Mathematics for Natural Scientists II

Iosif Petrakis

Fakultät für Geowissenschaften, Ludwig-Maximilians-Universität München

Summer term 2019

Contents

CHAPTER 1

Linear spaces and linear maps

In this chapter we study the basic properties of the linear spaces–also called vector spaces–and of the linear maps between them. A linear space is a set endowed with a linear structure, and a linear map between linear spaces is a function between their corresponding sets that preserves their linear structure.

1.1. Linear spaces

DEFINITION 1.1.1. A *linear space*, or a *vector space*, over $\mathbb{R}$ is a structure $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$, where X is a set, $\mathbf{0} \in X$, and $+, \cdot$ are functions

$$+ : X \times X \to X, \qquad \cdot : \mathbb{R} \times X \to X$$
$$(x, y) \mapsto x + y, \qquad (a, x) \mapsto a \cdot x,$$

such that the following conditions are satisfied:

(LS_1) $\forall_{x,y,z \in X} \big((x + y) + z = x + (y + z)\big)$.

(LS_2) $\forall_{x \in X} \big(x + \mathbf{0} = \mathbf{0} + x = x\big)$.

(LS_3) $\forall_{x \in X} \exists_{y \in X} \big(x + y = \mathbf{0}\big)$.

(LS_4) $\forall_{x,y \in X} \big(x + y = y + x\big)$.

(LS_5) $\forall_{x,y \in X} \forall_{a \in \mathbb{R}} \big(a \cdot (x + y) = a \cdot x + a \cdot y\big)$.

(LS_6) $\forall_{x \in X} \forall_{a,b \in \mathbb{R}} \big((a + b) \cdot x = a \cdot x + b \cdot x\big)$.

(LS_7) $\forall_{x \in X} \forall_{a,b \in \mathbb{R}} \big((ab) \cdot x = a \cdot (b \cdot x)\big)$.

(LS_8) $\forall_{x \in X} \big(1 \cdot x = x\big)$.

For simplicity, we may write ax instead of $a \cdot x$. The triple $(+, \mathbf{0}, \cdot)$ is called the *signature* of the linear space $\mathcal{V}$. If, instead of $\mathbb{R}$, we consider any field[1] $\mathbb{F}$, the

[1] A *field* is a structure $(\mathbb{F}; +, \mathbf{0}, \cdot, \mathbf{1})$, where $\mathbb{F}$ is a set, $\mathbf{0}, \mathbf{1} \in \mathbb{F}$, $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$, and $\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ such that together with (LS_1) − (LS_4) the following conditions are satisfied:

$\forall_{x,y,z \in \mathbb{F}} \big(x \cdot (y \cdot z) = (x \cdot y) \cdot z\big)$.

$\forall_{x,y,z \in \mathbb{F}} \big(x \cdot (y + z) = x \cdot y + x \cdot z\big)$.

$\forall_{x,y \in \mathbb{F}} \big(x \cdot y = y \cdot x\big)$.

$\forall_{x \in \mathbb{F}} \big(\mathbf{1} \cdot x = x\big)$.

corresponding structure is called a *linear space over* $\mathbb{F}$. A linear space over $\mathbb{R}$ is also called a *real* linear space, and a linear space over the field of complex numbers $\mathbb{C}$ is called a *complex* linear space. If $\mathcal{V}$ is a linear space, the elements of X are traditionally called *vectors*. A linear space is called *non-trivial*, if it contains a vector x such that $x \neq \mathbf{0}$. Unless stated otherwise, *the linear spaces considered here are going to be real*. When the linear structure on X is clear from the context, we use for simplicity X to denote the vector space $\mathcal{V}$.

Recall that if X, Y are sets, then

$$X \times Y := \{(x, y) \mid x \in X \ \& \ y \in Y\},$$

and if $(x, y), (x', y') \in X \times Y$, then

$$(x, y) = (x', y') \Leftrightarrow (x = x' \ \& \ y = y'.$$

EXAMPLE 1.1.2. Let the structure $\mathcal{R}^n := (\mathbb{R}^n; +, \mathbf{0}, \cdot)$, where

$$\mathbb{R}^n := \{(x_1, \ldots, x_n) \mid x_1 \in \mathbb{R} \ \& \ \ldots \ \& \ x_n \in \mathbb{R}\},$$

$$(x_1, \ldots, x_n) = (y_1, \ldots, y_n) \Leftrightarrow x_1 = y_1 \ \& \ \ldots \ \& \ x_n = y_n,$$

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) := (x_1 + y_1, \ldots, x_n + y_n),$$

$$\mathbf{0} := (0, \ldots, 0),$$

$$a \cdot (x_1, \ldots, x_n) := (ax_1, \ldots, ax_n).$$

Clearly, $\mathcal{R}^n$ a linear space over $\mathbb{R}$, and, similarly, $\mathcal{Q}^n := (\mathbb{Q}^n; +, \mathbf{0}, \cdot)$ is linear space over $\mathbb{Q}$, and $\mathcal{C}^n := (\mathbb{C}^n; +, \mathbf{0}, \cdot)$ is a linear space over $\mathbb{C}$.

If $\mathbb{F}(X, Y)$ is the set of all functions from X to Y, and $f, g \in \mathbb{F}(X, Y)$, then

$$f = g \Leftrightarrow \forall_{x \in X}\big(f(x) = g(x)\big).$$

EXAMPLE 1.1.3. If X is a set, $\mathbb{F}(X)$ is the set of all functions $f : X \to \mathbb{R}$, and if we define the functions $f + g$, $\overline{0}^X$ and $a \cdot f$, where $a \in \mathbb{R}$, by

$$(f + g)(x) := f(x) + g(x),$$

$$\overline{0}^X(x) := 0,$$

$$(a \cdot f)(x) := af(x),$$

for every $x \in X$, then $\mathcal{F}(X) := (\mathbb{F}(X); +, \overline{0}^X, \cdot)$ is a linear space over $\mathbb{R}$.

$\forall_{x \in \mathbb{F}}\big(x \neq \mathbf{0} \Rightarrow \exists_{y \in \mathbb{F}}(x \cdot y = \mathbf{1})\big)$.

It is immediate to see that the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$ have a field structure. Actually, $\mathbb{Q}$ is a *subfield* of $\mathbb{R}$ and $\mathbb{R}$ is a subfield of $\mathbb{C}$ i.e., the field-signature $(+, \mathbf{0}, \cdot, \mathbf{1})$ of $\mathbb{Q}$ is inherited from the field-signature of $\mathbb{R}$, which, in turn, can be inherited from the field-signature of $\mathbb{C}$.

The Example 1.1.3 shows that a mathematical object can be viewed as a vector, although no immediate geometric intuition is associated with it. If

$$\boldsymbol{n} := \{0, 1, \ldots, n-1\}$$

though, an element of $\mathbb{R}^n$ can be identified with a function $f : \boldsymbol{n} \to \mathbb{R}$, and then the Example 1.1.2 is a special case of the Example 1.1.3. If $f, g \in \mathbb{F}(X)$ and $a \in \mathbb{R}$,

$$f \leq g \Leftrightarrow \forall_{x \in X}\big(f(x) \leq g(x)\big),$$

$$f \leq a :\Leftrightarrow f \leq \overline{a}^X \Leftrightarrow \forall_{x \in X}\big(f(x) \leq a)\big),$$

where $\overline{a}^X(x) := a$, for every $x \in X$.

Remark 1.1.4. Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, $a, b \in \mathbb{R}$, and $x, y, z, w \in X$. The following hold:

(*i*) If $z = w$ and $x = y$, then $z + x = w + y$.

(*ii*) If $x = y$ and $a = b$, then $a \cdot x = b \cdot y$.

(*iii*) If $x + y = x + z = \mathbf{0}$, then $y = z$.

(*iv*) $0 \cdot x = \mathbf{0}$.

(*v*) $(-1) \cdot x = -x$, where, because of case (*iii*), $-x$ is the unique element y of X in condition (LS_3) such that $x + y = \mathbf{0}$.

(*vi*) If $x \neq \mathbf{0}$ and $a \cdot x = \mathbf{0}$, then $a = 0$.

Proof. Exercise. □

Definition 1.1.5. Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, and $Y \subseteq X$ such that the following conditions are satisfied:

(*i*) $\forall_{y, y' \in Y}\big(y + y' \in Y\big)$,

(*ii*) $\mathbf{0} \in Y$,

(*iii*) $\forall_{y \in Y}\forall_{a \in \mathbb{R}}\big(a \cdot y \in Y\big)$.

Then the structure

$$\mathcal{V}_{|Y} := (Y, +_{|Y \times Y}, \mathbf{0}, \cdot_{|\mathbb{R} \times Y}),$$

where $+_{|Y \times Y}$ is the restriction of $+$ to $Y \times Y$ and $\cdot_{|\mathbb{R} \times Y}$ is the restrictions of $\cdot$ to $\mathbb{R} \times Y$, is called a *linear subspace* of $\mathcal{V}$, or, simpler, a *subspace* of $\mathcal{V}$. We write $\mathcal{V}_{|Y} \preceq \mathcal{V}$ to denote that $\mathcal{V}_{|Y}$ is a linear subspace of $\mathcal{V}$, although, for simplicity, we refer to a linear subspace $\mathcal{V}_{|Y}$ mentioning only the set Y, and we write $Y \preceq X$. We denote by $\mathtt{Sub}(\mathcal{V})$ the set of all subspaces of $\mathcal{V}$.

Clearly, $\{0\}$ and X are linear subspaces of X.

Example 1.1.6. If $\mathbb{F}^*(X)$ is the set of all bounded functions in $\mathbb{F}(X)$ i.e.,

$$\mathbb{F}^*(X) = \big\{f \in \mathbb{F}(X) \mid \exists_{M > 0}\forall_{x \in X}\big(|f(x)| \leq M\big)\big\},$$

then $\mathbb{F}^*(X)$ is a linear subspace of $\mathbb{F}(X)$ (see Example 1.1.3). To see this let $f, g \in \mathbb{F}(X)$ and $M_f > 0, M_g > 0$, such that $|f| \leq M_f$ and $|g| \leq M_g$. Then

$|f+g| \leq M_f + M_g$ and $|af| \leq (1+|a|)M_f$, where $M_f + M_g > 0$ and $(1+|a|)M_f > 0$. Recall that $|f| \in \mathbb{F}(X)$ is defined by $|f|(x) := |f(x)|$, for every $x \in X$.

EXAMPLE 1.1.7. If $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ is a linear space, $n \geq 1$, and $x_1, \ldots, x_n \in X$, the set

$$\langle\{x_1, \ldots, x_n\}\rangle := \big\{a_1 \cdot x_1 + \ldots + a_n \cdot x_n \mid a_1 \in \mathbb{R} \ \& \ \ldots \ \& \ a_n \in \mathbb{R}\big\}$$

is a linear subspace of $\mathcal{V}$. We call an element

$$\sum_{i=1}^{n} a_i x_i := a_1 \cdot x_1 + \ldots + a_n x_n$$

of $\langle\{x_1, \ldots, x_n\}\rangle$ a *linear combination* of $x_1, \ldots, x_n$, and the space $\langle\{x_1, \ldots, x_n\}\rangle$ the *linear span* of $x_1, \ldots, x_n$. We may write $\langle x_1, \ldots, x_n\rangle$ instead of $\langle\{x_1, \ldots, x_n\}\rangle$.

If $e_1 := (1,0), e_2 := (0,1), (x,y) \in \mathbb{R}^2$, we get $\mathbb{R}^2 = \langle e_1, e_2\rangle$, since

$$(x.y) := x(1,0) + (0,1)y := xe_1 + ye_2.$$

PROPOSITION 1.1.8. *Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, $Y \subseteq X$, and let $U, V \preceq X$.*

(i) If $U + V := \{u + v \mid u \in U \ \& \ v \in V\}$, then $U + V \preceq X$.

(ii) If $U \cap V := \{x \in X \mid x \in U \ \& \ x \in V\}$, then $U \cap V \preceq X$.

(iii) If we define

$$\langle Y\rangle := \bigcap \big\{U \preceq X \mid Y \subseteq U\big\} := \big\{x \in X \mid \forall_{U \preceq X}(Y \subseteq U \Rightarrow x \in U)\big\},$$

then $\langle Y\rangle$ is well-defined (i.e., the set $\{U \preceq X \mid Y \subseteq Y\}$ is non-empty) and it is the least linear subspace of X that includes Y.

(iv) If $Y \neq \emptyset$, then

$$\langle Y\rangle = \left\{\sum_{i=1}^{n} a_i y_i \mid n \geq 1 \ \& \ \forall_{i \in \{1, \ldots, n\}}\big(a_i \in \mathbb{R} \ \& \ y_i \in Y\big)\right\}.$$

PROOF. Exercise. □

Since $\emptyset \subseteq \{\mathbf{0}\}$, we have that $\langle\emptyset\rangle = \{\mathbf{0}\}$. The subspace $U + V$ of $\mathcal{X}$ is called the *sum* of U and V. By Proposition 1.1.8 the linear span $\langle x_1, \ldots, x_n\rangle$ of $x_1, \ldots, x_n \in X$ is the least linear space containing $x_1, \ldots, x_n$. If $X = \langle Y\rangle$, we say that Y *generates* the linear space $\mathcal{V}$ (or X), and the elements of Y are called *generators* of $\mathcal{V}$.

1.2. Finite-dimensional linear spaces

DEFINITION 1.2.1. Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, $n \geq 1$, and let $x_1, \ldots, x_n \in X$. We say that the vectors $x_1, \ldots, x_n$ are *linearly dependent*, or that their set $\{y_1, \ldots, y_n\}$ is a linearly dependent subset of X, if

$$\exists_{a_1,\ldots,a_n\in\mathbb{R}}\bigg(\exists_{i\in\{1,\ldots,n\}}\big(a_i \neq 0\big) \ \& \ \sum_{i=1}^{n} a_i x_i = \mathbf{0}\bigg).$$

We say that $x_1, \ldots, x_n$ are *linearly independent*, if they are *not* linearly dependent. A subset Y of X is called *linearly dependent*, if

$$\exists_{n\geq 1}\exists_{y_1,\ldots,y_n\in Y}\bigg(\{y_1, \ldots, y_n\} \text{ is linearly dependent}\bigg),$$

while it is called *linearly independent*, if it is *not* a linearly dependent subset of X.

If $x_1, \ldots, x_n$ are linearly dependent, $a_1x_1 + \ldots + a_nx_n = \mathbf{0}$, and $a_i \neq 0$, then

$$x_i = \bigg(\frac{-a_1}{a_i}\bigg)x_1 + \ldots + \bigg(\frac{-a_{i-1}}{a_i}\bigg)x_{i-1} + \bigg(\frac{-a_{i+1}}{a_i}\bigg)x_{i+1} + \ldots + \bigg(\frac{-a_n}{a_i}\bigg)x_n$$

i.e., x_i is a linear combination of $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$.

Remark 1.2.2. Let X be a linear space and $Y, Z \subseteq X$.

(i) If $x_1, \ldots, x_n \in X$, then $x_1, \ldots, x_n$ are linearly independent if and only if

$$\forall_{a_1,\ldots,a_n\in\mathbb{R}}\bigg(\sum_{i=1}^{n} a_i x_i = \mathbf{0} \Rightarrow \forall_{i\in\{1,\ldots,n\}}(a_i = 0)\bigg).$$

(ii) Y is linearly independent if and only if

$$\forall_{n\geq 1}\forall_{y_1,\ldots,y_n\in Y}\bigg(\{y_1, \ldots, y_n\} \text{ is linearly independent}\bigg).$$

(iii) $\{\mathbf{0}\}$ and X are linearly dependent subsets of X.

(iv) If $x \neq \mathbf{0}$, then $\{x\}$ is a linearly independent subset of X.

(v) The empty set $\emptyset$ is a linearly independent subset of X.

(vi) If Y is linearly dependent and $Y \subseteq Z$, then Z is linearly dependent.

(vii) If Y is linearly independent and $Z \subseteq Y$, then Z is linearly independent.

Proof. (i) and (ii) By negating the corresponding defining formulas.
(iii) $1 \cdot \mathbf{0} = \mathbf{0}$, and $\{\mathbf{0}\}$ is a linearly dependent subset of X.
(iv) It follows immediately by Remark 1.1.4(vi).
(v) If we suppose that $\emptyset$ is a linearly dependent subset of X i.e.,

$$\exists_{n\geq 1}\exists_{y_1,\ldots,y_n}\bigg(y_1 \in \emptyset \ \& \ \ldots \ \& \ y_n \in \emptyset \ \& \ \{y_1, \ldots, y_n\} \text{ is linearly dependent}\bigg),$$

it is immediate that we get a contradiction from it.
(vi) and (vii) are immediate to show. □

EXAMPLE 1.2.3. The following n-vectors in $\mathbb{R}^n$

$$e_1 := (1, 0, \ldots, 0), \quad e_2 := (0, 1, 0, \ldots, 0), \ \ldots, \ e_n := (0, \ldots, 0, 1)$$

are linearly independent, since for every $a_1, \ldots, a_n \in \mathbb{R}$ we have that

$$\sum_{i=1}^{n} a_i e_i = 0 \Leftrightarrow (a_1, \ldots, a_n) = \mathbf{0} \Leftrightarrow a_1 = \ldots = a_n = 0.$$

EXAMPLE 1.2.4. For every $n \geq 1$, the following n-vectors in $\mathbb{F}(\mathbb{R})$

$$f_1(t) := e^t, \ \ldots, \ f_n(t) := e^{nt}$$

are linearly independent (Exercise).

REMARK 1.2.5. Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, $n \geq 1$, and $x_1, \ldots, x_n \in X$ linearly independent. If $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$, then

$$\sum_{i=1}^{n} a_i x_i = \sum_{i=1}^{n} b_i x_i \Rightarrow \big(a_1 = b_1 \ \& \ \ldots \ \& \ a_n = b_n\big).$$

Moreover, $x_i \neq \mathbf{0}$, for every $i \in \{1, \ldots, n\}$.

PROOF. It follows from the Definition 1.2.1 and the equivalence

$$\sum_{i=1}^{n} a_i x_i = \sum_{i=1}^{n} b_i x_i \Leftrightarrow \sum_{i=1}^{n} (a_i - b_i) x_i = \mathbf{0}.$$

If there is $i \in \{1, \ldots, n\}$ such that $x_i = 0$, then $0x_1 + 0x_{i-1} + 1x_i + 0x_{i+1} + \ldots + 0x_n = \mathbf{0}$, which is impossible. □

DEFINITION 1.2.6. If $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ is linear space, a subset B of X is called a *basis* of $\mathcal{V}$ (or, for simplicity a basis of X), if B is linearly independent, and $\langle B \rangle = X$. If $\mathcal{V}$ has a finite basis B, it is called a *finite-dimensional* linear space, while if it has an infinite basis, it is called *infinite-dimensional*.

Consequently, the subspace $\{\mathbf{0}\}$ has as a basis the empty set.

EXAMPLE 1.2.7. The set $E_n := \{e_1, \ldots, e_n\}$ of the linearly independent elements in $\mathbb{R}^n$ that were defined in the Example 1.2.3 is the *standard* basis of $\mathbb{R}^n$. Hence, $\mathcal{R}^n$ is finite-dimensional. It is easy to see that $\mathbb{R}^n$ has more than one bases. E.g., $B := \{(1, 1), (-1, 2)\}$ is another basis of $\mathbb{R}^2$.

EXAMPLE 1.2.8. Since the set $E := \{e^{nt} \mid n \geq 1\}$ is a linearly independent subset of $\mathbb{F}(\mathbb{R})$, the set E is a basis of the linear subspace $\langle E \rangle$ of $\mathbb{F}(\mathbb{R})$, and $\langle E \rangle$ is infinite-dimensional.

COROLLARY 1.2.9. *Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, and $x \in X$. If $B := \{v_1, \ldots, v_n\}$ is a basis of $\mathcal{V}$, there are unique $a_1, \ldots, a_n \in \mathbb{R}$ such that*

$$x = \sum_{i=1}^{n} a_i v_i.$$

PROOF. It follows by the definition of a basis and the Remark 1.2.5. □

These unique $a_1, \ldots, a_n \in \mathbb{R}$ are called the *coordinates* of x with respect to B.

DEFINITION 1.2.10. Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, $\{v_1, \ldots, v_n\} \subseteq X$ and $m \leq n$. The set $\{v_1, \ldots v_m\}$ is a *maximal subset of linearly independent elements of* X, if it is a linearly independent subset of X, and for every $k \in \mathbb{N}$, such that $m < k \leq n$, the set $\{v_1, \ldots, v_m, v_k\}$ is a linearly dependent subset of X.

