5^n}} \cdot \frac{\cancel{5^n}}{n} = \lim_{n\to+\infty} \frac{1}{5}\left(\frac{n+1}{n}\right) = \frac{1}{5} \lim_{n\to+\infty} 1 + \frac{1}{n} = \frac{1}{5}(1+0) = \frac{1}{5}$$

□

Theorem 1.2.8 (Root test): Let $\sum_{n=1}^{+\infty} a_n$ be a series. Suppose that:

$$\lim_{n\to+\infty} \sqrt[n]{|a_n|} = L$$

Then:

- If $L < 1$, the series is (absolutely) convergent;
- If $L > 1$, the series is divergent;
- If $L = 1$, the test is inconclusive.

Exercise 1.2.8: Prove that $\displaystyle\sum_{n=1}^{+\infty} \left(\frac{1}{n!}\right)^n = 0$

Solution: Applying Theorem 1.2.8:

$$\lim_{n\to+\infty} \sqrt[n]{\left|\left(\frac{1}{n!}\right)^n\right|} = \lim_{n\to+\infty} \left(\left(\frac{1}{n!}\right)^{\cancel{n}}\right)^{\cancel{\frac{1}{n}}} = \lim_{n\to+\infty} \frac{1}{n!} = 0$$

□

Theorem 1.2.9 (Leibniz's test): Consider a series of the form $\sum_{n=0}^{+\infty} (-1)^n a_n$. Suppose that:

1. a_n is always non negative;
2. $a_{n+1} \leq a_n$ for any $n \in \mathbb{N}$;
3. $\lim_{n \to +\infty} a_n = 0$.

Then, the series is convergent.

Exercise 1.2.9: Prove that $\sum_{n=1}^{+\infty} \frac{(-1)^n}{n}$ is convergent.

Solution: Rewriting the series as $\sum_{n=1}^{+\infty} (-1)^n \left(\frac{1}{n}\right)$ and applying Theorem 1.2.9:

1. $\frac{1}{n}$ is always non negative;
2. $\frac{1}{n+1} \leq \frac{1}{n}$ for any $n \in \mathbb{N}$;
3. As shown in Exercise 1.1.3, $\lim_{n \to +\infty} \frac{1}{n} = 0$.

Therefore, the series is convergent. □

Theorem 1.2.10 (Integral test): Let f be a function that is continuous, positive and decreasing in the interval $[N, +\infty)$. Then, the series on the left and the integral on the right have the same behaviour:

$$\sum_{n=N}^{+\infty} f(n) \qquad\qquad \int_N^{+\infty} f(x) dx$$

2. Linear Algebra

2.1. Matrices

2.1.1. Definition

A **matrix** is a bidimensional mathematical object, represented as follows:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} \qquad A = (a_{i,j}) \ i \in \{1, ..., m\}, j \in \{1, \ n\}$$

The real numbers $a_{i,j}$ are called the **entries** or the **elements** of the matrix, while the integers i and j are the **indexes**. The numbers $a_{i,1}, a_{i,2}, ..., a_{i,n}$ with $i \in \{1, ..., m\}$ costitute a **row** of the matrix, while the numbers $a_{1,j}, a_{2,j}, ..., a_{m,j}$ with $j \in \{1, ..., n\}$ constitute a **column**. The number of rows and columns of a matrix is called its **order** or **dimension**, and is denoted as $m \times n$. Two matrices are **equal** if they have the same order and if $(a_{i,j}) = (b_{i,j}) \forall i, j$.

A matrix that has $m = n$ is a **square matrix of order** n (or order m). The elements $a_{1,1}, a_{2,2}, ..., a_{n,n}$ are called the **diagonal elements** and constitute the **diagonal** of the matrix.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix} \qquad A = (a_{i,j}) \ i \in \{1, ..., n\}, j \in \{1, \ n\}$$

A matrix is called a **diagonal matrix** if $(a_{i,j}) = 0$ with $i \neq j$. In other words, a matrix is diagonal if all of its non-diagonal elements are 0.

$$A = \begin{pmatrix} a_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{n,n} \end{pmatrix}$$

A peculiar diagonal matrix is the **identity matrix**, denoted as I_n (where n is its order), whose diagonal elements are all 1. Another peculiar diagonal matrix is the **null matrix**, whose diagonal elements are all 0.

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \qquad O = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

A matrix constituted by a single row and n columns (that is, a $1 \times n$ matrix) is also called a **row matrix**. On the other hand, a matrix constituted by a single column and n rows (that is, a $n \times 1$ matrix) is also called a **column matrix**.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \end{pmatrix}$$

$$A = \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \end{pmatrix}$$

2.1.2. Operations

The **sum** between two matrices A and B having the same dimension $m \times n$ is defined as the sum, entry by entry, of A and B. Two matrices that have the same dimension are said to be *sum-conformant*. The sum between two matrices that are not sum-conformant is undefined.

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} \quad B = \begin{pmatrix} b_{1,1} & \dots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \dots & b_{m,n} \end{pmatrix} \quad A + B = \begin{pmatrix} a_{1,1} + b_{1,1} & \dots & a_{1,n} + b_{1,n} \\ a_{2,1} + b_{2,1} & \dots & a_{2,n} + b_{2,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} + b_{m,1} & \dots & a_{m,n} + b_{m,n} \end{pmatrix}$$

Given a matrix A, the **opposite matrix** of A is the matrix $-A$ such that $A + (-A) = \left(0_{i,j}\right)$.

The sum between matrices possesses the following properties:

- It is commutative;
- Obeys the cancellation law, meaning that $A + B = B + C$ can be simplified as $A = C$;
- It is associative;
- The null matrix is the identity element for matrix sum, since $A + O = O + A = A$ for any matrix A.

Given a $m \times p$ matrix A and a $p \times n$ matrix B, the **product** between two matrices (also called **row-by-column product**) is the $m \times n$ matrix $C = AB$ given by:

$$C = AB = \left(c_{i,j}\right) = \begin{pmatrix} \sum_{i=1}^{p} a_{1,p} \cdot b_{p,1} & \sum_{i=1}^{p} a_{1,p} \cdot b_{p,2} & \dots & \sum_{i=1}^{p} a_{1,p} \cdot b_{p,n} \\ \sum_{i=1}^{p} a_{2,p} \cdot b_{p,1} & \sum_{i=1}^{p} a_{2,p} \cdot b_{p,2} & \dots & \sum_{i=1}^{p} a_{2,p} \cdot b_{p,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{p} a_{m,p} \cdot b_{p,1} & \sum_{i=1}^{p} a_{m,p} \cdot b_{p,2} & \dots & \sum_{i=1}^{p} a_{m,p} \cdot b_{p,n} \end{pmatrix}$$

That is, the i, j-th entry of AB is given by the sum between the products corresponding entries of the i-th row of A and the j-th column of B. Two matrices that possess this property are called *product-conformant*. The product between two matrices that are not product-conformant is undefined.

Exercise 2.1.2.1: Compute the product of the following matrices:

$$A = \begin{pmatrix} -1 & 4 \\ 6 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 3 & 2 & -4 \\ 5 & 0 & 2 \end{pmatrix}$$

Solution:

$$AB = \begin{pmatrix} -1 & 4 \\ 6 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 & -4 \\ 5 & 0 & 2 \end{pmatrix} = \begin{pmatrix} (-1) \cdot 3 + 4 \cdot 5 & (-1) \cdot 2 + 4 \cdot 0 & (-1) \cdot (-4) + 4 \cdot 2 \\ 6 \cdot 3 + 1 \cdot 5 & 6 \cdot 2 + 1 \cdot 0 & 6 \cdot (-4) + 1 \cdot 2 \end{pmatrix} =$$
$$= \begin{pmatrix} -3 + 20 & -2 + 0 & 4 + 8 \\ 18 + 5 & 12 + 0 & -24 + 2 \end{pmatrix} = \begin{pmatrix} 17 & -2 & 12 \\ 23 & 12 & -22 \end{pmatrix}$$

□

The product between matrices possesses the following properties:

- It is not commutative, therefore $AB \neq BA$;
- Cancellation law does not hold. If $AB = BC$ holds, it does not necessarely holds that $A = C$;
- The product of two non null matrices can result in the null matrix. In other words, if $AB = 0$, neither A or B have to be the null matrix;
- The null matrix is the absorbing element for matrix multiplication. Multiplying a non null matrix by the null matrix results in the null matrix;
- It is associative;
- It is distributive with respect to the sum;
- The identity matrix is the identity element for matrix multiplication, since $AI = IA = A$ for any matrix A.

Consider the product AB between the two matrices A and B. If B is a column matrix, it is sometimes useful to write the product as:

$$AB = b_1 \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + b_2 \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{pmatrix} + \ldots + b_n \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{pmatrix}$$

Given a matrix A, the matrix A^T that has the rows of A as its columns and the columns of A as its rows is called the **transposed** of A. If a matrix is equal to its transposed, it is said to be **symmetric**.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \qquad A^T = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{n,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{n,n} \end{pmatrix}$$

Lemma 2.1.2.1: Let A and B be two product-conformant matrices. Then $(AB)^T = B^T A^T$.

The **product between a matrix and a scalar** is an operation that has as input a matrix A and a scalar k and has as output a matrix kA that has each entry multiplied by k.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \qquad kA = \begin{pmatrix} ka_{1,1} & ka_{1,2} & \cdots & ka_{1,n} \\ ka_{2,1} & ka_{2,2} & \cdots & ka_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m,1} & ka_{m,2} & \cdots & ka_{m,n} \end{pmatrix}$$

The product between a matrix and a scalar possesses the following properties:

- It is associative;
- It is commutative with respect to the matrix multiplication;
- It is distributive with respect to the matrix sum;
- It is distributive with respect to the matrix transposition;

2.2. Vector Spaces

2.2.1. Definition of vector space

Let V be a set, whose elements are called **vectors**. A vector $\underline{v}$ is denoted as $\underline{v} = (v_1, v_2, ..., v_n)$, where each v_i with $1 \leq i \leq n$ is called the i-th **component** of $\underline{v}$.

Let $+$ be an operation on such set, a *sum* of vectors, that has two vectors as arguments and returns another vector. That is, foreach $\left(\underline{x}, \underline{y}\right) \in V \times V$ there exists a vector $\underline{v} \in V$ such that $\underline{x} + \underline{y} = \underline{v}$.

Let $\cdot$ be another operation, a *product* between a vector and a real number, that has a real number and a vector as argument and returns another vector. That is, foreach $\lambda \in \mathbb{R}$ and $\underline{v} \in V$ there exists a vector $\underline{w} \in V$ such that $\lambda \cdot \underline{v} = \underline{w}$.

Suppose those operations possess the following properties:

- $(V, +)$ is an Abelian group;
- The product has the distributive property, such that for every $\lambda \in \mathbb{R}$ and for every $\underline{x}, \underline{y} \in V$ it is true that $\lambda \cdot \left(\underline{x} + \underline{y}\right) = \lambda \cdot \underline{x} + \lambda \cdot \underline{y}$;
- The product has the associative property, such that for every $\lambda, \mu \in \mathbb{R}$ and for every $\underline{x} \in V$ it is true that $(\lambda + \mu) \cdot \underline{x} = \lambda \cdot \underline{x} + \mu \cdot \underline{x}$;
- For every vector $\underline{v} \in V$, it is true that $1 \cdot \underline{v} = \underline{v}$.

If that is the case, the set V is called **vector space**. It should be noted that it does not matter what the elements of a vector space actually are (be they numbers, functions, polynomals, etcetera); as long as the aforementioned properties hold for the two operations, such set shares all of the properties that a vector space possesses.

For the sake of readability, the product between a real number and a vector is often represented without the dot. Hence, the expressions $\lambda \cdot \underline{x}$ and $\lambda\underline{x}$ have the same meaning.

Exercise 2.2.1.1: Denote as $\mathbb{R}^n$ the set containing all vectors of real components[1] in the n-dimensional plane. Prove that $\mathbb{R}^n$ constitutes a vector space.

Solution: It is possible to define both a sum between two vectors in the n-dimensional plane and a product between a vector in the n-dimensional space and a real number. To sum two vectors in the n-dimensional space, it suffices to sum each component with each component. To multiply a vector in the n-dimensional space with a real number it suffices to multiply each component by that number:

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} \qquad \lambda \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \lambda v_2 \\ \vdots \\ \lambda v_n \end{pmatrix}$$

Both operations obey the properties stated:

- $(\mathbb{R}^n, +)$ constitutes an Abelian group. Infact:
  - The sum has the associative property:

[1] This is a frustrating misnomer.

$$\begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 + u_1 \\ v_2 + w_2 + u_2 \\ \vdots \\ v_n + w_n + u_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 + u_1 \\ w_2 + u_2 \\ \vdots \\ w_n + u_n \end{pmatrix}$$

- There exists an identity element, in the form of the vector whose components are all zero:

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 + 0 \\ v_2 + 0 \\ \vdots \\ v_n + 0 \end{pmatrix} = \begin{pmatrix} 0 + v_1 \\ 0 + v_2 \\ \vdots \\ 0 + v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

- Each vector in the n-dimensional space has an inverse element with respect to the sum, that is the same vector multiplied by -1:

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + (-1)\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} -v_1 \\ -v_2 \\ \vdots \\ -v_n \end{pmatrix} = \begin{pmatrix} v_1 + (-v_1) \\ v_2 + (-v_2) \\ \vdots \\ v_n + (-v_n) \end{pmatrix} = \begin{pmatrix} -v_1 + v_1 \\ -v_2 + v_2 \\ \vdots \\ -v_n + v_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- The sum has the commutative property:

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} = \begin{pmatrix} w_1 + v_1 \\ w_2 + v_2 \\ \vdots \\ w_n + v_n \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

- The product has the associative property:

$$(\lambda + \mu)\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} (\lambda + \mu)v_1 \\ (\lambda + \mu)v_2 \\ \vdots \\ (\lambda + \mu)v_n \end{pmatrix} = \begin{pmatrix} \lambda v_1 + \mu v_1 \\ \lambda v_2 + \mu v_2 \\ \vdots \\ \lambda v_n + \mu v_n \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \lambda v_2 \\ \vdots \\ \lambda v_n \end{pmatrix} + \begin{pmatrix} \mu v_1 \\ \mu v_2 \\ \vdots \\ \mu v_n \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \mu \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

- The product has the distributive property:

$$\lambda\left(\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}\right) = \begin{pmatrix} \lambda(v_1 + w_1) \\ \lambda(v_2 + w_2) \\ \vdots \\ \lambda(v_n + w_n) \end{pmatrix} = \begin{pmatrix} \lambda v_1 + \lambda w_1 \\ \lambda v_2 + \lambda w_2 \\ \vdots \\ \lambda v_n + \lambda w_n \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \lambda \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

- Multiplying a vector by the number 1 leaves the vector unchanged:

$$1\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 1v_1 \\ 1v_2 \\ \vdots \\ 1v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

□

Exercise 2.2.1.2: Denote as $\mathbb{P}_n$ the set containing all polynomials with real coefficients and degree less than or equal to n. Prove that $\mathbb{P}_n$ constitutes a vector space.

Solution: It is possible to define both a sum between two polynomials with real coefficients and degree $\leq n$ and a product between a polynomial with real coefficients and degree $\leq n$ and a real number. To sum two such polynomials it suffices to sum the coefficients of their monomials having the same degree:

$$(a_n x^n + a_{n-1} x^{n-1} + \dots + a_0) + (b_m x^m + b_{m-1} x^{m-1} + \dots + b_0) =$$
$$a_n x^n + a_{n-1} x^{n-1} + \dots + (a_m + b_m) x^m + (a_{m-1} + b_{m-1}) x^{m-1} + \dots + (a_0 + b_0)$$

To multiply a polynomial with real coefficients and degree $\leq n$ with a real number it suffices to multiply each coefficient of its monomials by such number:

$$\lambda(a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0) = (\lambda a_n) x^n + (\lambda a_{n-1}) x^{n-1} + \dots + (\lambda a_1) x + (\lambda a_0)$$

Both operations satisfy the properties required.

□

Given a vector space V, a set W is said to be a **subspace** of V if it's a subset of V and it's itself a vector space with respect to the same operations defined for V. That is to say, W is a subspace of V if $W \subseteq V$ and W is algebraically closed with respect to the vector sum and the multiplication by a scalar as they are defined for V.

Exercise 2.2.1.3: Consider the vector space $\mathbb{R}^3$. Prove that the set $W_1 \subseteq \mathbb{R}^3$ is a subspace of $\mathbb{R}^3$ while $W_2 \subseteq \mathbb{R}^3$ is not.

$$W_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \middle| \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3, x_1 + x_2 = 0 \right\} \quad W_2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \middle| \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3, x_2 = 2x_3 + 1 \right\}$$

Solution: The sum between two elements in W_1 is closed:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_1 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ -y_1 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ -x_1 - y_1 \\ x_3 + y_3 \end{pmatrix} \Rightarrow$$
$$x_2 + y_2 = -x_1 - y_1 \Rightarrow x_2 + y_2 + (x_1 + y_1) = 0$$

So it is the multiplication by a scalar:

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ -x_1 \\ x_3 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ -\lambda x_1 \\ \lambda x_3 \end{pmatrix} \Rightarrow \lambda x_2 = -\lambda x_1 \Rightarrow \lambda(x_1 + x_2) = 0$$

Therefore, the first set is a subset of $\mathbb{R}^3$. On the other hand, the sum between two elements in W_2 is not closed:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ 2x_3 + 1 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ 2y_3 + 1 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ 2x_3 + 2y_3 + 2 \\ x_3 + y_3 \end{pmatrix} \Rightarrow$$
$$2x_3 + 2y_3 + 2 \neq 2(x_3 + y_3) + 1$$

□

2.3. Linear Dependence and Independence

2.3.1. Linear combinations, linear dependence and independence

Let $\{\underline{v_1}, \underline{v_2}, ..., \underline{v_n}\}$ be a set of n vectors of a vector space V, and let $\lambda_1, \lambda_2, ..., \lambda_n$ be n real numbers (not necessarely distinct). Every summation defined as such:

$$\sum_{i=1}^{n} \lambda_i \underline{v_i} = \lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} + ... + \lambda_i \underline{v_i} + ... + \lambda_n \underline{v_n}$$

Is called a **linear combination** of the vectors $\{\underline{v_1}, \underline{v_2}, ..., \underline{v_n}\}$, with $\lambda_1, \lambda_2, ..., \lambda_n$ as **coefficients**.

A set of vectors $\{\underline{v_1}, \underline{v_2}, ..., \underline{v_n}\}$ is said to be **linearly independent** if the only linear combination of such vectors that equals the null vector is the one that has 0 as every coefficient. If there exists a linear combination of such vectors that is equal to the null vector and that has at least a non-zero coefficient, those vectors are said to be **linearly independent**.

Lemma 2.3.1.1: Let $\{\underline{v_1}, \underline{v_2}, ..., \underline{v_n}\}$ be a set of n vectors of a vector space V. If those vectors are linearly dependent, there exists at least one vector of such set that can be expressed as a linear combination of the remaining vectors, and vice versa.

Proof: If such set of vectors is linearly dependent, there must exist a linear combination of the set that equals the null vector, be it:

$$\sum_{i=1}^{n} \lambda_i \underline{v_i} = \lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} + ... + \lambda_j \underline{v_j} + ... + \lambda_n \underline{v_n} = \underline{0}$$

Where there's at least a non-zero coefficient. Let's assume, without loss of generality, that the j-th coefficient is non-zero (whether such coefficient is the only one to be non-zero is irrelevant). The product of such coefficient with the corresponding vector can be moved to the right side of the equation like so:

$$\lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} + ... + \lambda_n \underline{v_n} = -\lambda_j \underline{v_j}$$

Dividing both sides by $-\lambda_j$ gives:

$$-\frac{\lambda_1}{\lambda_j} \underline{v_1} - \frac{\lambda_2}{\lambda_j} \underline{v_2} - ... - \frac{\lambda_n}{\lambda_j} \underline{v_n} = \underline{v_j}$$

Each $-\frac{\lambda_i}{\lambda_j}$ is itself a real number, and therefore the expression above is a linear combination that is equal to a vector of the set.

