Mathematics for Physicists II

Iosif Petrakis

Fakultät für Physik, Ludwig-Maximilians-Universität München
Summer term 2019

Acknowledgement

These lecture notes owe a lot to the excellent book of Lang [**17**]. I would like to thank Leonid Kolesnikov, Felix Palm, Anne Reif, Jago Silberbauer, Christian Koke and Florian Ingerl for giving the tutorials of this lecture course and for their comments and suggestions. Last, but not least, I would like to thank the students who followed the course.

Contents

CHAPTER 1

Introduction: Notions of space and categories

There is a big variety of notions of space in mathematics. For example, there are Hilbert spaces, which are a special case of inner product spaces, Banach spaces, which are a special case of normed spaces, metric spaces, topological spaces, measure spaces, manifolds and many more. Roughly speaking, an S-*space*, is a pair $\mathcal{S} := (X; S)$, where X is a set, which is called the *carrier set* of $\mathcal{S}$, and S is a *space-structure* associated to X. Usually, S is a family of functions associated to X and some fundamental sets in mathematics, like the set of real numbers $\mathbb{R}$. If a set X can be equipped with many different space-structures, then X is a very "interesting" set. E.g., if $n \geq 1$, the set

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

of n-tuples of real numbers is the carrier set of a Hilbert space, and hence of a Banach space, of a metric space, and of a topological space, but also of an n-manifold, and of a measure space. All these space-structures, which often are interconnected, shed a different light in the mathematical "nature" of $\mathbb{R}^n$. A central tool in the study of S-spaces is the notion of an S-map. Roughly speaking, if $\mathcal{S} := (X; S)$ and $\mathcal{T} := (Y; T)$ are S-spaces, an S-*map* is a function $f : X \to Y$ that "preserves" the corresponding space-structures. Let $C(\mathcal{S}, \mathcal{T})$ be the set of all S-maps from $\mathcal{S}$ to $\mathcal{T}$. It is expected that the identity map id_X on X, where

$$\mathrm{id}_X : X \to X, \qquad x \mapsto x,$$

is an S-map. Moreover, if $\mathcal{U} := (Z; U)$ is an S-space, and if $f : X \to Y$ is in $C(\mathcal{S}, \mathcal{T})$ and $g : Y \to Z$ is in $C(\mathcal{T}, \mathcal{U})$, their composition

$$g \circ f : X \to Z, \qquad x \mapsto g(f(x)),$$

is in $C(\mathcal{S}, \mathcal{U})$

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & Y \arrow[r, "g"] & Z. \end{tikzcd}$$

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

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

By the definition of composition of functions we have that $f \circ \mathrm{id}_X = f$, or that the following diagram *commutes*

$$\begin{tikzcd} X \arrow[r, "\mathrm{id}_X"] \arrow[rr, bend right, "f"'] & X \arrow[r, "f"] & Y, \end{tikzcd}$$

and that $\mathrm{id}_Y \circ f = f$, or that the following diagram commutes

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "f"'] & Y \arrow[r, "\mathrm{id}_Y"] & Y. \end{tikzcd}$$

Note that the composition of functions is associative i.e., if

$$X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} W,$$

then $h \circ (g \circ f) = (h \circ g) \circ f$, or the following outer diagram commutes

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rrr, bend left, "(h \circ g) \circ f"] \arrow[rr, bend right, "g \circ f"'] \arrow[rrr, bend right=40, "h \circ (g \circ f)"'] & Y \arrow[r, "g"] \arrow[rr, bend left, "h \circ g"] & Z \arrow[r, "h"] & W. \end{tikzcd}$$

If $f \in C(\mathcal{S}, \mathcal{T})$ and $g \in C(\mathcal{T}, \mathcal{S})$ such that the following diagrams commute

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "\mathrm{id}_X"'] & Y \arrow[r, "g"] \arrow[rr, bend left, "\mathrm{id}_Y"] & X \arrow[r, "f"'] & Y \end{tikzcd}$$

i.e., $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$, the corresponding S-spaces are considered to be the "same" S-spaces. We call then f, or g, an S-*isomorphism* between $\mathcal{S}$ and $\mathcal{T}$, while the S-spaces $\mathcal{S}$ and $\mathcal{T}$ are called S-*isomorphic*. In this case we write $\mathcal{S} \simeq \mathcal{T}$, and we also write $(f, g) : \mathcal{S} \simeq \mathcal{T}$ to express that f and g "prove" $\mathcal{S} \simeq \mathcal{T}$. It is easy to see that if $(f, g) : \mathcal{S} \simeq \mathcal{T}$ and $(f, g') : \mathcal{S} \simeq \mathcal{T}$, then $g = g'$.

Remark 1.0.1. Let $\mathcal{S} := (X; S), \mathcal{T} := (Y; T)$, and $\mathcal{U} := (Z; U)$ be S-spaces.
(*i*) $\mathcal{S} \simeq \mathcal{S}$.
(*ii*) If $\mathcal{S} \simeq \mathcal{T}$, then $\mathcal{T} \simeq \mathcal{S}$.
(*iii*) If $\mathcal{S} \simeq \mathcal{T}$ and $\mathcal{T} \simeq \mathcal{U}$, then $\mathcal{S} \simeq \mathcal{U}$.

Proof. Exercise. □

An important theme in the study of S-spaces is the construction of new S-spaces from given ones. A notion of an S-space is mathematically fruitful, if many such constructions are possible. It is desirable to have a notion of S-*subspace*, namely, if $Y \subseteq X$ and $\mathcal{S} := (X; S)$ is an S-space, an S-space $\mathcal{S}_{|Y} := (Y; S_{|Y})$ is defined. Moreover, if $\mathcal{S}$ and $\mathcal{T}$ are S-spaces, their S-*product* $\mathcal{S} \times \mathcal{T}$ is also expected to be defined. It is also useful to be able to "glue" together S-spaces, or to add an S-structure on the set $C(\mathcal{S}, \mathcal{T})$ of S-maps from $\mathcal{S}$ to $\mathcal{T}$. Usually, some specific S-spaces have a special role among other S-spaces, and they can be used to classify many other S-spaces. A "classification theorem" for S-spaces determines a large class of S-spaces that are S-isomorphic to some distinguished S-space. E.g., in the theory of linear spaces and linear maps, the linear space $\mathbb{R}^n$ has a very special role among the so-called finite-dimensional linear spaces. Usually, the set of S-maps from an S-space $\mathcal{S}$ to such a distinguished S-space provides information on the original space $\mathcal{S}$. The linear space of linear maps from a linear space $\mathcal{V}$ to $\mathbb{R}$ is such an example.

The collection of all S-spaces and S-maps between them forms the *category* of S-spaces. The study of categories of mathematical objects and abstract "maps" between them is the subject matter of *Category Theory* (see e.g. [**3**]). The theory of sets (see e.g., [**10**]), and the category theory are the most popular "dialects" in the language of modern mathematics.

Noe that, although the actual constructions of new S-spaces from given ones depend on the specific S-structure under study, their abstract properties are common to all categories of S-spaces. E.g., the product of two Hilbert spaces $\mathcal{H}_1 \times \mathcal{H}_2$ is a different mathematical object from the product of two topological spaces $\mathcal{X}_1 \times \mathcal{X}_2$, but the basic behaviour of the product $\mathcal{H}_1 \times \mathcal{H}_2$ in the category of Hilbert spaces is the same to the basic behaviour of the product $\mathcal{X}_1 \times \mathcal{X}_2$ in the category of topological spaces. As we say in category theory, both objects satisfy the *universal property* of the products. According to it, if $\mathcal{S} \times \mathcal{T} := \big(X \otimes Y; S \times T\big)$ is the product of the S-spaces $\mathcal{S}$ and $\mathcal{T}$, there are S-maps $\mathtt{pr}_X : X \otimes Y \to X$ and $\mathtt{pr}_Y : X \otimes Y \to Y$ such that for every S-space $\mathcal{U} := (Z, U)$ and every S-maps $f : Z \to X$ and $g : Z \to Y$, there is a unique S-map $h : Z \to X \otimes Y$ such that the following inner diagrams commute

$$\begin{tikzcd} & Z \arrow[dl, "f"'] \arrow[d, dashed, "h"] \arrow[dr, "g"] & \\ X & X \otimes Y \arrow[l, "\mathtt{pr}_X"] \arrow[r, "\mathtt{pr}_Y"'] & Y \end{tikzcd}$$

i.e., $f = \mathtt{pr}_X \circ h$ and $g = \mathtt{pr}_Y \circ h$. Quite often, but not always, the set $X \otimes Y$ is the set-theoretic product $X \times Y$ of the sets X and Y i.e., the set