THEOREM 1.2.11 (Finite basis-criterion I). *Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space, $n \geq 1$, and $\{v_1, \ldots, v_n\} \subseteq X$ such that $X = \langle\{v_1, \ldots, v_n\}\rangle$. If $\{v_1, \ldots, v_r\}$ is a maximal subset of linearly independent elements of X, where $1 \leq r \leq n$, then $\{v_1, \ldots, v_r\}$ is a basis of $\mathcal{V}$.*

PROOF. If $r = n$, then $\{v_1, \ldots, v_r\}$ is a linearly independent subset generating X i.e., it is a basis of $\mathcal{V}$. If $r < n$, by the maximality of $\{v_1, \ldots, v_r\}$ the sets

$$\{v_1, \ldots, v_r, v_{r+1}\},\ \{v_1, \ldots, v_r, v_{r+2}\},\ \ldots,\ \{v_1, \ldots, v_r, v_n\}$$

are linearly dependent subsets of X. We show that

$$v_{r+1} \in \langle\{v_1, \ldots, v_r\}\rangle \ \&\ v_{r+2} \in \langle\{v_1, \ldots, v_r\}\rangle \ \&\ \ldots\ \&\ v_n \in \langle\{v_1, \ldots, v_r\}\rangle.$$

We show this only for v_{r+1}, and for $v_{r+2}, \ldots, v_n$ we proceed similarly. Since $\{v_1, \ldots, v_n, v_{r+1}\}$ is linearly dependent, there are $a_1, \ldots, a_r, a_{r+1} \in \mathbb{R}$ such that

$$a_1 v_1 + \ldots + a_r v_r + a_{r+1} v_{r+1} = \mathbf{0},$$

and not all of them are equal to 0. If $a_{r+1} = 0$, then $a_1 v_1 + \ldots + a_r v_r = \mathbf{0}$, hence $a_1 = \ldots = a_r = a_{r+1} = 0$, which is a contradiction. Hence $a_{r+1} \neq 0$, and hence v_{r+1} can be written as a linear combination of $v_1, \ldots, v_r$. Since an element x of X is a linear combination of $v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$ and $v_{r+1}, \ldots, v_n$ are linear combinations of $v_1, \ldots, v_r$, then x is a linear combination of $v_1, \ldots, v_r$. □

Next we show that we can replace any number of elements of a finite basis by an equal number of any linearly independent vectors.

LEMMA 1.2.12 (Exchange lemma (Steinitz)). *Let $n, m \geq 1$, $\{v_1, \ldots v_n\}$ a basis of the linear space $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$, and let $w_1, \ldots, w_m \in X$ be linearly independent.*
(i) If $m < n$, there are $u_{m+1}, \ldots, u_n \in \{v_1, \ldots v_n\}$ such that

$$\langle\{w_1, \ldots, w_m, u_{m+1}, \ldots, u_n\}\rangle = X.$$

(ii) If $m = n$, then $\langle\{w_1, \ldots, w_n\}\rangle = X$.

PROOF. (i) By the definition of a basis there are $a_1, \ldots a_n \in \mathbb{R}$ such that

$$w_1 = a_1 v_1 + \ldots + a_n v_n.$$

Since by Remark 1.2.5 $w_1 \neq \mathbf{0}$, there is some $a_i \neq 0$, where $i \in \{1, \ldots, n\}$. Without loss of generality we can take $i = 1$ (if $a_1 = 0$, we can re-enumerate the elements of the set $\{v_1, \ldots v_n\}$ so that the first coefficient in the writing of w_1 as a linear combination of the elements of the set $\{v_1, \ldots v_n\}$ is non-zero). Hence

$$a_1 v_1 = w_1 - \sum_{i=2}^{n} a_i v_i \Leftrightarrow v_1 = \frac{1}{a_1} w_1 - \sum_{i=2}^{n} \frac{a_i}{a_1} v_i,$$

and consequently

$$v_1 \in \Big\langle \big\{ w_1, v_2, \ldots, v_n \big\} \Big\rangle,$$

and

$$\Big\langle \big\{ w_1, v_2, \ldots, v_n \big\} \Big\rangle = X.$$

By the inductive hypothesis, if $1 \leq r < m$ we get (possibly after a re-enumeration of the set $\{v_1, \ldots v_n\}$)

$$\Big\langle \{ w_1, \ldots, w_r, v_{r+1}, \ldots, v_n \} \Big\rangle = X.$$

Hence,

$$w_{r+1} = b_1 w_1 + \ldots + b_r w_r + c_{r+1} v_{r+1} + \ldots + c_n v_n.$$

Not all the terms $c_{r+1}, \ldots, c_n$ are equal to 0, since then w_{r+1} would be a linear combination of $w_1, \ldots, w_r$, something that contradicts the hypothesis of linear independence of the vectors $w_1, \ldots, w_m$. Without loss of generality, let $c_{r+1} \neq 0$, hence

$$c_{r+1} v_{r+1} = w_{r+1} - \big[\sum_{i=1}^{r} b_i w_i + \sum_{j=r+2}^{n} c_j v_j \big] \Leftrightarrow$$

$$v_{r+1} = \frac{1}{c_{r+1}} w_{r+1} - \sum_{i=1}^{r} \frac{b_i}{c_{r+1}} w_i - \sum_{j=r+2}^{n} \frac{c_j}{c_{r+1}} v_j,$$

and consequently

$$v_{r+1} \in \Big\langle \big\{ w_1, \ldots, w_r, w_{r+1}, v_{r+2}, \ldots, v_n \big\} \Big\rangle,$$

and

$$\Big\langle \{ w_1, \ldots, w_r, w_{r+1}, v_{r+2}, \ldots, v_n \} \Big\rangle = X.$$

After m-number of steps, we get $\langle \{ w_1, \ldots, w_m, u_{m+1}, \ldots, u_n \} \rangle = X$.
(ii) It follows immediately by (i). □

Theorem 1.2.13. *Let $0 < n < m$, and let $\{v_1, \ldots v_n\}$ be a basis of the linear space $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$. If $w_1, \ldots, w_m \in X$, then $w_1, \ldots, w_m$ are linearly dependent.*

Proof. Suppose that the vectors $w_1, \ldots, w_m$ are linearly independent. Since then the vectors $w_1, \ldots, w_n$ are also linearly independent, by the Lemma 1.2.12(ii) we have that $w_1, \ldots, w_n$ is a basis of X. By the hypothesis of linear independence we have that $w_{n+1} \neq \mathbf{0}$, hence it is also a non-trivial linear combination of $w_1, \ldots, w_n$. By this contradiction we conclude that the vectors $w_1, \ldots, w_m$ are linearly dependent. □

COROLLARY 1.2.14. *If B_1, B_2 are finite bases of a linear space $\mathcal{V}$, then B_1 and B_2 have the same number of elements.*

PROOF. If $\mathcal{V}$ is a trivial linear space, then the two bases are equal to the empty set, and $|B_1| = |B_2| = 0$, where $|I|$ denotes the number of elements, or the *cardinality*, of a set I. Let $\mathcal{V}$ be non-trivial, and let $n, m \geq 1$ such that $|B_1| = n$ and $|B_2| = m$. If $n < m$, then by the Theorem 1.2.13 we have that B_2 is linearly dependent, which is a contradiction. Hence $n \geq m$. Similarly we get $m \geq n$. □

Because of the Corollary 1.2.14 the following concept is well-defined.

DEFINITION 1.2.15. If $n \geq 1$ and $\{v_1, \ldots, v_n\}$ is a basis of a linear space $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$, we call $\mathcal{V}$ an *n-dimensional space*, and we write $\dim(X) := n$. A trivial linear space has dimension 0.

Clearly, $\dim(\mathbb{R}^n) := n$.

COROLLARY 1.2.16. *Let $n \geq 1$, and let $v_1, \ldots, v_n$ be linearly independent elements of a linear space X.*

(i) (Finite basis-criterion II) If their set $M := \{v_1, \ldots, v_n\}$ is a maximal set of linearly independent elements of X i.e., for every $x \in X$ we have that

$$x, v_1, \ldots, v_n$$

are linearly dependent elements of X, then M is a basis of X.

(ii) If $\dim(X) = n$, and $w_1, \ldots, w_n$ are linearly independent elements of X, then $B := \{w_1, \ldots, w_n\}$ is a basis of X.

(iii) If Y is a subspace of X with $\dim(Y) = \dim(X) = n$, then $Y = X$.

(iv) If $\dim(X) = n$, $1 \leq r < n$, and $w_1, \ldots, w_r$ are linearly independent elements of X, then there are elements $v_{r+1}, \ldots, v_n$ of X such that the set

$$\{w_1, \ldots, w_r, v_{r+1}, \ldots, v_n\}$$

is a basis of X.

PROOF. Exercise. □

Next we show that the existence of a basis of a linear space X implies the existence of a basis of any subspace of X.

COROLLARY 1.2.17. *Let $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$ be a linear space with $\dim(X) = n$. If $Y \preceq X$, then Y has a basis and $\dim(Y) \leq \dim(X)$.*

PROOF. If $Y := \{\mathbf{0}\}$, then $\emptyset$ is a basis of Y and $\dim(Y) = 0 \leq \dim(X)$. If Y is non-trivial, then either $Y = X$, or Y is a proper subspace of X. In the first case what we want to show follows trivially. If Y is a proper, non-trivial subspace of X, then there is some $y_1 \in Y$ such that $y_1 \neq \mathbf{0}$, and by the Remark 1.1.4(vi) $M_1 := \{y_1\}$ is linearly independent. By the principle of the excluded middle[2]

[2]This is the logical principle $P \vee \neg P$, where P is any well-formed formula.

(PEM), we have that M_1 is either a maximal set of linearly dependent elements of Y, hence by the Corollary 1.2.16(i) it is also a basis of Y, and hence $\dim(Y) = 1$, or there is $y_2 \in Y$ such that $M_2 := \{y_1, y_2\}$ is linearly independent. Proceeding similarly, we can repeat the same argument at most $(n-1)$ number of times, in order to reach the required conclusion. $\square$

Next we write the expression that abbreviates the *unique existence* of an element of a set X satisfying a formula $\phi(x)$:

$$\exists_{!x\in X}\big(\phi(x)\big) :\Leftrightarrow \exists_{x\in X}\bigg(\phi(x) \ \& \ \forall_{y\in X}\big(\phi(y) \Rightarrow y = x\big)\bigg).$$

PROPOSITION 1.2.18. *If X is a linear space, and $Y, Z \preceq X$, such that*

$$\forall_{x\in X}\exists_{!y\in Y}\exists_{!z\in Z}\big(x = y + z\big),$$

we write $X := Y \oplus Z$. The following are equivalent:
(i) $X = Y \oplus Z$.
(ii) $X = Y + Z$ and $Y \cap Z = \{\mathbf{0}\}$.

PROOF. Exercise. $\square$

PROPOSITION 1.2.19. *Let X be a linear space, $n \in \mathbb{N}$, and $\dim(X) = n$.*
(i) If $Y \preceq X$, there is some $Z \preceq X$ such that $X = Y \oplus Z$.
(ii) If $Y, Z \preceq X$ such that $X = Y \oplus Z$, then $\dim(X) = \dim(Y) + \dim(Z)$.

PROOF. Exercise. $\square$

Next we give a condition under which, a linearly independent subset of a linear space X can be extended to a larger linearly independent subset of X.

LEMMA 1.2.20. *Let Y be a linearly independent subset of a linear space X, and $x_0 \in X$. If $x_0 \notin \langle Y\rangle$, then $Y \cup \{x_0\}$ is a linearly independent subset of X.*

PROOF. Exercise. $\square$

1.3. Linear maps

DEFINITION 1.3.1. If X and Y are linear spaces, a function $f : X \to Y$ is called *linear*, or a *linear map*, if it satisfies the following conditions:
(i) $\forall_{x,x'\in X}\big(f(x + x') = f(x) + f(x')\big)$.
(ii) $\forall_{x\in X}\forall_{a\in\mathbb{R}}\big(f(a \cdot x) = a \cdot f(x)\big)$.
Moreover, we define the following sets:

$$\mathcal{L}(X, Y) := \{f : X \to Y \mid f \text{ is linear}\},$$

$$\mathcal{L}(X) := \mathcal{L}(X, X) := \{f : X \to X \mid f \text{ is linear}\},$$

$$X^* := \mathcal{L}(X, \mathbb{R}) := \{f : X \to \mathbb{R} \mid f \text{ is linear}\}.$$

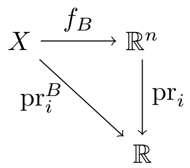
The elements of $\mathcal{L}(X)$ are called *operators* on X, or *linear transformations* on X, while X^* is called the *dual space* of X.

EXAMPLE 1.3.2. If X is a linear space with $\dim(X) = n$, for some $n \geq 1$, and $B := \{v_1, \ldots, v_n\}$ is a fixed basis of X, then the function $f_B : X \to \mathbb{R}^n$, defined by

$$f_B(x) := (a_1, \ldots, a_n), \qquad x = \sum_{i=1}^{n} a_i v_i,$$

is a linear map. Moreover, if $i \in \{1, \ldots, n\}$, the function $\mathrm{pr}_i^B : X \to \mathbb{R}$, defined by

$$\mathrm{pr}_i^B(x) := a_i, \qquad x = \sum_{i=1}^{n} a_i v_i,$$



is a linear map. If $n > m \geq 1$, the function $g : \mathbb{R}^n \to \mathbb{R}^m$ is linear, where

$$g(a_1, \ldots, a_m, a_{m+1}, \ldots, a_n) := (a_1, \ldots, a_m).$$

REMARK 1.3.3. The set $\mathcal{L}(X, Y)$ is equipped with the following linear structure

$$(f + g)(x) := f(x) + g(x), \qquad x \in X,$$

$$(a \cdot f)(x) := a \cdot f(x), \qquad a \in \mathbb{R}, \quad x \in X,$$

$$\mathbf{0}(x) := \mathbf{0}, \qquad x \in X.$$

PROOF. Exercise. □

DEFINITION 1.3.4. If $m, n \geq 1$, an array of real numbers

$$A := \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} =: [a_{ij}].$$

is called a *matrix* of *m-rows* and *n-columns.* If $1 \leq i \leq m$, the i-th row of A is the array

$$A_i := \begin{bmatrix} a_{i1} & \ldots & a_{in} \end{bmatrix} := [a_{ij}]_i,$$

and if $1 \leq j \leq n$, the j-th column of A is the array

$$A^j := \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} := [a_{ij}]^j.$$

The set of $m \times n$-matrices is denoted by $M_{m,n}(\mathbb{R})$, while the set of *square* matrices $M_{n,n}(\mathbb{R})$ is also denoted by $M_n(\mathbb{R})$. If $[a_{ij}], [b_{ij}] \in M_{m,n}(\mathbb{R})$, and $a \in \mathbb{R}$, we define

$$[a_{ij}] = [b_{ij}] :\Leftrightarrow \forall_{i \in \{1,\ldots,m\}} \forall_{j \in \{1,\ldots,n\}} \big(a_{ij} = b_{ij}\big).$$

$$[a_{ij}] + [b_{ij}] := [a_{ij} + b_{ij}],$$

$$a \cdot [b_{ij}] := [ab_{ij}],$$

$$\mathbf{0}_{mn} := [0],$$

and if $m = n$, we denote $\mathbf{0}_{nn}$ by $\mathbf{0}_n$, or, if n is clear from the context, by $\mathbf{0}$.

If $m = n = 2$, the above definitions take the form

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \Leftrightarrow a = a' \ \& \ b = b' \ \& \ c = c' \ \& \ d = d',$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} a + a' & b + b' \\ c + c' & d + d' \end{bmatrix},$$

$$\lambda \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{bmatrix}, \qquad \lambda \in \mathbb{R},$$

$$\mathbf{0}_2 := \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

It is easy to see that $M_{m,n}(\mathbb{R})$, and as a special case $M_2(\mathbb{R})$, equipped with the above operations, is a linear space.

Example 1.3.5. If

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

let $f_A : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$f_A(x, y) := A \begin{bmatrix} x \\ y \end{bmatrix} := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} := \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}.$$

Since

$$\begin{aligned} f_A\big((x, y) + (x', y')\big) &:= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x + x' \\ y + y' \end{bmatrix} \\ &= \begin{bmatrix} a(x + x') + b(y + y') \\ c(x + x') + d(y + y') \end{bmatrix} \\ &= \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} + \begin{bmatrix} ax' + by' \\ cx' + dy' \end{bmatrix} \end{aligned}$$

$$= A \begin{bmatrix} x \\ y \end{bmatrix} + A \begin{bmatrix} x' \\ y' \end{bmatrix}$$
$$= f_A\big((x,y)\big) + f_A\big((x',y')\big).$$

Similarly we show that $f_A\big(\lambda(x,y)\big) = \lambda f_A\big((x,y)\big)$, for every $\lambda \in \mathbb{R}$.

REMARK 1.3.6. Let X, Y, Z be linear spaces, $f \in \mathcal{L}(X,Y)$ and $g \in \mathcal{L}(Y,Z)$.
(i) The composite function $g \circ f$ is in $\mathcal{L}(X,Z)$, where $g \circ f : X \to Z$ is defined by

$$(g \circ f)(x) := g(f(x)), \qquad x \in X.$$

(ii) $\mathrm{id}_X \in \mathcal{L}(X)$.
(iii) $f(\mathbf{0}) = \mathbf{0}$.
(iv) if $x \in X$, then $f(-x) = -f(x)$.
(v) If $n \geq 1$, $a_1, \ldots a_n \in \mathbb{R}$, and $x_1, \ldots x_n \in X$, then

$$f\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i f(x_i).$$

PROOF. Exercise. For the inductive proof of the case (vi), use the following recursive definition of $\sum_{i=1}^{n} x_i$, where $x_1, \ldots, x_n \in X$ and $n \geq 1$:

$$\sum_{i=1}^{n} x_i := \begin{cases} x_1 & , n = 1 \\ \left(\sum_{i=1}^{n-1} x_i\right) + x_n & , n > 1 \end{cases}$$

□

A linear map preserves linear dependence, but not necessarily linear independence. The latter holds if a linear map is injective. If it is a bijection i.e., an injection and a surjection, it sends a basis of its domain to a basis of its codomain.

PROPOSITION 1.3.7. *If X, Z are linear spaces, $Y \subseteq X$, $f \in \mathcal{L}(X,Z)$, and $x_1, \ldots, x_n \in X$, the following hold.*

(i) *If $x_1, \ldots x_n$ are linearly dependent in X, then $f(x_1), \ldots, f(x_n)$ are linearly dependent in Z.*

(ii) *If Y is a linearly dependent subset of X, then $f(Y) := \{f(y) \mid y \in Y\}$ is a linearly dependent subset of Z.*

(iii) *If $x_1, \ldots x_n$ are linearly independent in X, then there is a linear map $g : X \to Z$ such that $g(x_1), \ldots, g(x_n)$ are linearly dependent in Z.*

(iv) *If $x_1, \ldots x_n$ are linearly independent in X, and if f is an injection, then $f(x_1), \ldots, f(x_n)$ are linearly independent in Z.*

(v) *If Y is a linearly independent subset of X, and if f is an injection, then $f(Y)$ is a linearly independent subset of Z.*

(vi) *If $X = \langle Y \rangle$, and if f is a surjection, then $Z = \langle f(Y) \rangle$.*

(vii) If Y is a basis of X, and if f is a bijection, then $f(Y)$ is a basis of Z.

Proof. (i) Let $a_1, \ldots a_n \in \mathbb{R}$, where $a_i \neq 0$, for some $i \in \{1, \ldots, n\}$ such that $\sum_{i=1}^n a_i x_i = \mathbf{0}$. Then what we want follows from the equalities

$$\mathbf{0} = f(\mathbf{0}) = f\bigg(\sum_{i=1}^n a_i x_i\bigg) = \sum_{i=1}^n a_i f(x_i).$$

(ii) It follows immediately from the case (i).

(iii) For example, we can take g to be the zero map.

(iv) By the injectivity of f, if $a_1, \ldots, a_n \in \mathbb{R}$, we have that

$$\begin{aligned}\sum_{i=1}^n a_i f(x_i) = \mathbf{0} &\Leftrightarrow f\bigg(\sum_{i=1}^n a_i x_i\bigg) = f(\mathbf{0}) \\ &\Leftrightarrow \sum_{i=1}^n a_i x_i = \mathbf{0} \\ &\Leftrightarrow a_1 = \ldots = a_n = 0.\end{aligned}$$

(v) It follows immediately from the case (iv).

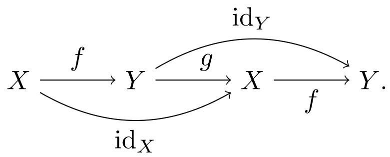
(vi) If X is trivial, then $Y = \emptyset$ or $Y = X$. In both cases what we want follows immediately. Let X be non-trivial, and let $z \in Z$. Then there is $x \in X$ such that $f(x) = z$. If $a_1, \ldots, a_n \in \mathbb{R}$ and $y_1, \ldots, y_n \in Y$ such that $x = \sum_{i=1}^n a_i y_i$, then

$$z = f(x) = f\bigg(\sum_{i=1}^n a_i y_i\bigg) = \sum_{i=1}^n a_i f(y_i) \in \langle f(Y) \rangle.$$

(vii) By the case (v) we have that $f(Y)$ is a linearly independent subset of Z, and by the case (vi) we have that $Z = \langle f(Y) \rangle$. □

A linear map $f : X \to Y$, which is is a linear isomorphism guarantees that the two linear spaces X and Y are the "same" from the linear-structure point of view.