On the other hand, assume that the j-th vector of the set is equal to a linear combination of the remaining vectors like so:

$$\lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} + ... + \lambda_n \underline{v_n} = \underline{v_j}$$

Moving $\underline{v_j}$ to the left gives:

$$\lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} + ... + (-1) \underline{v_j} + ... + \lambda_n \underline{v_n} = \underline{0}$$

Since -1 is a real number, the expression on the left side of the equation is indeed a linear combination of the whole set, that is equal to the null vector. $\square$

Exercise 2.3.1.1: Consider the vector space $\mathbb{R}^2$. Check if the vectors $\underline{v_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\underline{v_2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ are linearly independent or linearly dependent.

Solution: Consider such linear combination:

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

To test whether $\underline{v_1}$ and $\underline{v_2}$ are linearly dependent or independent, it suffices to solve this system of equations:

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 2\lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} 2\lambda_1 = 0 \\ \lambda_2 = 0 \end{cases} \Rightarrow \begin{cases} \lambda_1 = 0 \\ \lambda_2 = 0 \end{cases}$$

Since the only solution to such system is $\lambda_1 = 0, \lambda_2 = 0$, $\underline{v_1}$ and $\underline{v_2}$ are linearly independent. □

Exercise 2.3.1.2: Consider the vector space $\mathbb{R}^2$. Check if the vectors $\underline{v_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\underline{v_2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\underline{v_3} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ are linearly independent or linearly dependent.

Solution: Consider such linear combination:

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

To test whether $\underline{v_1}$, $\underline{v_2}$ and $\underline{v_3}$ are linearly dependent or independent, it suffices to solve this system of equations:

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} \lambda_1 + \lambda_2 + 2\lambda_3 = 0 \\ \lambda_2 = 0 \end{cases} \Rightarrow \begin{cases} \lambda_1 = -2\lambda_3 \\ \lambda_2 = 0 \end{cases}$$

Since there are infinite solutions to such system, including non-zero solutions, $\underline{v_1}$, $\underline{v_2}$ and $\underline{v_3}$ are linearly dependent. For example, setting $\lambda_1 = 1$ results in:

$$1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Which is, indeed, a correct identity. □

2.3.2. Span, bases, dimensions

A set of vectors $S = \{\underline{s_1}, ..., \underline{s_n}\}$ of a vector space V is said to **generate** V if every vector of V can be written as a linear combination of the vectors in S. That is to say, S generates V if for every $\underline{v} \in V$ there exist a set of coefficients $\lambda_1, ..., \lambda_n$ such that:

$$\underline{v} = \sum_{i=1}^{n} \lambda_i \underline{s_i} = \lambda_1 \underline{s_1} + ... + \lambda_n \underline{s_n}$$

The set that contains all possible linear combinations from vectors of a set $S = \{\underline{s_1}, ..., \underline{s_n}\}$ is also referred to as the **span** of S, denoted as $\operatorname{span}\{S\}$. Along this line of thought, if a set V is generated by S, it also said to be *spanned* by S.

$$\text{span}\{S\} = \left\{ \underline{v} \,\middle|\, \underline{v} = \sum_{i=1}^{n} \lambda_i \underline{s_i} \right\}$$

A set of vectors that can generate a vector space and is itself linearly independent is called a **basis** for such vector space.

The cardinality of a basis is called the **dimension** of the corresponding vector space. If a vector space contains just the null vector, such vector space is said to have dimension 0. The dimension of a vector space V is denoted as $\dim(V)$.

A vector space whose dimension is a finite number is said to be a **finite-dimensional** vector space. Otherwise, it is said to be an **infinite-dimensional** vector space.

Theorem 2.3.2.1 (Steinitz exchange lemma): Let U and W be finite subsets of a vector space V. If U is a set of linearly independent vectors and V is spanned by W, then:

1. $|U| \leq |W|$;
2. It's always possible to construct a (potentially empty) set $W' \subseteq W$ with $|W'| = |W| - |U|$ such that V is spanned by $U \cup W'$.

Proof: Suppose $U = \{u_1, ..., u_m\}$ and $W = \{w_1, ..., w_n\}$. Since both sets are finite, it's possible to prove the result by induction over $m = |U|$.

Consider the base case $m = 0$. If so, U is just the empty set, and therefore $|U|$ is certainly less than $|W|$. Moreover, a set $W' \subseteq W$ with $|W'| = |W| - |U|$ such that V is spanned by $U \cup W'$ exists and is the set W itself. This is because $W = W'$ is a subset of itself, $|W'| = |W| - 0 = |W|$ and $U \cup W'$ spans V because $W = W'$ spans V by hypothesis.

For the inductive step, suppose that the statement is true up to $k - 1 \in \mathbb{N}$. The set $U \cup W'$ can be written as $\{u_1, ..., u_{k-1}, w_k, ..., w_n\}$, rearranging the w_i vectors if necessary. The inductive hypothesis states that $k \leq n$ and that it's always possible to construct a set $W' \subseteq W$ with $|W'| = |W| - |U|$ such that V is spanned by $U \cup W'$. By hypothesis $U \subseteq V$, which means that the k-th vector of U is in V. It's then possible to write it as a linear combination of vectors in $U \cup W'$:

$$u_k = \sum_{i=1}^{k-1} \lambda_i u_i + \sum_{j=k}^{n} \lambda_j w_j$$

For some set of coefficients $\{\lambda_1, ..., \lambda_n\}$. Note how there's at least one λ_j with $j \geq k$ that is non zero, because otherwise the equality would read $u_k = \sum_{i=1}^{k-1} \lambda_i u_i$ and therefore U would not be a linearly independent set. This also implies that $k \leq n$. Suppose, without loss of generality, that λ_k is non zero (the vectors can be rearranged if necessary). Then:

$$w_k = \frac{1}{\lambda_k} \left(u_k - \sum_{j=1}^{k-1} \lambda_j u_j - \sum_{j=k+1}^{n} \lambda_j w_j \right)$$

Which means that it's possible to write w_k as a linear combination of the vectors $\{u_1, ..., u_{k-1}, u_k, w_{k+1}, ..., w_n\}$. This set contains all the sets that the inductive hypothesis assumes can generate V, hence this set can also generate V. □

Theorem 2.3.2.2 (Dimension theorem for vector spaces): All bases of any vector space have the same cardinality; this cardinality is the smallest number of linearly independent vectors needed to generate it.

Proof: Let V be a vector space and let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two bases for V. If V is finite-dimensional, this is a direct consequence of Theorem 2.3.2.1. This is because a basis is, by definition, a set of vectors that is linearly independent and that can generate its entire vector space. Therefore, $\mathcal{B}_1$ and $\mathcal{B}_2$ can appear interchangeably in the hypothesis of Theorem 2.3.2.1. In one case one has $|\mathcal{B}_1| \leq |\mathcal{B}_2|$, in the other $|\mathcal{B}_2| \leq |\mathcal{B}_1|$, hence $|\mathcal{B}_1| = |\mathcal{B}_2|$.

□

Exercise 2.3.2.1: Consider the vector space $\mathbb{P}_2$. Knowing that the sets $\mathcal{B}_1 = \{1, x, x^2\}$ and $\mathcal{B}_2 = \{5, x+1, 2x^2\}$ are both bases for $\mathbb{P}_2$, write the polynomial $p(x) = 3x^2 + 2x - 5$ as a linear combination of each.

Solution: It's trivial to see that, for the first basis, such linear combination is $p(x)$ itself:

$$\lambda_0 1 + \lambda_1 x + \lambda_2 x^2 = -5 + 2x + 3x^2 \Rightarrow \begin{cases} \lambda_0 = -5 \\ \lambda_1 = 2 \\ \lambda_2 = 3 \end{cases}$$

Regarding the second basis, it can be rewritten as:

$$\lambda_0 5 + \lambda_1 (x+1) + \lambda_2 2x^2 = -5 + 2x + 3x^2 \Rightarrow$$
$$\lambda_0 5 + \lambda_1 x + \lambda_1 + \lambda_2 2x^2 = -5 + 2x + 3x^2 \Rightarrow$$
$$(5\lambda_0 + \lambda_1) + (\lambda_1)x + (2\lambda_2)x^2 = -5 + 2x + 3x^2 \Rightarrow$$
$$\begin{cases} 5\lambda_0 + \lambda_1 = -5 \\ \lambda_1 = 2 \\ 2\lambda_2 = 3 \end{cases} \Rightarrow \begin{cases} \lambda_0 = -\frac{7}{5} \\ \lambda_1 = 2 \\ \lambda_2 = \frac{3}{2} \end{cases}$$

□

The basis of a vector space that renders calculations the most "comfortable" is called the **canonical basis** for such vector space. Such basis is different from vector space to vector space.

Exercise 2.3.2.2: Determine the dimension of $\mathbb{R}^n$

Solution: Consider any n-dimensional vector of coordinates $a_1, a_2, \dots, a_n$. It's easy to see that such vector is equal to the following linear combination:

$$\begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ \dots \\ 0 \end{pmatrix} + \dots + a_n \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}$$

A set containing such vectors in linearly independent. Infact:

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \dots + \lambda_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow \begin{cases} \lambda_1 + 0 + \dots + 0 = 0 \\ 0 + \lambda_2 + \dots + 0 = 0 \\ \vdots \\ 0 + 0 + \dots + \lambda_n = 0 \end{cases} \Rightarrow \begin{cases} \lambda_1 = 0 \\ \lambda_2 = 0 \\ \vdots \\ \lambda_n = 0 \end{cases}$$

This set of vectors is linearly independent and can generate $\mathbb{R}^n$, therefore it's a basis for $\mathbb{R}^n$. The dimension of $\mathbb{R}^n$ is then n, since such set has cardinality n. In particular, this specific basis is the canonical basis for $\mathbb{R}^n$. □

2.3.3. Vector representation

Bases can be employed to represent vectors, no matter their nature (polynomials, tuples, matrices, ecc...), in a unique and standardised form.

Lemma 2.3.3.1: Let V be a vector space of dimension n, and let $B = \{\underline{b_1}, \dots, \underline{b_n}\}$ be a basis. Given a generic vector $\underline{v}$, let $\sum_{i=1}^{n} \lambda_i \underline{b_i} = \underline{v}$ be a linear combination employing B that is equal to $\underline{v}$, with coefficients $\lambda_1, \dots, \lambda_n$. These coefficients are unique.

Proof: Suppose that this is not the case, and that exists instead a set of coefficients $\lambda'_1, \dots, \lambda'_n$ such that $\sum_{i=1}^{n} \lambda'_i \underline{b_i} = \underline{v}$. This means that:

$$\sum_{i=1}^{n} \lambda_{'(i)} \underline{b_i} = \sum_{i=1}^{n} \lambda_i \underline{b_i} \Rightarrow \sum_{i=1}^{n} \lambda_{'(i)} \underline{b_i} - \sum_{i=1}^{n} \lambda_i \underline{b_i} = 0 \Rightarrow \sum_{i=1}^{n} \left(\lambda_{'(i)} - \lambda_i\right) \underline{b_i} = 0$$

Since the vectors $\underline{b_i}$ are linearly independent, the only way for this linear combination to be meaningful is to have $\left(\lambda_{'(i)} - \lambda_i\right) = 0$ for all $i \in \{1, \dots, n\}$. However, this means that $\lambda_{'(i)} = \lambda_i$ for all $i \in \{1, \dots, n\}$, which is a contradiction. □

Lemma 2.3.3.1 implies that a vector in a vector space can be uniquely identified, once a certain basis is fixed, by the coordinates of the linear combination employing said basis used to construct it.

Let V be a vector space and let $\mathcal{B} = \{\mathscr{b}_1, \dots, \mathscr{b}_n\}$ be a basis for V. Any vector $\underline{v}$ has two, equivalent and unique, representations with respect to $\mathcal{B}$:

$$\sum_{i=1}^{n} \lambda_i \underline{\mathscr{b}_i} = \lambda_1 \underline{\mathscr{b}_1} + \lambda_2 \underline{\mathscr{b}_2} + \dots + \lambda_n \underline{\mathscr{b}_n} \qquad \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}_{\mathcal{B}}$$

In particular, the representation on the right is also referred to as its **coordinate vector representation**, or just **coordinate representation**, with respect to $\mathcal{B}$. Technically, the suffix $\mathcal{B}$ is necessary, otherwise it would be impossible to tell apart representations with respect to different bases. However, in general, the basis employed is either known from context or is not relevant, therefore the suffix is often omitted.

Clearly, employing different bases to represent the same vector will give different coordinate representations. Different representations of the same vector can be converted into another through a simple matrix multiplication.

Theorem 2.3.3.1 (Existence of the basis change matrix): Let V be a vector space, and let $\mathcal{B} = \{\underline{b_1}, \underline{b_2}, ..., \underline{b_n}\}$ and $\mathcal{B}' = \{\underline{b'_1}, \underline{b'_2}, ..., \underline{b'_n}\}$ be two bases of V. Consider a generic vector $\underline{x} \in V$, and suppose that its coordinate representations with respect to $\mathcal{B}$ and $\mathcal{B}'$ are, respectively:

$$\underline{x} \Longleftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{\mathcal{B}} \qquad \underline{x} \Longleftrightarrow \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}_{\mathcal{B}'}$$

There exists an invertible matrix P, independent of $\underline{x}$, such that:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = P \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} \qquad \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = P^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Proof: Both $\mathcal{B}$ and $\mathcal{B}'$ are bases of the same vector space. This means that it's possible to write the vectors of one as linear combinations of the vectors of the other:

$$\begin{cases} \underline{b'_1} = p_{1,1}\underline{b_1} + p_{1,2}\underline{b_2} + ... + p_{1,n}\underline{b_n} = \sum_{j=1}^{n} p_{1,j}\underline{b_j} \\ \underline{b'_2} = p_{2,1}\underline{b_1} + p_{2,2}\underline{b_2} + ... + p_{2,n}\underline{b_n} = \sum_{j=1}^{n} p_{2,j}\underline{b_j} \\ \vdots \\ \underline{b'_n} = p_{n,1}\underline{b_1} + p_{n,2}\underline{b_2} + ... + p_{n,n}\underline{b_n} = \sum_{j=1}^{n} p_{n,j}\underline{b_j} \end{cases}$$

Consider a generic vector $\overline{x} \in V$. It can be written as a linear combination of vectors of $\mathcal{B}'$, which gives:

$$\underline{x} = \sum_{i=1}^{n} x'_i \underline{b'_i} = \sum_{i=1}^{n} x'_i \left(\sum_{j=1}^{n} p_{i,j}\underline{b_j} \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_{i,j} x'_i \right) \underline{b_j}$$

This means that $\underline{x}$ can be represented with respect to $\mathcal{B}$ as:

$$\underline{x} \Longleftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{\mathcal{B}} = \begin{pmatrix} \sum_{j=1}^{n} p_{1,j} x'_1 \\ \sum_{j=1}^{n} p_{2,j} x'_2 \\ \vdots \\ \sum_{j=1}^{n} p_{n,j} x'_n \end{pmatrix} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}_{\mathcal{B}'} = P \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}_{\mathcal{B}'}$$

The matrix P, that has the coordinate representation of one basis with respect to the other as its columns, is the desired matrix. Multiplying both sides of the equation with P^{-1} gives the second equality. □

Exercise 2.3.3.1: Consider Exercise 2.3.2.1. What is the matrix that changes bases representations?

Solution: Each member of the first basis, in coordinate representation with respect to the second basis, constitutes one of the columns of the matrix:

$$\begin{cases} 1 = \lambda_{1,1}5 + \lambda_{1,2}(x+1) + \lambda_{1,3}2x^2 \\ x = \lambda_{2,1}5 + \lambda_{2,2}(x+1) + \lambda_{2,3}2x^2 \\ x^2 = \lambda_{3,1}5 + \lambda_{3,2}(x+1) + \lambda_{3,3}2x^2 \end{cases} \Rightarrow \begin{cases} 1 = \lambda_{1,1}5 + \lambda_{1,2}(x+1) + \lambda_{1,3}2x^2 \\ x = \left(\lambda_{2,2} + \lambda_{2,1}5\right) + \lambda_{2,2}x + \lambda_{2,3}2x^2 \Rightarrow \\ x^2 = \lambda_{3,1}5 + \lambda_{3,2}(x+1) + \lambda_{3,3}2x^2 \end{cases}$$

$$\begin{cases} 1 = \lambda_{1,1}5 \\ 0 = \lambda_{1,2} \\ 0 = \lambda_{1,3} \\ 0 = \lambda_{2,2} + \lambda_{2,1}5 \\ 0 = \lambda_{2,2} \\ x = \lambda_{2,2}x \\ 0 = \lambda_{3,1} \\ 0 = \lambda_{3,2} \\ x^2 = \lambda_{3,3}2x^2 \end{cases} \Rightarrow \begin{cases} \frac{1}{5} = \lambda_{1,1} \\ 0 = \lambda_{1,2} \\ 0 = \lambda_{1,3} \\ -\frac{1}{5} = \lambda_{2,1} \\ 1 = \lambda_{2,2} \\ 0 = \lambda_{2,3} \\ 0 = \lambda_{3,1} \\ 0 = \lambda_{3,2} \\ \frac{1}{2} = \lambda_{3,3} \end{cases} \Rightarrow \begin{pmatrix} \frac{1}{5} & -\frac{1}{5} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Considering, as an example, the two representations of $p(x) = 3x^2 + 2x - 5$:

$$\begin{pmatrix} \frac{1}{5} & -\frac{1}{5} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} -5 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} \cdot (-5) - \frac{1}{5} \cdot 2 + 0 \cdot 3 \\ 0 \cdot (-5) + 1 \cdot 2 + 0 \cdot 3 \\ 0 \cdot (-5) + 0 \cdot 2 + \frac{1}{2} \cdot 3 \end{pmatrix} = \begin{pmatrix} -\frac{7}{5} \\ 2 \\ \frac{3}{2} \end{pmatrix}$$

Which is correct. □

2.4. Determinant and Rank

2.4.1. Determinant

The **determinant** is a function that associates a number to a square matrix. Given a $n \times n$ matrix A, its determinant, denoted as $\det(A)$ or $|\ A\ |$, is defined recursively as follows:

$$\det(A) = \begin{cases} \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det\left(M_{i,j}\right) & \text{if } n > 1 \\ a_{11} & \text{otherwise} \end{cases}$$

Where j is any column of the matrix A chosen at random and $M_{i,j}$ is the matrix obtained by removing the i-th row and j-th column from A. The formula can also be applied with respect to rows instead of columns.

When the matrix has dimension $n = 2$, the formula can actually be simplified as follows:

$$\det(A) = \left(a_{1,1} \cdot a_{2,2}\right) - \left(a_{2,1} \cdot a_{1,2}\right)$$

A matrix whose determinant is equal to 0 is called a **singular matrix**.

Exercise 2.4.1.1: Given the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 1 & 1 & 0 \end{pmatrix}$, compute its determinant.

Solution: The fastest way to compute a determinant is to pick the row/column that has the most zeros, because the number of $\det\left(M_{i,j}\right)$ to compute is the smallest. In the case of A, the best choices are: the second row, the first column, the third row and the third column. Suppose the first column is chosen:

$$
\begin{aligned}
\det(A) &= \sum_{i=1}^{3} (-1)^{i+j} a_{i,j} \det\left(M_{i,j}\right) = \\
&= (-1)^{1+1} a_{1,1} \det\left(M_{1,1}\right) + (-1)^{2+1} a_{2,1} \det\left(M_{2,1}\right) + (-1)^{3+1} a_{3,1} \det\left(M_{3,1}\right) = \\
&= 1 \cdot \begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} - \cancel{0 \cdot \begin{vmatrix} 2 & 3 \\ 1 & 0 \end{vmatrix}} + 1 \cdot \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = \\
&= (1 \cdot 0 - 1 \cdot 2) + (2 \cdot 2 - 3 \cdot 1) = 0 - 2 + 4 - 3 = -1
\end{aligned}
$$

□

Peculiar matrices can have their determinant computed faster.

Lemma 2.4.1.1: The determinant of a triangular matrix is equal to the product of the elements on its diagonal.

Proof: Consider an upper triangular matrix A and pick the first column to apply the formula:

$$
\det(A) = \begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ 0 & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{n,n} \end{vmatrix} = a_{1,1} \begin{vmatrix} a_{2,2} & \dots & a_{2,n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_{n,n} \end{vmatrix} = \dots = \prod_{i=1}^{n} a_{i,i}
$$

The same is achieved for a lower triangular matrix by picking the first row. Note that this holds for diagonal matrices as well, since diagonal matrices are a special case of triangular matrices.
□

Lemma 2.4.1.2: The determinant is invariant with respect to transposition.