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

Definition 1.0.2 (Eilenberg, Mac Lane (1945)). A *category* $\boldsymbol{C}$ is a structure $(C_0, C_1, \mathtt{dom}, \mathtt{cod}, \circ, \mathbf{1})$, where

(i) C_0 is the collection of the *objects* of $\boldsymbol{C}$,

(ii) C_1 is the collection of the *arrows* of $\boldsymbol{C}$,

(iii) For every f in C_1, $\mathtt{dom}(f)$, the *domain* of f, and $\mathtt{cod}(f)$, the *codomain* of f, are objects in C_0, and we write $f : A \to B$, where $A = \mathtt{dom}(f)$ and $B = \mathtt{cod}(f)$,

(iv) If $f : A \to B$ and $g : B \to C$ are arrows of $\boldsymbol{C}$ i.e., $\mathtt{dom}(g) = \mathtt{cod}(f)$, there is an arrow $g \circ f : A \to C$, which is called the *composite* of f and g,

(v) For every A in C_0, there is an arrow $\mathbf{1}_A : A \to A$, the *identity arrow* of A,

such that the following conditions are satisfied:

(a) If $f : A \to B$, then $f \circ \mathbf{1}_A = f = \mathbf{1}_B \circ f$.

(b) If $f : A \to B$, $g : B \to C$ and $h : C \to D$, then $h \circ (g \circ f) = (h \circ g) \circ f$.

If A, B are in C_0, we denote by $\mathrm{Hom}_{\boldsymbol{C}}(A, B)$, or simply by $\mathrm{Hom}(A, B)$, if $\boldsymbol{C}$ is clear from the context, the collection of arrows f in C_1 with $\mathtt{dom}(f) = A$ and $\mathtt{cod}(f) = B$.

The objects of a category are not necessarily sets. E.g., in the next chapter we will study certain properties of the category of (real) linear spaces $\mathbf{Lin}_{\mathbb{R}}$, or simpler $\mathbf{Lin}$, that has as objects the (real) *linear spaces*, which are sets equipped with a linear structure. The arrows of $\mathbf{Lin}$ are certain functions between the carrier sets of the corresponding linear spaces, which are called *linear maps*.

Example 1.0.3. If X is not equipped with some S-structure, or, equivalently, if it is equipped with the empty structure, the corresponding category of S-spaces is the *category of sets* $\mathbf{Set}$. Its objects are sets, and its arrows are functions between sets. If A is a set, then $\mathbf{1}_A := \mathrm{id}_A$, and the composition of arrows is the composition of functions.

Next we give an example of a category the arrows of which are *not* functions.

Example 1.0.4. The category $\mathbf{Rel}$ has objects sets, and an arrow $f : A \to B$ is any subset of $A \times B$ i.e., any binary relation on A, B. If A is a set, let

$$\mathbf{1}_A := \big\{(a, a') \in A \times A \mid a = a'\big\},$$

while, if $R \subseteq A \times B$ and $S \subseteq B \times C$, let

$$S \circ R := \big\{(a, c) \in A \times C \mid \exists_{b \in B}\big((a, b) \in R \ \&\ (b, c) \in S\big)\big\}.$$

Definition 1.0.5. A *partially ordered set*, or a *poset*, is a pair $(I, \preceq)$, where I is a set, and $\preceq\ \subseteq I \times I$ satisfying the following conditions:

(i) $\forall_{i \in I}\big(i \preceq i\big)$.

(ii) $\forall_{i,j \in I}\big(i \preceq j \ \&\ j \preceq i \Rightarrow i = j\big)$.

(iii) $\forall_{i,j,k \in I}\big(i \preceq j \ \&\ j \preceq k \Rightarrow i \preceq k\big)$.

If $(J, \leq)$ is any poset, a function $m : I \to J$ is called *monotone*, if

$$\forall_{i,i' \in I}\big(i \preceq i' \Rightarrow m(i) \leq m(i')\big).$$

Clearly, $(\mathbb{R}, \leq)$ is a poset, which is *linearly ordered* i.e., $\forall_{a,b\in\mathbb{R}}\big(a \leq b \ \vee \ b \leq a\big)$.

Example 1.0.6. The category **Pos** has objects posets and an arrow $m : I \to J$ is any monotone function. If $(I, \preceq)$ is a poset, we have $\mathbf{1}_A := \mathrm{id}_A$, and $\circ$ in **Pos** is the composition of functions.

Categories can be used to describe various physical phenomena (see [**8**], [**9**]). If we consider a physical system of some type A (e.g., an electron), and if an operation (e.g., a measurement) is performed on it, which results in a system of some type B, this situation can be described by an arrow $f : A \to B$. An operation g on the system that follows f can be described by the arrow $g : B \to C$, and $g \circ f$ denotes the consecutive application of f and g. The trivial operation of "no operation" on a system of type A is denoted by $\mathbf{1}_A$. For many non-trivial applications of category theory to mathematical physics see [**21**].

Next we define the right notion of "map" between categories.

Definition 1.0.7. Let $\boldsymbol{C}$ and $\boldsymbol{D}$ be categories. A *covariant functor*, or simply a *functor* from $\mathcal{C}$ to $\mathcal{D}$ is a pair $F = (F_0, F_1)$, where:

(i) F_0 maps an object A of $\boldsymbol{C}$ to an object $F_0(A)$ of $\mathcal{D}$,

(ii) F_1 maps an arrow $f : A \to B$ of $\boldsymbol{C}$ to an arrow $F_1(f) : F_0(A) \to F_0(B)$ of $\boldsymbol{D}$,

such that the following conditions are satisfied:

(a) For every A in C_0 we have that $F_1(\mathbf{1}_A) = \mathbf{1}_{F_0(A)}$

$$\begin{tikzcd}
F_0(A) \arrow[d, bend right, "\mathbf{1}_{F_0(A)}"'] \arrow[d, bend left, "F_1(\mathbf{1}_A)"] \\
F_0(A).
\end{tikzcd}$$

(b) If $f : A \to B$ and $g : B \to C$, then $F_1(g \circ f) = F_1(g) \circ F_1(f)$ i.e., the following diagram commutes

$$\begin{tikzcd}
F_0(A) \arrow[r, "F_1(f)"] \arrow[rr, bend right, "F_1(g \circ f)"'] & F_0(B) \arrow[r, "F_1(g)"] & F_0(C),
\end{tikzcd}$$

where for simplicity we use the same symbol for the operation of composition in the categories $\boldsymbol{C}$ and $\boldsymbol{D}$. In this case we write[1] $F : \boldsymbol{C} \to \boldsymbol{D}$.

A *contravariant functor* from $\mathcal{C}$ to $\mathcal{D}$ is a pair $F := (F_0, F_1)$, where:

[1]In the literature it is often written $F(C)$ and $F(f)$, instead of $F_0(C)$ and $F_1(f)$.

(i) F_0 maps an object A of $\boldsymbol{C}$ to an object $F_0(A)$ of $\mathcal{D}$,
(ii') F_1 maps an arrow $f : A \to B$ of $\boldsymbol{C}$ to an arrow $F_1(f) : F_0(B) \to F_0(A)$ of $\boldsymbol{D}$, such that the following conditions are satisfied:
(a) $F_1(\mathbf{1}_A) = \mathbf{1}_{F_0(A)}$, for every A in C_0.
(b') If $f : A \to B$ and $g : B \to C$, then $F_1(g \circ f) = F_1(f) \circ F_1(g)$ i.e., the following diagram commutes

$$\begin{tikzcd} F_0(C) \arrow[r, "F_1(g)"] \arrow[rr, bend right, "F_1(g \circ f)"'] & F_0(B) \arrow[r, "F_1(f)"] & F_0(A), \end{tikzcd}$$

In this case we write $F : \boldsymbol{C}^{\mathrm{op}} \to \boldsymbol{D}$.