Definition 1.3.8. If X, Y are linear spaces, an $f \in \mathcal{L}(X, Y)$ is a *linear isomorphism* between X, Y, if there is $g : Y \to X$ with $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$



In this case, we write $f : X \simeq Y$, and we say that the linear spaces X and Y are (linearly) *isomorphic*.

Next we see that two isomorphic finite-dimensional linear spaces have the same dimension.

PROPOSITION 1.3.9. *Let* X, Y *be linear spaces, and* $f \in \mathcal{L}(X, Y)$ *a linear isomorphism.*

(i) f *is a bijection (i.e., an injection and a surjection).*

(ii) If $g : Y \to X$ *such that* $f \circ g = \mathrm{id}_Y$ *and* $g \circ f = \mathrm{id}_X$*, then* $g \in \mathcal{L}(Y, X)$*.*

(iii) If $n \in \mathbb{N}$*, and* $\dim(X) = n$*, then* $\dim(Y) = n$*.*

(iv) If $h : X \to Y$ *is a linear map, which is a bijection, then* h *is a linear isomorphism.*

PROOF. Exercise. □

The condition (iv) above could be taken as the definition of a linear isomorphism. If $n \geq 1$, an n-dimensional linear space is isomorphic to $\mathbb{R}^n$.

COROLLARY 1.3.10. *If* X *is a linear space, and* $n \geq 1$*, then* $\dim(X) = n$ *if and only if* X *is isomorphic to* $\mathbb{R}^n$*.*

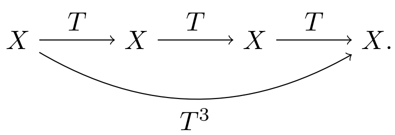
PROOF. Exercise. □

The set of operators $\mathcal{L}(X)$ of a linear space X is algebraically more interesting than $\mathcal{L}(X, Y)$, since a "multiplication", the composition of functions, is defined between its elements.

DEFINITION 1.3.11. If X is a linear space, and $T \in \mathcal{L}(X)$, we define

$$T^n := \begin{cases} \mathrm{id}_X & , n = 0 \\ T \circ T^{n-1} & , n > 0. \end{cases}$$

E.g., $T^3 = T \circ T \circ T$



REMARK 1.3.12. If X is a linear space, and $P \in \mathcal{L}(X)$, such that $P^2 = P$, then

$$X = \mathtt{Ker}(P) \oplus \mathtt{Im}(P).$$

PROOF. Exercise. □

REMARK 1.3.13. Let X be a linear space, $T \in \mathcal{L}(X)$, with $T^2 = \mathrm{id}_X$, and let

$$P := \frac{1}{2}(\mathrm{id}_X + T) \quad \& \quad Q := \frac{1}{2}(\mathrm{id}_X - T).$$

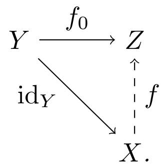
(i) $P + Q = \mathrm{id}_X$.

(ii) $P^2 = P$, and $Q^2 = Q$.

(iii) $PQ = QP = \mathbf{0}$.

PROOF. Exercise. □

PROPOSITION 1.3.14. *Let $n \geq 1$, X, Z be linear spaces, $Y \subseteq X$, and let the function $f_0 : Y \to Z$.*
(i) If $X = \langle Y \rangle$, there is at most one linear map $f : X \to Z$ that extends f_0 i.e., $f(y) = f_0(y)$, for every $y \in Y$, or, in other words, the following diagram commutes



(ii) If $Y = \{v_1, \ldots, v_n\}$ is a basis of X, there is a unique linear map $f : X \to Z$ that extends f_0, and hence, if $g, h : X \to Z$ are linear maps, we have that[3]

$$g_{|Y} = h_{|Y} \Rightarrow g = h.$$

PROOF. (i) If X is a trivial linear space, then $Y = \emptyset$ or $Y = X$. In the first case, f_0 is the empty set (as a set of pairs), and the only linear map that extends f_0 is the constant zero linear map. If $Y = X$, the only extension of f_0 is f_0 itself. If X is non-trivial, let $f, g : X \to Z$ be linear maps such that their restrictions $f_{|Y}, g_{|Y}$ to Y are equal to f_0, i.e.,

$$\forall_{y \in Y}\big(f(y) = f_0(y) = g(y)\big).$$

If $x \in X$, let $a_1, \ldots, a_n \in \mathbb{R}$ and $y_1, \ldots y_n \in Y$ such that $x = \sum_{i=1}^{n} a_i y_i$. By the Remark 1.3.6(v) we have that

$$f(x) = f\left(\sum_{i=1}^{n} a_i y_i\right) = \sum_{i=1}^{n} a_i f(y_i) = \sum_{i=1}^{n} a_i g(y_i) = g\left(\sum_{i=1}^{n} a_i y_i\right) = g(x).$$

(ii) If $x \in X$, then x has a unique writing as $x = \sum_{=1}^{n} a_i v_i$, for some $a_1, \ldots, a_n \in \mathbb{R}$. We define $f : X \to Z$ by

$$f\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i f_0(v_i).$$

It is easy to check that f is a linear map that extends f_0. Since Y generates X, by the case (i) we get that f is the unique extension of f_0. Moreover, if g and h are equal on the basis Y, then they are equal as functions from X to Z, since there is a unique extension of the restriction $g_{|Y}$ of g to Y.

□

[3]The *restriction* $g_{|Y}$ of g is the function $g_{|Y} : Y \to Z$, where $g_{|Y}(y) := g(y)$, for every $y \in Y$. Clearly, if Y is a subspace of a linear space X and $f \in \mathcal{L}(X, Z)$, then $f_Y \in \mathcal{L}(Y, Z)$.

1.4. The space of matrices

The set of $m \times n$-matrices $M_{m,n}(\mathbb{R})$, and the set of square matrices $M_n(\mathbb{R}) := M_{n,n}(\mathbb{R})$ was defined in the Definition 1.3.4.

REMARK 1.4.1. $M_{m,n}(\mathbb{R})$ is a linear space of dimension mn.

PROOF. The fact that $M_{m,n}(\mathbb{R})$ is a linear space is immediate from the Definition 1.3.4. To determine the dimension of $M_{m,n}(\mathbb{R})$, we associate to an $m\times n$-matrix

$$A := \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

the following element of $\mathbb{R}^{mn}$

$$\big(a_{11}, \dots, a_{1n}, \dots, a_{i1}, \dots, a_{in}, \dots, a_{m1}, \dots, a_{mn}\big).$$

E.g., to the 2×2-matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

we associate the 4-tuple

$$(a, b, c, d).$$

It is easy to see that this mapping $e : M_{m,n}(\mathbb{R}) \to \mathbb{R}^{mn}$ is a linear isomorphism, hence by the Proposition 1.3.9(iii) we get $\dim\big(M_{m,n}(\mathbb{R})\big) = \dim(\mathbb{R}^{mn}) = mn$. ☐

DEFINITION 1.4.2. Let the mapping ${}^t : M_{m,n}(\mathbb{R}) \to M_{n,m}(\mathbb{R})$, defined by

$$[a_{ij}] \mapsto [a_{ij}]^t,$$

where

$$[a_{ij}]^t := [b_{ji}], \quad b_{ji} := a_{ij}.$$

The matrix $[a_{ij}]^t$ is called the *transpose* of $[a_{ij}]$, and it has columns the rows of $[a_{ij}]$ and rows the columns of $[a_{ij}]$. If $A \in M_n(\mathbb{R})$ with $A^t = A$, we say that A is *symmetric*, and we denote their set by $\mathtt{Sym}_n(\mathbb{R})$. A *diagonal* matrix in $M_n(\mathbb{R})$ has the form

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} := \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} =: \mathtt{Diag}(\lambda_1, \dots, \lambda_n).$$

We denote by I_n the *unit* matrix in $M_n(\mathbb{R})$, defined by

$$I_n := \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} =: [\delta_{ij}],$$

where[4]

$$\delta_{ij} := \begin{cases} 1 & , \text{ if } i = j \\ 0 & , \text{ if } i \neq j. \end{cases}$$

E.g., if we consider the 2×3-matrix

$$A := \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 5 \end{bmatrix},$$

then its transpose A^t is the following 3×2-matrix

$$A^t := \begin{bmatrix} 2 & 1 \\ 1 & 3 \\ 0 & 5 \end{bmatrix}.$$

An example of a symmetric matrix is the following:

$$A = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 5 & 4 \\ -2 & 4 & -8 \end{bmatrix} = A^t.$$

REMARK 1.4.3. Let $A, B \in M_{m,n}(\mathbb{R})$, $C \in M_n(\mathbb{R})$, and $a \in \mathbb{R}$.
(i) $(A + B)^t = A^t + B^t$.
(ii) $(a \cdot B)^t = a \cdot B^t$.
(iii) $\left(A^t\right)^t = A$.
(iii) $C + C^t$ is symmetric.

PROOF. Exercise. □

Next we define the multiplication between matrices, an operation which, as we shall see later, is related to the composition of linear maps. To define the multiplication AB the number of columns of A has to be the number of rows of B!

DEFINITION 1.4.4. If $A := [a_{ij}] \in M_{m,n}(\mathbb{R})$ and $B := [b_{jk}] \in M_{n,l}(\mathbb{R})$, their product $AB \in M_{m,l}(\mathbb{R})$ is defined by

$$AB := [a_{ij}][b_{jk}] := [c_{ik}],$$

$$c_{ik} := \sum_{j=1}^{n} a_{ij} b_{jk},$$

[4]The symbol δ_{ki} is known as Kronecker's delta.

for every $1 \leq i \leq m$ and $1 \leq k \leq l$. If $A \in M_n(\mathbb{R})$, let

$$A^n := \begin{cases} I_n & , n = 0 \\ AA^{n-1} & , n > 0 \end{cases}$$

A matrix $A \in M_n(\mathbb{R})$ is *invertible*, if there is $B \in M_n(\mathbb{R})$ such that $AB = BA = I_n$. We denote by $\mathtt{Inv}_n(\mathbb{R})$ the set of invertible matrices in $M_n(\mathbb{R})$.

E.g., if

$$A := \begin{bmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{bmatrix} \quad \& \quad B := \begin{bmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{bmatrix},$$

then

$$AB := \begin{bmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 15 & 15 \\ 4 & 12 \end{bmatrix}.$$

It is not always true that $AB = BA$. E.g.,

$$\begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 0 & 5 \end{bmatrix},$$

and

$$\begin{bmatrix} 2 & -1 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & -3 \\ 0 & 5 \end{bmatrix}.$$

If $a, b \in \mathbb{R}$, and

$$A := \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \quad \& \quad B := \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix},$$

then

$$AB := \begin{bmatrix} 1 & a+b \\ 0 & 1 \end{bmatrix}.$$

Hence

$$\begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} = I_2.$$

Notice that, in contrast to what happens in $\mathbb{R}$, there are non-zero square matrices that are not invertible, like the matrix

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

PROPOSITION 1.4.5. *Let $A \in M_{m,n}(\mathbb{R})$, $B, C \in M_{n,l}(\mathbb{R})$, and $D \in M_{l,s}(\mathbb{R})$.*

(i) $AI_n = A$ and $I_m A = A$.

(ii) $A(B + C) = AB + AC$.

(iii) If $a \in \mathbb{R}$, then $A(a \cdot B) = a \cdot (AB)$.

(iv) $A(BD) = (AB)D$.

(v) The multiplication $B^t A^t$ is well-defined, and $(AB)^t = B^t A^t$.

PROOF. Exercise. □

COROLLARY 1.4.6. *Let $A, B, C \in M_n(\mathbb{R})$.*
(i) If $AB = BA = I_n = AC = CA$, then $B = C$. We denote the unique matrix B such that $AB = BA = I_n$ by A^{-1}, and we call it the inverse of A.
(ii) $I_n^t = I_n$.
(ii) If A is invertible, then $(A^{-1})^t = (A^t)^{-1}$.

PROOF. (i) $C = I_nC = (AB)C = (BA)C = B(AC) = BI_n = B$.
(ii) $[\delta_{ij}]^t := [d_{ij}]$, where $d_{ij} := \delta_{ij}$, and what we want follows from the obvious equality $\delta_{ij} = \delta_{ji}$.
(iii) By the Proposition 1.4.5(v) and the case (ii) we have that $I_n = I_n^t = (AA^{-1})^t = (A^{-1})^t A^t$, and $I_n = I_n^t = (A^{-1}A)^t = A^t(A^{-1})^t$. Since $I_n = (A^t)^{-1}A^t = A^t(A^t)^{-1}$, by the case (i) we get $(A^{-1})^t = (A^t)^{-1}$. □

One can show that if $A, B \in M_n(\mathbb{R})$, then

$$AB = I_n \Rightarrow BA = I_n,$$

hence we do not need to check both equalities in order to show that a matrix A is invertible. Note that this is the case only when the product AB is equal to I_n. If $A, B \in M_n(\mathbb{R})$ are invertible, then AB is also invertible and $(AB)^{-1} = B^{-1}A^{-1}$, since

$$(AB)(B^{-1}A^{-1}) = A[B(B^{-1}A^{-1})] = A[(BB^{-1})A^{-1}] = A[I_nA^{-1}] = AA^{-1} = I_n.$$

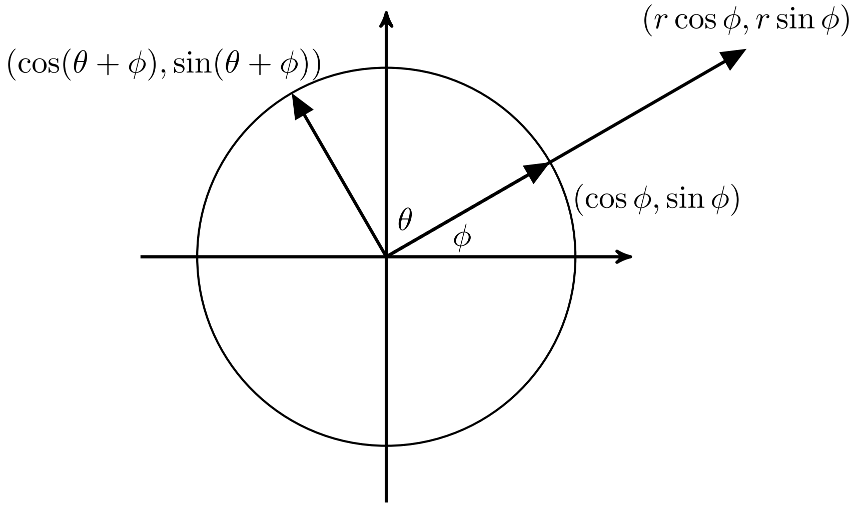
1.5. Matrices and linear maps

Matrices can be used to represent linear maps. Let's see the following important example. If $\theta \in \mathbb{R}$, let the matrix

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

Let the map $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$\begin{aligned} R_\theta(x, y) &:= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r\cos\phi \\ r\sin\phi \end{bmatrix} \\ &= r \begin{bmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi \\ \sin\theta\cos\phi + \cos\theta\cos\phi \end{bmatrix} \\ &= r \begin{bmatrix} \cos(\theta + \phi) \\ \sin(\theta + \phi) \end{bmatrix}, \end{aligned}$$



where $r := \sqrt{x^2+y^2}$. Hence, R_θ is the anti-clockwise θ-rotation of the vector (x,y). If $\theta_1, \theta_2 \in \mathbb{R}$, it is easy to see that

$$R(\theta_1)R(\theta_2) = R(\theta_1+\theta_2).$$

From that we can infer that the matrix $R(\theta)$ has an inverse.

Definition 1.5.1. If $A := [a_{ij}] \in M_{m,n}(\mathbb{R})$, the *linear map of* A is the mapping

$$T_A : \mathbb{R}^n \to \mathbb{R}^m$$

$$T_A(X) := AX,$$

where we view an arbitrary element $x := (x_1, \ldots, x_n) \in \mathbb{R}^n$ as an $n \times 1$-matrix X and the output $m \times 1$-matrix represents a vector in $\mathbb{R}^m$. I.e., we have

$$\begin{bmatrix} T_A(X)_1 \\ \vdots \\ T_A(X)_m \end{bmatrix} := \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

In a non-matrix form we write

$$T_A(x) := \left(\sum_{j=1}^n a_{1j}x_j, \ldots, \sum_{j=1}^n a_{ij}x_j, \ldots, \sum_{j=1}^n a_{mj}x_j\right).$$

If $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{R}^n$, and $l \in \{1, \ldots, n\}$, then

$$T_A(e_l) := (a_{1l}, \ldots, a_{ml}) = A^l,$$

where A^l is the l-column of the matrix A. and hence

$$T_A(e_l)_i = a_{il},$$

for every $i \in \{1, \ldots, m\}$. Using the Proposition 1.4.5 we can show the following.

Proposition 1.5.2. *If $A, B \in M_{m,n}(\mathbb{R})$, and $a \in \mathbb{R}$, the following hold:*
(i) $T_A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.
(ii) If $T_A = \mathbf{0}$, *then* $A = \mathbf{0}_{mn}$, *and if* $T_A = T_B$, *then* $A = B$.
(iii) $T_{A+B} = T_A + T_B$.
(iv) $T_{a \cdot A} = aT_A$.
(v) $T_{I_n} = \mathrm{id}_{\mathbb{R}_n}$ *and* $T_{\mathbf{0}_{mn}} = \mathbf{0}$.
(vi) If $C \in M_{n,l}(\mathbb{R})$, *then* $T_{AC} = T_A \circ T_C$

$$\mathbb{R}^l \xrightarrow{T_C} \mathbb{R}^n \xrightarrow{T_A} \mathbb{R}^m$$
$$\mathbb{R}^l \xrightarrow{T_{AC}} \mathbb{R}^m$$

(vii) If A is invertible, then T_A *is invertible and* $T_A^{-1} = T_{A^{-1}}$.
(viii) The function $\mathcal{T} : M_{m,n}(\mathbb{R}) \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, *defined by* $A \mapsto T_A$, *is a linear map.*

Proof. Exercise. □

So far we defined a a linear map $T_A : \mathbb{R}^n \to \mathbb{R}^m$, given a matrix $A \in M_{m,n}(\mathbb{R})$. Next we define a matrix $A_T \in M_{m,n}(\mathbb{R})$, given a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$. The two constructions are inverse to each other.

Theorem 1.5.3. *Let $n, m \geq 1$. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, there is a matrix $A_T \in M_{m,n}(\mathbb{R})$ such that $T = T_{A_T}$ i.e., for every $x \in \mathbb{R}^n$ we have that*

$$T(x) = T_{A_T}(x) := A_T x.$$

The matrix A_T is called the matrix of the linear map T.

Proof. If $B := \{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{R}^n$, then for every $i \in \{1, \ldots, n\}$ we write $T(e_1)$ a linear combination of the standard basis of $\mathbb{R}^m$ i.e.,

$$T(e_i) := \big(T(e_i)_1, \ldots, T(e_i)_m\big).$$

The matrix A_T is formed by taking these m-tuples as its columns i.e., we define

$$A_T := \begin{bmatrix} T(e_1)_1 & \ldots & T(e_n)_1 \\ \vdots & \vdots & \vdots \\ T(e_1)_j & \ldots & T(e_n)_j \\ \vdots & \vdots & \vdots \\ T(e_1)_m & \ldots & T(e_n)_m \end{bmatrix} =: [a_{ji}] = \big[T(e_i)_j\big].$$

By the Proposition 1.3.14, to show that the linear maps T and T_{A_T} are equal, it suffices to show that they are equal on the elements of B. Since

$$T_{A_T}(e_i) := A_T e_i := \big[T(e_i)_j\big] e_i = [a_{ji}] e_i = [c_{j1}],$$

where

$$c_{j1} = \sum_{i=1}^{n} a_{ji} b_{i1} = a_{ji} := T(e_i)_j,$$

we get[5] the required equality with the vector $T(e_i) := \big(T(e_i)_1, \ldots, T(e_i)_m\big)$. $\square$

For example, if $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map such that

$$T(0,1) := (a,c) \quad \& \quad T(1,0) := (b,d),$$

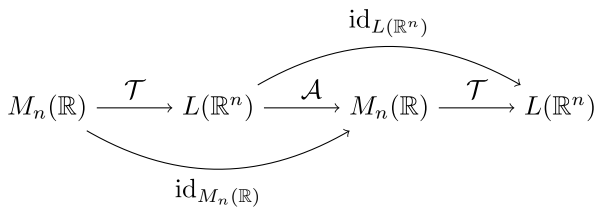
then we have that

$$A_T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

PROPOSITION 1.5.4. *Let the function* $\mathcal{A} : L(\mathbb{R}^n) \to M_n(\mathbb{R})$

$$T \mapsto A_T := \mathcal{A}(T).$$

(i) The mappings $\mathcal{T}$ *and* $\mathcal{A}$ *satisfy the following conditions:*
(i) $\mathcal{A} \circ \mathcal{T} = \mathrm{id}_{M_n(\mathbb{R})}$ *and* $\mathcal{T} \circ \mathcal{A} = \mathrm{id}_{L(\mathbb{R}^n)}$



(ii) $A_{S \circ T} = A_S A_T$.
(iii) $A_{I_n} = I_n$ *and* $A_{O_n} = O_n$.
(ix) $A_{S+T} = A_S + A_T$.
(x) $A_{\lambda T} = \lambda A_T$.
(xi) If T *is invertible, then* A_T *is invertible and* $A_T^{-1} = A_{T^{-1}}$.