Lemma 2.4.1.2 implies that the determinant of a matrix can be computed by applying the formula with respect to the columns, not only with respect to rows.

Theorem 2.4.1.1 (Binet's Theorem): Given two matrices A and B, $\det(AB) = \det(A)\det(B)$ (that is, the determinant is a multiplicative function).

Lemma 2.4.1.3: It two rows/columns of a matrix are swapped, its determinant changes sign.

Lemma 2.4.1.4: Given a $n \times n$ matrix A and a scalar k, $\det(kA) = k^n \det(A)$.

2.4.2. Rank

The number of linearly independent columns of a matrix is referred to as its **rank by column**, whereas the number of linearly independent rows of a matrix is referred to as its **rank by row**.

Since the rank by row of a matrix is always equal to its rank by column, it is possible to simply refer to the **rank** of matrix meaning either one or the other. Given a matrix A, its rank is denoted as $\operatorname{rank}(A)$.

Clearly, the rank of a matrix cannot be greater than the minimum between its number of rows and its number of columns. In particular, only a square matrix can have its rank exactly equal to its number of rows/columns. If this is the case, the matrix is said to be **full rank**. Not all square matrices are full rank, but all full rank matrices are square.

Lemma 2.4.2.1: The rank of a matrix is invariant with respect to transposition.

Lemma 2.4.2.2: A matrix has full rank if and only if it's not singular.

A matrix is in **row echelon form** if all rows having only zero entries are at the bottom and the left-most nonzero entry of every nonzero row, called the **pivot**, is on the right of the leading entry of every row above.

Exercise 2.4.2.1: Provide some examples of matrices in row echelon form.

Solution:

$$\begin{pmatrix} 0 & 4 & 1 & 5 & 2 \\ 0 & 0 & 6 & 1 & 9 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 & 3 \\ 0 & 4 & 2 \\ 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 & 5 & 0 \\ 0 & 9 & 0 & 1 \\ 0 & 0 & -4 & 2 \end{pmatrix}$$

□

These matrices are relevant because their rank is particularly easy to compute.

Lemma 2.4.2.3: The rank of a matrix in row echelon form is equal to the number of its pivots.

Any matrix can be converted in row echelon form by employing **Gaussian moves**, which are are special operations that can be performed on matrices. Said operations are as follows:

- Swapping two rows/columns;
- Multiplying a row/column by a scalar;
- Summing a row/column to another row/column multiplied by a scalar.

Lemma 2.4.2.4: The application of Gaussian moves to a matrix does not change its rank.

Lemma 2.4.2.4 allows one to convert a matrix in row echelon form while keeping the rank equal to the original.

Exercise 2.4.2.2: Compute the rank of the matrix $A = \begin{pmatrix} 1 & -1 & 3 & 2 \\ 3 & 2 & 7 & 6 \\ 1 & 4 & 1 & 2 \end{pmatrix}$.

Solution: Lemma 2.4.2.4 guarantees that applying the third Gaussian move to A renders a matrix with the same rank. A can therefore be converted into a matrix in row echelon form as follows:

1. Substituting the third row with itself summed to the first multiplied by -1;
2. Substituting the second row with itself summed to the first multiplied by -3;
3. Substituting the second row with itself summed to the third multiplied by -1.

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

As stated in Lemma 2.4.2.3, the rank of A is 2. □

2.4.3. Inverse matrix

Given a square matrix A, the matrix A^{-1} (if it exists) such that $AA^{-1} = A^{-1}A = I$ is called the **inverse matrix** of A. Such matrix is given by:

$$\left(a_{i,j}^{-1}\right) = \frac{(-1)^{i+j} \det\left(M_{j,i}\right)}{\det(A)}$$

Where $M_{j,i}$ is the matrix A with the j-th row and the i-th column removed. If A^{-1} exists, A is said to be **invertible**.

Lemma 2.4.3.1: A matrix is invertible if and only if it's not singular.

Proof: If A is a singular matrix, its determinant is 0. Therefore, the expression for $\left(a_{i,j}^{-1}\right)$ would involve a division by 0, which is not admissible.

□

Lemma 2.4.3.2: A matrix is invertible if and only if it has full rank.

Lemma 2.4.3.3: A matrix is invertible if and only if the set of its rows/columns forms a linearly independent set.

Lemma 2.4.3.4: Given an invertible matrix A, $\det\left(A^{-1}\right) = (\det(A))^{-1}$.

Proof: Applying Theorem 2.4.1.1 gives:

$$\det(AA^{-1}) = \det(A) \cdot \det(A^{-1}) = 1 \Rightarrow \det(A) = \frac{1}{\det(A^{-1})} \Rightarrow \det(A^{-1}) = \frac{1}{\det(A)}$$

□

Exercise 2.4.3.1: Compute the inverse of the following matrix:

$$A = \begin{pmatrix} 2 & 2 \\ 1 & 0 \end{pmatrix}$$

Solution: The determinant of A is $2 \cdot 0 - 1 \cdot 2 = -2$. Therefore, an inverse exists:

$$a_{1,1} = \frac{(-1)^{1+1} \det(M_{1,1})}{\det(A)} = \frac{0}{-2} = 0 \qquad a_{1,2} = \frac{(-1)^{1+2} \det(M_{2,1})}{\det(A)} = \frac{-2}{-2} = 1$$

$$a_{2,1} = \frac{(-1)^{2+1} \det(M_{1,2})}{\det(A)} = \frac{-1}{-2} = \frac{1}{2} \qquad a_{2,2} = \frac{(-1)^{2+2} \det(M_{2,2})}{\det(A)} = \frac{2}{-2} = -1$$

□

2.4.4. Equivalent and similar matrices

Two matrices A and B are said to be **equivalent** if there exist two invertible matrices S and T such that $B = T^{-1}AS$. If $T = S$, the two matrices are said to be **similar**. Matrix similarity is a special case of matrix equivalence.

Lemma 2.4.4.1: Matrix equivalence (and similarity) is an equivalence relation.

Lemma 2.4.4.2: If two matrices are similar, they have the same determinant and the same rank.

Proof: Suppose A and B are similar. Then:

$$\det(A) = \det(PBP^{-1}) = \det(P)\det(B)\det(P^{-1}) = \cancel{\det(P)}\det(B)\frac{1}{\cancel{\det(P)}} = \det(B)$$

□

2.4.5. Trace

The **trace** of a square matrix is defined as the sum of the elements on its diagonal:

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{i,i}$$

2.5. Systems of Linear Equations

Any equation constituted by real numbers in the form:

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$$

Is called a **linear equation**. Each x_i with $i = 1, 2, \ldots, n$ is called a **variable** or an **unknown**. If the number of unknowns is small, variables are denoted with $x, y, z, \ldots$. The b term and ach a_i with $i = 1, 2, \ldots, n$ is instead called a **coefficient**.

Any ordered n-ple of real numbers $(k_1, k_2, \ldots, k_n)$ is said to be a **solution** of the previous equation if the following holds:

$$a_1k_1 + a_2k_2 + \ldots + a_nk_n = b$$

A linear equation does not necessarely have a solution

Exercise 2.5.1: Consider the following equations:

$$3x = 5 \qquad\qquad 2x - y = 1 \qquad\qquad 0x = 1$$

Do they have any solutions?

Solution:
- The equation $3x = 5$ has a single solution, that can be found by rearranging the terms and obtaining $x = \frac{5}{3}$;
- The equation $2x - y = 1$ has infinite solutions. Indeed, any pair of real numbers $(k, 2k - 1)$ with $k \in \mathbb{R}$ is a solution for said equation;
- The equation $0x = 1$ has no solution, since any number multiplied by 0 is always 0, which is different from 1.

□

Any set of m linear equations of n unknowns $x_1, x_2, \ldots, x_n$ taken into account at the same time is called a **system of linear equations of m equations of n unknowns**, or simply a **linear system**. A linear system has the following form:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

If each b_i with $i = 1, 2, \ldots, m$ is zero, the system is said to be an **homogeneous linear system**.

Exercise 2.5.2: What are some examples of linear equations?

Solution:

$$\begin{cases} 2x + 3y - 5z = 1 \\ 4x + y - 2z = 3 \end{cases} \qquad \begin{cases} 2x + 3y - 2z = 0 \\ 4x + 3y - 2z = 0 \\ x + 2y - z = 0 \end{cases} \qquad \begin{cases} 2x + 2y = 1 \\ 4x + y = -1 \end{cases}$$

□

Any ordered n-ple of real numbers $(k_1, k_2, \ldots, k_n)$ is said to be a **solution** of the linear system if it is a solution of each of the equations that constitute that system at the same time.

The set of all solutions of a linear system is called the **general solution**. *Solving* a linear system means finding, if it exists, its general solution. A linear system having at least one solution (that is, whose general solution is not the empty set), is said to be **solvable**. A linear system that is not solvable (whose general solution is the empty set) is said to be **unsolvable**.

The easiest way to solve a linear system, though not necessarely the most efficient, consists in trying to isolate each unknown and substituting its value in the other. This is allowed, since each equation ought to give the same solution at the same time. This method is informally called the *substitution method*.

Exercise 2.5.3: Consider the following systems of equations:

$$\begin{cases} x + y = 0 \\ 2x + 2y = 1 \end{cases} \qquad \begin{cases} 2x + y = 2 \\ 4x + 2y = 4 \end{cases} \qquad \begin{cases} 3x + 2y = 4 \\ 5x + y = 1 \end{cases}$$

Are they solvable? If they are, what is their general solution?

Solution: The first linear system is unsolvable:

$$\begin{cases} x + y = 0 \\ 2x + 2y = 1 \end{cases} \Rightarrow \begin{cases} x = -y \\ 2(-y) + 2y = 1 \end{cases} \Rightarrow \begin{cases} x = -y \\ -2y + 2y = 1 \end{cases} \Rightarrow \begin{cases} x = -y \\ 0 = 1 \end{cases}$$

The second linear system is solvable and has infinite solutions. More specifically, its general solution is the set of all pair of numbers in the form $(k, 2 - 2k)$ with $k \in \mathbb{R}$:

$$\begin{cases} 2x + y = 2 \\ 4x + 2y = 4 \end{cases} \Rightarrow \begin{cases} y = 2 - 2x \\ 4x + 2(2 - 2x) = 4 \end{cases} \Rightarrow \begin{cases} y = 2 - 2x \\ 4x + 4 - 4x = 4 \end{cases} \Rightarrow \begin{cases} y = 2 - 2x \\ 0 = 0 \end{cases}$$

The third linear system is solvable. Its general solution is the set containing as its single member the pair $\left(-\frac{2}{7}, \frac{17}{7}\right)$.

$$\begin{cases} 3x + 2y = 4 \\ 5x + y = 1 \end{cases} \Rightarrow \begin{cases} y = 1 - 5x \\ 3x + 2(1 - 5x) = 4 \end{cases} \Rightarrow \begin{cases} y = 1 - 5x \\ 3x + 2 - 10x = 4 \end{cases} \Rightarrow \begin{cases} x = -\frac{2}{7} \\ y = \frac{17}{7} \end{cases}$$

□

The term "linear system" suggests that it is possible to encode a linear system of equations into a matrix. Indeed, this is the case; for each system of equations, it is possible write a matrix product that is equivalent to said system:

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,n}x_n = b_2 \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \dots + a_{m,n}x_n = b_m \end{cases} \Longleftrightarrow \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Where the $a_{i,j}$ coefficients with $i \in \{1, \dots, n\}, j \in \{1, \dots, m\}$ are collected into the matrix A, the x_i unknowns into a vector $\underline{x}$ and the b_i known terms into a vector $\underline{b}$. In compact form, an equivalent matrix product for the system is $A\underline{x} = \underline{b}$. Since $\underline{x}$ is a vector, a third way to write the matrix product is:

$$x_1 \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + x_2 \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Exercise 2.5.4: Write the following linear system of equations as a matrix product.

$$\begin{cases} 2x - y + z = 4 \\ x + 3y - 4z = 1 \\ -x + 5y + 6z = 0 \end{cases}$$

Solution:

$$AX = B \Rightarrow \begin{pmatrix} 2 & -1 & 1 \\ 1 & 3 & -4 \\ -1 & 5 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix}$$

□

The A matrix is called the **coefficient matrix**. The $A \mid \underline{b}$ matrix, costructed by justapposing $\underline{b}$ vector on the right side of A, is called the **augmented matrix**.

$$A \mid \underline{b} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} & b_m \end{pmatrix}$$

Theorem 2.5.1 (Rouché-Capelli Theorem): A linear system of equations is solvable if and only if its coefficient matrix and its augmented matrix have the same rank.

Proof: Consider the following linear system of equations $A\underline{x} = \underline{b}$:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

Suppose that the n-tuple $\underline{x}^p = (x_1^p, x_2^p, \dots, x_n^p)$ is a solution to the system. This means that:

$$x_1^p \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + x_2^p \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{pmatrix} + \dots + x_n^p \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

This means that the vector $\underline{b}$ is a linear combination of the column vectors of A, with $x_1^p, \dots, x_n^p$ being the coefficients. This in turn means that $\underline{b}$ belongs to the vector space generated by the column vectors of A. The rank of A is precisely the dimension of the vector space generated by A, hence $\text{rank}(A) = \text{rank}(A \mid \underline{b})$ because $\underline{b}$ is linearly dependent on the columns of A.

Suppose instead that $\text{rank}(A) = \text{rank}(A \mid \underline{b})$. It must then mean that $\underline{b}$ is linearly dependent on the columns of A. Then, there must exist a n-tuple of real numbers $(\lambda_1, \dots, \lambda_n)$ such that:

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \lambda_1 \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + \lambda_2 \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{pmatrix} + \dots + \lambda_n \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$

But this is precisely the definition of the system of equations having the n-tuple $(\lambda_1, \dots, \lambda_n)$ as a solution. Hence, the system is solvable. □

Exercise 2.5.5: Consider the following systems of equations:

$$\begin{cases} x - y + 2z = 1 \\ 3x + y + 3z = 6 \\ x + 3y - z = -1 \end{cases} \qquad \begin{cases} x + 3y - z = -2 \\ 4x + y + z = 1 \\ 2x - 5y + 3z = 5 \end{cases}$$

Are they solvable?

Solution:

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 3 & 1 & 3 \\ 1 & 3 & -1 \end{pmatrix} \qquad A \mid \underline{b} = \begin{pmatrix} 1 & -1 & 2 & 1 \\ 3 & 1 & 3 & 6 \\ 1 & 3 & -1 & -1 \end{pmatrix}$$

Since $\text{rank}(A) = 2$ and $\text{rank}(A \mid \underline{b}) = 3$, Theorem 2.5.1 states that the linear system is unsolvable.

$$A = \begin{pmatrix} 1 & 3 & -1 \\ 4 & 1 & 1 \\ 2 & -5 & 3 \end{pmatrix} \qquad A \mid \underline{b} = \begin{pmatrix} 1 & 3 & -1 & -2 \\ 4 & 1 & 1 & 1 \\ 2 & -5 & 3 & 5 \end{pmatrix}$$

Since $\text{rank}(A) = 2$ and $\text{rank}(A \mid \underline{b}) = 2$, Theorem 2.5.1 states that the linear system is solvable.
□

Lemma 2.5.1: An homogeneous linear system $A\underline{x} = \underline{0}$ is always solvable.

Proof: Let A and $A \mid \underline{0}$ be the coefficient and augmented matrices of an homogeneous linear system. Being homogeneous, the $\underline{0}$ vector is entirely constituted by zeros, and therefore A and $A \mid \underline{0}$ have the same rank. Applying Theorem 2.5.1, it is guaranteed that said linear system is solvable.

It is also easy to see that the *null solution*, constituted by all zeros, is certainly a solution for the system, since $A\underline{0} = \underline{0}$. This may or may not be the only one. □

Notice how Theorem 2.5.1 only deals with the solvability of a linear system of equations, and is not concerned with how many solutions exist. However, it is possible to know the number of possible solutions for any system.

Theorem 2.5.2: Any linear system of equations either has no solution, exactly one solution or infinitely many solutions.

Proof: To convince oneself that there exists at least one linear system of equations that has zero, one or infinitely many solutions, Exercise 2.5.3 shows some examples. What has to be shown is that if a linear system of equations has more than one solution, then it must have infinitely many solutions.

Consider a solvable linear system of equations $A\underline{x} = \underline{b}$ that has more than one solution. Suppose that two of its solutions are $\underline{x}^p$ and $\underline{x}^q$, with $\underline{x}^p \neq \underline{x}^q$. This means that $A\underline{x}^p = \underline{b}$ and $A\underline{x}^q = \underline{b}$, which in turn implies $A\underline{x}^p = A\underline{x}^q$. Grouping the terms, one has $A(\underline{x}^p - \underline{x}^q) = \underline{0}$.

Consider $\underline{x}^* = \underline{x}^p + k(\underline{x}^p - \underline{x}^q)$, with $k \in \mathbb{R}$. Since $\underline{x}^p \neq \underline{x}^q$, $k(\underline{x}^p - \underline{x}^q) \neq \underline{0}$, which means that $\underline{x}^* \neq \underline{x}^p$. $\underline{x}^*$ is another solution to $A\underline{x} = \underline{b}$, since:

$$A\underline{x}^* = A(\underline{x}^p + k(\underline{x}^p - \underline{x}^q)) = A\underline{x}^p + kA(\underline{x}^p - \underline{x}^q) = \underline{b} + kA(\underline{x}^p - \underline{x}^q) = \underline{b} + k\underline{0} = \underline{b}$$

However, since there is no restriction on which k should be chosen, $\underline{x}^*$ represents a family of solutions of infinite size. That is to say, if a linear system of equations has at least two solutions, it's always possible to construct infinitely many solutions from such solutions. □

A linear system of equations is said to be **determined** if it has as many equations as unknowns. It is instead said to be **overdetermined** if it has more equations than unknowns or **underdetermined** if it has more unknowns than equations. Determined, overdetermined or underdetermined are either solvable or unsolvable.

Lemma 2.5.2: Consider a solvable linear system of equations that is determined. The system has one and only solution if and only if the rank of the coefficient matrix equals the number of equations in the system. On the other hand, the system has infinitely many solutions if and only if the rank of the coefficient matrix is smaller than the number of equations in the system.

Lemma 2.5.3: Consider a solvable linear system of equations that is overdetermined. The system has one and only solution if and only if the rank of the coefficient matrix equals the number of equations in the system. On the other hand, the system has infinitely many solutions if and only if the rank of the coefficient matrix is smaller than the number of equations in the system.

Lemma 2.5.4: An underdetermined linear system of equations that is solvable always has infinitely many solutions. In particular, the number of free variables in the system is given by the difference between the number of equations and the rank of the coefficient matrix.

The previous lemmas characterize the number of solutions that different system of equations can have, but do not specify how to find them. Theorem 2.5.3 instructs how to find solutions for a determined system.

Theorem 2.5.3 (Cramer Theorem): Consider a determined system of linear equations $A\underline{x} = \underline{b}$. If A is not singular, then it is solvable. Moreover, its one and only solution is $\underline{x}^p = A^{-1}\underline{b}$.

Proof: To prove that $A\underline{x} = \underline{b}$ can have at most one solution, suppose by contradiction that this is not the case, and therefore that the system has two or more solutions. Suppose that two of its solutions are $\underline{x}^p$ and $\underline{x}^q$, with $\underline{x}^p \neq \underline{x}^q$. As it was the case in Theorem 2.5.2, it's possible to write $A(\underline{x}^p - \underline{x}^q) = \underline{0}$. Since A is assumed to be invertible:

$$A(\underline{x}^p - \underline{x}^q) = \underline{0} \Rightarrow \cancel{A^{-1}}\cancel{A}(\underline{x}^p - \underline{x}^q) = A^{-1}\underline{0} \Rightarrow \underline{x}^p - \underline{x}^q = \underline{0} \Rightarrow \underline{x}^p = \underline{x}^q$$

But it was assumed that $\underline{x}^p \neq \underline{x}^q$, which leads to a contradiction. It has to be true then that the system has one and only solution.