Example 1.0.8. If $\boldsymbol{C}$ is a category, the *identity functor* on $\boldsymbol{C}$ is the pair $\mathrm{Id}_{\boldsymbol{C}} := (\mathrm{Id}_0^{\boldsymbol{C}}, \mathrm{Id}_1^{\boldsymbol{C}}) : \boldsymbol{C} \to \boldsymbol{C}$, where $\mathrm{Id}_0^{\boldsymbol{C}}(X) := X$, for every X in C_0, and if $f : X \to Y$, then $\mathrm{Id}_1^{\boldsymbol{C}}(f) := f$.

Example 1.0.9. The pair $F := (F_0, F_1) : \mathbf{Set} \to \mathbf{Rel}$, where $F_0(X) := X$, and if $f : X \to Y$, then $F_1(f) := \{(a, b) \in A \times B \mid b = f(a)\} := \mathrm{Gr}(f)$, is a covariant functor from $\mathbf{Set}$ to $\mathbf{Rel}$.

Example 1.0.10. The pair $(G_0, G_1) : \mathbf{Set} \to \mathbf{Set}$, where $G_0(X) := \mathbb{F}(X) := \{\phi : X \to \mathbb{R}\}$, and if $f : X \to Y$, then $G_1(f) : \mathbb{F}(Y) \to \mathbb{F}(X)$ is defined by

$$[G_1(f)])(\theta) := \theta \circ f$$

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "\theta \circ f"'] & Y \arrow[d, "\theta"] \\ & \mathbb{R}, \end{tikzcd}$$

for every $\theta \in \mathbb{F}(Y)$, is a contravariant functor from $\mathbf{Set}$ to $\mathbf{Set}$. If X is a set, then

$$[G_1(\mathrm{id}_X)])(\phi) := \phi \circ \mathrm{id}_X = \phi$$

and since $\phi \in \mathbb{F}(X)$ is arbitrary, we conclude that $G_1(\mathrm{id}_X) = \mathrm{id}_{\mathbb{F}(X)} := \mathrm{id}_{G_0(X)}$. If $f : X \to Y$ and $g : Y \to Z$, then $G_1(f) : \mathbb{F}(Y) \to \mathbb{F}(X)$, $G_1(g) : \mathbb{F}(Z) \to \mathbb{F}(Y)$ and $G_1(g \circ f) : \mathbb{F}(Z) \to \mathbb{F}(X)$. Moreover, if $\eta \in \mathbb{F}(Z)$, we have that

$$\begin{aligned} [G_1(g \circ f)](\eta) &:= \eta \circ (g \circ f) \\ &= (\eta \circ g) \circ f \\ &:= [G_1(f)](\eta \circ g) \\ &:= G_1(f)\big([G_1(g)](\eta)\big) \\ &:= [G_1(f) \circ G_1(g)](\eta). \end{aligned}$$

The next most important concept is that of a natural transformation.

DEFINITION 1.0.11. Let $\boldsymbol{C}, \boldsymbol{D}$ be categories and $F := (F_0, F_1), G := (G_0, G_1)$ functors from $\boldsymbol{C}$ to $\boldsymbol{D}$. A *natural transformation* from F to G is a family of arrows in $\boldsymbol{D}$ of the form

$$\tau_C : F_0(C) \to G_0(C)$$

for every C in C_0, and every $f : C \to C'$ in $\boldsymbol{C}$, the following diagram commutes

$$\begin{tikzcd} F_0(C) \arrow[r, "F_1(f)"] \arrow[d, "\tau_C"'] & F_0(C') \arrow[d, "\tau_{C'}"] \\ G_0(C) \arrow[r, "G_1(f)"'] & G_0(C'). \end{tikzcd}$$

We denote a natural transformation τ from F to G by $\tau : F \Rightarrow G$.

EXAMPLE 1.0.12. Let $\mathrm{Id}_{\mathbf{Set}} := (\mathrm{Id}_0^{\mathbf{Set}}, \mathrm{Id}_1^{\mathbf{Set}})$ be the identity functor on **Set** (Example 1.0.8), and let the functor $H := (H_0, H_1) : \mathbf{Set} \to \mathbf{Set}$, defined by

$$H_0(X) := \mathbb{F}(\mathbb{F}(X)) := \{\Phi : \mathbb{F}(X) \to \mathbb{R}\},$$

and if $f : X \to Y$, then $H_1(f) : \mathbb{F}(\mathbb{F}(X)) \to \mathbb{F}(\mathbb{F}(Y))$ is defined by

$$\big[H_1(f)\big](\Phi) := \Phi \circ G_1(f)$$

$$\begin{tikzcd} \mathbb{F}(Y) \arrow[r, "G_1(f)"] \arrow[dr, "\Phi \circ G_1(f)"'] & \mathbb{F}(X) \arrow[d, "\Phi"] \\ & \mathbb{R}, \end{tikzcd}$$

where G_1 is defined in the Example 1.0.10. Then the family of arrows in **Set**

$$\tau_X : X \to \mathbb{F}(\mathbb{F}(X))$$

$$\tau_X(x) := \hat{x},$$

$$\hat{x}(\phi) := \phi(x),$$

for every $x \in X$, $\phi \in \mathbb{F}(X)$, and X in **Set**, is a natural transformation from $\mathrm{Id}_{\mathbf{Set}}$ to H i.e., the following diagram commutes

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\tau_X"'] & Y \arrow[d, "\tau_Y"] \\ \mathbb{F}(\mathbb{F}(X)) \arrow[r, "H_1(f)"'] & \mathbb{F}(\mathbb{F}(Y)). \end{tikzcd}$$

EXAMPLE 1.0.13. If $\boldsymbol{C}, \boldsymbol{D}$ are categories the *functor category* $\mathrm{Fun}(\boldsymbol{C}, \boldsymbol{D})$ has objects the functors from $\boldsymbol{C}$ to $\boldsymbol{D}$, and if $F, G : \boldsymbol{C} \to \boldsymbol{D}$, an arrow from F to G is a natural transformation from F to G. The identity arrow $\mathbf{1}_F : F \Rightarrow F$ is the family of arrows $(\mathbf{1}_F)_C : F_0(C) \to F_0(C)$, where $(\mathbf{1}_F)_C := \mathbf{1}_{F_0(C)}$, and the following diagram trivially commutes

$$\begin{tikzcd} F_0(C) \arrow[r, "F_1(f)"] \arrow[d, "\mathbf{1}_{F_0(C)}"'] & F_0(C') \arrow[d, "\mathbf{1}_{F_0(C')}"] \\ F_0(C) \arrow[r, "F_1(f)"'] & F_0(C'). \end{tikzcd}$$

If $F, G, H : \boldsymbol{C} \to \boldsymbol{D}$, $\tau : F \Rightarrow G$ and $\sigma : G \Rightarrow H$, the composite arrow $\sigma \circ \tau$ is defined by

$$(\sigma \circ \tau)_C := \sigma_C \circ \tau_C : F_0(C) \to H_0(C),$$

for every C in C_0, and, if $f : C \to C'$, the required commutativity of the following outer diagram

$$\begin{tikzcd} F_0(C) \arrow[r, "F_1(f)"] \arrow[d, "\tau_C"] \arrow[dd, bend right, "(\sigma \circ \tau)_C"'] & F_0(C') \arrow[d, "\tau_{C'}"'] \arrow[dd, bend left, "(\sigma \circ \tau)_{C'}"] \\ G_0(C) \arrow[r, "G_1(f)"'] \arrow[d, "\sigma_C"] & G_0(C') \arrow[d, "\sigma_{C'}"'] \\ H_0(C) \arrow[r, "H_1(f)"'] & H_0(C') \end{tikzcd}$$

is shown by the commutativity of the inner diagrams as follows:

$$\begin{aligned} (\sigma \circ \tau)_{C'} \circ F_1(f) &:= \big(\sigma_{C'} \circ \tau_{C'}\big) \circ F_1(f) \\ &= \sigma_{C'} \circ \big(\tau_{C'} \circ F_1(f)\big) \\ &= \sigma_{C'} \circ \big(G_1(f) \circ \tau_C\big) \\ &= \big(\sigma_{C'} \circ G_1(f)\big) \circ \tau_C \\ &= \big(H_1(f) \circ \sigma_C\big) \circ \tau_C \\ &= H_1(f) \circ \big(\sigma_C \circ \tau_C\big) \\ &:= H_1(f) \circ (\sigma \circ \tau)_C. \end{aligned}$$

It is straightforward to show now that $\mathrm{Fun}(\boldsymbol{C}, \boldsymbol{D})$ is indeed a category.