PROOF. Exercise. $\square$

1.6. Determinants

DEFINITION 1.6.1. If

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

[5] A simpler argument is the following. As we have shown after the Definition 1.5.1, $T_A(e_i)$ is the i-column of A. Hence, $T_{A_T}(e_i)$ is the i-column of A_T, which is exactly $T(e_i)$ by the definition of A_T.

is a 2×2-matrix, its *determinant* $\mathtt{Det}(A)$ is defined by

$$\mathtt{Det}(A) := \begin{vmatrix} a & b \\ c & d \end{vmatrix} := ad - bc.$$

If

$$A^1 := \begin{bmatrix} a \\ c \end{bmatrix} \quad \& \quad A^2 := \begin{bmatrix} b \\ d \end{bmatrix}$$

are the columns of A, we use the notation

$$\mathtt{Det}(A) = \mathtt{Det}(A^1, A^2).$$

We have that

$$\mathtt{Det}(I_2) := \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} := 1 - 0 = 1.$$

It is also clear that

$$\mathtt{Det}(A) := \begin{vmatrix} a & b \\ c & d \end{vmatrix} := ad - bc =: \begin{vmatrix} a & c \\ b & d \end{vmatrix} =: \mathtt{Det}(A^t).$$

Remark 1.6.2. Let the following 2×1 matrices:

$$A^1 := \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad C^1 := \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad B^2 := \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad D^2 := \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}.$$

The following hold.
(i) $\mathtt{Det}(A^1 + C^1, B^2) = \mathtt{Det}(A^1, B^2) + \mathtt{Det}(C^1, B^2)$.
(ii) $\mathtt{Det}(A^1, B^2 + D^2) = \mathtt{Det}(A^1, B^2) + \mathtt{Det}(A^1, D^2)$.
(iii) If $\lambda \in \mathbb{R}$, then $\mathtt{Det}(\lambda A^1, B^2) = \lambda \mathtt{Det}(A^1, B^2) = \mathtt{Det}(A^1, \lambda B^2)$.
(iv) If $A^1 = B^2$, then $\mathtt{Det}(A^1, B^2) = 0$.

Proof. We prove only (i), and the rest is an exercise.

$$\begin{aligned} \mathtt{Det}(A^1 + C^1, B^2) &:= \begin{vmatrix} a_1 + c_1 & b_1 \\ a_2 + c_2 & b_2 \end{vmatrix} \\ &:= (a_1 + c_1)b_2 - b_1(a_2 + c_2) \\ &= (a_1 b_2 - b_1 a_2) + (c_1 b_2 - b_1 c_2) \\ &:= \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} + \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} \\ &:= \mathtt{Det}(A^1, B^2) + \mathtt{Det}(C^1, B^2). \end{aligned}$$

□

Although one can use the definition of $\mathtt{Det}(A)$ to show the following corollary, its proof is simpler, if we use the fundamental properties of the Remark 1.6.2.

Corollary 1.6.3. *Let the following* 2×1 *matrices:*

$$A^1 := \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad B^2 := \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

The following hold.
(i) If $\lambda \in \mathbb{R}$, *then* $\mathtt{Det}(A^1 + \lambda B^2, B^2) = \mathtt{Det}(A^1, B^2)$.
(ii) If $\lambda \in \mathbb{R}$, *then* $\mathtt{Det}(A^1, B^2 + \lambda A^1) = \mathtt{Det}(A^1, B^2)$.
(iii) $\mathtt{Det}(A^1, B^2) = -\mathtt{Det}(B^2, A^1)$.

Proof. Exercise. □

The determinant of a matrix A provides non-trivial information on vectors related to A. We have seen that $\mathtt{Det}(I_2) = 1 \neq 0$, and we know that the columns $e_1 := (1, 0)$ and $e_2 := (0, 1)$ of the matrix I_2 are linearly independent elements. This is a special case of the following general fact.

Proposition 1.6.4. *Let the following* 2×1 *matrices:*

$$A := \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad B := \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

The vectors (a_1, a_2) *and* (b_1, b_2) *are linearly independent in* $\mathbb{R}^2$ *if and only if*

$$\mathtt{Det}(A, B) \neq 0.$$

Proof. ($\Rightarrow$) Suppose that (a_1, a_2) and (b_1, b_2) are linearly independent in $\mathbb{R}^2$, and suppose that

$$\mathtt{Det}(A, B) := \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} := a_1 b_2 - b_1 a_2 = 0.$$

Since then we have that

$$b_2(a_1, a_2) + (-a_2)(b_1, b_2) = (b_2 a_1 - a_2 b_1, b_2 a_2 - a_2 b_2) = (0, 0),$$

by the hypothesis of linear independence of (a_1, a_2) and (b_1, b_2) we get

$$b_2 = 0 = -a_2 = a_2.$$

Hence the two vectors take the form $(a_1, 0)$ and $(b_1, 0)$. Since they are linearly independent, these are non-zero vectors, hence $a_1 \neq 0$ and $b_1 \neq 0$. Consequently, we have that $(a_1, 0) = \frac{a_1}{b_1}(b_1, 0)$ i.e., the vectors (a_1, a_2) and (b_1, b_2) are linearly dependent, which is a contradiction. Hence, $\mathtt{Det}(A, B) \neq 0$.
($\Leftarrow$) Suppose that $\mathtt{Det}(A, B) \neq 0$, and let $\lambda, \mu \in \mathbb{R}$ such that

$$\lambda(a_1, a_2) + \mu(b_1, b_2) = (0, 0) \Leftrightarrow (\lambda a_1 + \mu b_1, \lambda a_2 + \mu b_2) = (0, 0),$$

hence

$$\lambda a_1 = -\mu b_1 \;\;\&\;\; \lambda a_2 = -\mu b_2.$$

Suppose that $\lambda \neq 0$ (if we suppose that $\mu \neq 0$. we proceed similarly). By the Remark 1.6.2 we have that

$$\begin{aligned}\mathtt{Det}(A,B) &= \begin{vmatrix}(\frac{-\mu}{\lambda})b_1 & b_1\\ (\frac{-\mu}{\lambda})b_2 & b_2\end{vmatrix}\\ &= \left(\frac{-\mu}{\lambda}\right)\begin{vmatrix}b_1 & b_1\\ b_2 & b_2\end{vmatrix}\\ &= \left(\frac{-\mu}{\lambda}\right)0\\ &= 0,\end{aligned}$$

which is a contradiction. Hence $\lambda = 0 = \mu$, and the vectors $(a_1, a_2), (b_1, b_2)$ are linearly independent. □

Proposition 1.6.5. *Let $A, B \in M_2(\mathbb{R})$.*
(i) $\mathtt{Det}(AB) = \mathtt{Det}(A)\mathtt{Det}(B)$.
(ii) A is invertible if and only if $\mathtt{Det}(A) \neq 0$.

Proof. (i) Exercise.
(ii) If $AA^{-1} = I_2$, then by the case (i) we have that

$$1 = \mathtt{Det}(I_2) = \mathtt{Det}(AA^{-1}) = \mathtt{Det}(A)\mathtt{Det}(A^{-1}),$$

hence $\mathtt{Det}(A) \neq 0$, $\mathtt{Det}(A^{-1}) \neq 0$, and

$$\mathtt{Det}(A^{-1}) = \frac{1}{\mathtt{Det}(A)}.$$

For the converse let

$$A = \begin{bmatrix} a & b\\ c & d\end{bmatrix}$$

and suppose that

$$\mathtt{Det}(A) := \begin{vmatrix} a & b\\ c & d\end{vmatrix} := ad - bc \neq 0.$$

We show that the system

$$\begin{bmatrix} a & b\\ c & d\end{bmatrix}\begin{bmatrix} x & y\\ z & w\end{bmatrix} = \begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix} \Leftrightarrow$$

$$\begin{bmatrix} ax+bz & ay+bw\\ cx+dz & cy+dw\end{bmatrix} = \begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix} \Leftrightarrow$$

$$ax + bz = 1 \ \ \& \ \ cx + dz = 0,$$

and

$$ay + bw = 0 \ \ \& \ \ cy + dw = 1,$$

has a solution. If we multiply the equation $ax + bz = 1$ by d and the equation $cx + dz = 0$ by b, and we subtract them we get

$$dax + dbz - bcx - bdz = d \Leftrightarrow x = \frac{d}{ad - bc}.$$

Working similarly, we get

$$A^{-1} := \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \frac{1}{\mathtt{Det}(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

□

DEFINITION 1.6.6. If

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is a 3×3-matrix, its *determinant* $\mathtt{Det}(A)$ is defined by

$$\mathtt{Det}(A) := \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} := a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

As expected, we have that

$$\mathtt{Det}(I_3) := \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} := 1 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} - 0 \begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix} + 0 \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} = 1.$$

More generally, if we consider a matrix in diagonal form, then for the corresponding determinant we have that

$$\begin{vmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{vmatrix} := \lambda_1 \begin{vmatrix} \lambda_2 & 0 \\ 0 & \lambda_3 \end{vmatrix} - 0 \begin{vmatrix} 0 & 0 \\ 0 & \lambda_3 \end{vmatrix} + 0 \begin{vmatrix} 0 & \lambda_2 \\ 0 & 0 \end{vmatrix} = \lambda_1 \lambda_2 \lambda_3.$$

All results we showed for the determinant of a matrix in $M_2(\mathbb{R})$ hold also for the determinant of a matrix in $M_3(\mathbb{R})$.

1.7. The inner product on $\mathbb{R}^n$

DEFINITION 1.7.1. Let X be a linear space. An *inner product* on X is a mapping $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ such that for every $x, y, z \in X$ and $\lambda \in \mathbb{R}$ the following conditions hold:

(i) $\langle x, x \rangle \geq 0$ (positivity).
(ii) $\langle x, x \rangle = 0 \Rightarrow x = \mathbf{0}$ (definiteness).
(iii) $\langle x, y \rangle = \langle y, x \rangle$ (symmetry).
(iv) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ (left additivity).
(v) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ (left homogeneous).

If $\langle \cdot, \cdot \rangle$ is an inner product on X, the pair $(X, \langle \cdot, \cdot \rangle)$ is called an *inner product space*. A *norm* on X is a mapping $||.|| : X \to \mathbb{R}$ such that for every $x, y \in X$ and $\lambda \in \mathbb{R}$ the following hold:
(i) $||x|| \geq 0$ (positivity).
(ii) $||x|| = 0 \Rightarrow x = \mathbf{0}$ (definiteness).
(iii) $||x + y|| \leq ||x|| + ||y||$ (triangle inequality).
(iv) $||\lambda x|| = |\lambda|||x||$.
If $||.||$ is a norm on X, the pair $(X, ||.||)$ is called a *normed space*.

Because of symmetry an inner product is bilinear i.e., it is also right additive and right homogeneous:

(iv') $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ (right additivity).
(v') $\langle x, \lambda y \rangle = \lambda \langle x, y \rangle$ (right homogeneous).

Notice also that

$$|| - x|| = ||(-1)x|| = | - 1|||x|| = 1||x|| = ||x||.$$

Definition 1.7.2. If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are in $\mathbb{R}^n$, their *Euclidean inner product* is defined by

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i.$$

If $n = 1$, then the Euclidean inner product on $\mathbb{R}$ is the standard product on $\mathbb{R}$. By definition we have that

$$\langle x, x \rangle := \sum_{i=1}^{n} x_i x_i = \sum_{i=1}^{n} x_i^2 = x_1^2 + \ldots + x_n^2.$$

It is easy to see that the Euclidean inner product is an inner product on $\mathbb{R}^n$. Next we show that an inner product is determined by its diagonal entries.

Proposition 1.7.3. *Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space and $x, y \in X$.*
(i) (*Polarization identity*) $\langle x, y \rangle = \frac{1}{4}\big(\langle x + y, x + y \rangle - \langle x - y, x - y \rangle\big)$.
(ii) $x = \mathbf{0} \Leftrightarrow \forall_{z \in X}\big(\langle x, z \rangle = 0\big)$.
(iii) $\forall_{z \in X}\big(\langle x, z \rangle = \langle y, z \rangle\big) \Rightarrow x = y$.

Proof. Exercise. □

If $x = \mathbf{0}$, then $||x|| = 0$, since

$$||\mathbf{0}|| = ||0 \cdot \mathbf{0}|| = |0|||\mathbf{0}|| = 0||\mathbf{0}|| = 0.$$

Moreover, if $x = \mathbf{0}$, or $y = \mathbf{0}$, or $y = \lambda x$, for some $\lambda > 0$, then the equality holds in the triangle inequality $||x + y|| \leq ||x|| + ||y||$.

DEFINITION 1.7.4. If $x \in \mathbb{R}^n$, the *Euclidean norm* $|x|$ of x is defined by

$$|x| := \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = \sqrt{\langle x, x\rangle}.$$

Geometrically, if $x \in \mathbb{R}^n$, then $|x|$ is the *length* of the vector x. To show that the Euclidean norm is a norm we need the following inequality.

PROPOSITION 1.7.5 (Inequality of Cauchy). *If $x, y \in \mathbb{R}^n$, then*

$$|\langle x, y\rangle| \leq |x||y|.$$

PROOF. By definition we need to show

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leq \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} y_i^2\right)^{\frac{1}{2}},$$

which is equivalent to

$$A := \left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) =: B.$$

This we get as follows:

$$\begin{aligned}
B - A &= \sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} y_j^2 - \sum_{i=1}^{n} x_i y_i \sum_{j=1}^{n} x_j y_j \\
&= \frac{1}{2}\sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} y_j^2 + \frac{1}{2}\sum_{j=1}^{n} x_j^2 \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{j=1}^{n} x_j y_j \\
&= \sum_{i,j=1}^{n} \frac{1}{2}\left(x_i^2 y_j^2 + x_j^2 y_i^2 - 2x_i y_i x_j y_j\right) \\
&= \sum_{i,j=1}^{n} \frac{1}{2}\left(x_i y_j - x_j y_i\right)^2 \\
&\geq 0.
\end{aligned}$$

□

An inner product on X always induces a norm on X, which is defined by

$$||x|| = \langle x, x\rangle^{\frac{1}{2}},$$

for every $x \in X$. To show that $||.||$ is a norm on X we need the inequality

$$\left|\langle x, y\rangle\right| \leq ||x|| \, ||y||,$$

which generalizes the inequality of Cauchy.

DEFINITION 1.7.6. A *metric* d on a set X is a function $d : X \times X \to \mathbb{R}$ such that for every $x, y, z \in X$ the following hold:
(*i*) $d(x, y) \geq 0$.
(*ii*) $d(x, y) = 0 \Leftrightarrow x = y$.
(*iii*) $d(x, y) = d(y, x)$.
(*iv*) $d(x, y) \leq d(x, z) + d(z, y)$.
If d is a metric on X, the pair (X, d) is called a *metric space.*

A norm $||.||$ on a linear space X induces a metric on X defined by

$$d(x, y) := ||x - y||.$$

DEFINITION 1.7.7. The *Euclidean metric* d on $\mathbb{R}^n$ is the metric induced by the Euclidean norm on $\mathbb{R}^n$ i.e.,

$$d(x, y) := |x - y| := \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}} =$$

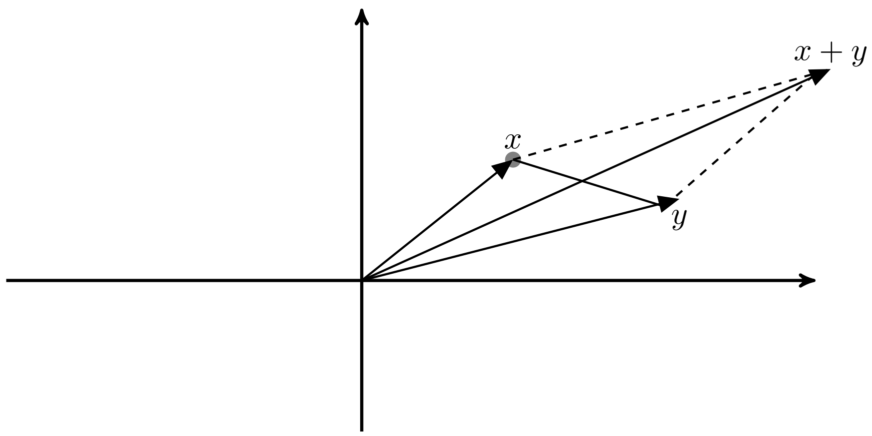
$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} = \sqrt{\langle x - y, x - y \rangle},$$

for every $x, y \in \mathbb{R}^n$.

The Euclidean norm is the norm induced by the Euclidean inner product. To understand the geometric meaning of the Euclidean inner product we first see that a vector $x \in \mathbb{R}^n$ is *orthogonal* to a vector $y \in \mathbb{R}^n$, in symbols $x \perp y$, if and only if $\langle x, y \rangle = 0$. To explain this we work as follows. It is easy to see geometrically[6] that

$$x \perp y \Leftrightarrow |x - y| = |x + y|,$$

since the diagonals of the parallelogram are equal only if x is perpendicular to y.



[6]The following figure also explains why $|x + y| \leq |x| + |y|$.

We show that

$$|x - y| = |x + y| \Leftrightarrow \langle x, y\rangle = 0.$$

Since $|x| \geq 0$, we have that

$$\begin{aligned} |x - y| = |x + y| &\Leftrightarrow |x - y|^2 = |x + y|^2 \\ &:\Leftrightarrow \langle x - y, x - y\rangle = \langle x + y, x + y\rangle \\ &\Leftrightarrow \langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \\ &\Leftrightarrow 4\langle x, y\rangle = 0 \\ &\Leftrightarrow \langle x, y\rangle = 0. \end{aligned}$$

By the last two equivalences we get the required equivalence

$$x \perp y \Leftrightarrow \langle x, y\rangle = 0.$$

Corollary 1.7.8 (Pythagoras theorem). *If $x, y \in \mathbb{R}^n$, such that $x \perp y$, then*

$$|x + y|^2 = |x|^2 + |y|^2.$$

Proof. Exercise. □

By the inequality of Cauchy we have that

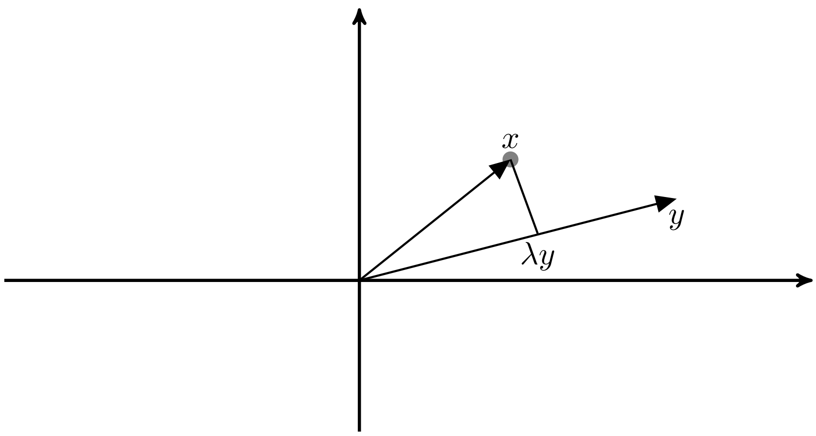
$$\left| \frac{|\langle x, y\rangle|}{|x||y|} \right| = \frac{|\langle x, y\rangle|}{|x||y|} \leq 1 \Leftrightarrow -1 \leq \frac{\langle x, y\rangle}{|x||y|} \leq 1.$$

hence, there exists a unique angle $\theta \in [0, \pi]$ such that

$$\cos\theta = \frac{\langle x, y\rangle}{|x||y|},$$

and we call θ the *angle between x and y*. Clearly, if $\langle x, y\rangle = 0$, then $\theta = \frac{\pi}{2}$.