The fact that $\underline{x}^p = A^{-1}\underline{b}$ is a possible solution to the system stems from the obvious fact that substituting $\underline{x}$ with $\underline{x}^p$ gives $AA^{-1}\underline{b} = \underline{b}$. However, since $x^p = A^{-1}\underline{b}$, the existence of x^p necessitates A to be invertible, which is to say that A has to be not singular. Also, since it has just been shown that such system has a single solution, this solution also happens to be the one and only. □

Theorem 2.5.3 is powerful, but holds only for determined system. This is because the coefficient matrix, to be inverted, has to be a square matrix, which happens only in determined systems. Moreover, for large matrices, computing the inverse matrix can be very expensive. A more generic approach, which also happens to be faster, is given in Theorem 2.5.4.

Theorem 2.5.4 (Principle of Gaussian elimination): Let A be the coefficient matrix of a linear system of equations S. Applying any Gauss move one or more times to A gives a new matrix A' associated to a new linear system of equations S' whose general solution is the same of S.

2.6. Linear Transformations

2.6.1. Definition

A transformation $\phi : V \mapsto W$, with both V and W being vector spaces, is called a **linear transformation** or **vector space homomorphism** if and only if:

$$\phi\left(\underline{v_1} + \underline{v_2}\right) = \phi\left(\underline{v_1}\right) + \phi\left(\underline{v_2}\right) \ \forall \underline{v_1}, \underline{v_2} \in V \qquad \phi(\lambda\underline{v}) = \lambda\phi(\underline{v}) \ \forall \underline{v} \in V, \lambda \in \mathbb{R}$$

Exercise 2.6.1.1: Consider the vector space $\mathbb{R}$ (that is, the set of real numbers). Check whether the transformations $\phi_1(x) = 2x$ and $\phi_2(x) = x + 1$ are linear or not.

Solution:

- The transformation $\phi_1(x) = 2x$ is linear. Infact, given two real numbers a and b, is indeed true that $2(a + b) = 2a + 2b$, since the product between real numbers has the distributive property. Similarly, given a real number a and a real number λ, it is true that $2(\lambda a) = 2\lambda a$, since the product between real numbers has the associative property;

- The transformation $\phi_2(x) = x + 1$ is not linear. Given two real numbers a and b, it results in $\phi_2(a + b) = (a + b) + 1 = a + b + 1$, while $\phi_2(a) + \phi_2(b) = a + 1 + b + 1 = a + b + 2$.

$\square$

A linear transformation $\phi : V \mapsto W$ is said to be:

- **Injective** if, for any two distinct $\underline{x}, \underline{y} \in V$, $\phi(x) \neq \phi(y)$;
- **Surjective** if $\phi(V) = W$;
- **Bijective** if it's both injective and surjective.

With respect to these definitions, it is possible to classify a linear transformation $\phi : V \mapsto W$ as follows:

- If it's bijective, it's called an **isomorphism**;
- If $V = W$, it's called an **endomorphism**;
- If it's both an isomorphism and an endomorphism, it's called an **automorphism**.

A remarkable property of linear transformations is that they are equivalent to matrices. That is, for any linear transformation there exist matrices that, when multiplied, have the same effect as applying T.

Theorem 2.6.1.1 (Equivalence between linear transformations and matrices): Let $\phi : V \mapsto W$ be a linear transformation between two vector spaces V and W. Let $\mathcal{B} = \left\{\underline{b_1}, ..., \underline{b_n}\right\}$ be a basis for V and $\mathcal{B}' = \left\{\underline{c_1}, ..., \underline{c_m}\right\}$ a basis for W. Let C and C' be the coordinate representation of, respectively, a vector of V and the result of applying ϕ to said vector. There exist a matrix M_ϕ, dependent both on $\mathcal{B}$ and on $\mathcal{B}'$, such that $M_\phi C = C'$.

Proof: Consider a generic vector $\underline{x} \in V$. This vector can be written as a linear combination of the vectors of $\mathcal{B}$:

$$\underline{x} = \sum_{i=1}^{n} \lambda_i \underline{b_i}$$

Applying ϕ to $\underline{x}$ gives:

$$\phi(\underline{x}) = \phi\left(\sum_{i=1}^{n} \lambda_i \underline{b_i}\right) = \sum_{i=1}^{n} \phi\left(\lambda_i \underline{b_i}\right) = \sum_{i=1}^{n} \lambda_i \phi\left(\underline{b_i}\right)$$

The two rightmost equalities stem from the fact that ϕ is linear.

Each i-th vector $\phi\left(\underline{b_i}\right)$ is a member of W, since it's the result of applying ϕ to a vector of V. Therefore, each $\phi\left(\underline{b_i}\right)$ can itself be written as a linear combination of the vectors of $\mathcal{B}'$:

$$\phi\left(\underline{b_i}\right) = \sum_{j=1}^{m} \gamma_{j,i} \underline{c_j}$$

Substituting back in the previous expression gives:

$$\phi(\underline{x}) = \sum_{i=1}^{n} \lambda_i \left(\sum_{j=1}^{m} \gamma_{j,i} \underline{c_j}\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_i \gamma_{j,i} \underline{c_j}$$

Recall that it's always possible to encode the information regarding a vector into a column vector containing the coefficients of the linear combination used to generate it. In particular, consider the matrix for $\phi(\underline{x})$:

$$\phi(\underline{x}) \Longleftrightarrow \begin{pmatrix} \sum_{j=1}^{m} \lambda_1 \gamma_{j,1} \\ \vdots \\ \sum_{j=1}^{m} \lambda_n \gamma_{j,n} \end{pmatrix}_{\mathcal{B}'} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,n} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,1} & \gamma_{m,2} & \cdots & \gamma_{m,n} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

However, the matrix on the right is just the coordinate representation for $\underline{x}$. Renaming the entries of the matrix on the left with $\mu_1, ..., \mu_n$, one gets:

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}_{\mathcal{B}'} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,n} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,1} & \gamma_{m,2} & \cdots & \gamma_{m,n} \end{pmatrix}_{\mathcal{B}'}^{\mathcal{B}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}_{\mathcal{B}}$$

The $\gamma_{i,j}$ matrix is the desired matrix. The i-th column is composed by the coordinate representation of the result of applying ϕ to the i-th vector of the basis $\mathcal{B}$. □

Exercise 2.6.1.2: Consider the linear transformation $T : \mathbb{R}^3 \mapsto \mathbb{R}^2$ defined below. Write it as a matrix multiplication with respect to the two bases $\mathcal{B}_1$ and $\mathcal{B}_2$.

$$T(x, y, z) = (x - z, y + z) \quad \mathcal{B}_1 = \{(1, 0, 1), (2, 1, -1), (-2, 1, 4)\} \quad \mathcal{B}_2 = \{(1, 2), (2, 1)\}$$

Solution: Applying the transformation to the vectors in $\mathcal{B}_1$ gives:

$$\begin{cases} T(1, 0, 1) &= (1 - 1, 0 + 1) = (0, 1) = \gamma_{1,1}(1, 2) + \gamma_{2,1}(2, 1) \\ T(2, 1, -1) &= (2 - (-1), 1 + (-1)) = (3, 0) = \gamma_{1,2}(1, 2) + \gamma_{2,2}(2, 1) \\ T(-2, 1, 4) &= (-2 - 4, 1 + 4) = (-6, 5) = \gamma_{1,3}(1, 2) + \gamma_{2,3}(2, 1) \end{cases}$$

The coefficients are:

$$\begin{cases} 0 = \gamma_{1,1} + 2\gamma_{2,1} \\ 1 = 2\gamma_{1,1} + \gamma_{2,1} \\ 3 = \gamma_{1,2} + 2\gamma_{2,2} \\ 0 = 2\gamma_{1,2} + \gamma_{2,2} \\ -6 = \gamma_{1,3} + 2\gamma_{2,3} \\ 5 = 2\gamma_{1,3} + \gamma_{2,3} \end{cases} = \begin{cases} \gamma_{1,1} = -2\gamma_{2,1} \\ 1 = 2(-2\gamma_{2,1}) + \gamma_{2,1} \\ 3 = \gamma_{1,2} + 2(-2\gamma_{1,2}) \\ \gamma_{2,2} = -2\gamma_{1,2} \\ \gamma_{1,3} = -2\gamma_{2,3} - 6 \\ 5 = 2(-2\gamma_{2,3} - 6) + \gamma_{2,3} \end{cases} = \begin{cases} \gamma_{1,1} = -2\gamma_{2,1} \\ 1 = -3\gamma_{2,1} \\ 3 = -3\gamma_{1,2} \\ \gamma_{2,2} = -2\gamma_{1,2} \\ \gamma_{1,3} = -2\gamma_{2,3} - 6 \\ 17 = -3\gamma_{2,3} \end{cases} = \begin{cases} \gamma_{1,1} = \frac{2}{3} \\ \gamma_{2,1} = -\frac{1}{3} \\ \gamma_{1,2} = -1 \\ \gamma_{2,2} = 2 \\ \gamma_{1,3} = \frac{16}{3} \\ \gamma_{2,3} = -\frac{17}{3} \end{cases}$$

Which gives the matrix

$$M_T = \begin{pmatrix} \frac{2}{3} & -1 & \frac{16}{3} \\ -\frac{1}{3} & 2 & -\frac{17}{3} \end{pmatrix}_{\mathcal{B}_2}^{\mathcal{B}_1}$$

Consider, as an example, the triple $(3, 1, 0) \in V$. Its coordinate representation with respect to $\mathcal{B}_1$ is:

$$(3,1,0) = 1(1,0,1) + 1(2,1,-1) + 0(-2,1,4) \Longleftrightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}_{\mathcal{B}_1}$$

Applying ϕ to $(3,1,0)$ gives $\phi(3,1,0) = (3-0, 1+0) = (3,1)$. Its coordinate representation with respect to $\mathcal{B}_2$ is:

$$(3,1) = -\frac{1}{3}(1,2) + \frac{5}{3}(2,1) \Longleftrightarrow \begin{pmatrix} -\frac{1}{3} \\ \frac{5}{3} \end{pmatrix}_{\mathcal{B}_2}$$

Indeed:

$$\begin{pmatrix} \frac{2}{3} & -1 & \frac{16}{3} \\ -\frac{1}{3} & 2 & -\frac{17}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \cdot 1 + (-1) \cdot 1 + \frac{16}{3} \cdot 0 \\ -\frac{1}{3} \cdot 1 + 2 \cdot 1 + \left(-\frac{17}{3}\right) \cdot 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} - 1 + 0 \\ -\frac{1}{3} + 2 + 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} \\ \frac{5}{3} \end{pmatrix}$$

□

Note how Theorem 2.3.3.1 is just a special case of Theorem 2.6.1.1 when both vector spaces are the same and the linear transformation is the one that maps every vector to itself, the **identity transformation**. Also, the same result allows one to construct the matrix representation of the same linear transformation employing different bases.

Lemma 2.6.1.1: Let $\phi : V \mapsto W$ be a linear transformation between two vector spaces V and W. Let $\mathcal{B}$ and $\mathcal{B}'$ be two distinct bases for V and let $\mathcal{C}$ and $\mathcal{C}'$ be two distinct bases for W. Given a generic vector $\underline{x} \in V$, let B, B' be the coordinate representation of $\underline{x}$ with respect to $\mathcal{B}$ and $\mathcal{B}'$, respectively. Given $\phi(\underline{x}) \in W$, the mapping of $\underline{x}$ through ϕ, let C, C' be the coordinate representation of $\phi(\underline{x})$ with respect to $\mathcal{C}$ and $\mathcal{C}'$, respectively. Let M_ϕ be the matrix representation of ϕ with respect to $\mathcal{B}$ and $\mathcal{C}$, and let M'_ϕ be the matrix representation of ϕ with respect to $\mathcal{B}'$ and $\mathcal{C}'$. Let P be the matrix of basis change from $\mathcal{B}$ to $\mathcal{B}'$, and let Q be the matrix of basis change from $\mathcal{C}$ to $\mathcal{C}'$. Then $M_\phi = Q^{-1} M'_\phi P$.

Proof: By definition, $M_\phi B = C$ and $M'_\phi B' = C'$. But $PB = B'$ and $QC = C'$, which gives:

$$M'_\phi B' = C' \Rightarrow M'_\phi P B = QC \Rightarrow Q^{-1} M'_\phi P B = C \Rightarrow Q^{-1} M'_\phi P B = M_\phi B \Rightarrow Q^{-1} M'_\phi P = M_\phi$$

□

2.7. Kernel and Image

2.7.1. Image

Let $T : V \mapsto W$ a linear transformation between vector spaces V and W. The set of all vectors of W that have a correspondant in V through T is called the **image** of the transformation T, and is denoted as $\mathfrak{I}(T)$. It may or may not coincide with W.

$$\mathfrak{I}(T) = \{\underline{w} \in W : \exists \underline{v} \in V \quad \text{s.t.} \quad T(\underline{v}) = \underline{w}\}$$

Exercise 2.7.1.1: Given the linear transformation $T : \mathbb{R}_2[x] \mapsto \mathbb{R}_2[x]$ defined as $T(p(x)) = xp(x) - \frac{1}{2}x^2 \frac{d}{dx}p(x)$, compute its image.

Solution: The image of T is the set of all polynomials $q(x) = a_1 + a_2x + a_3x^2$ such that $q(x) = T(p(x))$. That is:

$$q(x) = xp(x) - \frac{1}{2}x^2 p(x) \Rightarrow$$

$$a_1 + a_2x + a_3x^2 = x(b_1 + b_2x + b_3x^2) - \frac{1}{2}x^2 \frac{d}{dx}(b_1 + b_2x + b_3x^2) \Rightarrow$$

$$a_1 + a_2x + a_3x^2 = xb_1 + b_2x^2 + \cancel{b_3x^3} - \frac{1}{2}x^2 b_2 - \cancel{b_3x^3} \Rightarrow$$

$$a_1 + (a_2 - b_1)x + \left(a_3 - \frac{1}{2}b_2\right)x^2 = 0 \Rightarrow \begin{cases} a_1 = 0 \\ a_2 - b_1 = 0 \\ a_3 - \frac{1}{2}b_2 = 0 \end{cases} \Rightarrow \begin{cases} a_1 = 0 \\ a_2 = b_1 \\ a_3 = \frac{1}{2}b_2 \end{cases}$$

This means that $\Im(T)$ is the set:

$$\Im(T) = \left\{q(x) \in \mathbb{R}_2[x] \,\middle|\, q(x) = hx + \frac{1}{2}kx^2, h \in \mathbb{R}, k \in \mathbb{R}\right\}$$

$\square$

Lemma 2.7.1.1: Let $T : V \mapsto W$ be a linear transformation between vector spaces V and W. $\Im(W)$ is a subspace of W.

Proof: To show that $\Im(W)$ is a subspace of W, it has to be proven that $\underline{w_1} + \underline{w_2} \in \Im(W)$ holds for all $\underline{w_1}, \underline{w_2} \in \Im(W)$ and that $\lambda\underline{w} \in \Im(W)$ holds for all $\underline{w} \in \Im(W)$ and $\lambda \in \mathbb{R}$.

By definition, if $\underline{w} \in \Im(W)$ then there exists $\underline{v} \in V$ such that $T(\underline{v}) = \underline{w}$. Therefore:

$$\underline{w_1} + \underline{w_2} = T\left(\underline{v_1}\right) + T\left(\underline{v_2}\right) \qquad \lambda\underline{w} = \lambda T(\underline{v})$$

By virtue of T being linear:

$$\underline{w_1} + \underline{w_2} = T\left(\underline{v_1}\right) + T\left(\underline{v_2}\right) = T\left(\underline{v_1} + \underline{v_2}\right) \qquad \lambda\underline{w} = \lambda T(\underline{v}) = T(\lambda\underline{v})$$

In both cases, there exists a vector in V such that the application of T gives such vector, therefore $\Im(W)$ is algebraically closed with respect to the operations defined for W. $\square$

Since it's possible to represent any linear transformation employing a matrix, it's also possible to extend the notion of image to matrices as well. Consider the matrix M_ϕ associated to the linear transformation $\phi : V \mapsto W$ with respect to two bases $\mathcal{B}$ and $\mathcal{C}$. The image of M_ϕ is the set of all coordinate vector representations Y of vectors of W such that it exists a coordinate representation vector X of a vector in V such that $MX = Y$.

Exercise 2.7.1.2: What is the image of $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 5 & 12 \end{pmatrix}$?

Solution:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 5 & 12 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \cdot x + 2 \cdot y + 3 \cdot z \\ 1 \cdot x + 4 \cdot y + 9 \cdot z \\ 1 \cdot x + 5 \cdot y + 12 \cdot z \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} x + 2y + 3z \\ x + 4y + 9z \\ x + 5y + 12z \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \Rightarrow \begin{cases} x + 2y + 3z = l_1 \\ x + 4y + 9z = l_2 \\ x + 5y + 12z = l_3 \end{cases} \Rightarrow \begin{cases} l_3 = 12z + 5y + x \\ l_2 = x + 4y + 9z \\ l_1 = 3l_2 - 2l_3 \end{cases}$$

Which gives:

$$\mathfrak{I}(A) = \left\{ X \,\middle|\, X = \begin{pmatrix} 3h - 2k \\ h \\ k \end{pmatrix}, h \in \mathbb{R}, k \in \mathbb{R} \right\}$$

□

2.7.2. Kernel

Let $T : V \mapsto W$ a linear transformation between vector spaces V and W. The set of all vectors of V such that the application of T to those vectors gives the null vector (of W) is called the **kernel** of T, and is denoted as $\ker(T)$.

$$\ker(T) = \{\underline{v} \in V : T(\underline{v}) = \underline{0}\}$$

Exercise 2.7.2.1: Given the linear transformation $T : \mathbb{R}_2[x] \mapsto \mathbb{R}_2[x]$ defined as $T(p(x)) = xp(x) - \frac{1}{2}x^2 \frac{d}{dx}p(x)$, compute its rank.

Solution: The kernel of T is the set of all polynomials $p(x) = a_1 + a_2x + a_3x^2$ such that $T(p(x)) = 0$. That is:

$$T(p(x)) = 0 \Rightarrow xp(x) - \frac{1}{2}x^2 p(x) = 0 \Rightarrow$$

$$x(a_1 + a_2x + a_3x^2) - \frac{1}{2}x^2 \frac{d}{dx}(a_1 + a_2x + a_3x^2) = 0 \Rightarrow$$

$$a_1x + a_2x^2 + a_3x^3 - \frac{1}{2}x^2(a_2 + 2a_3x) = 0 \Rightarrow$$

$$a_1x + a_2x^2 + \cancel{a_3x^3} - \frac{1}{2}a_2x^2 - \cancel{a_3x^3} = 0 \Rightarrow$$

$$a_1x + \frac{1}{2}a_2x^2 = 0 \Rightarrow a_1 + \frac{1}{2}a_2x = 0 \Rightarrow \begin{cases} a_1 = 0 \\ \frac{1}{2}a_2 = 0 \\ a_3 = a_3 \end{cases} \Rightarrow \begin{cases} a_1 = 0 \\ a_2 = 0 \\ a_3 = a_3 \end{cases}$$

This means that $\ker(T)$ is the set:

$$\ker(T) = \{p(x) \in \mathbb{R}_2[x] \mid p(x) = hx^2, h \in \mathbb{R}\}$$

Indeed:

$$T(hx^2) = x(hx^2) - \frac{1}{2}x^2 \frac{d}{dx}(hx^2) = hx^3 - \frac{1}{2}x^2(2hx) = \cancel{hx^3} - \cancel{hx^3} = 0$$

□

Lemma 2.7.2.1: Let $T : V \mapsto W$ be a linear transformation between vector spaces V and W. $\ker(V)$ is a subspace of V.

Proof: To show that $\ker(W)$ is a subspace of W, it has to be proven that $\underline{w_1} + \underline{w_2} \in \ker(W)$ holds for all $\underline{w_1}, \underline{w_2} \in \ker(W)$ and that $\lambda\underline{w} \in \ker(W)$ holds for all $\underline{w} \in \ker(W)$ and $\lambda \in \mathbb{R}$.

By definition, if $\underline{v} \in \ker(V)$ holds, then $T(\underline{v}) = 0$. By virtue of T being linear:

$$T\left(\underline{v_1} + \underline{v_2}\right) = T\left(\underline{v_1}\right) + T\left(\underline{v_2}\right) = \underline{0} + \underline{0} = \underline{0} \qquad T(\lambda\underline{v}) = \lambda T(\underline{v}) = \lambda(\underline{0}) = \underline{0}$$

□

As it was done for the image, it's possible to define the kernel of a matrix. Consider the matrix M_ϕ associated to the linear transformation $\phi : V \mapsto W$ with respect to two bases $\mathcal{B}$ and $\mathcal{C}$. The kernel of M_ϕ is the set of all coordinate vector representations X of vectors of V such that $MX = \underline{0}$.