CHAPTER 2

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 carrier sets that preserves their linear structure. Both, the inner product spaces and the normed spaces, are linear spaces with some extra topological structure. Hence, the linear spaces are instrumental in the mathematical description of physical reality. The linear structure of $\mathbb{R}^3$ is a fundamental component of the geometric representation of the classical physical world. Throughout these lecture notes, when we write $\mathbb{R}^n$, we mean that $n \geq 1$.

2.1. Linear spaces and linear subspaces

Definition 2.1.1. 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 $(\mathrm{LS}_1) - (\mathrm{LS}_4)$ the following conditions are satisfied:

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 $\mathcal{V}$.

EXAMPLE 2.1.2. 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 $\mathbb{C}$.

EXAMPLE 2.1.3. 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 $\mathbb{R}$.

The Example 2.1.3 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 2.1.2 is a special case of the Example 2.1.3. If $f, g \in \mathbb{F}(X)$ and $a \in \mathbb{R}$,

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

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

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

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

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

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

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 $\mathbb{C}$.

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

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

REMARK 2.1.4. 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 2.1.5. 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 2.1.6. 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 2.1.3). 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 2.1.7. 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 $e_1 := (1, 0), e_2 := (0, 1)$, and $(x, y) \in \mathbb{R}^2$, we get $\mathbb{R}^2 = \langle e_1, e_2\rangle$, since

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

PROPOSITION 2.1.8. *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. $\square$

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 2.1.8 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 $\mathcal{V}$.

2.2. Finite-dimensional linear spaces

DEFINITION 2.2.1. 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

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

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, $a_1 x_1 + \ldots + a_n x_n = \mathbf{0}$, and $a_i \neq 0$, then

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

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

Remark 2.2.2. 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

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

(ii) Y is linearly independent if and only if

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

(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. E.g., for (i) we use the Corollary 6.1.3 and the Lemma 6.1.1 in the Appendix.
(iii) $1 \cdot \mathbf{0} = \mathbf{0}$, and $\{\mathbf{0}\}$ is a linearly dependent subset of X.
(iv) It follows immediately by Remark 2.1.4(vi).
(v) If we suppose that $\emptyset$ is a linearly dependent subset of X i.e.,

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

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

Example 2.2.3. The following n-vectors in $\mathbb{R}^n$

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

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 2.2.4. 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 2.2.5. 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 2.2.1 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 $i \in \{1, \ldots, n\}$ such that $x_i = \mathbf{0}$, then $0x_1 + 0x_{i-1} + 1x_i + 0x_{i+1} + \ldots + 0x_n = \mathbf{0}$, which by the hypothesis of linear independence is impossible[2]. $\square$

DEFINITION 2.2.6. 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*.

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

EXAMPLE 2.2.7. The set $E_n := \{e_1, \ldots, e_n\}$ of the linearly independent elements in $\mathbb{R}^n$ that were defined in the Example 2.2.3 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 2.2.8. 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 2.2.9. *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 2.2.5. $\square$

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

DEFINITION 2.2.10. 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.

[2]This also follows from the Remark 2.2.2(vii)

THEOREM 2.2.11 (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 $v_1, \ldots, v_r$. $\square$

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 2.2.12 (Exchange lemma (Steinitz)). *Let $n, m \geq 1$, let $\{v_1, \ldots v_n\}$ be 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 2.2.5 $w_1 \neq \mathbf{0}$, 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 2.2.13. *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 2.2.12(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 2.2.14. *If B_1, B_2 are finite bases of a linear space $\mathcal{V} := (X; +, \mathbf{0}, \cdot)$, 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 2.2.13 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 2.2.14 the following concept is well-defined.

Definition 2.2.15. 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 2.2.16. *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. $\square$

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 2.2.17. *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 2.1.4(vi) $M_1 := \{y_1\}$ is linearly independent. By the principle of the excluded middle[3] (PEM), we have that M_1 is either a maximal set of linearly independent elements of Y, hence by the Corollary 2.2.16(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. $\square$

Proposition 2.2.18. *If X is a linear space, and $Y, Z \preceq X$, such that*[4]

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

[3] See section 6.1 of the Appendix.

[4] The unfolding of a "unique existence"-formula $\exists_{!x \in X} \phi(x)$ is found in the section 6.1 of the Appendix.

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 2.2.19. *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 2.2.20. *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. $\square$

2.3. Existence of a basis

A trivial linear space has the empty set as a basis. In this section we show that a non-trivial linear space has always a basis. The proof of this fact requires the use of *Zorn's lemma* (see section 6.2 of the Appendix).

DEFINITION 2.3.1. A subset C of a poset $(I, \preceq)$ (see Definition 1.0.5) is called a *chain* in I, or a *totally ordered* subset of I, if

$$\forall_{c,c' \in C}\big(c \preceq c' \ \vee \ c' \preceq c\big).$$

A subset J of I is *bounded* in I, if there is $i_0 \in I$ such that $\forall_{j \in J}\big(j \preceq i_0\big)$. In this case i_0 is called a *bound* of J. An element i_0 of I is called *maximal* in I, if

$$\forall_{i \in I}\big(i_0 \preceq i \Rightarrow i = i_0\big).$$

A bound i_0 of I itself is called the *maximum* element[5] of I. If the poset $(I, \preceq)$ is clear from the context, we just say C is a chain, J is bounded, and i_0 is a maximal element. As usual, for simplicity we say that I is a poset, and we do not write the whole structure $(I, \preceq)$, when $\preceq$ is clear from the context.

[5]A maximum element i_0 is uniquely determined i.e., if j_0 is also bound of I, then $j_0 = i_0$. The maximum element is also a maximal element, while the converse is not generally the case.

The powerset $\mathcal{P}(I)$ of a set I is the set of all subsets of I, and it is partially ordered by the relation $A \subseteq B$, "the subset A is included in the subset B",

$$A \subseteq B :\Leftrightarrow \forall_{i \in I}\big(i \in A \Rightarrow i \in B\big).$$

An infinite, countable chain C in $\mathcal{P}(X)$ can take the form

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots,$$

and

$$\bigcup_{n=1}^{\infty} A_n := \big\{i \in I \mid \exists_{n \geq 1}\big(i \in A_n\big)\big\}$$

is a bound of C. Clearly, X is the maximum element of $\mathcal{P}(X)$, and $\emptyset$ is the minimum element of $\mathcal{P}(X)$, where the notions of minimal and minimum are dual to those of maximal and maximum.

Zorn's lemma (ZL): If I is a non-empty[6] poset, such that every chain in I is bounded, then I has a maximal element.

Theorem 2.3.2 (ZL). *A non-trivial linear space X has a basis.*

Proof. If we define the set

$$I(X) := \{Y \subseteq X \mid Y \text{ is linearly independent}\},$$

then $\emptyset \in I(X)$, hence $I(X)$ is non-empty. If $Y, Z \in I(X)$, we define $Y \preceq Z :\Leftrightarrow Y \subseteq Z$, which is a partial order on $I(X)$. If $C \subseteq I(X)$ is a chain in $I(X)$, then

$$\bigcup C := \{x \in X \mid \exists_{A \in C}(x \in A)\}$$

is a bound of C in $\mathcal{P}(X)$, and it is also a bound in $I(X)$ i.e., $\bigcup C \in I(X)$. To show this, we use the Remark 2.2.2(ii). Let $x_1, \ldots, x_n \in \bigcup C$, for some $n \geq 1$. By the definition of $\bigcup C$ there are $Y_1, \ldots, Y_n \in C$ such that $x_1 \in Y_1 \ \& \ \ldots \ \& \ x_n \in Y_n$. Since C is a chain there is some $i \in \{1, \ldots, n\}$ such that $Y_1 \subseteq Y_i \ \& \ \ldots \ \& \ Y_n \subseteq Y_i$. Hence, $\{x_1, \ldots, x_n\} \subseteq Y_i$, and since Y_i is linearly independent, $\{x_1, \ldots, x_n\}$ is also linearly independent. Since the set $\{x_1, \ldots, x_n\}$ is an arbitrary finite subset of $\bigcup C$, we conclude that $\bigcup C$ is in $I(X)$. Since C is an arbitrary chain in $I(X)$, the hypothesis of ZL is satisfied. Hence, by ZL the poset $I(X)$ has a maximal element B. We show that B is a basis of X. Since $B \in I(X)$, it is a linearly independent subset of X. It remains to show that B generates X. Let $x \in X$, and suppose that $x \notin \langle B \rangle$. By the Lemma 2.2.20 we have that $B \cup \{x\}$ is linearly independent, which contradicts the hypothesis of maximality of B. Hence, $x \in \langle B \rangle$. $\square$

[6] If $I = \emptyset$, then it is easy to show that the rule Efq (see section 6.1 of the Appendix) implies that if every chain of $\emptyset$ is bounded, then $\emptyset$ has a maximal element.