Proposition 1.7.9. *If $x, y \in \mathbb{R}^n$, and $y \neq \mathbf{0}$, then the projection* $\mathrm{pr}_y(x)$ *of x on y is given by*



$$\mathrm{pr}_y(x) := \lambda y \quad \& \quad \lambda := \frac{\langle x, y\rangle}{\langle y, y\rangle}.$$

Proof. Since $(x - \lambda y) \perp y$, and $y \neq \mathbf{0}$, we have that

$$\begin{aligned} \langle (x - \lambda y), y\rangle = 0 &\Leftrightarrow \langle x, y\rangle - \langle \lambda y, y\rangle = 0 \\ &\Leftrightarrow \langle x, y\rangle - \lambda\langle y, y\rangle = 0 \\ &\Leftrightarrow \lambda = \frac{\langle x, y\rangle}{\langle y, y\rangle}. \end{aligned}$$

□

CHAPTER 2

Functions of several variables

2.1. Curves in $\mathbb{R}^n$

DEFINITION 2.1.1. Let I be an *interval* of $\mathbb{R}$ of the form

$$(-\infty, a),\ (-\infty, a],\ (a, +\infty),\ [a, +\infty),\ \mathbb{R},\ (a, b),\ (a, b],\ [a, b),\ [a, b],$$

where $a, b \in \mathbb{R}$ such that $a \leq b$. A *curve* in $\mathbb{R}^n$ is a function

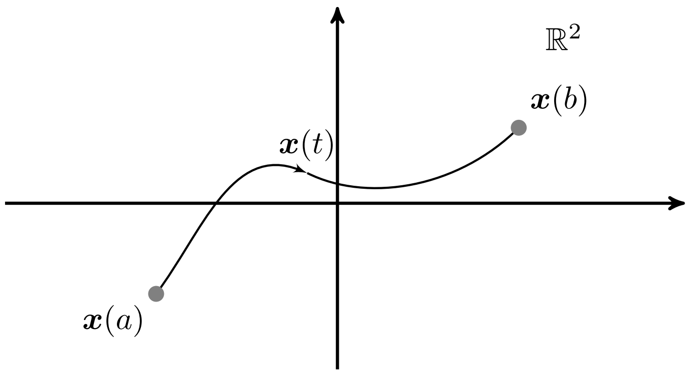
$$\boldsymbol{x} : I \to \mathbb{R}^n \qquad I \ni t \mapsto \boldsymbol{x}(t) \in \mathbb{R}^n, \qquad t \in I.$$

We also write

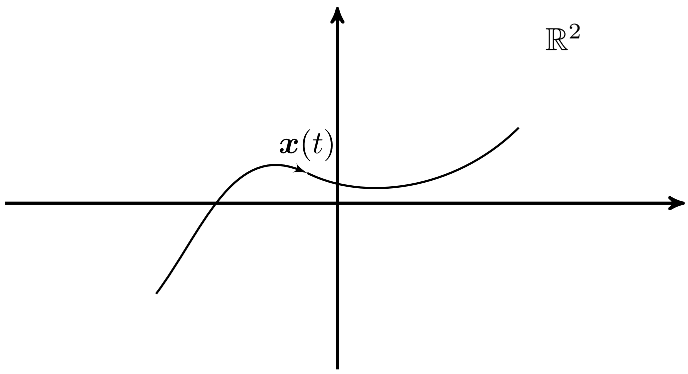
$$\boldsymbol{x}(t) = \big(x_1(t), \ldots, x_n(t)\big), \qquad t \in I,$$

where $x_i : I \to \mathbb{R}$ is the *i-coordinate function* of $\boldsymbol{x}$, for every $i \in \{1, \ldots, n\}$. We also call $\boldsymbol{x}(t)$ the *position vector* of $\boldsymbol{x}$ at time t. We call $\boldsymbol{x}$ *differentiable* on (every element of) I, if the coordinate functions $x_1(t), \ldots, x_n(t)$ of $\boldsymbol{x}$ are differentiable on (every element of) I. A point $P \in \mathbb{R}^n$ *belongs to* $\boldsymbol{x}$, if there is some $t \in I$ such that $P = \boldsymbol{x}(t)$.

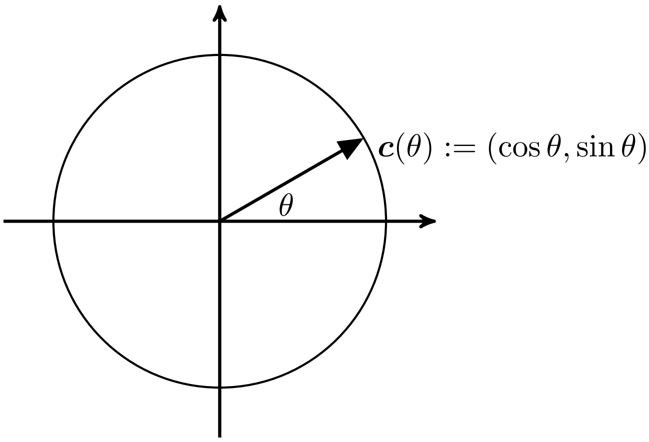
Next we draw the image of a differentiable curve $\boldsymbol{x} : [a, b] \to \mathbb{R}^2$



and the image of a differentiable curve $\boldsymbol{x} : (a, b) \to \mathbb{R}^2$.



Let also the curve $\boldsymbol{c} : [0, 2\pi] \to \mathbb{R}^2$, defined by $\theta \mapsto (\cos\theta, \sin\theta)$, for every $\theta \in [0, 2\pi]$, the image of which is the unit circle in $\mathbb{R}^2$.



This is a differentiable curve, since $\boldsymbol{c}(\theta) := \big(c_1(\theta), c_2(\theta)\big)$, and its coordinate functions $c_1(\theta) := \cos\theta$, and $c_2(\theta) := \sin\theta$ are differentiable on $[0, 2\pi]$, since $\cos'\theta = -\sin\theta$, and $\sin'\theta = \cos\theta$, for every $\theta \in [0, 2\pi]$. Moreover, $\boldsymbol{c}$ is a *closed* curve, since $\boldsymbol{c}(0) = \boldsymbol{c}(2\pi)$.

If $\boldsymbol{x}(t) : I \to \mathbb{R}^n$ is a differentiable curve in $\mathbb{R}^n$, $t_0 \in I$, and $h \in \mathbb{R}$, then

$$\begin{aligned}\frac{\boldsymbol{x}(t_0 + h) - \boldsymbol{x}(t_0)}{h} &= \frac{1}{h}\Big[\big(x_1(t_0 + h), \ldots, x_n(t_0 + h)\big) - \big(x_1(t_0), \ldots, x_n(t_0)\big)\Big] \\ &= \frac{1}{h}\big(x_1(t_0 + h) - x_1(t_0), \ldots, x_n(t_0 + h) - x_n(t_0)\big) \\ &= \left(\frac{x_1(t_0 + h) - x_1(t_0)}{h}, \ldots, \frac{x_n(t_0 + h) - x_n(t_0)}{h}\right),\end{aligned}$$

and hence

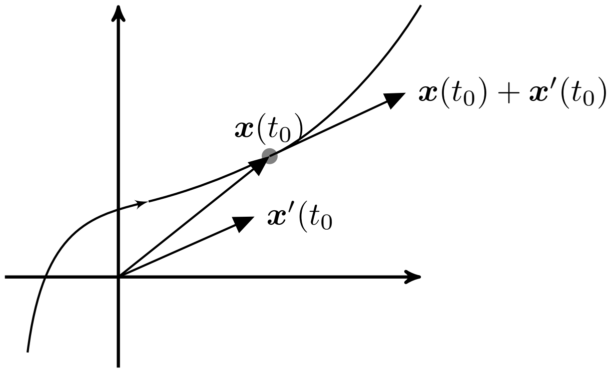
$$\lim_{h \to 0} \frac{\boldsymbol{x}(t_0 + h) - \boldsymbol{x}(t_0)}{h} = \big({x_1}'(t_0), \ldots, {x_n}'(t_0)\big).$$

DEFINITION 2.1.2. If $\boldsymbol{x} : I \to \mathbb{R}^n$ is a differentiable curve, its *derivative* is the curve $\boldsymbol{x}' : I \to \mathbb{R}^n$ defined, for every $t_0 \in I$, by

$$\boldsymbol{x}'(t_0) := \frac{d\boldsymbol{x}}{dt}(t_0) := \big(x_1{}'(t_0), \ldots, x_n{}'(t_0)\big) := \left(\frac{dx_1}{dt}(t_0), \ldots, \frac{dx_n}{dt}(t_0)\right).$$

We call $\boldsymbol{x}'(t_0)$ the *velocity vector* of $\boldsymbol{x}(t)$ at time t_0.

The velocity vector $\boldsymbol{x}'(t_0)$ is located at the origin of the Euclidean plane, but we view it as a vector tangent to the curve at t_0 and parallel to it.



DEFINITION 2.1.3. Let $\boldsymbol{x} : I \to \mathbb{R}^n$ be a differentiable curve. Its *speed* $v_{\boldsymbol{x}} : I \to [0, +\infty)$ is defined, for every $t \in I$, by

$$v_{\boldsymbol{x}}(t) := |\boldsymbol{x}'(t)|,$$

where $|\boldsymbol{x}'(t)|$ is the Euclidean norm of the vector $\boldsymbol{x}'(t)$. If the derivative $\boldsymbol{x}' : I \to \mathbb{R}^n$ of $\boldsymbol{x}$ is differentiable, the *acceleration vector* of $\boldsymbol{x}(t)$ at time $t_0 \in I$ is defined by

$$\boldsymbol{x}''(t_0) := \frac{d\boldsymbol{x}'}{dt}(t_0) := \frac{d^2\boldsymbol{x}}{dt}(t_0).$$

Notice that by the definition of the Euclidean norm $|.|$ we have that

$$v_{\boldsymbol{x}}(t)^2 := |\boldsymbol{x}'(t)|^2 = \langle x'(t), x'(t)\rangle.$$

PROPOSITION 2.1.4. *Let $\boldsymbol{x}, \boldsymbol{y} : I \to \mathbb{R}^n$ be differentiable curves, $\lambda \in \mathbb{R}$, and $f : I \to \mathbb{R}$ a differentiable function.*

(i) The sum $\boldsymbol{x} + \boldsymbol{y} : I \to \mathbb{R}^n$, defined by

$$(\boldsymbol{x} + \boldsymbol{y})(t) := \boldsymbol{x}(t) + \boldsymbol{y}(t),$$

for every $t \in I$, is a differentiable curve, and, for every $t_0 \in I$, we have that

$$(\boldsymbol{x} + \boldsymbol{y})'(t_0) = \boldsymbol{x}'(t_0) + \boldsymbol{y}'(t_0).$$

(ii) The product $\lambda\boldsymbol{x} : I \to \mathbb{R}^n$, defined by

$$(\lambda\boldsymbol{x})(t) := \lambda\boldsymbol{x}(t),$$

for every $t \in I$, is a differentiable curve, and, for every $t_0 \in I$, we have that

$$(\lambda \boldsymbol{x})'(t_0) = \lambda \boldsymbol{x}'(t_0).$$

(iii) The product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle : I \to \mathbb{R}$, defined by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle(t) := \langle \boldsymbol{x}(t), \boldsymbol{y}(t) \rangle,$$

for every $t \in I$, where $\langle \boldsymbol{x}(t), \boldsymbol{y}(t) \rangle$ is the Euclidean inner product of $\boldsymbol{x}(t), \boldsymbol{y}(t)$, is a differentiable function, and, for every $t_0 \in I$, we have that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle'(t_0) = \langle \boldsymbol{x}'(t_0), \boldsymbol{y}(t_0) \rangle + \langle \boldsymbol{x}(t_0), \boldsymbol{y}'(t_0) \rangle.$$

(iv) The product $\boldsymbol{x}^2 : I \to \mathbb{R}$, defined by

$$(\boldsymbol{x}^2)(t) := \langle \boldsymbol{x}(t), \boldsymbol{x}(t) \rangle,$$

for every $t \in I$, is a differentiable function, and, for every $t_0 \in I$, we have that

$$(\boldsymbol{x}^2)'(t_0) = 2\langle \boldsymbol{x}(t_0), \boldsymbol{x}'(t_0) \rangle.$$

(v) The product $f\boldsymbol{x} : I \to \mathbb{R}^n$, defined by

$$(f\boldsymbol{x})(t) := f(t)\boldsymbol{x}(t),$$

for every $t \in I$, is a differentiable curve, and, for every $t_0 \in I$, we have that

$$(f\boldsymbol{x})'(t_0) = f'(t_0)\boldsymbol{x}(t_0) + f(t_0)\boldsymbol{x}'(t_0).$$

Proof. We prove only the case (iii), and the rest is an exercise. By the definition of the Euclidean inner product we have that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle(t) := \langle \boldsymbol{x}(t), \boldsymbol{y}(t) \rangle := \sum_{i=1}^{n} x_i(t) y_i(t) = x_1(t)y_1(t) + \ldots + x_n(t)y_n(t),$$

hence we have that

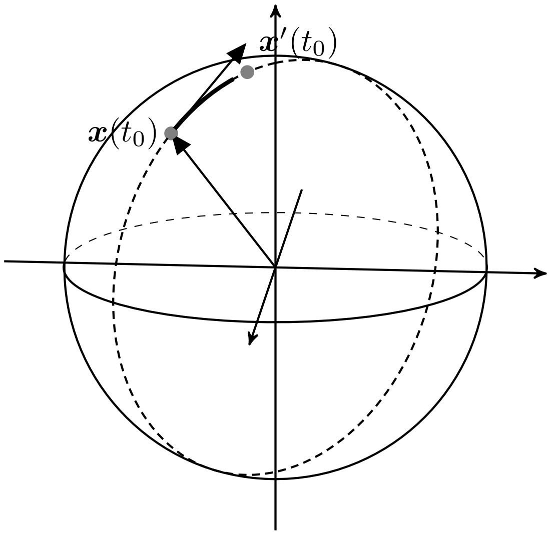
$$\begin{aligned}
&\langle \boldsymbol{x}, \boldsymbol{y} \rangle'(t_0) = \\
&= [x_1(t)y_1(t)]'(t_0) + \ldots + [x_n(t)y_n(t)]'(t_0) \\
&= [{x_1}'(t_0)y_1(t_0) + x_1(t_0){y_1}'(t_0)] + \ldots + [{x_n}'(t_0)y_n(t_0) + x_n(t_0){y_n}'(t_0)] \\
&= [{x_1}'(t_0)y_1(t_0) + \ldots + {x_n}'(t_0)y_n(t_0)] + [x_1(t_0){y_1}'(t_0) + \ldots + x_n(t_0){y_n}'(t_0)] \\
&= \sum_{i=1}^{n} {x_i}'(t_0)y_i(t_0) + \sum_{i=1}^{n} x(t_0){y_i}'(t_0) \\
&:= \langle \boldsymbol{x}'(t_0), \boldsymbol{y}(t_0) \rangle + \langle \boldsymbol{x}(t_0), \boldsymbol{y}'(t_0) \rangle.
\end{aligned}$$

□

Corollary 2.1.5. *Let $\boldsymbol{x} : I \to \mathbb{R}^n$ be a differentiable curve such that for every $t \in I$ the distance of $\boldsymbol{x}(t)$ from the origin remains constant i.e.,*

$$|\boldsymbol{x}(t)| = r > 0,$$

for every $t \in I$. Then for every $t_0 \in I$ the position vector $\boldsymbol{x}(t_0)$ of $\boldsymbol{x}$ at t_0 is orthogonal to the velocity vector $\boldsymbol{x}'(t_0)$ of $\boldsymbol{x}$ at t_0.



PROOF. If $|\boldsymbol{x}(t)| = r > 0$, for every $t \in I$, then $\boldsymbol{x}(t)$ lies on the sphere of radius r. Moreover,

$$r^2 = |\boldsymbol{x}(t)|^2 = \langle \boldsymbol{x}(t), \boldsymbol{x}(t) \rangle := \langle \boldsymbol{x}, \boldsymbol{x} \rangle(t),$$

hence by the Proposition 2.1.4(iv), and since $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$ is a constant function on I, we have that

$$0 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle'(t_0) = 2\langle \boldsymbol{x}(t_0), \boldsymbol{x}'(t_0) \rangle \Leftrightarrow 0 = \langle \boldsymbol{x}(t_0), \boldsymbol{x}'(t_0) \rangle \Leftrightarrow \boldsymbol{x}(t_0) \perp \boldsymbol{x}'(t_0).$$

□

DEFINITION 2.1.6. If $\boldsymbol{x} : I \to \mathbb{R}$ is a differentiable curve with continuous derivative $\boldsymbol{x}'$, its *length* $L_{ab}(\boldsymbol{x})$ between two values $a, b \in I$, where $a \leq b$, is defined by the corresponding integral of its speed i.e.,

$$L_{a,b}(\boldsymbol{x}) := \int_a^b v_{\boldsymbol{x}}(t)dt := \int_a^b |\boldsymbol{x}'(t)|dt.$$

By the definition of the Euclidean norm we have that

$$L_{a,b}(\boldsymbol{x}) = \int_a^b \sqrt{\left(\frac{dx_1}{dt}(t)\right)^2 + \left(\frac{dx_2}{dt}(t)\right)^2} dt,$$

if $\boldsymbol{x}(t) := \big(x_1(t), x_2(t)\big)$, and

$$L_{a,b}(\boldsymbol{x}) = \int_a^b \sqrt{\left(\frac{dx_1}{dt}(t)\right)^2 + \left(\frac{dx_2}{dt}(t)\right)^2 + \left(\frac{dx_3}{dt}(t)\right)^2} dt,$$

if $\boldsymbol{x}(t) := \big(x_1(t), x_2(t), x_3(t)\big)$. In the general case, where $\boldsymbol{x}(t) := \big(x_1(t), \ldots, x_n(t)\big)$, we have that

$$L_{a,b}(\boldsymbol{x}) = \int_a^b \sqrt{\left(\frac{dx_1}{dt}(t)\right)^2 + \ldots + \left(\frac{dx_3}{dt}(t)\right)^2}\, dt.$$

If for example, we consider the unit circle $\boldsymbol{c}(\theta) := (\cos\theta, \sin\theta)$, where $\theta \in [0, 2\pi]$, then we have that

$$\begin{aligned} v_{\boldsymbol{c}}(\theta) &:= |\boldsymbol{c}'(\theta)| \\ &:= \sqrt{{c_1}'(\theta)^2 + {c_2}'(\theta)^2} \\ &= \sqrt{(-\sin\theta)^2 + (\cos\theta)^2} \\ &= \sqrt{\sin^2\theta + \cos^2\theta} \\ &= \sqrt{1} \\ &= 1, \end{aligned}$$

and hence we get the expected value for the length of $\boldsymbol{c}$ between 0 and 2π:

$$L_{0,2\pi}(\boldsymbol{c}) := \int_0^{2\pi} v_{\boldsymbol{c}}(\theta) d\theta := \int_0^{2\pi} 1 d\theta = \int_0^{2\pi} d\theta = 2\pi - 0 = 2\pi.$$

Let the differentiable curve $\boldsymbol{x} : \mathbb{R} \to \mathbb{R}^2$ defined by

$$\boldsymbol{x}(t) := (e^t \cos t, e^t \sin t),$$

for every $t \in \mathbb{R}$. Its derivative $\boldsymbol{x}'$ is given by

$$\boldsymbol{x}'(t) := (e^t \cos t - e^t \sin t, e^t \sin t + e^t \cos t),$$

for every $t \in \mathbb{R}$. After some calculations we get

$$|\boldsymbol{x}(t)| = e^t \quad \& \quad |\boldsymbol{x}'(t)| = \sqrt{2} e^t \quad \& \quad \langle \boldsymbol{x}'(t), \boldsymbol{x}(t) \rangle = e^{2t},$$

for every $t \in \mathbb{R}$. Hence,

$$\frac{\langle \boldsymbol{x}'(t), \boldsymbol{x}(t) \rangle}{|\boldsymbol{x}'(t)||\boldsymbol{x}(t)|} = \frac{e^{2t}}{\sqrt{2} e^t e^t} = \frac{1}{\sqrt{2}},$$

for every $t \in \mathbb{R}$ i.e., the angle between $\boldsymbol{x}'(t)$ and $\boldsymbol{x}(t)$ is constant $\frac{\pi}{4}$, for every $t \in \mathbb{R}$. Moreover,

$$L_{0,1}(\boldsymbol{x}) = \int_0^1 \sqrt{2} e^t dt = \sqrt{2}(e - 1).$$

2.2. Open sets in $\mathbb{R}^n$

We consider vector-valued functions defined on appropriate subsets of $\mathbb{R}^n$ that we call open.