Exercise 2.7.2.2: What is the kernel of $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 5 & 12 \end{pmatrix}$?

Solution:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 5 & 12 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \cdot x + 2 \cdot y + 3 \cdot z \\ 1 \cdot x + 4 \cdot y + 9 \cdot z \\ 1 \cdot x + 5 \cdot y + 12 \cdot z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} x + 2y + 3z \\ x + 4y + 9z \\ x + 5y + 12z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} x + 2y + 3z = 0 \\ x + 4y + 9z = 0 \\ x + 5y + 12z = 0 \end{cases} \Rightarrow \begin{cases} x = -2y - 3z \\ y + 3z = 0 \\ 3y + 9z = 0 \end{cases} \Rightarrow \begin{cases} x = 3z \\ y = -3z \\ z = z \end{cases}$$

Which gives:

$$\ker(A) = \left\{ X \,\middle|\, X = t\begin{pmatrix} 3 \\ -3 \\ 1 \end{pmatrix}, t \in \mathbb{R} \right\}$$

□

2.7.3. Rank and nullity

Let $T : V \mapsto W$ be a linear transformation between vector spaces V and W. The dimension of the image of T is called the **rank** of T, and denoted as $\operatorname{rank}(T)$, while the dimension of the kernel of T

is called the **nullity** of T, and denoted as $\mathrm{null}(T)$. Again, the notions of rank and nullity extend to matrices in the natural way.

Theorem 2.7.3.1 (Rank-nullity theorem): For any linear transformation $T : V \mapsto W$, $\dim(V) = \mathrm{rank}(T) + \mathrm{null}(T)$.

Exercise 2.7.3.1: Consider the linear transformation $T : \mathrm{Mat}(2 \times 2) \mapsto \mathrm{Mat}(2 \times 2)$ defined as:

$$T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} a+b+c & 2b \\ b & 3d+2c \end{pmatrix}$$

What are its rank and nullity?

Solution: The kernel of T is given by:

$$\begin{pmatrix} a+b+c & 2b \\ b & 3d+2c \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow \begin{cases} a+b+c = 0 \\ 2b = 0 \\ b = 0 \\ 3d+2c = 0 \end{cases} \Rightarrow \begin{cases} a = -c \\ b = 0 \\ b = 0 \\ d = -\frac{2}{3}c \end{cases}$$

Which means that $\ker(T)$ is the set:

$$\ker(T) = \left\{ M \in \mathrm{Mat}(2 \times 2) \,\middle|\, M = h\begin{pmatrix} 1 & 0 \\ 0 & \frac{2}{3} \end{pmatrix}, h \in \mathbb{R} \right\}$$

$\ker(T)$ is spanned by a single matrix, which means that its dimension is 1. The image of T is given by:

$$\begin{pmatrix} a+b+c & 2b \\ b & 3d+2c \end{pmatrix} = \begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix} \Rightarrow \begin{cases} a+b+c = l_1 \\ 2b = l_2 \\ b = l_3 \\ 3d+2c = l_4 \end{cases} \Rightarrow \begin{cases} l_1 = a+b+c \\ l_2 = 2l_3 \\ l_3 = b \\ l_4 = 3d+2c \end{cases}$$

Which means that $\mathfrak{I}(T)$ is the set:

$$\mathfrak{I}(T) = \left\{ M \in \mathrm{Mat}(2 \times 2) \,\middle|\, M = \alpha\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \beta\begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} + \gamma\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R} \right\}$$

$\mathfrak{I}(T)$ is spanned by three matrices, which means that its dimension is 3. Note that the dimension of $\mathrm{Mat}(2 \times 2)$ is 4, and indeed $\mathrm{rank}(T) + \mathrm{null}(T) = 3 + 1 = 4$. □

Lemma 2.7.3.1: The rank of a linear transformation is equal to the rank of its matrix representation (with respect to any basis).

2.7.4. Inverse

Let $T : V \mapsto W$ be a linear transformation between vector spaces V and W. The linear transformation $T^{-1} : W \mapsto V$ is said to be the **inverse** of T if:

$$T^{-1}(T(\underline{v})) = T(T^{-1}(\underline{v})) = \underline{v}, \ \forall \underline{v} \in V$$

As for any function, a linear transformation T has an inverse if and only if it's bijective. For this reason, bijective linear transformations are also referred to as **invertible** transformations.

Lemma 2.7.4.1: Let $T : V \mapsto W$ be a linear transformation. T is injective if and only if $\text{null}(T) = 0$.

Proof: If T is injective then, for any distinct $\underline{v_1}, \underline{v_2} \in V$, $T\left(\underline{v_1}\right) \neq T\left(\underline{v_2}\right)$, which is to say $T\left(\underline{v_1}\right) - T\left(\underline{v_2}\right) \neq \underline{0}$. But T is linear by definition, therefore $T\left(\underline{v_1}\right) - T\left(\underline{v_2}\right) = T\left(\underline{v_1} - \underline{v_2}\right)$. Being V a vector space, it algebraically closed with respect to the sum of vectors, therefore $\left(\underline{v_1} - \underline{v_2}\right)$ is itself a member of V distinct from $\underline{0}$, be it $\underline{v}$. In other words, if T is injective, $T(\underline{v})$ has to be different from $\underline{0}$ for any $\underline{v} \in V$, that isn't the null vector, that is to say that the kernel is only composed of the null vector, which is the definition of the nullity of a linear transformation to be 0. $\square$

Corollary 2.7.4.1: Let $T : V \mapsto W$ be a linear transformation. T is invertible if and only if $\dim(V) = \dim(W)$.

Proof: By Theorem 2.7.3.1, $\dim(V) = \dim(\ker(T)) + \dim(\mathfrak{I}(T))$. Being T invertible, the dimension of the image equals the dimension of the codomain W. By Lemma 2.7.4.1, $\dim(\ker(T)) = 0$. Therefore, $\dim(V) = 0 + \dim(\mathfrak{I}(T)) = \dim(W)$. $\square$

Exercise 2.7.4.1: Consider the invertible linear transformation $T : \text{Mat}(2 \times 2) \mapsto \text{Mat}(2 \times 2)$ defined as:

$$T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} a + b + c & 2b \\ b + c & 3d + 2c \end{pmatrix}$$

What is its inverse?

Solution: Reversing the equality and solving for $\{l_1, l_2, l_3, l_4\}$:

$$\begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix} = \begin{pmatrix} a+b+c & 2b \\ b+c & 3d+2c \end{pmatrix} \Rightarrow \begin{cases} l_1 = a+b+c \\ l_2 = 2b \\ l_3 = b+c \\ l_4 = 3d+2c \end{cases} \Rightarrow \begin{cases} l_1 = a+\frac{1}{2}l_2+c \\ b = \frac{1}{2}l_2 \\ l_3 = \frac{1}{2}l_2+c \\ l_4 = 3d+2c \end{cases} \Rightarrow$$

$$\begin{cases} l_1 = a+l_3 \\ b = \frac{1}{2}l_2 \\ c = l_3 - \frac{1}{2}l_2 \\ l_4 = 3d+2l_3-l_2 \end{cases} \Rightarrow \begin{cases} a = l_1 - l_3 \\ b = \frac{1}{2}l_2 \\ c = l_3 - \frac{1}{2}l_2 \\ l_4 = 3d+2l_3-l_2 \end{cases} \Rightarrow \begin{cases} a = l_1 - l_3 \\ b = \frac{1}{2}l_2 \\ c = l_3 - \frac{1}{2}l_2 \\ d = \frac{1}{3}l_4 - \frac{2}{3}l_3 + \frac{1}{3}l_2 \end{cases}$$

Which gives:

$$T^{-1}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} a-c & \frac{1}{2}b \\ c-\frac{1}{2}b & \frac{1}{3}d-\frac{2}{3}c+\frac{1}{3}b \end{pmatrix}$$

Indeed:

$$T^{-1}\left(T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)\right) = T^{-1}\left(\begin{pmatrix} a+b+c & 2b \\ b+c & 3d+2c \end{pmatrix}\right) =$$

$$= \begin{pmatrix} (a+b+c)-(b+c) & \frac{1}{2}(2b) \\ (b+c)-\frac{1}{2}(2b) & \frac{1}{3}(3d+2c)-\frac{2}{3}(b+c)+\frac{1}{3}2b \end{pmatrix} =$$

$$= \begin{pmatrix} a+\cancel{b}+\cancel{c}-\cancel{b}-\cancel{c} & b \\ \cancel{b}+c-\cancel{b} & d+\cancel{\frac{2}{3}c}-\cancel{\frac{2}{3}b}+\cancel{\frac{2}{3}c}+\cancel{\frac{2}{3}b} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$\square$

3. Analytic Geometry

3.1. Inner product and cross product

3.1.1. Inner product

Vector spaces, to be qualified as such, must provide the notion of a sum between two vectors and the notion of a multiplication between a vector and a scalar. However, some vector spaces support operations that go beyond these two.

One such example is the **inner product**: given a vector space V, the inner product of two vectors $\underline{v_1}$ and $\underline{v_2}$ of V, denoted as $\langle \underline{v_1}, \underline{v_2} \rangle$ or as $\underline{v_1} \cdot \underline{v_2}$, is an operation that returns a scalar and that possessing these properties:

- **Symmetry**: for any vectors $\underline{v_1}, \underline{v_2}$, $\langle \underline{v_1}, \underline{v_2} \rangle = \langle \underline{v_2}, \underline{v_1} \rangle$
- **Linearity of the first term**: For any two scalars a, b and for any vectors $\underline{v_1}, \underline{v_2}, \underline{v_3}$, $\langle a\underline{v_1} + b\underline{v_2}, \underline{v_3} \rangle = a\langle \underline{v_1}, \underline{v_3} \rangle + b\langle \underline{v_2}, \underline{v_3} \rangle$
- **Positive-definiteness**: for any non-null vectors $\underline{v_1}, \underline{v_2}$, $\langle \underline{v_1}, \underline{v_2} \rangle \geq 0$.

Vector spaces having a well defined notion of an inner product are called **inner product vector spaces**. Since it always return a scalar, the inner product is sometimes also referred to as the **scalar product**. Since it is sometimes represented as a dot, it is sometimes also referred to as the **dot product**. Even though *scalar product*, *dot product* and *inner product* are all synonims, the first two are mostly used when considering the vector space $\mathbb{R}^n$, whereas the third is more generic.

Exercise 3.1.1.1: Is $\mathbb{R}^n$ an inner product vector space?

Solution: Yes. Given any two vectors $\underline{x}$ and $\underline{y}$ in $\mathbb{R}^n$, the standard inner product over $\mathbb{R}^n$ is defined as:

$$\langle \underline{x}, \underline{y} \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_{n-1} y_{n-1} + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

All three properties are satisfied:

- $\langle \underline{x}, \underline{y} \rangle = x_1 y_1 + \ldots + x_n y_n = y_1 x_1 + \ldots + y_n x_n = \langle \underline{y}, \underline{x} \rangle$
- $\langle a\underline{x} + b\underline{y}, \underline{z} \rangle = (a x_1 + b y_1) z_1 + \ldots + (a x_n + b y_n) z_n = a(x_1 z_1) + b(y_1 z_1) + \ldots + a(x_n z_n) + b(y_n z_n) = a\langle \underline{x}, \underline{z} \rangle + b\langle \underline{y}, \underline{z} \rangle$
- $\langle \underline{x}, \underline{x} \rangle = x_1 x_1 + \ldots + x_n x_n = x_1^2 + \ldots + x_n^2 \geq 0$

□

Any inner product induces a definition of **norm** of a vector, which generalizes the intuitive notion of “length”:

$$\| \underline{v} \| = \sqrt{\langle \underline{v}, \underline{v} \rangle}$$

The inner product of a vector with itself is always non-negative, therefore the square root is non problematic. Since it’s always possible to define a norm for any inner product vector space, they are also called **normed vector spaces**.

Exercise 3.1.1.2: Consider the vector space $\mathbb{R}^3$. Compute the norm of $(1, 2, 3)$.

Solution:

$$\|(1,2,3)\| = \sqrt{\langle (1,2,3), (1,2,3) \rangle} = \sqrt{1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3} = \sqrt{1 + 4 + 9} = \sqrt{13}$$

□

The inner product of two vectors is also related to the **angle** that is formed between them.

Theorem 3.1.1.1: Given two vectors $\underline{a}$ and $\underline{b}$, let $0 \leq \theta \leq \pi$ be the angle between them. Then:

$$\cos(\theta) = \frac{\langle \underline{a}, \underline{b} \rangle}{\|\underline{a}\| \|\underline{b}\|}$$

Solution: Given the two vectors $\underline{a}$ and $\underline{b}$, let $\underline{a} - \underline{b}$ be the difference between the two:

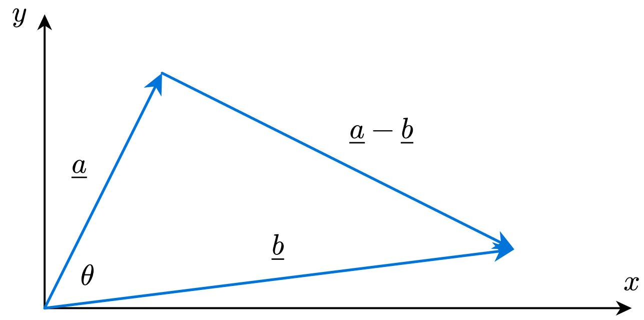

Figure 1: Given $\underline{a}$ and $\underline{b}$, the vector $\underline{a} \pm \underline{b}$ is given by applying the Parallelogram Law.

Applying the Law of Cosines:

$$\|\underline{a} - \underline{b}\|^2 = \|\underline{a}\|^2 + \|\underline{b}\|^2 - 2\|\underline{a}\|\|\underline{b}\| \cos(\theta) \Rightarrow -2\|\underline{a}\|\|\underline{b}\| \cos(\theta) = \|\underline{a} - \underline{b}\|^2 - \|\underline{a}\|^2 - \|\underline{b}\|^2$$

Expanding the right hand side:

$$\begin{aligned}
-2\|\underline{a}\|\|\underline{b}\| \cos(\theta) &= \|\underline{a} - \underline{b}\|^2 - \|\underline{a}\|^2 - \|\underline{b}\|^2 = \\
&= \left((a_x - b_x)^2 + (a_y - b_y)^2 \right) - \left(a_x^2 + a_y^2 \right) - \left(b_x^2 + b_y^2 \right) = \\
&= (a_x - b_x)^2 + (a_y - b_y)^2 - a_x^2 - a_y^2 - b_x^2 - b_y^2 = \\
&= \cancel{a_x^2} + \cancel{b_x^2} - 2a_x b_x + \cancel{a_y^2} + \cancel{b_y^2} - 2a_y b_y - \cancel{a_x^2} - \cancel{a_y^2} - \cancel{b_x^2} - \cancel{b_y^2} = \\
&= -2\left(a_x b_x + a_y b_y \right) = -2\langle \underline{a}, \underline{b} \rangle
\end{aligned}$$

Simplifying the -2 factor on both sides gives the desired result. □

If the cosine of the angle between two vectors is 1, said vectors are said to be **parallel**, while if it is 0 they are said to be **orthogonal**. In particular:

$$1 = \frac{\langle \underline{x}, \underline{y} \rangle}{\|\underline{x}\| \|\underline{y}\|} \Rightarrow \langle \underline{x}, \underline{y} \rangle = \|\underline{x}\| \|\underline{y}\| \qquad\qquad 0 = \frac{\langle \underline{x}, \underline{y} \rangle}{\|\underline{x}\| \|\underline{y}\|} \Rightarrow \langle \underline{x}, \underline{y} \rangle = 0$$

3.1.2. Cross product

Given two vectors $\underline{a} = (a_1, a_2, a_3), \underline{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$, the **cross product** of $\underline{a}$ and $\underline{b}$, denoted as $\underline{a} \times \underline{b}$, is given by:

$$\underline{a} \times \underline{b} = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)$$

Unlike the inner product, the result of the cross product is a vector. For this reason, the cross product is also called **vector product**. Also, while it is perfectly valid to define the inner product for any number of dimensions, the cross product really makes sense only in the context of 3-dimensional vectors.

Lemma 3.1.2.1: Given two vectors $\underline{a}, \underline{b} \in \mathbb{R}^3$, their cross product $\underline{a} \times \underline{b}$ is perpendicular to both $\underline{a}$ and $\underline{b}$.

Proof: This can be checked explicitly:

$$\begin{aligned}
\langle (\underline{a} \times \underline{b}), \underline{a} \rangle &= \langle (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1), \underline{a} \rangle = \\
&= \sqrt{a_1 (a_2 b_3 - a_3 b_2) + a_2 (a_3 b_1 - a_1 b_3) + a_3 (a_1 b_2 - a_2 b_1)} = \\
&= \sqrt{\cancel{a_1 a_2 b_3} - \cancel{a_1 a_3 b_2} + \cancel{a_2 a_3 b_1} - \cancel{a_2 a_1 b_3} + \cancel{a_3 a_1 b_2} - \cancel{a_3 a_2 b_1}} = \sqrt{0} = 0
\end{aligned}$$

$$\begin{aligned}
\langle (\underline{b} \times \underline{a}), \underline{b} \rangle &= \langle (b_2 a_3 - b_3 a_2, b_3 a_1 - b_1 a_3, b_1 a_2 - b_2 a_1), \underline{b} \rangle = \\
&= \sqrt{b_1 (b_2 a_3 - b_3 a_2) + b_2 (b_3 a_1 - b_1 a_3) + b_3 (b_1 a_2 - b_2 a_1)} = \\
&= \sqrt{\cancel{b_1 b_2 a_3} - \cancel{b_1 b_3 a_2} + \cancel{b_2 b_3 a_1} - \cancel{b_2 b_1 a_3} + \cancel{b_3 b_1 a_2} - \cancel{b_3 b_2 a_1}} = \sqrt{0} = 0
\end{aligned}$$

□

There is also a relationship between the cross product of two vectors and the angle between them.

Theorem 3.1.2.1: Given two vectors $\underline{a}$ and $\underline{b}$, let $0 \leq \theta \leq \pi$ be the angle between them. Then:

$$\sin(\theta) = \frac{\|\underline{a} \times \underline{b}\|}{\|\underline{a}\| \|\underline{b}\|}$$

Solution: Given the two vectors $\underline{a}$ and $\underline{b}$, let $\underline{a} \times \underline{b}$ be their cross product:

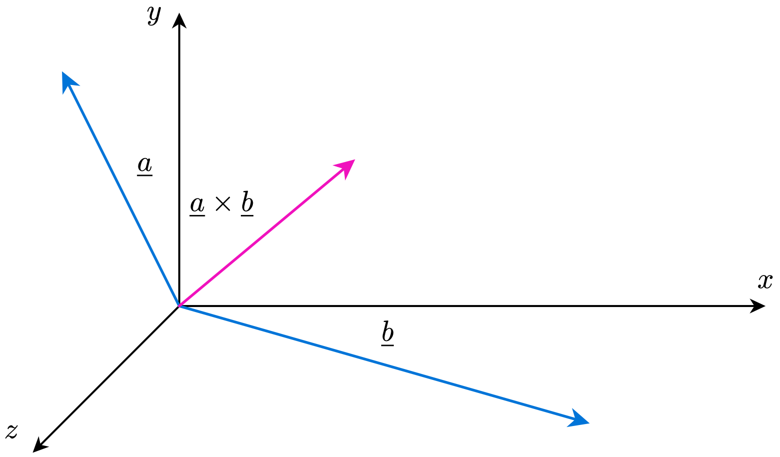

Figure 2: Cross product of $\underline{a}$ and $\underline{b}$. The resulting vector is $\underline{a} \times \underline{b}$.