Notice that the previous proof of existence of a basis is very "indirect", as it provides no method, or algorithm, to find, or construct a basis. One can show[7] that the Theorem 2.3.2 implies ZL, hence "the existence of a base of a linear space" and ZL are equivalent (over ZF).

One can show similarly the following stronger version of Theorem 2.3.2. Note that this version generalises the Corollary 2.2.16(iv), without using the hypothesis of the existence of a (finite) basis, while if $Y = \emptyset$, it implies the Theorem 2.3.2.

THEOREM 2.3.3 (ZL). *If Y is a linearly independent subset of a non-trivial linear space X, there is a basis B of X, such that $Y \subseteq B$.*

PROOF. Exercise. $\square$

2.4. Linear maps

DEFINITION 2.4.1. 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 2.4.2. 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 $\texttt{pr}_i^B : X \to \mathbb{R}$, defined by

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

[7] For that see [**6**]. In [**19**] many statements from classical mathematics are shown to be equivalent to the axiom of choice.

$$\begin{tikzcd}
X \arrow[r, "f_B"] \arrow[dr, "\mathtt{pr}_i^B"'] & \mathbb{R}^n \arrow[d, "\mathtt{pr}_i"] \\
& \mathbb{R}
\end{tikzcd}$$

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 2.4.3. 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. If $f, g \in \mathcal{L}(X, Y)$, then $f + g \in \mathcal{L}(X, Y)$, since if $x, x' \in X$, then

$$\begin{aligned}
\big(f + g\big)(x + x') &:= f(x + x') + g(x + x') \\
&= \big(f(x) + f(x')\big) + \big(g(x) + g(x')\big) \\
&= \big(f(x) + g(x)\big) + \big(f(x') + g(x')\big) \\
&:= \big(f + g\big)(x) + \big(f + g\big)(x'),
\end{aligned}$$

and if $b \in \mathbb{R}$, then

$$(f+g)(b\cdot x) := f(b\cdot x)+g(b\cdot x) = b\cdot f(x)+b\cdot g(x) = b\cdot\big(f(x)+g(x)\big) := b\cdot\big[\big(f+g\big)(x)\big].$$

If $a \in \mathbb{R}$, and $f \in \mathcal{L}(X, Y)$, then $a \cdot f \in \mathcal{L}(X, Y)$, since

$$\begin{aligned}
(a \cdot f)(x + x') &:= a \cdot f(x + x') = a \cdot \big[f(x) + f(x')\big] = a \cdot f(x) + a \cdot f(x') \\
&:= (a \cdot f)(x) + (a \cdot f)(x'),
\end{aligned}$$

and

$$(a \cdot f)(bx) := a \cdot f(b \cdot x) = a \cdot \big[b \cdot f(x)\big] = (ab) \cdot f(x) = b \cdot \big[a \cdot f(x)\big] := b \cdot \big[(a \cdot f)(x)\big].$$

That the function $\overline{\mathbf{0}} : X \to Y$, $x \mapsto \mathbf{0}$, is in $\mathcal{L}(X, Y)$ is immediate to show. It is trivial to show that $\mathcal{L}(X, Y)$ satisfies the conditions of a linear space. $\square$

Clearly, X^* is a subspace of $\mathbb{F}(X)$. The dimension of $\mathcal{L}(X, Y)$ for finite-dimensional linear spaces X and Y is determined in the Theorem 2.4.17.

Remark 2.4.4. 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 $\mathcal{L}(X, Z)$.

(*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

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

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 $n \geq 1$:

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

□

By the previous remark the structure with objects the linear spaces and arrows the linear maps is a category, which we call the *category of (real) linear spaces* **Lin**.

Proposition 2.4.5. *Let $n \geq 1$, X, Z be linear spaces, $Y \subseteq X$, and $f_0 : Y \to Z$.*
(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

$$\begin{tikzcd} Y \arrow[r, "f_0"] \arrow[dr, "\mathrm{id}_Y"'] & Z \\ & X. \arrow[u, dashed, "f"'] \end{tikzcd}$$

(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[8]

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

(iii) If $1 \leq m \leq n$, $\dim(X) = n$, and if $Y = \{v_1, \ldots, v_m\}$ is a linearly independent subset of X, there is a linear map $f : X \to Z$ that extends f_0.

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 2.4.4(v) we have that

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

[8]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 $f_Y \in \mathcal{L}(Y, Z)$.

(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

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

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.
(iii) By the Corollary 2.2.16(iv) Y is extended to a finite basis B of X. We can extend f_0 to a function $f_1 : B \to Z$ (e.g., f_1 maps every element of $B \setminus Y := \{x \in X \mid x \in B \ \&\ x \notin Y\}$ to the zero element of Z). By the case (ii) we get a linear extension $f : X \to Z$ of f_1. Clearly, f is a linear extension of f_0 too. $\square$

Next we show that 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.

LEMMA 2.4.6. *Let X, Z be linear spaces, $Y \subseteq X$, $f \in \mathcal{L}(X, Z)$, and $x_1, \ldots, x_n \in X$.*

(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 $Z = \langle f(Y) \rangle$. $\square$

PROPOSITION 2.4.7. *If X, Y are linear spaces, and $f \in \mathcal{L}(X, Y)$, let*

$$\begin{aligned}\mathtt{Ker}(f) &:= \big\{x \in X \mid f(x) = \mathbf{0}\big\}, \\ \mathtt{Im}(f) &:= \big\{y \in Y \mid \exists_{x \in X}\big(f(x) = y\big)\big\}.\end{aligned}$$

(i) $\mathtt{Ker}(f) \preceq X$ *and* $\mathtt{Im}(f) \preceq Y$.

(ii) $\mathtt{Ker}(f) = \{\mathbf{0}\}$ *if and only if f is an injection.*

PROOF. Exercise. $\square$

THEOREM 2.4.8. *If X, Y are linear spaces,* $\dim(X) = n$*, for some $n \geq 1$, and $f \in \mathcal{L}(X, Y)$, then*

$$\dim(X) = \dim\big(\mathtt{Ker}(f)\big) + \dim\big(\mathtt{Im}(f)\big).$$

PROOF. By the Corollary 2.2.17 the subspace $\mathtt{Ker}(f)$ of X has a basis $B = \{v_1, \ldots, v_k\}$, for some $k \geq 0$, such that $k \leq n$. By the Corollary 2.2.16(iii) there are $e_{k+1}, \ldots, e_n \in X$ such that the set

$$B := \{v_1, \ldots, v_k, e_{k+1}, \ldots, e_n\}$$

is a basis of X. If $y \in \mathtt{Im}(f)$, there is $x \in X$ and $a_1, \ldots, a_k, b_{k+1}, \ldots, b_n \in \mathbb{R}$ with

$$\begin{aligned}y &= f(x) \\ &= f\bigg(\sum_{i=1}^{k} a_i v_i + \sum_{j=k+1}^{n} b_j e_j\bigg)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{k} a_i f(v_i) + \sum_{j=k+1}^{n} b_j f(e_j) \\
&= \mathbf{0} + \sum_{j=k+1}^{n} b_j f(e_j) \\
&= \sum_{j=k+1}^{n} b_j f(e_j).
\end{aligned}
$$