Definition 2.2.1. Let $x \in \mathbb{R}^n$ and $\epsilon > 0$. The *open ball* $\mathcal{B}(x, \epsilon)$ with center x and radius ϵ is defined by

$$\begin{aligned} \mathcal{B}(x, \epsilon) &:= \{y \in \mathbb{R}^n \mid d(x, y) < \epsilon\} \\ &:= \{y \in \mathbb{R}^n \mid |x - y| < \epsilon\} \\ &:= \{y \in \mathbb{R}^n \mid \sqrt{(x_1 - y_1)^2 + \ldots (x_n - y_n)^2} < \epsilon\}. \end{aligned}$$

We also say that $\mathcal{B}(x, \epsilon)$ is the open r-ball at x. The *closed ball* $\mathcal{B}(x, \epsilon]$ with center x and radius ϵ is defined by

$$\mathcal{B}(x, \epsilon] := \{y \in \mathbb{R}^n \mid d(x, y) \leq \epsilon\}.$$

If $U \subseteq \mathbb{R}^n$, we say that U is an *open* subset of $\mathbb{R}^n$, if

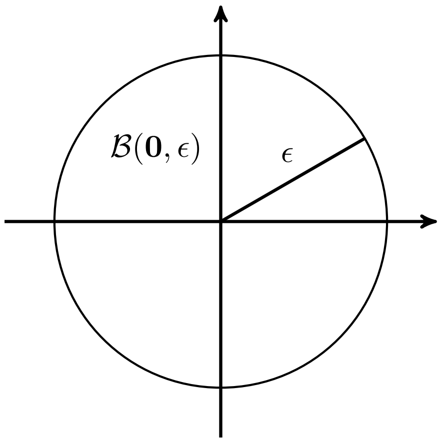
$$\forall_{x \in U} \exists_{\epsilon > 0} \big(\mathcal{B}(x, \epsilon) \subseteq U\big).$$

If $F \subseteq \mathbb{R}^n$, we say that F is a *closed* subset of $\mathbb{R}^n$, if its complement

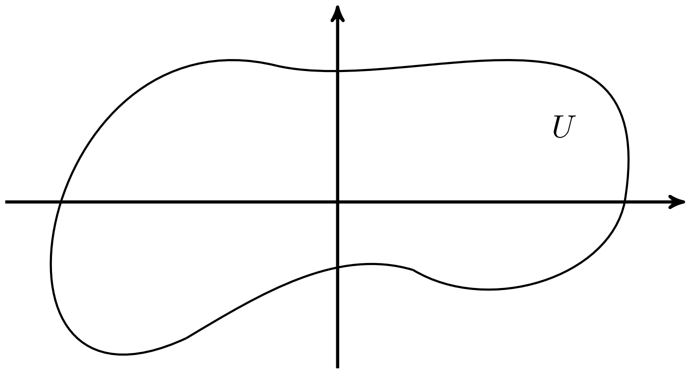
$$F^c := \{y \in \mathbb{R}^n \mid x \notin F\}$$

is open.

The open ϵ-ball $\mathcal{B}(\mathbf{0}, \epsilon)$ at the origin $(0, 0)$ is the open ϵ-disc around $(0, 0)$



and the closed ϵ-ball $\mathcal{B}(\mathbf{0}, \epsilon]$ at the origin $(0, 0)$ is the ϵ-disc around $(0, 0)$ *with* the ϵ-circle around the origin. It is easy to see that the open ϵ-ball $\mathcal{B}(\mathbf{0}, \epsilon)$, as any open ball, is an open set, since if we take any point in the disc, we can find a small disc around it that is included in the larger one. Note that the closed ϵ-ball $\mathcal{B}(\mathbf{0}, \epsilon]$ is *not* open, since any disc around a point at the ϵ-circle is not included in $\mathcal{B}(\mathbf{0}, \epsilon]$. It is clear though, that $\mathcal{B}(\mathbf{0}, \epsilon]$ is closed. Using a similar argument we can show that the interior U of the following curve in $\mathbb{R}^2$ is open in $\mathbb{R}^2$.



Note that the open ϵ-ball $\mathcal{B}(\mathbf{0}, \epsilon)$ in $\mathbb{R}$ at the origin 0 is the open interval $(-\epsilon, \epsilon)$.

PROPOSITION 2.2.2. *Let $n \geq 1$.*

(i) $\mathbb{R}^n$ and $\emptyset$ are both open and closed.

(ii) If $U \subseteq \mathbb{R}^n$, then U is open if and only if its complement U^c is closed.

(iii) If U, V are open in $\mathbb{R}^n$, then $U \cap V$ and $U \cup V$ are open in $\mathbb{R}^n$.

(iv) If F, K are closed in $\mathbb{R}^n$, then $F \cap K$ and $F \cup K$ are closed in $\mathbb{R}^n$.

(v) If $(U_i)_{i \in I}$ is a family of open sets in $\mathbb{R}^n$ i.e., U_i is open for every $i \in I$, then their union

$$\bigcup_{i \in I} U_i := \big\{x \in \mathbb{R}^n \mid \exists_{i \in I}\big(x \in U_i\big)\big\}$$

is open.

(vi) If $(F_i)_{i \in I}$ is a family of closed sets in $\mathbb{R}^n$ i.e., U_i is closed for every $i \in I$, then their intersection

$$\bigcap_{i \in I} F_i := \big\{x \in \mathbb{R}^n \mid \forall_{i \in I}\big(x \in F_i\big)\big\}$$

is closed.

PROOF. (i) If $x \in \mathbb{R}^n$, then $\mathcal{B}(x, 1) \subseteq \mathbb{R}^n$, and hence $\mathbb{R}^n$ is open. Consequently, $\emptyset$ is closed, since $\emptyset^c = \mathbb{R}^n$. The implication $x \in \emptyset \Rightarrow \mathcal{B}(x, 1) \subseteq \emptyset$ is trivially true, since its premise is false. Hence $\emptyset$ is also open, and $\mathbb{R}^n$ is also closed, since $(\mathbb{R}^n)^c = \emptyset$.

(ii) If U is open, then U^c is closed, since $(U^c)^c = U$ is open. If U^c is closed, then by definition $(U^c)^c = U$ is open.

(iii) First we show that $U \cap V$ is open. If $x \in U \cap V$, then $x \in U$ and $x \in V$. Since U is open, there is some $\epsilon_1 > 0$ such that $\mathcal{B}(x, \epsilon_1) \subseteq U$. Since V is open, there is some $\epsilon_2 > 0$ such that $\mathcal{B}(x, \epsilon_2) \subseteq Y$. If

$$\epsilon := \min\{\epsilon_1, \epsilon_2\},$$

then

$$\mathcal{B}(x, \epsilon) \subseteq V \cap U.$$

To show this, let $y \in \mathbb{R}^n$ such that $|y - x| < \epsilon \leq \epsilon_1$. Hence $y \in U$. Similarly, $|y - x| < \epsilon \leq \epsilon_2$, and hence $y \in Y$. Consequently, $y \in V \cap U$.

Next we show that $U \cup V$ is open. If $x \in U \cup V$, then $x \in U$, or $x \in V$. In the first case we have that $\mathcal{B}(x, \epsilon_1) \subseteq U \subseteq U \cup V$, and in the second we have that $\mathcal{B}(x, \epsilon_2) \subseteq V \subseteq U \cup V$.
(iv) We use the case (iii) and the equalities

$$(F \cap K)^c = F^c \cup K^c \quad \& \quad (F \cup K)^c = F^c \cap K^c.$$

(v) and (vi) is an exercise. □

The intersection of a countable family of open sets is not generally open. E.g.,

$$(0, 1] = \bigcap_{n \geq 1} \left(0, 1 + \frac{1}{n}\right),$$

and $(0, 1]$ is not open, as any non-trivial interval around 1 intersects $(1, +\infty)$. The union of a countable family of closed sets is not generally closed. E.g.,

$$(0, 1) = \bigcup_{n \geq 2} \left[\frac{1}{n}, 1 - \frac{1}{n}\right],$$

and $(0, 1)$ is not closed, since its complement $(-\infty, 0] \cup [1, +\infty)$ is not open. It is not hard to see that the cartesian product of open sets in $\mathbb{R}$ is an open set in the corresponding $\mathbb{R}^n$. E.g., the set

$$(0, 1) \times (-1, 1) := \{(x, y) \in \mathbb{R}^2 \mid x \in (0, 1) \ \& \ y \in (-1, 1)\}$$

is open in $\mathbb{R}^2$. Similarly the set

$$(0, 1) \times (-1, 1) \times \mathbb{R} := \{(x, y, z) \in \mathbb{R}^3 \mid x \in (0, 1) \ \& \ y \in (-1, 1)\}$$

is open in $\mathbb{R}^3$.

2.3. Partial derivatives

If U is an open subset of $\mathbb{R}^n$, and $x = (x_1, \ldots, x_n) \in U$, then for every $i \in \{1, \ldots, n\}$, there are appropriately small values of $h \in \mathbb{R}$ such that the point

$$(x_1, \ldots, x_i + h, \ldots, x_n) \in U,$$

and the following concept is well-defined.

DEFINITION 2.3.1. Let U be an open subset of $\mathbb{R}^n$, $x = (x_1, \ldots, x_n) \in U$, and $f : U \to \mathbb{R}$. If the following limit exists

$$\lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h},$$

we let

$$D_i f(x) := \frac{\partial f}{\partial x_i}(x) := \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h},$$

and we call $D_i f(x)$, or $\frac{\partial f}{\partial x_i}(x)$, the *$i$-th partial derivative* of f at x.

If $B_n := \{e_1, \ldots, e_i, \ldots, e_n\}$ is the standard basis of $\mathbb{R}^n$, we have that

$$D_i f(x) = \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}.$$

If for example $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$f(x, y) := x^2 y^3,$$

then

$$\begin{aligned} D_1 f(x) &:= \frac{\partial f}{\partial x}(x) \\ &:= \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h} \\ &= \lim_{h \to 0} \frac{(x + h)^2 y^3 - x^2 y^3}{h} \\ &= y^3 \lim_{h \to 0} \frac{(x + h)^2 - x^2}{h} \\ &= y^3 2x \\ &= 2xy^3, \end{aligned}$$

since the term in the right is the derivative of the function $g(x) = x^2$. I.e., to calculate $D_1 f(x)$ we treat y as a constant and we differentiate with respect to x. Similarly we have that

$$\begin{aligned} D_2 f(x) &:= \frac{\partial f}{\partial y}(x) \\ &:= \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h} \\ &= \lim_{h \to 0} \frac{x^2 (y + h)^3 - x^2 y^3}{h} \\ &= x^2 \lim_{h \to 0} \frac{(y + h)^3 - y^3}{h} \\ &= x^2 3y^2 \\ &= 3x^2 y^2, \end{aligned}$$

since the term in the right is the derivative of the function $h(y) = y^3$. I.e., to calculate $D_2 f(x)$ we treat x as a constant and we differentiate with respect to y.

If $f, g : U \to \mathbb{R}$, and $x \in U$ such that $D_i f(x)$ and $D_i g(x)$ exist, then by the properties of the derivative of real-valued functions on intervals of $\mathbb{R}$ we get immediately

$$D_i(f+g)(x) = D_i f(x) + D_i g(x),$$
$$D_i(\lambda f)(x) = \lambda D_i f(x),$$

for every $\lambda \in \mathbb{R}$.

Definition 2.3.2. Let U be an open subset of $\mathbb{R}^n$, $x; = (x_1, \ldots, x_n) \in U$, and $f : U \to \mathbb{R}$. If the partial derivatives at x

$$D_1 f(x) := \frac{\partial f}{\partial x_1}(x), \ldots, D_n f(x) := \frac{\partial f}{\partial x_n}(x)$$

exist, the *gradient* $(\mathrm{grad} f)(x)$ of f at x is the vector

$$(\mathrm{grad} f)(x) := \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right)$$
$$:= \Big(D_1 f(x), \ldots, D_n f(x)\Big).$$

E.g., if $f : \mathbb{R}^2 \to \mathbb{R}$ is defined as above by $f(x, y) := x^2 y^3$, then

$$(\mathrm{grad} f)(x) := (2xy^3, 3x^2y^2).$$

Because of the above linearity of D_i, we get immediately that if $f, g : U \to \mathbb{R}$, and $x \in U$ such that $D_i f(x)$ and $D_i g(x)$ exist, then

$$(\mathrm{grad}(f+g))(x) = (\mathrm{grad} f)(x) + (\mathrm{grad} g)(x),$$
$$(\mathrm{grad}(\lambda f))(x) = \lambda(\mathrm{grad} f)(x),$$

for every $\lambda \in \mathbb{R}$. If $D_i f(x)$ and $D_i g(x)$ exist, for every $x \in U$, we get

$$\mathrm{grad}(f+g) = \mathrm{grad} f + \mathrm{grad} g,$$
$$\mathrm{grad}(\lambda f) = \lambda \mathrm{grad} f,$$

for every $\lambda \in \mathbb{R}$. If $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(x, y) := x^2 y^3$, we showed that

$$D_1 f(x) := \frac{\partial f}{\partial x}(x) = 2xy^3 \quad \& \quad D_2 f(x) := \frac{\partial f}{\partial y}(x) = 3x^2y^2.$$

Since $D_1 f, D_2 f : \mathbb{R}^2 \to \mathbb{R}$, we can determine the *repeated partial derivatives*

$$D_1 D_1 f(x, y) := D_1^2 f(x, y) := \frac{\partial^2 f}{\partial x^2}(x, y) := (D_1(D_1 f))(x, y) =$$
$$= \frac{\partial(2xy^3)}{\partial x}(x, y) = 2y^3,$$
$$D_1 D_2 f(x, y) := \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(x, y) := \frac{\partial^2 f}{\partial x \partial y}(x, y) := (D_1(D_2 f))(x, y) =$$

$$= \frac{\partial(3x^2y^2)}{\partial x}(x,y) = 6xy^2,$$

$$D_2D_1f(x,y) := \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)(x,y) := \frac{\partial^2 f}{\partial y \partial x}(x,y) := (D_2(D_1f))(x,y) =$$

$$\frac{\partial(2xy^3)}{\partial y}(x,y) = 6xy^2,$$

$$D_2D_2f(x,y) := D_2^2f(x,y) := \frac{\partial^2 f}{\partial y^2}(x,y) := (D_2(D_2f))(x,y) =$$

$$= \frac{\partial(3x^2y^2)}{\partial y}(x,y) = 6x^2y.$$

Notice that

$$2y^3 = \frac{\partial^2 f}{\partial x^2}(x,y) \neq \left(\frac{\partial f}{\partial x}(x)\right)^2 = (2xy^3)^2 = 4x^2y^6.$$

But we have that

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(x,y) = 6xy^2 = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)(x,y).$$

This is not accident. One can show that if $U \subseteq \mathbb{R}^2$ is open and $f : U \to \mathbb{R}$ such that the partial derivatives

$$D_1f(x,y),\ D_2f(x,y),\ D_1D_2f(x,y),\ D_2D_1f(x,y)$$

exist and are continuous, then for every $(x,y) \in U$ we have that

$$D_1D_2f(x,y) = D_2D_1f(x,y).$$

We can have repeated partial derivatives for $n > 2$. If $f : \mathbb{R}^3 \to \mathbb{R}$ is defined by

$$f(x,y,z) = x^2yz^3,$$

then

$$D_1f(x,y,z) = 2xyz^3 \quad D_2D_1f(x,y,z) = 2xz^3 \quad D_3D_2D_1f(x,y,z) = 6xz^2,$$

and

$$D_3f(x,y,z) = 3x^2yz^2 \quad D_2D_3f(x,y,z) = 3x^2z^2 \quad D_1D_2D_3f(x,y,z) = 6xz^2$$

i.e.,

$$D_3D_2D_1f(x,y,z) = 6xz^2 = D_1D_2D_3f(x,y,z).$$

By our previous remark on the equality $D_1D_2f(x,y) = D_2D_1f(x,y)$, if all the related partial derivatives exist and are continuous we get

$$\begin{aligned} D_3D_2D_1f(x,y,z) &= D_3D_1D_2f(x,y,z) \\ &= D_1D_3D_2f(x,y,z) \\ &= D_1D_2D_3f(x,y,z). \end{aligned}$$

2.4. The chain rule

In this section we define when a function $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$, is differentiable at some point $x_0 \in U$. To motivate this definition we notice the following fact.

REMARK 2.4.1. Let U be an open subset of $\mathbb{R}$, $x_0 \in U$ and $f : U \to \mathbb{R}$. The following are equivalent:
(i) f is differentiable at x_0.
(ii) There are $\epsilon > 0, a \in \mathbb{R}$, and a function $g : (-\epsilon, \epsilon) \to \mathbb{R}$ such that

$$f(x_0 + h) - f(x_0) = ah + |h|g(h),$$

for every $h \in (-\epsilon, \epsilon)$, and

$$\lim_{h\to 0} g(h) = 0.$$

PROOF. If f is differentiable at x_0, then

$$a := f'(x_0) := \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} \in \mathbb{R},$$

and if $h \neq 0$, we define

$$\phi(h) = \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0),$$

while if $h = 0$, we define $\phi(0) := 0$. Clearly,

$$\lim_{h\to 0} \phi(h) = 0,$$

and for every h in some ϵ-interval around 0 we have that

$$f(x_0 + h) - f(x_0) = f'(x_0)h + h\phi(h).$$

If we define $g(h) := \phi(h)$, if $h \geq 0$, and $g(h) := -\phi(h)$, if $h < 0$, we have that

$$|h|g(h) = h\phi(h),$$

and we get the required equality

$$f(x_0 + h) - f(x_0) = ah + |h|g(h).$$

Of course,

$$\lim_{h\to 0} g(h) = 0.$$

For the converse, if $h \neq 0$, then

$$\frac{f(x_0 + h) - f(x_0)}{h} = \frac{ah + |h|g(h)}{h} = a + \frac{|h|}{h}g(h),$$

which converges to a, as h converges to 0 i.e., $a = f'(x_0)$. □

DEFINITION 2.4.2. Let U be an open subset of $\mathbb{R}^n$, $x_0 \in U$ and $f : U \to \mathbb{R}$. We say that f is *differentiable at* x_0, if

(a) The gradient of f at x_0

$$\mathrm{grad} f(x_0) := \big(D_1 f(x_0), \ldots, D_n f(x_0)\big) = \left(\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0)\right)$$

exists, and

(b) there is a function g defined on a small open ball around the origin $\mathbf{0}$ such that

$$\lim_{|h|\to 0} g(h) = 0,$$

and

$$\begin{aligned} f(x_0 + h) - f(x_0) &= \frac{\partial f}{\partial x_1}(x_0)h_1 + \ldots + \frac{\partial f}{\partial x_n}(x_0)h_n + |h|g(h) \\ &:= \big\langle (\mathrm{grad} f)(x_0), h \big\rangle + |h|g(h). \end{aligned}$$

We say that f is *differentiable on* U, if it is differentiable at every point of U.

To show that a function f as above is differentiable on U, it suffices to show that the gradient of f at every point of U exists, and that the partial derivatives on U are continuous functions (the proof is omitted).

PROPOSITION 2.4.3. *If U is an open subset of $\mathbb{R}^n$, $x_0 \in U$ and $f : U \to \mathbb{R}$, then f is differentiable at x_0, if all partial derivatives of f at x_0 exist in U and for each $i \in \{1, \ldots, n\}$ the function*

$$U \ni x \mapsto \frac{\partial f}{\partial x_i}(x)$$

is continuous at x_0.

PROOF. See [**4**], p. 322. □

In the one dimensional case the chain rule takes the form

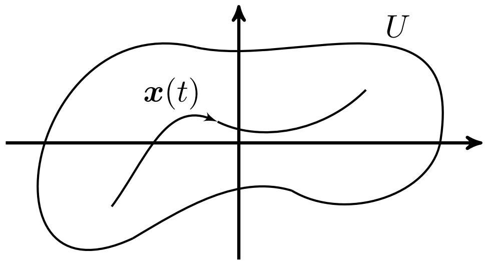
$$(f \circ g)'(t) = f'(g(t))g'(t),$$

where f and g are as indicated in the following diagram

$$\begin{array}{ccc} I & \xrightarrow{\;g\;} & U \subseteq \mathbb{R} \\ & {\scriptstyle f\circ g}\searrow & \downarrow {\scriptstyle f} \\ & & \mathbb{R}. \end{array}$$

Next we prove the generalisation of this rule.