Taking the square of $\|\underline{a} \times \underline{b}\|$ gives:

$$\begin{aligned}\|\underline{a} \times \underline{b}\|^2 &= (a_2 b_3 - a_3 b_2)^2 + (a_3 b_1 - a_1 b_3)^2 + (a_1 b_2 - a_2 b_1)^2 = \\ &= a_2^2 b_3^2 + a_3^2 b_2^2 - 2a_2 a_3 b_2 b_3 + a_3^2 b_1^2 + a_1^2 b_3^2 - 2a_1 a_3 b_1 b_3 + a_1^2 b_2^2 + a_2^2 b_1^2 - 2a_1 a_2 b_1 b_2\end{aligned}$$

Completing the square of the trinomial:

$$\begin{aligned}\|\underline{a} \times \underline{b}\|^2 &= a_2^2 b_3^2 + a_3^2 b_2^2 - 2a_2 a_3 b_2 b_3 + a_3^2 b_1^2 + a_1^2 b_3^2 - 2a_1 a_3 b_1 b_3 + a_1^2 b_2^2 + a_2^2 b_1^2 - 2a_1 a_2 b_1 b_2 + \\ &+ (a_1^2 b_1^2 - a_1^2 b_1^2) + (a_2^2 b_2^2 - a_2^2 b_2^2) + (a_3^2 b_3^2 - a_3^2 b_3^2) = \\ &= -(a_1 b_1 + a_2 b_2 + a_3 b_3)^2 + a_2^2 b_3^2 + a_3^2 b_2^2 + a_3^2 b_1^2 + a_1^2 b_3^2 + a_1^2 b_2^2 + a_2^2 b_1^2 + a_1^2 b_1^2 + a_2^2 b_2^2 + a_3^2 b_3^2 = \\ &= a_1^2 (b_1^2 + b_2^2 + b_3^2) + a_2^2 (b_1^2 + b_2^2 + b_3^2) + a_3^2 (b_1^2 + b_2^2 + b_3^2) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 = \\ &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2\end{aligned}$$

Substituting the explicit expression of the scalar product:

$$\|\underline{a} \times \underline{b}\|^2 = (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 = \|\underline{a}\|^2 \|\underline{b}\|^2 - \langle \underline{a}, \underline{b} \rangle^2$$

Applying Theorem 3.1.1.1:

$$\begin{aligned}\|\underline{a} \times \underline{b}\|^2 = \|\underline{a}\|^2 \|\underline{b}\|^2 - \langle \underline{a}, \underline{b} \rangle^2 &= \|\underline{a}\|^2 \|\underline{b}\|^2 - (\|\underline{a}\| \|\underline{b}\| \cos(\theta))^2 = \\ &= \|\underline{a}\|^2 \|\underline{b}\|^2 - \|\underline{a}\|^2 \|\underline{b}\|^2 \cos^2(\theta) = \|\underline{a}\|^2 \|\underline{b}\|^2 (1 - \cos^2(\theta)) = \\ &= \|\underline{a}\|^2 \|\underline{b}\|^2 \sin^2(\theta) = (\|\underline{a}\| \|\underline{b}\| \sin(\theta))^2\end{aligned}$$

Taking the square root on both sides, one obtains the desired result. □

Corollary 3.1.2.1: Two non-null vectors are parallel if and only if their cross product is the null vector.

Proof: For two non-null vectors $\underline{a}$ and $\underline{b}$ to be parallel, the angle θ between them has to be equal to π (or to 0). If so, then $\sin(\theta) = 0$, but since Theorem 3.1.2.1 states that $\sin(\theta) = \|\underline{a} \times \underline{b}\| / \|\underline{a}\| \|\underline{b}\|$, this means that $\|\underline{a} \times \underline{b}\| = 0$. The only vector that has norm equal to 0 is the norm vector, therefore $\underline{a} \times \underline{b} = \underline{0}$. □

The cross product also has an interesting geometric interpretation.

Corollary 3.1.2.2: The norm of the cross product $\underline{a} \times \underline{b}$ is equivalent to the area of the parallelogram traced by the vectors $\underline{a}$ and $\underline{b}$.

Proof:

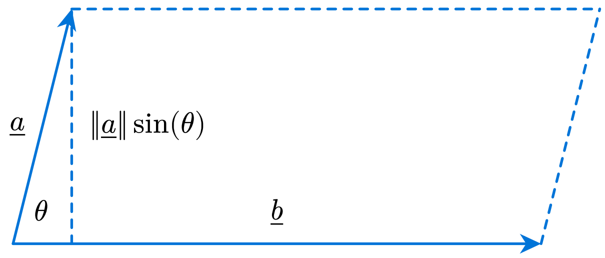

Figure 3: Extending $\underline{a}$ and $\underline{b}$ along each other's parallel direction, one obtains a parallelogram. Its area is equal to $\underline{a} \times \underline{b}$.

Consider two vectors $\underline{a}$ and $\underline{b}$, with an angle θ between them. The projection of $\underline{a}$ onto $\underline{b}$ has, by definition, norm equal to $\|\underline{a}\| \sin(\theta)$. The area of the traced parallelogram is base times height, that is, $\|\underline{b}\|\|\underline{a}\| \sin(\theta)$, which is also equal to $\|\underline{a} \times \underline{b}\|$ due to Theorem 3.1.2.1. □

Lemma 3.1.2.2: Given three vectors $\underline{a}, \underline{b}, \underline{c} \in \mathbb{R}^3$ and a scalar k:

1. $\underline{a} \times \underline{b} = -\underline{b} \times \underline{a}$;
2. $(k\underline{a}) \times \underline{b} = k(\underline{a} \times \underline{b}) = \underline{a} \times (k\underline{b})$;
3. $\underline{a} \times (\underline{b} + \underline{c}) = \underline{a} \times \underline{b} + \underline{a} \times \underline{c}$;
4. $(\underline{a} + \underline{b}) \times \underline{c} = \underline{a} \times \underline{c} + \underline{b} \times \underline{c}$;
5. $\underline{a} \cdot (\underline{b} \times \underline{c}) = (\underline{a} \times \underline{b}) \cdot \underline{c}$;
6. $\underline{a} \times (\underline{b} \times \underline{c}) = (\underline{a} \cdot \underline{c})\underline{b} - (\underline{a} \cdot \underline{b})\underline{c}$.

3.2. Eigenvalues and Eigenvectors

3.2.1. Definition

Let A be an $n \times n$ square matrix, and let λ be a real value. The n-dimensional vector $\underline{x}$ is said to be an **eigenvector** of A if it's not null and if:

$$A\underline{x} = \lambda\underline{x}$$

Where λ is the corresponding **eigenvalue** of A. The set of all the (distinct) eigenvalues of a matrix is called its **eigenspectrum**, or just **spectrum**. It is customary to sort the spectrum of a matrix in descending order.

Note that it's impossible to retrieve the eigenvectors of a matrix A by applying such definition directly. This is because the information contained in the equation is insufficient:

$$A\underline{x} = \lambda\underline{x} \Rightarrow \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \Rightarrow \begin{cases} a_{1,1} \cdot x_1 + \dots + x_n a_{1,n} = \lambda x_1 \\ a_{2,1} \cdot x_1 + \dots + x_n a_{2,n} = \lambda x_2 \\ \vdots \\ a_{n,1} \cdot x_1 + \dots + x_n a_{n,n} = \lambda x_n \end{cases}$$

Even if the entries of A were to be known, each equation has n known terms and $n + 1$ unknowns (the n components of $\underline{x}$ and λ).

3.2.2. Computing the eigenvalues

The correct way to obtain the eigenvectors of a matrix is to retrieve its eigenvalues first. Then, once known, apply the definition to retrieve the eigenvectors.

Given a square matrix A and a real value λ, the **characteristic polynomial** of A is defined as:

$$p_A(\lambda) = \det(A - \lambda I) = c_0 + c_1\lambda + c_2\lambda^2 + ... + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n$$

With $c_0, c_1, ..., c_{n-1} \in \mathbb{R}$. In particular:

$$c_0 = \det(A) \qquad c_{n-1} = (-1)^{n-1}\operatorname{tr}(A)$$

The characteristic polynomial of a matrix can be used to retrieve said spectrum.

Theorem 3.2.2.1: A real value is an eigenvalue for a given matrix if and only if it is a root of its characteristic polynomial.

Proof: First, suppose that $\lambda \in \mathbb{R}$ is an eigenvalue for a $n \times n$ square matrix A. By definition of eigenvalue, there must exist a non-null vector $\underline{x}$ such that $A\underline{x} = \lambda\underline{x}$. Then:

$$A\underline{x} = \lambda\underline{x} \Rightarrow A\underline{x} = \lambda I\underline{x} \Rightarrow A\underline{x} - \lambda I\underline{x} = \underline{0} \Rightarrow (A - \lambda I)\underline{x} = \underline{0}$$

This means that $\underline{x}$ is a vector that belongs to the kernel of the matrix $(A - \lambda I)$. Therefore, the nullity of $(A - \lambda I)$ can't be zero.

By Theorem 2.7.3.1, $\dim(A - \lambda I) = \operatorname{rank}(A - \lambda I) + \operatorname{null}(A - \lambda I)$. But $(A - \lambda I)$ and A have the same dimension, therefore $n = \operatorname{rank}(A - \lambda I) + \operatorname{null}(A - \lambda I)$. Since $\operatorname{null}(A - \lambda I)$ is non zero, for this equality to hold the rank of $(A - \lambda I)$ has to be less than n. By Lemma 2.4.3.2, the matrix $(A - \lambda I)$ cannot be invertible, and by Lemma 2.4.3.1 this must mean that the determinant of $(A - \lambda I)$ is 0.

Suppose then that λ is a root for the characteristic polynomial of A. This means that $\det(A - \lambda I)$ is equal to 0. By Lemma 2.4.3.1, this must mean that $(A - \lambda I)$ is not invertible, which in turn by Lemma 2.4.3.2 must mean that the rank of $(A - \lambda I)$ is less than n. By Theorem 2.7.3.1, $n = \operatorname{rank}(A - \lambda I) + \operatorname{null}(A - \lambda I)$, and being the rank less than n in turn implies that the kernel of $(A - \lambda I)$ does not contain just the null vector. This means that it exists a vector $\underline{x}$ such that $(A - \lambda I)\underline{x} = \underline{0}$. But then:

$$(A - \lambda I)\underline{x} = \underline{0} \Rightarrow A\underline{x} - \lambda I\underline{x} = \underline{0} \Rightarrow A\underline{x} = \lambda I\underline{x} \Rightarrow A\underline{x} = \lambda\underline{x}$$

Which is the definition of eigenvalue. □

Note that an eigenvalue might appear more than once as root of the characteristic polynomial. The number of times an eigenvalue figures as solution to the characteristic polynomial of a matrix is referred to as the **algebraic multiplicity** of the eigenvalue. The algebraic multiplicity of an eigenvalue λ is denoted as $m_a(\lambda)$.

Exercise 3.2.2.1: What is the spectrum of the following matrix?

$$A = \begin{pmatrix} 7 & \frac{10}{3} & -\frac{2}{3} \\ -1 & \frac{7}{3} & -\frac{2}{3} \\ 1 & \frac{2}{3} & \frac{11}{3} \end{pmatrix}$$

Solution: The determinant of A is 75. The $A - \lambda I$ matrix is:

$$A - \lambda I = \begin{pmatrix} 7 & \frac{10}{3} & -\frac{2}{3} \\ -1 & \frac{7}{3} & -\frac{2}{3} \\ 1 & \frac{2}{3} & \frac{11}{3} \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \begin{pmatrix} 7-\lambda & \frac{10}{3} & -\frac{2}{3} \\ -1 & \frac{7}{3}-\lambda & -\frac{2}{3} \\ 1 & \frac{2}{3} & \frac{11}{3}-\lambda \end{pmatrix}$$

The characteristic polynomial of A is:

$$\begin{aligned} p_A(\lambda) = \det(A - \lambda I) = \det\left(\begin{pmatrix} 7-\lambda & \frac{10}{3} & -\frac{2}{3} \\ -1 & \frac{7}{3}-\lambda & -\frac{2}{3} \\ 1 & \frac{2}{3} & \frac{11}{3}-\lambda \end{pmatrix}\right) = \\ = (7-\lambda)\left(\left(\frac{7}{3}-\lambda\right)\left(\frac{11}{3}-\lambda\right)+\frac{4}{9}\right) - \frac{10}{3}\left(-\frac{11}{3}+\lambda+\frac{2}{3}\right) - \frac{2}{3}\left(-\frac{2}{3}-\frac{7}{3}+\lambda\right) = \\ = (7-\lambda)\left(\frac{77}{9}-\frac{7}{3}\lambda-\frac{11}{3}\lambda+\lambda^2+\frac{4}{9}\right) - \frac{10}{3}(\lambda-3) - \frac{2}{3}(\lambda-3) = \\ = (7-\lambda)(\lambda^2-6\lambda+9) - 4(\lambda-3) = (7-\lambda)(\lambda-3)^2 - 4(\lambda-3) = \\ = (\lambda-3)((7-\lambda)(\lambda-3)-4) = (\lambda-3)(7\lambda-21-\lambda^2+3\lambda-4) = \\ = (\lambda-3)(-\lambda^2+10\lambda-25) = -(\lambda-3)(\lambda-5)^2 \end{aligned}$$

The two roots of $p_A(\lambda)$ are $\lambda_1 = 3$, with algebraic multiplicity equal to 1, and $\lambda_2 = 5$, with algebraic multiplicity equal to 2. Applying Theorem 3.2.2.1, the spectrum of A is the set $\{3, 5\}$.

□

Lemma 3.2.2.1: Similar matrices have the same spectrum.

Proof: Let A and B be two similar matrices. There exist then an invertible matrix P such that $A = PBP^{-1}$. Let $p_A(\lambda)$ and $p_B(\lambda)$ be the characteristic polynomial of A and B respectively. Applying Lemma 2.6.1.1 to $p_A(\lambda)$:

$$\begin{aligned} p_A(\lambda) &= \det(A - \lambda I) = \det(PBP^{-1} - \lambda I) = \det(PBP^{-1} - \lambda PIP^{-1}) = \\ &= \det(P(BP^{-1} - \lambda IP^{-1})) = \det(P(B - \lambda I)P^{-1}) = \\ &= \cancel{\det(P)}\det(B - \lambda I)\cancel{\det(P^{-1})} = \det(B - \lambda I) = p_B(\lambda) \end{aligned}$$

□

There is also an interesting connection between the eigenvalues of a matrix, its determinant and its trace.

Lemma 3.2.2.2: The determinant of a matrix is the product of its eigenvalues (counted with multiplicity), whereas the trace of a matrix is the sum of its eigenvalues (counted with multiplicity).

3.2.3. Computing the eigenvectors

Once Theorem 3.2.2.1 is applied, it is then possible to retrieve the eigenvectors of a matrix by applying the definition.

Note that the expression $\det(A - \lambda I) = 0$ entails that the system of equations associated to $A\underline{x} = \lambda\underline{x}$ has an infinite number of solutions. This means that for each eigenvalue there exist not just one eigenvector, but an infinite set of them.

The vector space spanned by each set of eigenvectors associated to a certain eigenvalue is called its **eigenspace**. The eigenspace associated to an eigenvalue λ is denoted as E_λ. The dimension of an eigenspace is referred to as the **geometric multiplicity** of the corresponding eigenvalue. The geometric multiplicity of an eigenvalue λ is denoted as $m_g(\lambda)$.

Exercise 3.2.3.1: What are the eigenspaces of the matrix of Exercise 3.2.2.1 ?

Solution: The eigenspace of λ_1 is:

$$A\underline{x} = \lambda_1\underline{x} \Rightarrow \begin{pmatrix} 7 & \frac{10}{3} & -\frac{2}{3} \\ -1 & \frac{7}{3} & -\frac{2}{3} \\ 1 & \frac{2}{3} & \frac{11}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 3 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} 7 \cdot x + \frac{10}{3} \cdot y - \frac{2}{3} \cdot z \\ -1 \cdot x + \frac{7}{3} \cdot y - \frac{2}{3} \cdot z \\ 1 \cdot x + \frac{2}{3} \cdot y + \frac{11}{3} \cdot z \end{pmatrix} = \begin{pmatrix} 3x \\ 3y \\ 3z \end{pmatrix} \Rightarrow$$

$$\begin{cases} 7x + \frac{10}{3}y - \frac{2}{3}z = 3x \\ -x + \frac{7}{3}y - \frac{2}{3}z = 3y \\ x + \frac{2}{3}y + \frac{11}{3}z = 3z \end{cases} \Rightarrow \begin{cases} 21x + 10y - 2z - 9x = 0 \\ 3x - 7y + 2z + 9y = 0 \\ 3x + 2y + 11z - 9z = 0 \end{cases} \Rightarrow \begin{cases} 6x + 5y - z = 0 \\ 3x + 2y + 2z = 0 \\ 3x + 2y + 2z = 0 \end{cases} \Rightarrow$$

$$\begin{cases} 5x + 4y = 0 \\ z = -\frac{3}{2}x - y \\ z = -\frac{3}{2}x - y \end{cases} \Rightarrow \begin{cases} x = x \\ y = -\frac{5}{4}x \\ z = -\frac{1}{4}x \end{cases}$$

The eigenspace of λ_2 is:

$$A\underline{x} = \lambda_2\underline{x} \Rightarrow \begin{pmatrix} 7 & \frac{10}{3} & -\frac{2}{3} \\ -1 & \frac{7}{3} & -\frac{2}{3} \\ 1 & \frac{2}{3} & \frac{11}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 5 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} 7 \cdot x + \frac{10}{3} \cdot y - \frac{2}{3} \cdot z \\ -1 \cdot x + \frac{7}{3} \cdot y - \frac{2}{3} \cdot z \\ 1 \cdot x + \frac{2}{3} \cdot y + \frac{11}{3} \cdot z \end{pmatrix} = \begin{pmatrix} 5x \\ 5y \\ 5z \end{pmatrix} \Rightarrow$$

$$\begin{cases} 7x + \frac{10}{3}y - \frac{2}{3}z = 5x \\ -x + \frac{7}{3}y - \frac{2}{3}z = 5y \\ x + \frac{2}{3}y + \frac{11}{3}z = 5z \end{cases} \Rightarrow \begin{cases} 21x + 10y - 2z - 15x = 0 \\ 3x - 7y + 2z + 15y = 0 \\ 3x + 2y + 11z - 15z = 0 \end{cases} \Rightarrow \begin{cases} 3x + 5y - z = 0 \\ 3x + 8y + 2z = 0 \\ 3x + 2y - 4z = 0 \end{cases} \Rightarrow$$

$$\begin{cases} z = 3x + 5y \\ x + 2y = 0 \\ x + 2y = 0 \end{cases} \Rightarrow \begin{cases} z = -y \\ y = y \\ x = -2y \end{cases}$$

This means that the two eigenspaces are $E_{\lambda_1} = \operatorname{span}\left\{\left(1, -\frac{5}{4}, -\frac{1}{4}\right)^T\right\}$ and $E_{\lambda_2} = \operatorname{span}\{(-1, 1, -2)^T\}$. Both have geometric multiplicity equal to 1. □

Lemma 3.2.3.1: For any eigenvalue λ, $1 \leq m_g(\lambda) \leq m_a(\lambda)$.

3.2.4. Eigenvectors and eigenvalues of linear transformations

Eigenvectors and eigenvalues can be defined with respect to linear transformations as well. Given an endomorphism $T : V \mapsto V$, a vector $\underline{v} \in V$ is an eigenvector for T if $T\underline{v} = \lambda\underline{v}$, where λ is an eigenvalue for T.

As stated in Theorem 3.2.2.1, the eigenvalues of a matrix can be computed employing its characteristic polynomial. Since any linear transformation can be represented using a matrix, to compute the eigenvalues of a linear transformation it is possible to compute the characteristic polynomial of the associated matrix of the linear transformation. Once the eigenvalues are known, the eigenvectors can be retrieved as usual.

Lemma 3.2.2.1 states that similar matrices have the same eigenvalues, and all matrix representations of the same linear transformations are always similar. This means that the choice of the bases of the matrix associated to the endomorphism, with respect to finding its eigenvalues, are irrelevant. For this reason, it is possible to refer to the characteristic polynomial of a linear transformation without having to specify the basis. Of course, from a practical standpoint, the most convenient choice of basis is most likely the canonical basis.

Exercise 3.2.4.1: What are the eigenvectors and eigenvalues of the linear transformation $T : \mathbb{R}_2[x] \mapsto \mathbb{R}_2[x]$ defined as $T(p(x)) = p(x) - 3x\frac{d}{dx}p(x) + 4\frac{d^2}{dx}p(x)$?