Hence, the set $\{f(e_{k+1}, \ldots, f(e_n)\}$ generates $\mathtt{Im}(f)$, and it is also linearly independent in Y, since the restriction $f_{|\langle\{e_{k+1},\ldots,e_n\}\rangle}$ of f to the subspace $\langle\{e_{k+1}, \ldots, e_n\}\rangle$ of X is an injection, and then we use the Lemma 2.4.6(iv). To show that $f_{|\langle\{e_{k+1},\ldots,e_n\}}$ is an injection, we suppose that there is $x \in \langle\{e_{k+1}, \ldots, e_n\}\rangle$, such that $f(x) = \mathbf{0}$. In this case there are $a_1, \ldots, a_k, b_{k+1}, \ldots, b_n \in \mathbb{R}$ such that

$$x = \sum_{i=1}^{k} a_i v_i = \sum_{j=k+1}^{n} b_j e_j.$$

If $x \neq \mathbf{0}$, there is some $i \in \{1, \ldots, k\}$ such that $a_i \neq 0$, and some $j \in \{k+1, \ldots, n\}$, such that $b_j \neq 0$. Hence e_j is written as a linear combination of the elements $v_1, \ldots, v_k, e_{k+1}, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n$ of B, which is a contradiction. Hence, $x = \mathbf{0}$. The required equality $\dim(X) = \dim\big(\mathtt{Ker}(f)\big) + \dim\big(\mathtt{Im}(f)\big)$ follows immediately from the fact that B is a basis of X. $\square$

Notice that in the previous result the cases $k = 0$, and $k = n$ i.e., f is an injection, and f is the constant map $\mathbf{0}$, respectively, follow as special cases. A nice consequence of the previous result is that there is no linear map from $\mathbb{R}^3$ to $\mathbb{R}^2$, which is an injection!

Proposition 2.4.9. *Let X, Y be linear spaces with $\dim(X) = n$ and $\dim(Y) = m$, for some $n, m \geq 1$. Let the following linear operations defined on $X \times Y$:*

$$
\begin{aligned}
(x, y) + (x', y') &:= (x + x', y + y'), \\
a \cdot (x, y) &:= (a \cdot x, a \cdot y), \\
\mathbf{0} &:= (\mathbf{0}, \mathbf{0}).
\end{aligned}
$$

(i) $X \times Y$ is a linear space, which we call the product linear space of X and Y.
(ii) If $\{v_1, \ldots, v_n\}$ is a basis of X and $\{w_1, \ldots, w_m\}$ is a basis of Y, then

$$\big\{(v_1, \mathbf{0}), \ldots, (v_n, \mathbf{0}), (\mathbf{0}, w_1), \ldots, (\mathbf{0}, w_m)\big\}$$

is a basis of $X \times Y$, and $\dim(X \times Y) = n + m$.
(iii) The projections $\mathtt{pr}_X : X \times Y \to X$ and $\mathtt{pr}_Y : X \times Y \to Y$, defined by

$$\mathtt{pr}_X(x, y) := x \ \ \& \ \ \mathtt{pr}_Y(x, y) := y,$$

are linear maps.

(iv) $X \times Y$ *satisfies the universal property of products i.e., for every linear space* Z*, and every linear map* $f : Z \to X$*, and every linear map* $g : Z \to Y$ *there is a unique linear map* $h : Z \to X \times Y$ *such that the following inner diagrams commute*

$$\begin{tikzcd} & Z \arrow[dl, "f"'] \arrow[d, dashed, "h"] \arrow[dr, "g"] & \\ X & X \times Y \arrow[l, "\mathtt{pr}_X"] \arrow[r, "\mathtt{pr}_Y"'] & Y \end{tikzcd}$$

i.e., $f = \mathtt{pr}_X \circ h$ *and* $g = \mathtt{pr}_Y \circ h$*.*

PROOF. Exercise. □

The "up to isomorphism"-uniqueness of $X \times Y$ is shown in the Appendix (section 6.3). Next we show that the direct sum $Y \oplus Z$ of the linear subspaces Y, Z of a linear space X behaves in a way dual to the product. Notice that the arrows in the next diagram are opposite to the arrows of the previous one.

PROPOSITION 2.4.10. *Let* X *be a linear space, and* $Y, Z \preceq X$*.*
(i) The injections $\mathtt{in}_Y : Y \to Y \oplus Z$ *and* $\mathtt{in}_Z : Z \to Y \oplus Z$*, defined by*

$$\mathtt{in}_Y(y) := y + \mathbf{0} = y \quad \& \quad \mathtt{in}_Z(z) := \mathbf{0} + z = z,$$

are linear maps.
(ii) $Y \oplus Z$ *satisfies the universal property of coproducts i.e., for every linear space* W*, and every linear map* $f : Y \to W$*, and every linear map* $g : Z \to W$ *there is a unique linear map* $h : Y \oplus Z \to W$ *such that the following inner diagrams commute*

$$\begin{tikzcd} & W & \\ Y \arrow[ur, "f"] \arrow[r, "\mathtt{in}_Y"'] & Y \oplus Z \arrow[u, dashed, "h"] & Z \arrow[ul, "g"'] \arrow[l, "\mathtt{in}_Z"] \end{tikzcd}$$

i.e., $f = h \circ \mathtt{in}_Y$ *and* $g = h \circ \mathtt{in}_Z$*.*

PROOF. Exercise. □

COROLLARY 2.4.11. *If* X *is a linear space, and* $n \in \mathbb{N}$*, such that* $\dim(X) = n$*, then if* $Y \preceq X$ *and* $Z \preceq X$*, we have that*

$$\dim(Y) + \dim(Z) = \dim(Y + Z) + \dim(Y \cap Z).$$

PROOF. Exercise [Hint: use the Proposition 2.4.9 and the Theorem 2.4.8]. □

Notice that the Proposition 2.2.19(ii) is a special case of the Corollary 2.4.11.

Corollary 2.4.12. *If X, Y are linear spaces, $n \in \mathbb{N}$, $\dim(X) = n = \dim(Y)$, and $f \in \mathcal{L}(X, Y)$, the following are equivalent:*

(i) $\mathtt{Ker}(f) = \{\mathbf{0}\}$.

(ii) $\mathtt{Im}(f) = Y$.

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

Proof. Exercise. $\square$

Definition 2.4.13. 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 $g \circ f = \mathrm{id}_X$

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "\mathrm{id}_X"'] & Y \arrow[r, "g"] \arrow[rr, bend left, "\mathrm{id}_Y"] & X \arrow[r, "f"'] & Y. \end{tikzcd}$$

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

If $(f, g) : X \simeq Y$, and $(f, h) : X \simeq Y$, then $g = h$. Since f is also a bijection (Exercise sheet 1, Exercise 1(i)), we write $g := f^{-1}$. The following converse to the Corollary 2.4.12 expresses the “invariance” of the finite dimension of a linear space under an isomorphism.

Remark 2.4.14. Let X, Y be linear spaces, and $f \in \mathcal{L}(X, Y)$ a linear isomorphism.

(i) If $(f, g) : X \simeq Y$, then $g \in \mathcal{L}(Y, X)$.

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

Proof. Exercise. $\square$

If $n \geq 1$, then an n-dimensional linear space is isomorphic to $\mathbb{R}^n$.

Corollary 2.4.15. *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. $\square$

Now we can determine the dimension of the dual of a finite-dimensional space.