PROPOSITION 2.4.4 (Chain rule). *Let I be an interval of $\mathbb{R}$, and $\boldsymbol{x} : I \to \mathbb{R}^n$ differentiable curve on I such that $\boldsymbol{x}(I) \subseteq U$,*



where U is an open subset of $\mathbb{R}^n$. If $f : U \to \mathbb{R}$ is differentiable on U, then the function

$$\begin{array}{ccc} I & \xrightarrow{\ \boldsymbol{x}\ } & U \subseteq \mathbb{R}^n \\ & {\scriptstyle f\circ \boldsymbol{x}} \searrow & \downarrow f \\ & & \mathbb{R}. \end{array}$$

$f \circ \boldsymbol{x} : I \to \mathbb{R}$ is differentiable, and for every $t \in I$ we have that

$$(f \circ \boldsymbol{x})'(t) = \big\langle (\mathrm{grad} f)(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle.$$

Proof. Let the quotient

$$\frac{f(\boldsymbol{x}(t+h)) - f(\boldsymbol{x}(t))}{h},$$

which, if we define

$$K := K(t,h) := \boldsymbol{x}(t+h) - \boldsymbol{x}(t),$$

and hence $\boldsymbol{x}(t+h) = K + \boldsymbol{x}(t)$, it becomes

$$\frac{f(\boldsymbol{x}(t) + K) - f(\boldsymbol{x}(t))}{h}.$$

Since f is differentiable on U, and $\boldsymbol{x}(t)$ is included in U, f is differentiable at $\boldsymbol{x}(t)$, for every $t \in I$. By the Definition 2.4.2 there is a function g such that

$$f(\boldsymbol{x}(t) + K) - f(\boldsymbol{x}(t)) = \big\langle (\mathrm{grad} f)(\boldsymbol{x}(t)), K \big\rangle + |K| g(K),$$

and

$$\lim_{|K| \to 0} g(K) = 0.$$

Hence,

$$\begin{aligned} \frac{f(\boldsymbol{x}(t+h)) - f(\boldsymbol{x}(t))}{h} &= \left\langle (\mathrm{grad} f)(\boldsymbol{x}(t)), \frac{\boldsymbol{x}(t+h) - \boldsymbol{x}(t)}{h} \right\rangle \\ &\quad + \frac{|\boldsymbol{x}(t+h) - \boldsymbol{x}(t)|}{h} g(K) \end{aligned}$$

$$= \left\langle (\mathrm{grad} f)(\boldsymbol{x}(t)), \frac{\boldsymbol{x}(t+h) - \boldsymbol{x}(t)}{h} \right\rangle$$
$$\pm \left| \frac{\boldsymbol{x}(t+h) - \boldsymbol{x}(t)}{h} \right| g(K).$$

If $h \to 0$, then

$$\left\langle (\mathrm{grad} f)(\boldsymbol{x}(t)), \frac{\boldsymbol{x}(t+h) - \boldsymbol{x}(t)}{h} \right\rangle \to \big\langle (\mathrm{grad} f)(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle,$$

and

$$\pm \left| \frac{\boldsymbol{x}(t+h) - \boldsymbol{x}(t)}{h} \right| g(K) \to \pm |\boldsymbol{x}'(t)| 0 = 0,$$

since if $h \to 0$, then $K := \boldsymbol{x}(t+h) - \boldsymbol{x}(t) \to 0$, and we use the fact that $\lim_{|K| \to 0} g(K) = 0$. □

Unfolding the chain rule we get

$$\begin{aligned}
(f \circ \boldsymbol{x})'(t) &= \left\langle \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}(t)), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}(t)) \right), \boldsymbol{x}'(t) \right\rangle \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial x_1}(\boldsymbol{x}(t)) {x_i}'(t) \\
&=: \sum_{i} \frac{\partial f}{\partial x_i}(\boldsymbol{x}(t)) \frac{dx_i}{dt}(t) \\
&:= \frac{\partial f}{\partial x_1}(\boldsymbol{x}(t)) \frac{dx_1}{dt}(t) + \ldots + \frac{\partial f}{\partial x_n}(\boldsymbol{x}(t)) \frac{dx_n}{dt}(t),
\end{aligned}$$

where $\boldsymbol{x}(t) = \big(x_1(t), \ldots, x_n(t)\big)$. For simplicity we also write

$$\frac{df(\boldsymbol{x}(t))}{dt} = \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \ldots + \frac{\partial f}{\partial x_n} \frac{dx_n}{dt}.$$

For example, let the following functions

$$\begin{array}{ccc}
\mathbb{R} & \xrightarrow{\ \boldsymbol{x}\ } & \mathbb{R}^3 \\
 & {\scriptstyle f \circ \boldsymbol{x}} \searrow & \downarrow {\scriptstyle f} \\
 & & \mathbb{R}
\end{array}$$

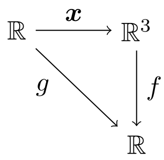
defined by $\boldsymbol{x}(t) := \big(e^t, t, t^2\big) = \big(x(t), y(t), z(t)\big)$ and $f(x, y, z) := x^2 yz$. Then f is differentiable on $\mathbb{R}^3$ by the Proposition 2.4.3, and by the chain rule we have that

$$\begin{aligned}
\frac{df(\boldsymbol{x}(t))}{dt} &= \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt} \\
&= 2xyze^t + x^2 z + x^2 y 2t.
\end{aligned}$$

As a simple example of applying the chain rule, let $f : \mathbb{R}^3 \to \mathbb{R}$ differentiable, and let $g : \mathbb{R} \to \mathbb{R}$, defined by

$$g(t) = f(P + tQ),$$

for every $t \in \mathbb{R}$, and some $P, Q \in \mathbb{R}^3$. In order to find $g'(t)$, let $\boldsymbol{x} : \mathbb{R} \to \mathbb{R}^3$



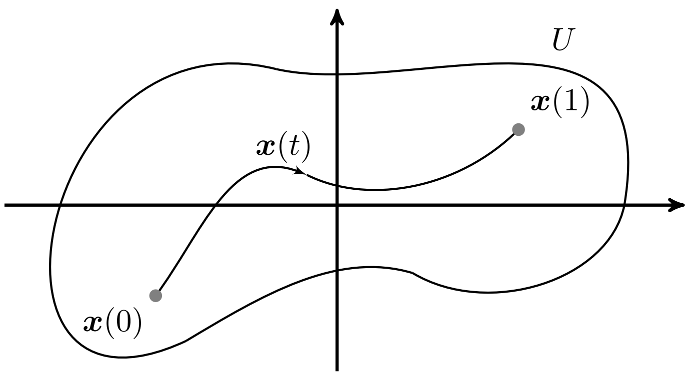
such that $g = f \circ \boldsymbol{x}$. Let

$$\boldsymbol{x}(t) = P + tQ = (p_1 + tq_1, p_2 + tq_2, p_3 + tq_3),$$

for every $t \in \mathbb{R}$. Since $\boldsymbol{x}'(t) = (q_1, q_2, q_3) = Q$, we get

$$g'(t) = (f \circ \boldsymbol{x})'(t) = \big\langle (\mathrm{grad} f)(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle = \big\langle (\mathrm{grad} f)(P + tQ), Q \big\rangle.$$

COROLLARY 2.4.5. *Let U be an open subset of $\mathbb{R}^n$ such that for every two points $x_0, x_1 \in U$ there is a differentiable curve $\boldsymbol{x} : [0, 1] \to U$ such that $\boldsymbol{x}(0) = x_0$ and $\boldsymbol{x}(0) = x_1$.*



If $f : U \to \mathbb{R}$ is differentiable on U, such that

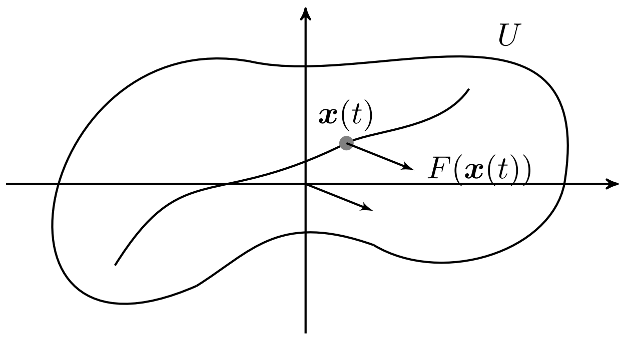
$$(\mathrm{grad} f)(x) = \mathbf{0},$$

for every $x \in U$, then f is constant on U.

PROOF. Exercise. □

2.5. Curve integrals

A vector field is a function $F : U \to \mathbb{R}^n$ that can be interpreted as a field of forces. If $\boldsymbol{x} : I \to U$ is a cure in U, the vector $\boldsymbol{x}(t)$ is interpreted as the position of the particle at time $t \in I$, and $F(\boldsymbol{x}(t))$ is the force acted upon the particle at position $\boldsymbol{x}(t)$. We may also say that the *particle is moving in the force field* F.



DEFINITION 2.5.1. Let U be an open subset of $\mathbb{R}^n$. A *vector field* on U is a function $F : U \to \mathbb{R}^n$. If F is represented by its coordinate functions i.e.,

$$F = (f_1, \dots, f_n),$$

F is *differentiable* on U, if each $f_i : U \to \mathbb{R}$ is differentiable on U. F is called *conservative*, if there is a differentiable function $V : U \to \mathbb{R}$ such that[1]

$$F = -\mathrm{grad} V.$$

In this case V is called a *potential energy* function for F.

If V is a potential energy function for F and $c \in \mathbb{R}$ is some constant, then

$$V + c$$

is also a potential energy function for F. If f is a differentiable function on U, then, by the Definition 2.4.2, we get the vector field on U defined by

$$U \ni x \mapsto (\mathrm{grad} f)(x).$$

Let $F : U \to \mathbb{R}^n$ be a differentiable vector field on U and $\boldsymbol{x} : I \to U$ a differentiable curve in U. Then the function $F \circ \boldsymbol{x} : I \to \mathbb{R}^n$ is well-defined

[1] The negative sign is only traditional, and it can be avoided.

$$\begin{array}{ccc} I & \xrightarrow{\ \boldsymbol{x}\ } & U \subseteq \mathbb{R}^n \\ & \underset{F \circ \boldsymbol{x}}{\searrow} & \downarrow F \\ & & \mathbb{R}^n, \end{array}$$

and let the function on I defined by

$$t \mapsto \big\langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle,$$

for every $t \in I$. E.g., let $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$F(x, y) := (e^{xy}, y^2),$$

for every $(x, y) \in \mathbb{R}^2$, and let $\boldsymbol{x} : \mathbb{R} \to \mathbb{R}^2$ be defined by

$$\boldsymbol{x}(t) := (t, \sin t),$$

for every $t \in \mathbb{R}$. Then

$$\boldsymbol{x}'(t) = (1, \cos t),$$
$$F(\boldsymbol{x}(t)) = \big(e^{t \sin t}, \sin^2 t\big),$$

and

$$\big\langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle = e^{t \sin t} + (\sin^2 t)(\cos t),$$

for every $t \in \mathbb{R}$.

Definition 2.5.2. Let $U \subseteq \mathbb{R}^n$ be open, $\boldsymbol{x} : [a, b] \to U$ a differentiable curve with a differentiable derivative curve $\boldsymbol{x}'$, and let $F : U \to \mathbb{R}^n$ be a differentiable vector field. The *curve integral* of F along $\boldsymbol{x}$ is defined by

$$\int_{\boldsymbol{x}} F := \int_a^b \big\langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle dt.$$

By the continuity of the inner product and our hypotheses on $\boldsymbol{x}$ and F the function on $[a, b]$ defined by

$$t \mapsto \big\langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle$$

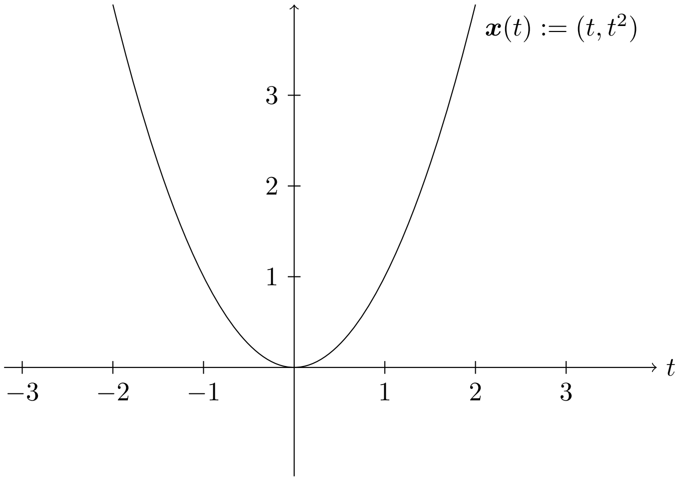
is continuous, hence Riemann-integrable. The above curve integral is a generalisation of the substitution method of the integral of functions in one variable:

$$\int_{u(a)}^{u(b)} f(u) du = \int_a^b f(u(t)) \frac{du}{dt} dt.$$

We use the following parametrisations of a linear, parabolic or circular segment:
(I) If $P, Q \in \mathbb{R}^n$, the linear segment "from P to Q" is parametrised by the curve $\boldsymbol{x} : [0, 1] \to \mathbb{R}^n$, defined by

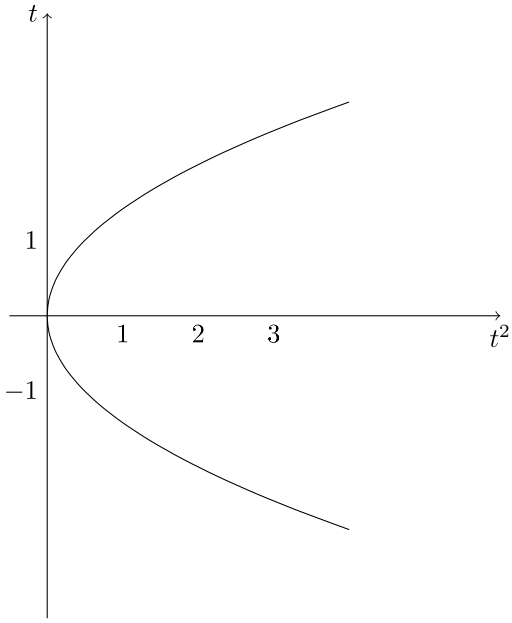
$$\boldsymbol{x}(t) := P + t(Q - P),$$

for every $t \in [0, 1]$. Clearly, $\boldsymbol{x}(0) = P$ and $\boldsymbol{x}(1) = Q$.
(II) A parabolic segment of the parabola $y = t^2$



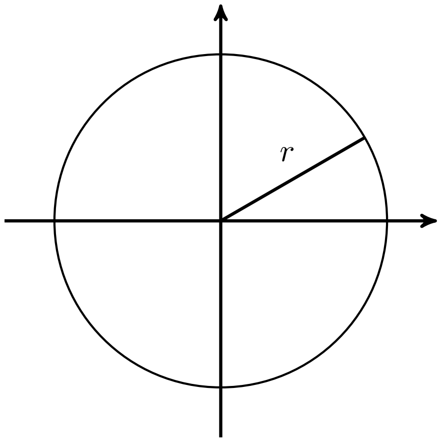
is parametrised by the curve $\boldsymbol{x}(t) := (t, t^2)$, where t is in a closed interval determined by the specifications of the respected problem.

(III) A parabolic segment of the parabola $x = t^2$



is parametrised by the curve $\boldsymbol{x}(t) := (t^2, t)$, where t is in a closed interval determined by the specifications of the respected problem.

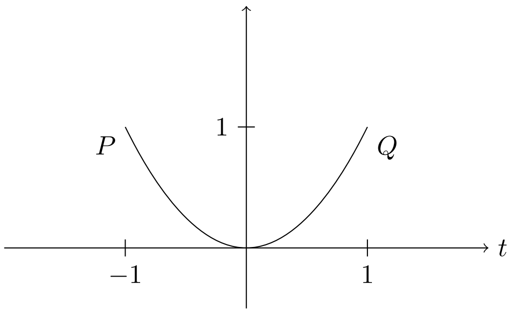
(IV) A circular segment of the circle of radius $r > 0$ centered at $(0, 0)$ in $\mathbb{R}^2$



is parametrised by the curve $\boldsymbol{x}(t) := (r\cos t, r\sin t)$, where t is in a closed interval determined by the specifications of the respected problem.
Let the vector filed $F : \mathbb{R}^2 \to \mathbb{R}^2$, defined by

$$F(x,y) := (x^2, xy),$$

for every $(x,y) \in \mathbb{R}^2$. To determine the integral of F over the parabolic segment



from $P := (-1,1)$ to $Q := (1,1)$ we have that $\boldsymbol{x}(t) = (t, t^2)$ and $\boldsymbol{x}'(t) = (1, 2t)$, and

$$F(\boldsymbol{x}(t)) = F(t, t^2) = (t^2, t^3),$$
$$\big\langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \big\rangle = t^2 + 2t^4,$$

hence, since $-1 \leq t \leq 1$,

$$\begin{aligned} \int_{\boldsymbol{x}} F &= \int_{-1}^{1} (t^2 + 2t^4) dt \\ &= \int_{-1}^{1} t^2 dt + \int_{-1}^{1} 2t^4 dt \\ &= \frac{2}{3} + \frac{4}{5}. \end{aligned}$$

To determine the curve integral of the vector field $F : \mathbb{R}^2 \to \mathbb{R}^2$, defined by

$$F(x, y) := (x^2 y, y^3),$$

for every $(x, y) \in \mathbb{R}^2$, over the line segment from $P := (0, 0)$ to $Q := (1, 1)$ we use the parametrisation of the segment

$$\boldsymbol{x}(t) = P + t(Q - P) = (0, 0) + t((1, 1) - (0, 0)) = t(1, 1) = (t, t),$$

where $t \in [0, 1]$, and hence $F(\boldsymbol{x}(t)) = F(t, t) = (t^3, t^3)$, $\boldsymbol{x}'(t) = (1, 1)$,

$$\big\langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t)\big\rangle = t^3 + t^3 = 2t^3,$$

and

$$\int_{\boldsymbol{x}} F = \int_0^1 2t^3 = 2\int_0^1 t^3 = 2\frac{1}{4} = \frac{1}{2}.$$

Let the vector field $F : \mathbb{R}^2 \setminus \{(0, 0)\} \to \mathbb{R}^2$, defined by

$$F(x, y) := \left(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2}\right),$$

for every (x, y) in the open subset $\mathbb{R}^2 \setminus \{(0, 0)\}$ of $\mathbb{R}^2$. To determine its integral over the circular segment of the circle of radius 3 around $(0, 0)$ from $P := (3, 0)$ to

$$Q := \left(\frac{3\sqrt{3}}{2}, \frac{3}{3}\right)$$

we consider the curve

$$\boldsymbol{x}(t) = (3\cos t, 3\sin t), \quad \boldsymbol{x}'(t) = (-3\sin t, 3\cos t), \qquad t \in \big[0, \frac{\pi}{6}\big],$$

since $\boldsymbol{x}(0) = P$ and $\boldsymbol{x}(\frac{\pi}{6}) = Q$. Since

$$\begin{aligned}
F(\boldsymbol{x}(t)) &= F(3\cos t, 3\sin t)\\
&= \left(\frac{-3\sin t}{(3\cos t)^2 + (-3\sin t)^2}, \frac{3\cos t}{(3\cos t)^2 + (-3\sin t)^2}\right)\\
&= \left(\frac{-3\sin t}{9}, \frac{3\cos t}{9}\right)\\
&= \frac{1}{3}(-\sin t, \cos t),
\end{aligned}$$

and

$$\big\langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t)\big\rangle = \sin^2 t + \cos^2 t = 1,$$

we get

$$\int_{\boldsymbol{x}} F = \int_0^{\frac{\pi}{6}} dt = \frac{\pi}{6}.$$

DEFINITION 2.5.3. A *path* in an open subset U of $\mathbb{R}^n$ is a finite sequence

$$p := (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m),$$

where $m \geq 1$, $\boldsymbol{x}_1 : [a_1, b_1] \to U, \ldots, \boldsymbol{x}_m : [a_m, b_m] \to U$ are curves in U such that

$$\boldsymbol{x}_1(b_1) = \boldsymbol{x}_2(a_2) \ \& \ \ldots \ \& \ \boldsymbol{x}_m(a_m) = \boldsymbol{x}_{m-1}(b_{m-1}).$$

A path p is called *differentiable* on U, if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are differentiable curves on U with differentiable derivative curves on U. We also say that p is *closed*, if

$$\boldsymbol{x}_1(a_1) = \boldsymbol{x}_m(b_m).$$

If $F : U \to \mathbb{R}^n$ is a differentiable vector field on U, and p is a differentiable path on U, the *path integral* of F over p is defined by

$$\int_p F := \int_{\boldsymbol{x}_1} F + \ldots + \int_{\boldsymbol{x}_m} F.$$

Clearly, a curve in U is a special case of a path in U.

2.6. Conservative vector fields

THEOREM 2.6.1. *Let $U \subseteq \mathbb{R}^n$ be open, and $F : U \to \mathbb{R}^n$ be a differentiable vector field on U.*

(I) Let $F = \mathrm{grad}V$, for some differentiable function $V : U \to \mathbb{R}$.

(a) If $\boldsymbol{x} : [a, b] \to U$ is a differentiable curve in U with $\boldsymbol{x}(a) = P$ and $\boldsymbol{x}(b) = Q$, then

$$\int_{\boldsymbol{x}} F = V(Q) - V(P).$$

(b) If $\boldsymbol{y} : [a, b] \to U$ is a differentiable curve in U with $\boldsymbol{y}(a) = P$ and $\boldsymbol{y}(b) = Q$, then

$$\int_{\boldsymbol{y}} F = \int_{\boldsymbol{x}} F.$$

(c) If $\boldsymbol{z} : [a, b] \to U$ is a closed differentiable curve in U i.e., $\boldsymbol{z}(a) = P = \boldsymbol{z}(b)$, then

$$\int_{\boldsymbol{z}} F = 0.$$

(II) Let $F = -\mathrm{grad}V$, for some differentiable function $V : U \to \mathbb{R}$.