Solution: First, it is necessary to construct the matrix representation of T. Evaluating the canonical basis $\{1, x, x^2\}$ with T gives:

$$\begin{cases} T(1) &= 1 - 3x\frac{d}{dx}(1) + 4\frac{d^2}{dx}(1) = 1 - 0 + 0 = 1 \\ T(x) &= x - 3x\frac{d}{dx}(x) + 4\frac{d^2}{dx}(x) = x - 3x = -2x \\ T(x^2) &= x^2 - 3x\frac{d}{dx}(x^2) + 4\frac{d^2}{dx}(x^2) = x^2 - 6x^2 + 8 = 8 - 5x^2 \end{cases}$$

The matrix A is therefore constituted by the following columns:

$$1 \Longleftrightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad -2x \Longleftrightarrow \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} \qquad 8 - 5x^2 \Longleftrightarrow \begin{pmatrix} 8 \\ 0 \\ -5 \end{pmatrix}$$

The characteristic polynomial of A is:

$$p_A(\lambda) = \det\left(\begin{pmatrix} 1 & 0 & 8 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) =$$

$$= \det\left(\begin{pmatrix} 1-\lambda & 0 & 8 \\ 0 & -2-\lambda & 0 \\ 0 & 0 & -5-\lambda \end{pmatrix}\right) = -(\lambda - 1)(\lambda + 2)(\lambda + 5)$$

Which means that the spectrum of A (and of T) is $\{-5, -2, 1\}$. The eigenspace of A is:

$$\begin{pmatrix} 1 & 0 & 8 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = -5\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \cdot x + 0 \cdot y + 8 \cdot z \\ 0 \cdot x - 2 \cdot y + 0 \cdot z \\ 0 \cdot x + 0 \cdot y - 5 \cdot z \end{pmatrix} = -5\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} x + 8z = -5x \\ -2y = -5y \\ -5z = -5z \end{cases} \Rightarrow \begin{cases} x = -\frac{4}{3}z \\ y = 0 \\ z = z \end{cases}$$

$$\begin{pmatrix} 1 & 0 & 8 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = -2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \cdot x + 0 \cdot y + 8 \cdot z \\ 0 \cdot x - 2 \cdot y + 0 \cdot z \\ 0 \cdot x + 0 \cdot y - 5 \cdot z \end{pmatrix} = -2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} x + 8z = -2x \\ -2y = -2y \\ -5z = -2z \end{cases} \Rightarrow \begin{cases} x = 0 \\ y = y \\ z = 0 \end{cases}$$

$$\begin{pmatrix} 1 & 0 & 8 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \cdot x + 0 \cdot y + 8 \cdot z \\ 0 \cdot x - 2 \cdot y + 0 \cdot z \\ 0 \cdot x + 0 \cdot y - 5 \cdot z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} x + 8z = x \\ -2y = y \\ -5z = z \end{cases} \Rightarrow \begin{cases} x = x \\ y = 0 \\ z = 0 \end{cases}$$

The eigenspaces are $E_{\lambda_1} = \text{span}\left\{\left(-\frac{4}{3}, 0, 1\right)^T\right\}$, $E_{\lambda_2} = \text{span}\{(0,1,0)^T\}$ and $E_{\lambda_3} = \text{span}\{(1,0,0)^T\}$. All have geometric multiplicity equal to 1.

The eigenspaces of T can be constructed by "undoing" the vector representation of the eigenspaces of A:

$$\begin{aligned} \left(-\frac{4}{3}t\right) \cdot 1 + 0 \cdot x + t \cdot x^2 &= -\frac{4}{3}t + x^2 t \\ 0 \cdot 1 + t \cdot x + 0 \cdot x^2 &= xt \\ t \cdot 1 + 0 \cdot x + 0 \cdot x^2 &= t \end{aligned}$$

Which gives: $E_{\lambda_1} = \text{span}\left\{-\frac{4}{3}t + x^2\right\}$, $E_{\lambda_2} = \text{span}\{x\}$, $E_{\lambda_3} = \text{span}\{1\}$. □

3.2.5. Diagonalization

Theorem 3.2.5.1 (Diagonalization theorem):

- (**With respect to matrices**) Let A be a $n \times n$ matrix that has n linearly independent eigenvectors $\underline{e_1}, \ldots, \underline{e_n}$. Then there exist a diagonal matrix D that is similar to A, meaning that there exist a matrix P such that $A = PDP^{-1}$. In particular, the matrix D has the eigenvalues of A as non-zero elements (counted with multiplicity) and P has the eigenvectors of A as columns.
- (**With respect to endomorphisms**) Let $T : V \mapsto V$ be an endomorphism mapping vector spaces of dimension n having n linearly independent eigenvectors $\underline{e_1}, \ldots, \underline{e_n}$. Let M be the matrix representation of T with respect to said eigenvectors. The matrix M is a diagonal matrix whose non-zero entries are the eigenvalues of T.

Proof:

Consider the two matrices P and D:

$$P = \begin{pmatrix} \underline{e_1} & \underline{e_2} & \ldots & \underline{e_n} \end{pmatrix} = \begin{pmatrix} e_{1,1} & e_{2,1} & \ldots & e_{n,1} \\ e_{1,2} & e_{2,2} & \ldots & e_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1,n} & e_{2,n} & \ldots & e_{n,n} \end{pmatrix} \qquad D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

The matrix multiplication AP is by definition equivalent to multiplying A with each column vector of P. That is, the i-th column of AP is given by multiplying the matrix A with the i-th column vector of P, giving $A\underline{e_i}$. But by definition multiplying the matrix representation of an endomorphism with one of his eigenvectors is equivalent to multiplying said eigenvector by its corresponding eigenvalue. Therefore:

$$AP = \begin{pmatrix} A\underline{e_1} & A\underline{e_2} & \ldots & A\underline{e_n} \end{pmatrix} = \begin{pmatrix} \lambda_1 \underline{e_1} & \lambda_2 \underline{e_2} & \ldots & \lambda_n \underline{e_n} \end{pmatrix}$$

Consider the matrix multiplication PD. By definition, the i-th element of such matrix is given by the sum of the products of the corresponding elements of the i-th row of P and the i-th column of D. By construction, the elements in D are zero except for the ones on its diagonal, therefore the i-th column of PD is just the i-th column vector of P multiplied by the i, i-th element of D, which is λ_i. Therefore:

$$PD = \left(\lambda_1 \underline{e_1} \;\; \lambda_2 \underline{e_2} \;\; \dots \;\; \lambda_n \underline{e_n}\right)$$

This shows that the two matrix products AP and PD are equivalent. Since by assumption the set of the eigenvectors of A form a basis, by Lemma 2.4.3.2 P has to be invertible. But then:

$$AP = PD \Rightarrow A\cancel{PP^{-1}} = PDP^{-1} \Rightarrow A = PDP^{-1}$$

□

If Theorem 3.2.5.1 holds for a certain matrix, said matrix is said to be **diagonalizable**. Note that, even though a matrix always has as many eigenvalues as its dimension, not all matrices are diagonalizable. This is because the Fundamental Theorem of Algebra states that the characteristic polynomial of a matrix of dimension n will always have n roots (albeit some might be complex numbers), but the corresponding eigenvectors might not form a basis.

A matrix that is not diagonalizable is said to be **defective**. Determining whether a matrix is diagonalizable or defective can be done either applying Theorem 3.2.5.1 directly, but there are equivalent necessary and sufficient conditions that can simplify the process.

Theorem 3.2.5.2: A matrix is diagonalizable if and only if, for each of its eigenvalues λ_i, $m_g(\lambda_i) = m_a(\lambda_i)$.

Corollary 3.2.5.1: If a matrix of dimension n has n distinct eigenvalues, then it's always diagonalizable.

Proof: Consider a matrix of dimension n having n distinct eigenvalues. Each eigenvalue, by definition, has both algebraic and geometric multiplicity equal to 1. Therefore, Theorem 3.2.5.2 always applies. □

The process of "breaking down" a diagonalizable matrix A into a triplet PDP^{-1} applying Theorem 3.2.5.1 is called **diagonalization**. Note that, for any matrix diagonalizable A, there exist an infinite amount of possible diagonalizations, since for any $k \in \mathbb{R}$, the matrix $(kP)D(k^{-1}P^{-1})$ would still be a valid diagonalization (multiplying an eigenvector by a scalar still gives an eigenvector).

Exercise 3.2.5.1: Consider the matrices in Exercise 3.2.3.1 and Exercise 3.2.4.1. Are they diagonalizable?

Solution: The matrix in Exercise 3.2.3.1 has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 5$. Their algebraic multiplicity are 2 and 1 respectively, whereas their geometric multiplicity are both 1. By virtue of Theorem 3.2.5.2, it is not diagonalizable.

On the other hand, the matrix in Exercise 3.2.4.1 has three eigenvalues, whose algebraic and geometric multiplicity is 1 in all of three cases. This means that, due to Corollary 3.2.5.1, it is diagonalizable. The diagonalization is:

$$\begin{pmatrix} 1 & 0 & 8 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{pmatrix} = \begin{pmatrix} -\frac{4}{3} & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} -5 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & \frac{4}{3} \end{pmatrix}$$

□

3.3. Spectral Theorem

3.3.1. Orthogonal sets

A set of vectors that are all orthogonal to each other is called an **orthogonal set**. A set of vectors $V = \{\underline{v_1}, ..., \underline{v_n}\}$ is orthogonal if, for any pair of vectors $\underline{v_i}$ and $\underline{v_j}$ that are distinct, $\langle \underline{v_i}, \underline{v_j} \rangle = 0$. An orthogonal set that also forms a basis is called an **orthogonal basis**.

A set of vectors that is both orthogonal and having all vectors with norm equal to 1 is called an **orthonormal set**; an orthonormal set that also forms a basis is called an **orthonormal basis**. A set V of vectors forms an orthonormal set if, for any $\underline{v_i}, \underline{v_j} \in V$, the following holds:

$$\langle \underline{v_i}, \underline{v_j} \rangle = \begin{cases} 1 \text{ if } \underline{v_i} = \underline{v_j} \\ 0 \text{ otherwise} \end{cases}$$

By extension, an **orthogonal matrix** is a matrix whose rows/columns, considered as vectors, form an orthogonal set. In the same way, an **orthonormal matrix** is a matrix whose rows/columns, considered as vectors, form an orthonormal set.

Lemma 3.3.1.1: An orthogonal matrix has its inverse equal to its transposed.

Lemma 3.3.1.2: Rotation matrices are orthonormal.

Proof: Consider, for the sake of simplicity, rotations in $\mathbb{R}^2$; the case for arbitrary dimensions follows the same idea. Let then $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ for a given angle θ. Its determinant is:

$$\det\left(\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}\right) = \cos(\theta)\cos(\theta) - \sin(\theta)(-\sin(\theta)) = \cos^2(\theta) + \sin^2(\theta) = 1$$

The inverse of R_θ is:

$$R_\theta^{-1} = \begin{pmatrix} \frac{\cos(\theta)}{1} & \frac{\sin(\theta)}{1} \\ \frac{-\sin(\theta)}{1} & \frac{\cos(\theta)}{1} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = R_\theta^T$$

Lemma 3.3.1.1 ensures that R_θ is orthogonal. As for the norm of its columns:

$$\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \cdot \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \cos(\theta)\cos(\theta) + \sin(\theta)\sin(\theta) = \cos^2(\theta) + \sin^2(\theta) = 1$$

$$\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \cdot \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} = \cancel{\cos(\theta)(-\sin(\theta))} + \cancel{\sin(\theta)\cos(\theta)} = 0$$

$$\begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \cancel{\sin(\theta)\cos(\theta)} + \cancel{\cos(\theta)(-\sin(\theta))} = 0$$

$$\begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} = (-\sin(\theta))(-\sin(\theta)) + \cos(\theta)\cos(\theta) = \sin^2(\theta) + \cos^2(\theta) = 1$$

Which means that R_θ is also orthonormal by definition. □

3.3.2. Spectral theorem

Theorem 3.2.5.1 holds an interesting result when applied to symmetric matrices.

Theorem 3.3.2.1 (Spectral theorem): If A is a symmetric matrix, there exists an orthonormal basis of the corresponding vector space V consisting of eigenvectors of A, and each eigenvalue is real.

Exercise 3.3.2.1: Diagonalize the following symmetric matrix $A \in \mathrm{Mat}(3 \times 3)$. Find an orthonormal basis out of its eigenvectors.

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

Solution: The characteristic polynomial of A is given by:

$$p_A(\lambda) = \det(A - \lambda I) = \det\left(\begin{pmatrix} -1-\lambda & 1 & 0 \\ 1 & -1-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{pmatrix}\right) =$$

$$= (3-\lambda)\det\left(\begin{pmatrix} -1-\lambda & 1 \\ 1 & -1-\lambda \end{pmatrix}\right) = (3-\lambda)((-1-\lambda)^2 - 1) =$$

$$= (3-\lambda)(\cancel{1} + \lambda^2 + 2\lambda \cancel{-1}) = \lambda(3-\lambda)(\lambda+2)$$

The spectrum of A is $\{-2, 0, 3\}$. As for the eigenvectors:

$$A\underline{x} = \lambda\underline{x} \Rightarrow \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} -1 \cdot x + 1 \cdot y + 0 \cdot z \\ 1 \cdot x - 1 \cdot y + \cdot 0 \cdot z \\ 0 \cdot x + 0 \cdot y + 3 \cdot z \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix} \Rightarrow \begin{cases} y - x = \lambda x \\ x - y = \lambda y \\ 3z = \lambda z \end{cases}$$

Solving for $\lambda_1 = -2, \lambda_2 = 0, \lambda_3 = 3$:

$$\begin{cases} y - x = -2x \\ x - y = -2y \\ 3z = -2z \end{cases} \Rightarrow \begin{cases} x = -y \\ y = y \\ z = 0 \end{cases} \qquad \begin{cases} y - x = 0 \\ x - y = 0 \\ 3z = 0 \end{cases} \Rightarrow \begin{cases} x = y \\ y = y \\ z = 0 \end{cases} \qquad \begin{cases} y - x = 3x \\ x - y = 3y \\ 3z = 3z \end{cases} \Rightarrow \begin{cases} x = 0 \\ y = 0 \\ z = z \end{cases}$$

Which gives: $E_{\lambda_1} = \text{span}\{(-1,1,0)^T\}$, $E_{\lambda_2} = \text{span}\{(1,1,0)^T\}$, $E_{\lambda_3} = \text{span}\{(0,0,1)^T\}$. Out of all the eigenvectors, the interest is to find the ones having norm equal to 1.

$$\|(-t,t,0)^T\| = 1 \Rightarrow \sqrt{(-t)\cdot(-t) + t\cdot t + 0\cdot 0} = 1 \Rightarrow \sqrt{2t^2} = 1 \Rightarrow t = \pm\frac{1}{\sqrt{2}}$$
$$\|(t,t,0)^T\| = 1 \Rightarrow \sqrt{t\cdot t + t\cdot t + 0\cdot 0} = 1 \Rightarrow \sqrt{2t^2} = 1 \Rightarrow t = \pm\frac{1}{\sqrt{2}}$$
$$\|(0,0,t)^T\| = 1 \Rightarrow \sqrt{0\cdot 0 + 0\cdot 0 + t\cdot t} = 1 \Rightarrow \sqrt{t^2} = 1 \Rightarrow t = \pm 1$$

The diagonalization of A is as follows:

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

Where the $\pm$ sign has been factored out. Note that, due to Lemma 3.3.1.1, the eigenvector matrices have their inverse equal to their transposed. □

3.3.3. Definite matrices

Let A be a symmetric matrix and let $\underline{x}$ be a column vector. A is said to be:

- **Definite positive** if, for any $\underline{x}$, $\langle \underline{x}, A\underline{x} \rangle > 0$;
- **Semidefinite positive** if, for any $\underline{x}$, $\langle \underline{x}, A\underline{x} \rangle \geq 0$;
- **Definite negative** if, for any $\underline{x}$, $\langle \underline{x}, A\underline{x} \rangle < 0$;
- **Semidefinite negative** if, for any $\underline{x}$, $\langle \underline{x}, A\underline{x} \rangle \leq 0$;
- **Indefinite** otherwise.

Lemma 3.3.3.1: Let A be a symmetric matrix. If A is:

- Definite positive, all of its eigenvalues are strictly positive;
- Semidefinite positive, all of its eigenvalues are non negative;
- Definite negative, all of its eigenvalues are strictly negative;
- Semidefinite negative, all of its eigenvalues are non positive;

Lemma 3.3.3.2: For any matrix A, the matrices $A^T A$ and AA^T are positive semidefinite.

Proof: For a matrix to be positive definite it also needs to be symmetric. Matrix $A^T A$ is indeed symmetric since $(A^T A)^T = A^T (A^T)^T = A^T A$. Let $\underline{y} = A\underline{x}$. Then $\underline{y}^T = (A\underline{x})^T = \underline{x}^T A^T$. This means that:

$$\langle \underline{x}, A^T A\underline{x} \rangle = \underline{x}^T A^T A\underline{x} = \underline{y}^T \underline{y} = \sum_{i=1}^{n} y_i^2$$

Which, by definition, is greater or equal than 0. AA^T can be proven to be positive semidefinite following a similar line of thought. □

3.4. Decompositions

Theorem 3.4.1 (Cholesky Decomposition): For any positive definite matrix A there exists a lower triangular matrix L such that $A = LL^T$.

Proof: The theorem can be proven in a constructive way by defining an algorithm that recursively retrieves said L matrix.

First, the three matrices at play ought to have such form:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{1,2} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{n,n} \end{pmatrix} \qquad L = \begin{pmatrix} l_{1,1} & 0 & \dots & 0 \\ l_{1,2} & l_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{1,n} & l_{2,n} & \dots & l_{n,n} \end{pmatrix} \qquad L^T = \begin{pmatrix} l_{1,1} & l_{1,2} & \dots & l_{1,n} \\ 0 & l_{2,2} & \dots & l_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & l_{n,n} \end{pmatrix}$$

The $(1, 1)$ entry of the product between L and L^T is given by the inner product of the first row of L and the first column of L^T:

$$l_{1,1} \cdot l_{1,1} + 0 \cdot 0 + 0 \cdot 0 + \dots + 0 \cdot 0 = l_{1,1}^2$$

This means that, for the equality $A = LL^T$ to be true, $l_{1,1}$ ought to be equal to $\sqrt{a_{1,1}}$.

The generic $(1, i)$ entry of the product between L and L^T is given by the inner product of the first row of L and the i-th column of L^T:

$$l_{1,1} \cdot l_{1,i} + 0 \cdot l_{2,i} + 0 \cdot l_{3,i} + \dots + 0 \cdot 0 = l_{1,1} l_{1,i}$$

This means that, for the equality $A = LL^T$ to be true, $a_{1,i}$ ought to be equal to $l_{1,1} l_{1,i}$, which in turn means that $l_{1,i}$ ought to be equal to $a_{1,i} / l_{1,1}$.

The $(2, 2)$ entry of the product between L and L^T is given by the inner product of the second row of L and the second column of L^T:

$$l_{1,2} \cdot l_{1,2} + l_{2,2} \cdot l_{2,2} + 0 \cdot 0 + \dots + 0 \cdot 0 = l_{1,2}^2 + l_{2,2}^2$$

This means that, for the equality $A = LL^T$ to be true, $l_{2,2}$ must be equal to $\sqrt{a_{2,2} - l_{1,2}^2}$.

□

Theorem 3.2.5.1 states that if a square matrix A possesses certain properties, it can be written as a product in the form PDP^{-1}. A more generic result can be achieved for non-square matrices.

Theorem 3.4.2 (Singular Value Decomposition): Any $m \times n$ matrix A can be written as the product $A = U \Sigma V^T$, where:

- U is a $m \times m$ orthogonal matrix whose column vectors $\left(\underline{u_1}, \dots, \underline{u_m}\right)$ belong to $\mathbb{R}^m$ and are called **left singular vectors**;
- Σ is a $m \times n$ matrix such that the $\sigma_{i,i}$ entries, called **singular values**, are greater or equal than 0 while the $\sigma_{i,j}$ entries with $i \neq j$ are exactly 0;
- V is a $n \times n$ orthogonal matrix whose column vectors $\left(\underline{v_1}, \dots, \underline{v_m}\right)$ belong to $\mathbb{R}^n$ and are called **right singular vectors**.

Proof: The theorem can be proven in a constructive way by defining an algorithm that generates said matrices, which in turn can prove that the equality holds. Assume, without loss of generality, that $m > n$.