Corollary 2.4.16. *If X is a linear space, $n \in \mathbb{N}$, and $\dim(X) = n$, then $\dim(X^*) = n$.*

Proof. If X is trivial i.e., $X := \{\mathbf{0}\}$, then $X^* := \{f : \{\mathbf{0}\} \to \mathbb{R} \mid f \text{ is linear}\}$. Hence, X^* contains only the zero map, and $\dim(X) = \dim(X^*) = 0$. If X is non-trivial and $B := \{v_1, \ldots, v_n\}$ is a basis of X, let $e_B : X^* \to \mathbb{R}^n$ defined by

$$e_B(f) := \big(f(v_1), \ldots, f(v_n)\big),$$

for every $f \in X^*$. It is easy to see that e_B is a linear map. Moreover, since

$$e_B(f) = \mathbf{0} \Leftrightarrow \big(f(v_1), \ldots, f(v_n)\big) = \big(0, \ldots, 0\big) \Leftrightarrow f(v_1) = 0 \ \& \ \ldots \ \& \ f(v_n) = 0,$$

we have that $f(x) = f\bigg(\sum_{i=1}^n a_i v_i\bigg) = \sum_{i=1}^n a_i f(v_i) = 0$, for every $x \in X$, hence $f = \overline{0}^X$ i.e., $\mathtt{Ker}(e_B) = \{\mathbf{0}\}$, or e_B is an injection. Next we show that e_B is a surjection. If $(a_1, \ldots, a_n) \in \mathbb{R}^n$, we define a function $f_0 : \{v_1, \ldots, v_n\} \to \mathbb{R}^n$ by $f_0(v_1) := a_1 \ \& \ \ldots \ \& \ f_0(v_n) := a_n$. By the Proposition 2.4.5(ii) there is a unique linear extension $f : X \to \mathbb{R}^n$ of f_0. Moreover, $e_B(f) := \big(f(v_1), \ldots, f(v_n)\big) = \big(f_0(v_1), \ldots, f_0(v_n)\big) = \big(a_1, \ldots, a_n\big)$. $\square$

The previous corollary is a special case of the following theorem.

Theorem 2.4.17. *Let X, Y be linear spaces, $m, n \geq 1$, $\{v_1, \ldots, v_m\}$ a basis of X, and $\{w_1, \ldots, w_n\}$ a basis of Y. If for every $i \in \{1, \ldots, m\}$ and every $j \in \{1, \ldots, n\}$ the function $f_{ij} : X \to Y$ is the unique linear extension of the function*[9]

$$f_{ij}^0 : \{v_1, \ldots, v_m\} \to Y$$

$$f_{ij}^0(v_k) := \delta_{ki} w_j,$$

$$\delta_{ki} := \begin{cases} 1 & , k = i \\ 0 & , k \neq i, \end{cases}$$

then the set

$$B := \big\{f_{ij} \mid i \in \{1, \ldots, m\} \ \& \ j \in \{1, \ldots, n\}\big\}$$

is a basis of $\mathcal{L}(X, Y)$, and $\dim\big(\mathcal{L}(X, Y)\big) = mn$.

Proof. By the definition of f_{ij} we have that

$$f_{ij}(v_k) = \begin{cases} w_j & , k = i \\ 0 & , k \neq i, \end{cases}$$

for every $k \in \{1, \ldots, m\}$. First we show that $\langle B \rangle = \mathcal{L}(X, Y)$. Let $h \in \mathcal{L}(X, Y)$. If $k \in \{1, \ldots, m\}$, there are $a_{k1}, \ldots, a_{kn} \in \mathbb{R}$ such that

$$h(v_k) = \sum_{j=1}^n a_{kj} w_j.$$

We show that

$$h = \sum_{i=1}^m \sum_{j=1}^n a_{ij} f_{ij} = \sum_{i=1, j=1}^{m \ \ n} a_{ij} f_{ij} \in \langle B \rangle.$$

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

By the Proposition 2.4.5(ii) it suffices to show that the two functions, h and $\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}f_{ij} \in \mathcal{L}(X,Y)$, are equal on the given basis of X. If $k \in \{1,\ldots,m\}$, by the definition of the linear operations on $\mathcal{L}(X,Y)$ we have that

$$\begin{aligned}\Bigg(\sum_{i=1,j=1}^{m\ \ n} a_{ij}f_{ij}\Bigg)(v_k) &:= \sum_{i=1,j=1}^{m\ \ n} a_{ij}f_{ij}(v_k)\\ &:= \sum_{i=1,j=1}^{m\ \ n} a_{ij}\delta_{ki}w_j\\ &= \sum_{j=1}^{n} a_{kj}w_j\\ &= h(v_k).\end{aligned}$$

Next we show that the elements of B are linearly independent. If $a_{ij} \in \mathbb{R}$, where $i \in \{1,\ldots,m\}$ and $j \in \{1,\ldots,n\}$, then

$$\begin{aligned}\sum_{i=1,j=1}^{m\ \ n} a_{ij}f_{ij} = \mathbf{0} &\Rightarrow \forall_{k\in\{1,\ldots,m\}}\Bigg(\Bigg(\sum_{i=1,j=1}^{m\ \ n} a_{ij}f_{ij}\Bigg)(v_k) = \mathbf{0}\Bigg)\\ &\Leftrightarrow \forall_{k\in\{1,\ldots,m\}}\Bigg(\sum_{i=1,j=1}^{m\ \ n} a_{ij}f_{ij}(v_k) = \mathbf{0}\Bigg)\\ &\Leftrightarrow \forall_{k\in\{1,\ldots,m\}}\Bigg(\sum_{i=1,j=1}^{m\ \ n} a_{ij}\delta_{ki}w_j = \mathbf{0}\Bigg)\\ &\Leftrightarrow \forall_{k\in\{1,\ldots,m\}}\Bigg(\sum_{j=1}^{n} a_{kj}w_j = \mathbf{0}\Bigg)\\ &\Leftrightarrow \forall_{k\in\{1,\ldots,m\}}\big(a_{k1} = \ldots = a_{kn} = 0\big)\\ &\Leftrightarrow \forall_{i\in\{1,\ldots,m\}}\forall_{j\in\{1,\ldots,n\}}\big(a_{ij} = 0\big).\end{aligned}$$

Since the cardinality of B is mn, we conclude that $\dim\big(\mathcal{L}(X,Y)\big) = mn$. $\square$

The basis of X^* determined from the proof of the Theorem 2.4.17, if $Y = \mathbb{R}$, is equal to the basis of X^* determined from the proof of the Corollary 2.4.16 i.e., the set $\{e_B^{-1}(e_1),\ldots,e_B^{-1}(e_m)\}$, where $\{e_1,\ldots,e_m\}$ is the standard basis of $\mathbb{R}^m$. According to the proof of the Theorem 2.4.17, X^* has as bases the m-functions $f_{11},\ldots,f_{m1}$, where f_{i1} is the unique linear extension of $f_{i1}^0 : \{v_1,\ldots v_m\} \to \mathbb{R}$, and

$$f_{i1}(v_k) = \begin{cases} 1 & , k = i\\ 0 & , k \neq i,\end{cases} =: \delta_{ki},$$

for every $k \in \{1,\ldots,m\}$. On the other hand, if $f_1 := e_B^{-1}(e_1),\ldots,f_m := e_B^{-1}(e_m)$, is the basis of X^* determined by the proof of the Corollary 2.4.16, then

$$e_B(f_i) = e_i \Leftrightarrow \big(f(v_1),\ldots,f(v_{i-1}),f(v_i),f(v_{i+1}),\ldots,f(v_n)\big) = (0,\ldots,0,1,0,\ldots,0)$$

i.e., $f_i(v_k) = \delta_{ki} = f_{i1}(v_k)$, and hence $f_i = f_{i1}$, for every $i \in \{1, \ldots, m\}$.

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 2.4.18. 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., $T^3 = T \circ T \circ T$

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[rrr, bend right, "T^3"'] & X \arrow[r, "T"] & X \arrow[r, "T"] & X. \end{tikzcd}$$

Remark 2.4.19. 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. $\square$

Remark 2.4.20. 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. $\square$

2.5. Quotient spaces

A subspace Y of a linear space X generates a new linear space by "identifying" the elements of Y. This identification is done by "quotienting" over Y with respect to some appropriate equivalence relation.

Remark 2.5.1. Let X be a linear space and $Y \preceq X$. If $x, x' \in X$, we define

$$x \sim x'(\mathrm{mod} Y) :\Leftrightarrow x' - x \in Y.$$

Then the relation $x \sim x'(\mathrm{mod} Y)$ is an *equivalence relation* on X i.e., for every $x, x', x'' \in X$ the following conditions are satisfied:

(i) $x \sim x(\mathrm{mod} Y)$.

(ii) If $x \sim x'(\mathrm{mod} Y)$, then $x' \sim x(\mathrm{mod} Y)$.