(a) If $\boldsymbol{x} : [a, b] \to U$ is a differentiable curve in U with $\boldsymbol{x}(a) = P$ and $\boldsymbol{x}(b) = Q$, then

$$\int_{\boldsymbol{x}} F = V(P) - V(Q).$$

(b) If $\boldsymbol{y} : [a, b] \to U$ is a differentiable curve in U with $\boldsymbol{y}(a) = P$ and $\boldsymbol{y}(b) = Q$, then

$$\int_{\boldsymbol{y}} F = \int_{\boldsymbol{x}} F.$$

(c) If $\boldsymbol{z} : [a, b] \to U$ is a closed differentiable curve in U i.e., $\boldsymbol{z}(a) = P = \boldsymbol{z}(b)$, then

$$\int_{\boldsymbol{z}} F = 0.$$

Proof. We prove only the first part of (i) and the rest is an exercise. By the definition of the curve integral of f and the chain rule on $V \circ \boldsymbol{x}$

$$\begin{array}{ccc} [a,b] & \xrightarrow{\ \boldsymbol{x}\ } & U \subseteq \mathbb{R}^n \\ & {\scriptstyle V \circ \boldsymbol{x}} \searrow & \downarrow V \\ & & \mathbb{R}, \end{array}$$

we have that

$$\begin{aligned} \int_{\boldsymbol{x}} F &:= \int_a^b \langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \rangle dt \\ &= \int_a^b \langle (\mathrm{grad} V)(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \rangle dt \\ &= \int_a^b (V \circ \boldsymbol{x})'(t) dt \\ &= \big[V \circ \boldsymbol{x}\big]_a^b \\ &= V(\boldsymbol{x}(b)) - V(\boldsymbol{x}(a)) \\ &= V(Q) - V(P). \end{aligned}$$

□

Because of the above independence of the integral $\int_{\boldsymbol{x}} F$ of a conservative vector field from the curve connecting the points P and Q in U, we write

$$\int_P^Q F := \int_{\boldsymbol{x}} F = V(Q) - V(P),$$

where $\boldsymbol{x}$ is any curve in U from P to Q. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ a vector field defined by

$$F(x, y, z) := \big(2xy^3z, 3x^2y^2z, x^2y^3\big),$$

for every $(x, y, z) \in \mathbb{R}^3$. If $V : \mathbb{R}^3 \to \mathbb{R}$ is defined by

$$V(x, y, z) := x^2y^3z,$$

it is easy to see that $F = \mathrm{grad} V$. If $P := (1, -1, 2)$ and $Q := (-3, 2, 5)$, then

$$\int_P^Q F = V(Q) - V(P) = V(-3, 2, 5) - V(1, -1, 2) = 360 - (-2) = 362.$$

Let $G : \mathbb{R}^3 \setminus \{(0,0,0)\} \to \mathbb{R}^3$ be defined by

$$G(x,y,z) := \frac{k(x,y,z)}{|(x,y,z)|^3},$$

for every $(x,y,z) \in \mathbb{R}^3 \setminus \{(0,0,0)\}$ and some $k \in \mathbb{R}$. If $V : \mathbb{R}^3 \setminus \{(0,0,0)\} \to \mathbb{R}$ is defined by

$$V(x,y,z) := -\frac{k}{|(x,y,z)|},$$

for every $(x,y,z) \in \mathbb{R}^3 \setminus \{(0,0,0)\}$, then one can show (exercise) that

$$\mathrm{grad} V = G$$

i.e.,

$$\left(\frac{\partial V}{\partial x}(x,y,z), \frac{\partial V}{\partial y}(x,y,z), \frac{\partial V}{\partial z}(x,y,z)\right) = \frac{k}{|(x,y,z)|^3}(x,y,z).$$

If $P := (1,1,1)$ and $Q := (1,2,-1)$, then

$$\begin{aligned}\int_P^Q G &= V(Q) - V(P)\\ &= -\frac{k}{|Q|} - \left(-\frac{k}{|P|}\right)\\ &= -k\left(\frac{1}{|Q|} - \frac{1}{|P|}\right)\\ &= -k\left(\frac{1}{\sqrt{6}} - \frac{1}{\sqrt{3}}\right).\end{aligned}$$

2.7. Green's theorem on rectangles

DEFINITION 2.7.1. An *rectangle* $\mathcal{R}$ in $\mathbb{R}^2$ is a set of the form

$$\mathcal{R} := [a,b] \times [c,d] := \{(x,y) \in \mathbb{R}^2 \mid a \leq x \leq b \;\&\; c \leq y \leq d\},$$

and an *open rectangle* is a set of the form

$$\mathcal{R}^o := (a,b) \times (c,d) := \{(x,y) \in \mathbb{R}^2 \mid a < x < b \;\&\; c < y < d\},$$

If $f : \mathcal{R} \to \mathbb{R}$ is a continuous function[2], the *double integral* of f on $\mathcal{R}$ is defined by

$$\iint_{\mathcal{R}} f := \int_a^b \left(\int_c^d f(x,y)dy\right)dx.$$

[2]All functions defined on a rectangle that we are going to study here are going to be continuous.

Let $\mathcal{R} := [1,2] \times [-3,4]$ and $f : \mathcal{R} \to \mathbb{R}$, defined by

$$f(x,y) := x^2 y,$$

for every $(x,y) \in \mathbb{R}^2$. Then

$$\begin{aligned}
\iint_{\mathcal{R}} f &= \int_1^2 \left(\int_{-3}^4 x^2 y dy \right) dx \\
&= \int_1^2 x^2 \left(\int_{-3}^4 y dy \right) dx \\
&= \int_1^2 x^2 \frac{1}{2}(16-9) dx \\
&= \frac{7}{2} \int_1^2 x^2 dx \\
&= \frac{7}{2}\frac{1}{3}(2^3 - 1^3) \\
&= \frac{49}{6}.
\end{aligned}$$

If $U \subseteq \mathbb{R}^2$ is open, and let $F : U \to \mathbb{R}^2$ be a differentiable vector field on U such that

$$F(x,y) := \big(p(x,y), q(x,y)\big),$$

where $p, q : U \to \mathbb{R}$ are the components of F. If $\boldsymbol{x} : [a,b] \to U$ is a differentiable curve in U, then

$$\begin{aligned}
\int_{\boldsymbol{x}} F &:= \int_a^b \langle F(\boldsymbol{x}(t)), \boldsymbol{x}'(t) \rangle dt \\
&= \int_a^b \left(p(x,y)\frac{dx}{dt} + q(x,y)\frac{dy}{dt} \right) dt \\
&= \int_a^b p(x,y)dx + q(x,y)dy.
\end{aligned}$$

According to the next theorem, if we want to find the path-integral

$$\int_{(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)} F$$

of a differentiable vector field $F = (p,q)$ defined on an open rectangle, where the path

$$(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)$$

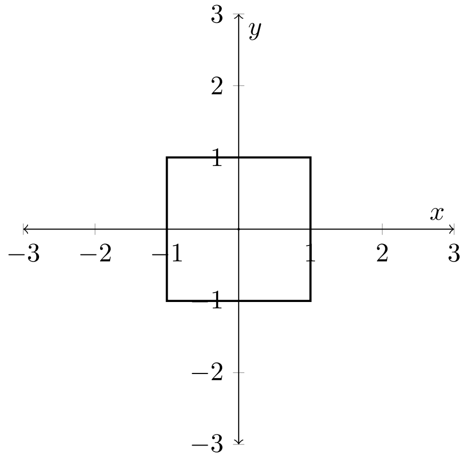
parametrises counterclockwise the rectangle $\mathcal{R}$, it suffices to calculate the double integral

$$\iint_{\mathcal{R}} \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y} \right).$$

Hence we do not need to calculate the curve integrals separately i.e., to use the equality

$$\int_{(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4)} F = \int_{\boldsymbol{x}_1} F + \int_{\boldsymbol{x}_2} F + \int_{\boldsymbol{x}_3} F + \int_{\boldsymbol{x}_4} F.$$

If e.g., we consider the rectangle $[-1,1]\times[-1,1]$



a path that parametrises it counterclockwise is the following sequence of linear segments

$$\big((-1,-1)\to(1,-1),(1,-1)\to(1,1),(1,1)\to(-1,1),(-1,1)\to(-1,-1)\big).$$

The proof of the next theorem is omitted.

THEOREM 2.7.2 (Green's theorem on rectangles). *Let $F:(a,b)\times(c,d)\to\mathbb{R}^2$ be a differentiable vector field on the open rectangle $(a,b)\times(c,d)$ such that*

$$F(x,y) := \big(p(x,y), q(x,y)\big),$$

for every $(x,y)\in(a,b)\times(c,d)$. Then

$$\int_{(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4)} p(x,y)dx + q(x,y)dy = \iint_{\mathcal{R}} \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right).$$

Let the vector field $F:\mathbb{R}^2\to\mathbb{R}^2$, defined by

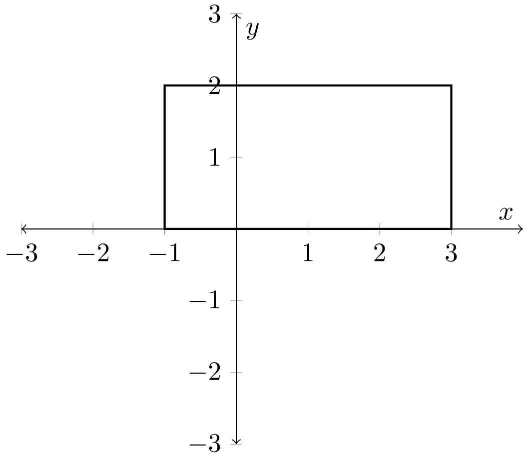
$$F(x,y) = \big(3xy, x^2\big),$$

for every $(x,y)\in\mathbb{R}^2$. Hence,

$$p(x,y) = 3xy, \qquad q(x,y) = x^2,$$

$$\frac{\partial q}{\partial x} = 2x \quad \& \quad \frac{\partial p}{\partial y} = 3x.$$

The integral of F around the following rectangle



is calculated with the use of Green's theorem as follows

$$\begin{aligned}
\int_{(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{x}_4)} p(x,y)dx + q(x,y)dy &= \iint_{\mathcal{R}} \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right) \\
&= \int_{-1}^{3} \left(\int_0^2 (2x - 3x)dy\right) dx \\
&= \int_{-1}^{3} \left(\int_0^2 (-x)dy\right) dx \\
&= \int_{-1}^{3} (-x)\left(\int_0^2 dy\right) dx \\
&= \int_{-1}^{3} (-x)2dx \\
&= -2\int_{-1}^{3} xdx \\
&= -8.
\end{aligned}$$

CHAPTER 3

Appendix

3.1. Solution to Exercise 2(ii), Sheet 1

If $f : \mathbb{R} \to \mathbb{R}$, we say that f is *differentiable* at $x_0 \in \mathbb{R}$, if there is some $l \in \mathbb{R}$ such that

$$l = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

where using the (ϵ-δ)-definition of the notion of limit, this means that

$$\forall_{\epsilon > 0} \exists_{\delta_f(\epsilon) > 0} \forall_{h \neq 0} \bigg(|h| < \delta_f(\epsilon) \Rightarrow \left| \frac{f(x_0 + h) - f(x_0)}{h} - l \right| \leq \epsilon \bigg).$$

This necessarily unique limit l is called the *derivative* of f at x_0, and it is denoted by $f'(x_0)$. The function f is called *continuous* at x_0, if

$$\lim_{h \to 0} f(x_0 + h) = f(x_0) \Leftrightarrow \lim_{h \to 0} \big[f(x_0 + h) - f(x_0) \big] = 0,$$

where using the (ϵ-δ)-definition of the notion of limit, this means that

$$\forall_{\epsilon > 0} \exists_{\delta_f(\epsilon) > 0} \forall_{h \in \mathbb{R}} \bigg(|h| < \delta_f(\epsilon) \Rightarrow \big| f(x_0 + h) - f(x_0) \big| \leq \epsilon \bigg).$$

If f is differentiable at x_0, then f is continuous at x_0. To show this we remark that the function

$$\phi(h) := \begin{cases} \frac{f(x_0+h) - f(x_0)}{h} & , h \neq 0 \\ f'(x_0) & , h = 0 \end{cases}$$

is continuous at 0. Since for $h \neq 0$ we have that

$$f(x_0 + h) - f(x_0) = h\phi(h) \Rightarrow f(x_0 + h) = h\phi(h) + f(x_0),$$

we get

$$\lim_{h \to 0} f(x_0 + h) = \lim_{h \to 0} \big[h\phi(h) + f(x_0) \big] = 0f'(x_0) + f(x_0) = f(x_0).$$

The function f is called *differentiable*, if it is differentiable at *every* $x_0 \in \mathbb{R}$, and it is called *continuous*, if it is continuous at *every* $x_0 \in \mathbb{R}$. If

$$C(\mathbb{R}) := \{ f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous} \},$$

$$D(\mathbb{R}) := \{ f : \mathbb{R} \to \mathbb{R} \mid f \text{ is differentiable} \},$$

the previous remark implies that

$$D(\mathbb{R}) \subseteq C(\mathbb{R}).$$

Next we show that $D(\mathbb{R})$ is a linear space, and for $C(\mathbb{R})$ we work similarly. Clearly, the constant function 0 is differentiable and its derivative is at every $x_0 \in \mathbb{R}$ again 0. Next we show that if $f, g \in D(\mathbb{R})$, then $f+g \in D(\mathbb{R})$. Let $x_0 \in \mathbb{R}$. Suppose that

$$l = \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h} \quad \& \quad \lim_{h\to 0} \frac{g(x_0+h)-g(x_0)}{h} = m.$$

Using the triangle inequality

$$|a+b| \leq |a| + |b|,$$

where $a, b \in \mathbb{R}$, we have that

$$\left|\frac{(f+g)(x_0+h)-(f+g)(x_0)}{h} - (l+m)\right| =$$

$$\left|\frac{f(x_0+h)+g(x_0+h)-f(x_0)-g(x_0)}{h} - l - m\right| =$$

$$\left|\frac{f(x_0+h)-f(x_0)}{h} - l + \frac{g(x_0+h)-g(x_0)}{h} - m\right| \leq$$

$$\left|\frac{f(x_0+h)-f(x_0)}{h} - l\right| + \left|\frac{g(x_0+h)-g(x_0)}{h} - m\right|,$$

and since these two terms become arbitrarily small, for appropriate h, we get that

$$(f+g)'(x_0) = l + m = f'(x_0) + g'(x_0),$$

and since x_0 is an arbitrary real number, we conclude that $f+g \in D(\mathbb{R})$. Finally, we show that if $a \in \mathbb{R}$ and $f \in D(\mathbb{R})$, then $a \cdot f \in D(\mathbb{R})$. Since $|ab| = |a||b|$, for every $a, b \in \mathbb{R}$, we have that

$$\left|\frac{(a\cdot f)(x_0+h)-(a\cdot f)(x_0)}{h} - al\right| = \left|a\left(\frac{f(x_0+h)-f(x_0)}{h} - l\right)\right|$$

$$= |a|\left|\frac{f(x_0+h)-f(x_0)}{h} - l\right|,$$

and since the right term becomes arbitrarily small for appropriate h, we get

$$(a\cdot f)'(x_0) = al = af'(x_0).$$

Since $x_0 \in \mathbb{R}$ is arbitrary, we conclude that $a \cdot f \in D(\mathbb{R})$.

3.2. On the solution of the Exercise 4(i), Sheet 3

The fact that $\int f \in D(\mathbb{R})$ is explained by the following fundamental result.

THEOREM 3.2.1. *Let $a, b, c, d \in \mathbb{R}$ such that $a \leq b \leq c \leq d$, and $f : [a, d] \to \mathbb{R}$ continuous. The function $\phi : [a, d] \to \mathbb{R}$, defined by*

$$\phi(x) := \int_a^x f(t)dt,$$

for every $x \in [a, d]$ is differentiable in $[a, d]$ and $\phi'(x) = f(x)$.

PROOF. We will use the following two basic properties of the Riemann integral.
(1) If $m \leq f(t) \leq M$, for every $t \in [b, c]$, then

$$m(c - b) \leq \int_b^c f(t)dt \leq M(c - b).$$

(2)
$$\int_a^c f(t)dt = \int_a^b f(t)dt + \int_b^c f(t)dt.$$

If $x_0 \in [a, d]$, then by (2) we have that

$$\begin{aligned}\frac{\phi(x_0 + h) - \phi(x_0)}{h} &:= \frac{\int_a^{x_0+h} f(t)dt - \int_a^{x_0} f(t)dt}{h} \\ &= \frac{\int_a^{x_0} f(t)dt + \int_{x_0}^{x_0+h} f(t)dt - \int_a^{x_0} f(t)dt}{h} \\ &= \frac{\int_{x_0}^{x_0+h} f(t)dt}{h}.\end{aligned}$$

Since f is continuous on the compact interval $[x_0, x_0 + h]$, let $s, s' \in [x_0, x_0 + h]$ such that

$$f(s) := \min\{f(t) \mid t \in [x_0, x_0 + h]\} := m,$$

$$f(s') := \max\{f(t) \mid t \in [x_0, x_0 + h]\} := M.$$

By (1) we have that

$$m(x_0 + h - x_0) \leq \int_{x_0}^{x_0+h} f(t)dt \leq M(x_0 + h - x_0) \Leftrightarrow$$

$$f(s)h \leq \int_{x_0}^{x_0+h} f(t)dt \leq f(s')h \overset{h \neq 0}{\Rightarrow}$$

$$f(s) \leq \frac{\int_{x_0}^{x_0+h} f(t)dt}{h} \leq f(s').$$

If $h \to 0$, then $s, s' \to x_0$, and by the continuity of f we get $f(s) \to f(x_0)$ and $f(s') \to f(x_0)$. By the sandwich lemma we get

$$\phi'(x_0) := \lim_{h\to 0} \frac{\phi(x_0+h) - \phi(x_0)}{h} = \lim_{h\to 0} \frac{\int_{x_0}^{x_0+h} f(t)dt}{h} = f(x_0).$$

□

3.3. Solution to Exercise 4(iv)-(v), Sheet 4

Let $A := [a_{ij}] \in M_{m,n}(\mathbb{R})$, $B := [b_{jk}] \in M_{n,l}(\mathbb{R})$ and $D := [d_{kr}] \in M_{l,s}(\mathbb{R})$.
(iv) By the definition of the multiplication of matrices we have that

$$AB := [a_{ij}][b_{jk}] := \left[\sum_{j=1}^{n} a_{ij}b_{jk}\right],$$

$$BD := [b_{jk}][d_{kr}] := \left[\sum_{k=1}^{l} b_{jk}d_{kr}\right],$$

$$(AB)D := \left[\sum_{j=1}^{n} a_{ij}b_{jk}\right][d_{kr}] := \left[\sum_{k=1}^{l}\left(\sum_{j=1}^{n} a_{ij}b_{jk}\right)d_{kr}\right],$$

$$A(BD) := [a_{ij}]\left[\sum_{k=1}^{l} b_{jk}d_{kr}\right] := \left[\sum_{j=1}^{n} a_{ij}\left(\sum_{k=1}^{l} b_{jk}d_{kr}\right)\right].$$

Since

$$\sum_{k=1}^{l}\left(\sum_{j=1}^{n} a_{ij}b_{jk}\right)d_{kr} = \sum_{j=1}^{n} a_{ij}\left(\sum_{k=1}^{l} b_{jk}d_{kr}\right) = \sum_{k=1}^{l}\sum_{j=1}^{n} a_{ij}b_{jk}d_{kr},$$

we get that $(AB)D = A(BD)$.
(v) Since

$$A^t := [\alpha_{ji}] \in M_{n,m}(\mathbb{R}), \qquad \alpha_{ji} := a_{ij},$$
$$B^t := [\beta_{kj}] \in M_{l,n}(\mathbb{R}), \qquad \beta_{kj} := b_{jk},$$

the product $B^tA^t \in M_{l,m}(\mathbb{R})$ is well-defined. Moreover, we have that

$$(AB)^t := \left[\sum_{j=1}^{n} a_{ij}b_{jk}\right]^t := [\gamma_{ki}],$$

where

$$\gamma_{ki} := \sum_{j=1}^{n} a_{ij}b_{jk}.$$

Since

$$B^tA^t := [\beta_{kj}][\alpha_{ji}] := \left[\sum_{j=1}^{n} \beta_{kj}\alpha_{ji}\right] = \left[\sum_{j=1}^{n} b_{jk}a_{ij}\right] = \left[\sum_{j=1}^{n} a_{ij}b_{jk}\right],$$

we get $B^tA^t = (AB)^t$.

Bibliography

[1] H. B. Enderton: *Elements of Set Theory*, Academic Press, 1977.
[2] O. Forster: *Analysis 2*, Vieweg + Teubner, 2007.
[3] W. Greub: *Linear Algebra*, Springer-Verlag, 1975.
[4] S. Lang: *Undergraduate Analysis*, Springer, 1983.
[5] S. Lang: *Linear Algebra*, Springer-Verlag, 1987.
[6] S. Lang: Calculus of Several Variables, Springer, 1987.
[7] J. E. Marsden, A. Tromba: Vector Calculus, W. H. Freeman and Company Publishers, 2012.