Suppose that $A = U\Sigma V^T$. This means that:

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = \left((V^T)^T \Sigma^T U^T\right) U\Sigma V^T = V\Sigma^T \cancel{U^T} \cancel{U} \Sigma V^T = V\Sigma^T \Sigma V^T$$

By Lemma 3.3.3.2, $A^T A$ is positive semidefinite. In turn, by Theorem 3.3.2.1, it can be diagonalized as PDP^T, where the eigenvalues along the diagonal of D are non negative as of Lemma 3.3.3.1.

Since the dimension of A is $m \times n$, the dimension of $A^T A$ ought to be $n \times n$. In turn, the dimension of D ought to be $n \times n$ as well. By how Σ has been defined, the product $\Sigma^T \Sigma$ ought to be:

$$\Sigma^T \Sigma = \begin{pmatrix} \sigma_{1,1} & 0 & \dots & 0 \\ 0 & \sigma_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{n,n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{1,1} & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_{2,2} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_{n,n} & \dots & 0 \end{pmatrix} = \begin{pmatrix} \sigma_{1,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{2,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{n,n}^2 \end{pmatrix}$$

Being $\Sigma^T \Sigma$ a diagonal matrix, the representations PDP^T and $V\Sigma^T \Sigma V^T$ can be equated as long as the non-zero values of $\Sigma^T \Sigma$ are the square roots of the eigenvalues of $A^T A$ and the column vectors of V are the normalized eigenvectors of $A^T A$.

Similarly:

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma V^T \left((V^T)^T \Sigma^T U^T\right) = U\Sigma \cancel{V^T} \cancel{V} \Sigma^T U^T = U\Sigma \Sigma^T U^T$$

By Lemma 3.3.3.2, AA^T is positive semidefinite. In turn, by Theorem 3.3.2.1, it can be diagonalized as QCQ^T, where the eigenvalues along the diagonal of C are non negative as of Lemma 3.3.3.1.

Since the dimension of A is $m \times n$, the dimension of AA^T ought to be $m \times m$. In turn, the dimension of C ought to be $m \times m$ as well. By how Σ has been defined, the product $\Sigma\Sigma^T$ ought to be:

$$\Sigma \Sigma^T = \begin{pmatrix} \sigma_{1,1} & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_{2,2} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_{m,m} & \dots & 0 \end{pmatrix} \begin{pmatrix} \sigma_{1,1} & 0 & \dots & 0 \\ 0 & \sigma_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{m,m} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \sigma_{1,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{2,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{m,m}^2 \end{pmatrix}$$

Being $\Sigma\Sigma^T$ a diagonal matrix, the representations QCQ^T and $U\Sigma\Sigma^T U^T$ can be equated as long as the non-zero values of $\Sigma\Sigma^T$ are the square roots of the eigenvalues of AA^T and the column vectors of U are the normalized eigenvectors of AA^T.

The next step consists of comparing the eigenvalues of the two matrices D and C.

By definition, an eigenvector $\underline{e_i}$ of $A^T A$ satisfies the equation $A^T A \underline{e_i} = \lambda_i \underline{e_i}$. But:

$$A^T A\underline{e_i} = \lambda_i \underline{e_i} \Rightarrow AA^T A\underline{e_i} = A\lambda_i \underline{e_i} \Rightarrow (AA^T)\left(A\underline{e_i}\right) = \lambda_i \left(A\underline{e_i}\right) \Rightarrow (AA^T)\underline{u_i} = \lambda_i \underline{u_i}$$

If $\underline{u_i}$ is guaranteed to be different from the null vector, then it is an eigenvector for AA^T. If it were to be the null vector, then:

$$A^T A\underline{e_i} = \lambda_i \underline{e_i} \Rightarrow A^T \underline{0} = \lambda_i \underline{e_i} \Rightarrow \underline{0} = \lambda_i \underline{e_i}$$

Which is true exclusively if λ_i is 0, because $\underline{e_i}$ can't be the null vector by definition of eigenvector.

This means that $A^T A$ and AA^T have the same eigenvalues (even though not necessarely the same eigenvectors) as long as said eigenvalues are not 0. In other words, if $d_{i,i}$ and $c_{i,i}$ are both different from 0, it is guaranteed that they are equal, while if one of them is equal to 0 the other one is different from 0.

To complete the proof, it is necessary to show that A is indeed equal to $U\Sigma V^T$. Consider the matrices AV and $U\Sigma$. Said matrices are constituted by the following column vectors:

$$AV = \left(\underline{A_1 v_1} \ \cdots \ \underline{A_n v_n}\right) \qquad U\Sigma = \left(\underline{u_1 \sigma_1} \ \cdots \ \underline{u_n \sigma_n}\right)$$

By the previous result, $\underline{A_i v_i}$ is equal to $\underline{u_i \sigma_i}$ whenever the corresponding eigenvalues are non zero and both zero otherwise. This means that the two matrices can be equated column by column. Therefore:

$$AV = U\Sigma \Rightarrow A\cancel{VV^T} = U\Sigma V^T \Rightarrow A = U\Sigma V^T$$

□

Summarizing, it is possible to perform the Singular Value Decomposition (SVD) of a matrix A by applying the algorithm stated in Theorem 3.4.2:

- Compute $A^T A$;
- Compute the eigenvalues of $A^T A$. The diagonal entries of Σ will be the square root of said eigenvalues;
- Compute the eigenvectors of $A^T A$. The normalized choice of eigenvectors will be the column vectors of V;
- If the i-th entry of Σ is non zero, the i-th column vector of U can be computed from the i-th column vector of V as $A\underline{e_i}/\sigma_{i,i}$;
- If the i-th entry of Σ is 0, the i-th column vector of U has to be computed from the kernel of $A\underline{x}$;
- The SVD of A is given by $U\Sigma V^T$.

It is customary to place the diagonal entries of Σ in decreasing order.

Exercise 3.4.1: Compute the Singular Value Decomposition of the following matrix:

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

Solution: First, it is necessary to compute $A^T A$:

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

Whose eigenvalues can be retrieved from its characteristic polynomial:

$$\det(A-\lambda I) = \begin{vmatrix} 2-\lambda & 1 & 0 \\ 1 & 1-\lambda & 1 \\ 0 & 1 & 2-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 1-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} - \begin{vmatrix} 1 & 1 \\ 0 & 2-\lambda \end{vmatrix} + 0\begin{vmatrix} 1 & 1-\lambda \\ 0 & 1 \end{vmatrix} =$$

$$(2-\lambda)((1-\lambda)(2-\lambda)-1\cdot 1)-(1(2-\lambda)-0\cdot 1) = (2-\lambda)(2-\lambda-2\lambda+\lambda^2-1)-(2-\lambda) =$$

$$(2-\lambda)(\cancel{1}-3\lambda+\lambda^2\cancel{-1}) = \lambda(2-\lambda)(\lambda-3) \Rightarrow \lambda_1 = 3, \lambda_2 = 2, \lambda_3 = 0$$

The diagonal elements of Σ are therefore, from top left to bottom right: $\sqrt{3}, \sqrt{2}, 0$.

The normalized eigenvectors of $A^T A$ will be the column vectors of V:

$$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 3\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} 2x+y=3x \\ x+y+z=3y \\ y+2z=3z \end{cases} \Rightarrow \begin{cases} y=x \\ x+z=2y \\ y=z \end{cases} \Rightarrow \begin{cases} x=y \\ y=z \\ z=x \end{cases} \Rightarrow \begin{pmatrix} k \\ k \\ k \end{pmatrix} \quad \forall k \in \mathbb{R}$$

$$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} 2x+y=2x \\ x+y+z=2y \\ y+2z=2z \end{cases} \Rightarrow \begin{cases} y=0 \\ x+z=y \\ y=0 \end{cases} \Rightarrow \begin{cases} z=-x \\ y=0 \\ y=0 \end{cases} \Rightarrow \begin{pmatrix} k \\ 0 \\ -k \end{pmatrix} \quad \forall k \in \mathbb{R}$$

$$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{cases} 2x+y=0x \\ x+y+z=0y \\ y+2z=0z \end{cases} \Rightarrow \begin{cases} y=-2x \\ x+z=-y \\ y=-2z \end{cases} \Rightarrow \begin{cases} y=-2x \\ z=x \\ x=z \end{cases} \Rightarrow \begin{pmatrix} k \\ -2k \\ k \end{pmatrix} \quad \forall k \in \mathbb{R}$$

Out of all eigenvectors, it is necessary to pick the ones whose norm is equal to 1:

$$\parallel e_1 \parallel = \sqrt{\langle \begin{pmatrix} k \\ k \\ k \end{pmatrix}, \begin{pmatrix} k \\ k \\ k \end{pmatrix} \rangle} = \sqrt{k\cdot k + k\cdot k + k\cdot k} = \sqrt{3k^2} = \sqrt{3}\ |k| = 1 \Rightarrow k = \pm\frac{1}{\sqrt{3}}$$

$$\parallel e_2 \parallel = \sqrt{\langle \begin{pmatrix} k \\ 0 \\ -k \end{pmatrix}, \begin{pmatrix} k \\ 0 \\ -k \end{pmatrix} \rangle} = \sqrt{k\cdot k + 0\cdot 0 + (-k)\cdot(-k)} = \sqrt{2k^2} = \sqrt{2}\ |k| = 1 \Rightarrow k = \pm\frac{1}{\sqrt{2}}$$

$$\parallel e_3 \parallel = \sqrt{\langle \begin{pmatrix} k \\ -k \\ k \end{pmatrix}, \begin{pmatrix} k \\ -k \\ k \end{pmatrix} \rangle} = \sqrt{k\cdot k + (-2k)\cdot(-2k) + k\cdot k} = \sqrt{6k^2} = \sqrt{6}\ |k| = 1 \Rightarrow k = \pm\frac{1}{\sqrt{6}}$$

Out of the two choices for the sign, the positive one is taken for the sake of simplicity.

$A^T A$ has three eigenvalues, and only one of those is 0. Since the dimension of V is 3, the dimension of U is necessarely 2 and both eigenvectors can be computed directly from the columns of V.

$$\underline{u_1} = \frac{A\underline{e_1}}{\sigma_1} = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} = \frac{1}{\sqrt{3}}\begin{pmatrix} 1\cdot\frac{1}{\sqrt{3}} + 0\cdot\frac{1}{\sqrt{3}} - 1\cdot\frac{1}{\sqrt{3}} \\ 1\cdot\frac{1}{\sqrt{3}} + 1\cdot\frac{1}{\sqrt{3}} + 1\cdot\frac{1}{\sqrt{3}} \end{pmatrix} = \frac{1}{\sqrt{3}}\begin{pmatrix} 0 \\ \frac{3}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\underline{u_2} = \frac{A\underline{e_2}}{\sigma_2} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1\cdot\frac{1}{\sqrt{2}} + 0\cdot 0 - 1\cdot\left(-\frac{1}{\sqrt{2}}\right) \\ 1\cdot\frac{1}{\sqrt{2}} + 1\cdot 0 + 1\cdot\left(-\frac{1}{\sqrt{2}}\right) \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{2}{\sqrt{2}} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

As expected, said vectors are already normalized.

The SVD of A is therefore as follows:

$$A = U\Sigma V^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

□

4. Multivariable calculus

4.1. Partial derivatives

Let $f(x_1, x_2, ..., x_n) : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function of n arguments that returns a single value. For each of its i-th argument it is possible to define the **partial derivative** with respect to x_i as:

$$\frac{\partial f}{\partial x_i} = \lim_{h \rightarrow 0} \frac{f(x_1, x_2, ..., x_i + h, ..., x_n) - f(x_1, x_2, ..., x_i, ..., x_n)}{h}$$

That is, a partial derivative of a $f : \mathbb{R}^n \rightarrow \mathbb{R}$ function is "regular" derivative that is computed with respect to a single variable, treating all other variables as they were constants. Of course, as any regular derivative, a partial derivative may or may not exist.

A partial derivative, as a "regular" derivative, describes the rate of change of the function. The difference is that a function of n variables has n distinct directions, and its i-th partial derivative describes the rate of change of the function along the i-th axis.

Exercise 4.1.1: Compute both partial derivatives of the function $f(x, y) = \sin\left(\frac{x}{1+y}\right)$.

Solution:

$$\frac{\partial f}{\partial x} \sin\left(\frac{x}{1+y}\right) = \cos\left(\frac{x}{1+y}\right)\left(\frac{\partial f}{\partial x}\left(\frac{x}{1+y}\right)\right) = \cos\left(\frac{x}{1+y}\right)\left(\frac{1}{1+y}\right)\left(\frac{\partial f}{\partial x}(x)\right) = \frac{\cos\left(\frac{x}{1+y}\right)}{1+y}$$

$$\frac{\partial f}{\partial y} \sin\left(\frac{x}{1+y}\right) = \cos\left(\frac{x}{1+y}\right)\left(\frac{\partial f}{\partial y}\left(\frac{x}{1+y}\right)\right) = \cos\left(\frac{x}{1+y}\right)x\left(\frac{\partial f}{\partial y}\left(\frac{1}{1+y}\right)\right) = \frac{-x\cos\left(\frac{x}{1+y}\right)}{(1+y)^2}$$

□

It is possible to compute a partial derivative more than one time. That is to say, a partial derivative can be computed with respect to the result of applying a partial derivative.

For example, to denote that to the function f is first applied a partial derivative with respect to the variable x_i and then, to the result, is applied a partial derivative with respect to another variable x_j, the notation $\partial^2 f / \partial x_j \partial x_i$ is used. Notice how, in accord to the way function composition works, the order of derivation is from right to left.

Said notation can be extended to the case of computing a partial derivative for k times. When a partial derivative is computed with respect to the same variable more than once, it is possible to use the shortand notation $\partial^k f / \partial x_i^k$, meaning that a partial derivative of f is taken with respect to x_i, to which a partial derivative with respect to x_i is applied, to which a partial derivative with respect to x_i is applied, ecc...

Exercise 4.1.2: Find all second partial derivatives of $f(x, y) = x^3 + x^2 y^3 - 2y^2$.

Proof: The two first partial derivatives are:

$$\frac{\partial f}{\partial x}(x^3 + x^2y^3 - 2y^2) = 3x^2 + 2xy^3 \qquad \frac{\partial f}{\partial y}(x^3 + x^2y^3 - 2y^2) = 3y^2x^2 - 4y$$

From those, it is possible to compute four second partial derivatives:

$$\frac{\partial^2 f}{\partial^2 x}(3x^2 + 2xy^3) = 6x + 2y^3 \qquad \frac{\partial^2 f}{\partial x \partial y}(3y^2x^2) - 4y = 6xy^2$$

$$\frac{\partial^2 f}{\partial y \partial x}(3x^2 + 2xy^3) = 6xy^2 \qquad \frac{\partial^2 f}{\partial^2 y}(3y^2x^2 - 4y) = 6x^2y - 4$$

□

Notice how in Exercise 4.1.2, taking the partial derivative of the function first with respect to x and then with respect to y is the same as taking the derivative with respect to y and then to x. This isn't always the case, and instead happens only when the function satisfies certain conditions.

Theorem 4.1.1 (Schwartz's theorem): Let f be a function defined as $f(x_1, ..., x_n) : A \subseteq \mathbb{R}^n \to \mathbb{R}$, and let $\boldsymbol{p} = (p_1, ..., p_n) \in \mathbb{R}^n$ be a point. If some neighborhood of $\boldsymbol{p}$ is contained in A and f has continuous partial derivatives in said neighborhood, then, for any $i, j \in \{1, ..., n\}$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} f(\boldsymbol{p}) = \frac{\partial^2 f}{\partial x_j \partial x_i} f(\boldsymbol{p})$$

Exercise 4.1.3: Given $f = \sin(3x + yz)$, compute $\partial^4 f / \partial z \partial y \partial^2 x$.

Solution:

$$\frac{\partial f}{\partial x}(\sin(3x + yz)) = \cos(3x + yz)\frac{\partial f}{\partial x}(3x + yz) = 3\cos(3x + yz)$$

$$\frac{\partial f}{\partial x}(3\cos(3x + yz)) = -3\sin(3x + yz)\frac{\partial f}{\partial x}(3x + yz) = -9\sin(3x + yz)$$

$$\frac{\partial f}{\partial y}(-9\sin(3x + yz)) = -9\cos(3x + yz)\frac{\partial f}{\partial y}(3x + yz) = -9z\cos(3x + yz)$$

$$\frac{\partial f}{\partial z}(-9z\cos(3x + yz)) = -9\cos(3x + yz) - 9z\frac{\partial f}{\partial z}(\cos(3x + yz)) = 9yz\sin(3x + yz) - 9\cos(3x + yz)$$

Summing up:

$$\frac{\partial^4 f}{\partial z \partial y \partial^2 x}(\sin(3x + yz)) = 9yz\sin(3x + yz) - 9\cos(3x + yz)$$

□

As stated, a partial derivative describes the rate of change of the function along the i-th axis in the n-dimensional plane. Since each axis is described by a unit vector, a derivative along the i-th axis can be conceived as the rate of change of the function along the direction described by the i-th unit vector. Since any direction can be described by a vector, it is possible to compute a derivative of a function along any arbitrary direction, not just the ones described by the n unit vectors.

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be a function, $\boldsymbol{v} \in \mathbb{R}^n$ a non-null vector and $\boldsymbol{c}$ a point in A. f is said to have **directional derivative** along $\boldsymbol{v}$ in $\boldsymbol{c}$, denoted as $D_{\boldsymbol{v}} f(\boldsymbol{c})$ if the following limit exists:

$$D_{\boldsymbol{v}} f(\boldsymbol{c}) = \lim_{h \to 0} \frac{f(\boldsymbol{c} + h\boldsymbol{v}) - f(\boldsymbol{c})}{h} = \lim_{h \to 0} \frac{f(c_1 + hv_1, c_2 + hv_2, ..., c_n + hv_n) - f(c_1, c_2, ..., c_n)}{h}$$

If $\boldsymbol{v}$ is the i-th unit vector, the directional derivative is just a partial derivative with respect to the i-th variable.

Let $f(x_1, ..., x_n) : A \subset \mathbb{R}^n \to \mathbb{R}$ be a function and let $\boldsymbol{p}$ be a point in A. The **gradient** of f in $\boldsymbol{p}$, denoted as $\nabla_f(\boldsymbol{p})$, is a column vector whose components are the first partial derivatives of f, arranged from the first to the n-th:

$$\nabla_f(\boldsymbol{p}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T = \frac{\partial f}{\partial x_1}\hat{\imath}_1 + \frac{\partial f}{\partial x_2}\hat{\imath}_n + ... + \frac{\partial f}{\partial x_n}\hat{\imath}_n$$

The gradient of a (n-valued) scalar function is actually a special case of a more generic matrix of a (n -valued) vectorial function, called **Jacobian matrix**:

$$\boldsymbol{J}_f(\boldsymbol{p}) = \begin{bmatrix} \frac{\partial \boldsymbol{f}}{\partial x_1} & \cdots & \frac{\partial \boldsymbol{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^T_{f_1}(\boldsymbol{p}) \\ \vdots \\ \nabla^T_{f_n}(\boldsymbol{p}) \end{bmatrix}$$

The gradient vector is strongly related to maxima and minima of a function.

Theorem 4.1.2: If a function has a local maximum or minimum in a point and all partial derivatives in said point exist, the gradient in said point is zero.

Any point whose gradient is zero is called a **critical point** or **stationary point**. Note that, while Theorem 4.1.2 guarantees that maxima and minima are also critical points, the converse is not necessarely true. In particular, a critical point that is neither a maximum or a minimum is called a **saddle point**.

Similarly to the gradient, the second partial derivatives of f can be arranged in a matrix, called **Hessian matrix** and denoted as H_f:

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial^2 x_n} \end{bmatrix}$$

The Hessian matrix is strongly related the natures of a critical point of a function.

Theorem 4.1.3 (Second derivative test for functions having more than one variable) : Consider the function $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$, twice (or more) differentiable. Let $\boldsymbol{p}$ be a critical point of f, and let $H_f(\boldsymbol{p})$ be the Hessian matrix of f at $\boldsymbol{p}$. Then:

- If $H_f(\boldsymbol{p})$ is positive definite, $\boldsymbol{p}$ is a local minimum;
- If $H_f(\boldsymbol{p})$ is negative definite, $\boldsymbol{p}$ is a local maximum;
- If $H_f(\boldsymbol{p})$ has at least a positive eigenvalue, a negative eigenvalue and no eigenvalue is zero, $\boldsymbol{p}$ is a saddle point;
- If none of the above is true, the test is inconclusive.