(*iii*) If $x \sim x'(\mathrm{mod}Y)$ and $x' \sim x''(\mathrm{mod}Y)$, then $x \sim x''(\mathrm{mod}Y)$.
Moreover, the relation $x \sim x'(\mathrm{mod}Y)$ is compatible with the operations of X i.e.,
(a) If $x \sim x'(\mathrm{mod}Y)$ and $y \sim y'(\mathrm{mod}Y)$, then $x + y \sim x' + y'(\mathrm{mod}Y)$.
(b) If $x \sim x'(\mathrm{mod}Y)$ and $a \in \mathbb{R}$, then $a \cdot x \sim a \cdot x'(\mathrm{mod}Y)$.

PROOF. Exercise. $\square$

DEFINITION 2.5.2. If $x \in X$, the *equivalence class* $[x]_\sim$ of x with respect to the equivalence relation $x \sim x'(\mathrm{mod}Y)$ is the set

$$
\begin{aligned}
[x]_{\sim(\mathrm{mod}Y)} &:= \big\{x' \in X \mid x' \sim x(\mathrm{mod}Y)\big\} \\
&:= \big\{x' \in X \mid x' - x \in Y)\big\} \\
&= \big\{x' \in X \mid \exists_{y \in Y}\big(x' - x = y\big)\big\} \\
&= \big\{x' \in X \mid \exists_{y \in Y}\big(x' = x + y\big)\big\} \\
&= \big\{x + y \mid y \in Y\big\} \\
&=: x + Y.
\end{aligned}
$$

The set of all these equivalence classes is denoted by X/Y i.e.,

$$X/Y := \{x + Y \mid x \in X\},$$

and it is called the *quotient space*, or the *factor space* of X with respect to Y. If $x, x' \in X$ and $a \in \mathbb{R}$, and using for simplicity the same symbols to the symbols of the signature of X, let

$$
\begin{gathered}
(x + Y) + (x' + Y) := (x + x') + Y, \\
a \cdot (x + Y) := (a \cdot x) + Y, \\
\mathbf{0} := \mathbf{0} + Y = Y.
\end{gathered}
$$

The *canonical projection* of X onto Y is the surjection

$$\pi_Y : X \to X/Y, \quad x \mapsto x + Y, \quad x \in X.$$

By the above definitions we get $x + Y = x' + Y \Leftrightarrow x \sim x'(\mathrm{mod}Y) \Leftrightarrow x - x' \in Y$.

REMARK 2.5.3. Let X be a linear space and $Y \preceq X$. If $x \sim x'(\mathrm{mod}Y)$ is defined as above, then the quotient space $\mathcal{V}/Y := (X/Y; +, \mathbf{0}, \cdot)$ is a linear space, the canonical projection π_Y of X onto Y is a linear map, and this linear structure on X/Y is the unique linear structure on X/Y that makes π_Y a linear map.

PROOF. It is immediate to show that X/Y is a linear space. Next we show that π_Y is a linear map. If $x, x' \in Y$ and $a \in \mathbb{R}$, we have that

$$
\begin{gathered}
\pi_Y(x + x') := (x + x') + Y := (x + Y) + (x' + Y) := \pi_Y(x) + \pi_Y(x'), \\
\pi_Y(ax) := (ax) + Y := a(x + Y) := a\pi_Y(x).
\end{gathered}
$$

If $(\oplus, \mathbf{0}, \odot)$ is a linear structure on X/Y that makes π_Y a linear map, then

$$\pi_Y(x + x') = \pi_Y(x) \oplus \pi_Y(x') :\Leftrightarrow (x + x') + Y = (x + Y) \oplus (x' + Y),$$

$$\pi_Y(a \cdot x) = a \odot \pi_Y(x) :\Leftrightarrow ax + Y = a \odot (x + Y)$$

i.e., the linear structure $(\oplus, \mathbf{0}, \odot)$ on X/Y is the one given in the Definition 2.5.2. $\square$

The quotient space X/X is a trivial linear space, since

$$x \sim x'(\text{mod}X) :\Leftrightarrow x' - x \in X,$$

and hence any two elements of X are equivalent, or, the only equivalence class is X itself i.e., $X/X = \{X\}$. The quotient space $X/\{\mathbf{0}\}$ is isomorphic to X, since

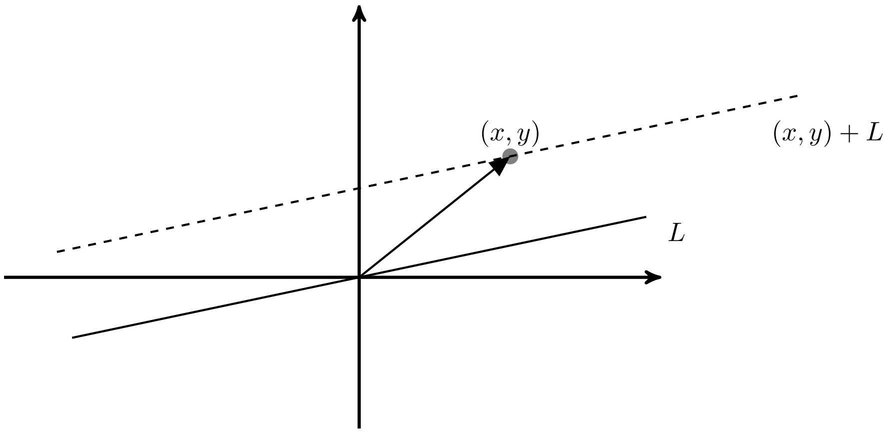
$$x \sim x'(\text{mod}\{\mathbf{0}\}) :\Leftrightarrow x' - x \in \{\mathbf{0}\} \Leftrightarrow x' - x = \mathbf{0} \Leftrightarrow x' = x,$$

and hence $x + \{\mathbf{0}\} = \{x\} \in X/\{\mathbf{0}\}$. Consequently, the mapping $e : X \to X/\{\mathbf{0}\}$, defined by $x \mapsto \{x\}$, for every $x \in X$, is a linear isomorphism. Next follows a more interesting example of a quotient space.

Example 2.5.4. If L is a line in $\mathbb{R}^2$ that passes through the origin $(0, 0)$, then it is easy to see that if $(x, y) \in \mathbb{R}^2$, the equivalence class

$$(x, y) + L$$

is the line that passes through (x, y) and it is parallel to L



or, in other words, $(x', y') \sim (x, y)(\text{mod}L)$ if and only if (x', y') is in the line that passes through (x, y) and it is parallel to L. We can now give a geometrical interpretation of the condition

$$(x, y) \sim (x', y')(\text{mod}L) \ \ \& \ \ (u, w) \sim (u', w')(\text{mod}L)$$

$$\Rightarrow (x, y) + (u, w) \sim (x', y') + (u', w')(\text{mod}L).$$

If $M := (x, y), M' := (x', y'), N := (u, w)$, and $N' := (u', w')$, then $\overrightarrow{OM'} + \overrightarrow{ON'}$ is in the line that passes from $\overrightarrow{OM} + \overrightarrow{ON}$ and it is parallel to L. To see this, if L is represented by the equation $y = ax$, for some $a \in \mathbb{R}$, then

$$\begin{aligned}\frac{(y'+w')-(y+w)}{(x'+u')-(x+u)} &= \frac{(y'-y)+(w'-w)}{(x'-x)+(u'-u)} \\ &= \frac{a(x'-x)+a(u'-u)}{(x'-x)+(u'-u)} \\ &= a\frac{(x'-x)+u'-u)}{(x'-x)+(u'-u)} \\ &= a.\end{aligned}$$

The following condition is interpreted similarly

$$(x,y) \sim (x',y')(\mathrm{mod}L) \;\;\&\;\; \lambda \in \mathbb{R} \;\Rightarrow\; \lambda(x,y) \sim \lambda(x',y')(\mathrm{mod}L),$$

since

$$\frac{\lambda y' - \lambda y}{\lambda x' - \lambda x} = \frac{y'-y}{x'-x} = a.$$

The quotient space X/L is the set of all equivalence classes $(x,y)+L$, hence, according to the above geometrical interpretation of $(x,y)+L$, it is the set of all lines parallel to L. Each such line is determined from the point of its intersection with the axis $y'y$, which is a real number $a_{(x,y)+L}$, where

$$\frac{y - a_{(x,y)+L}}{x-0} = a \Leftrightarrow a_{(x,y)+L} = y - ax.$$

It is easy to see that the addition $[(x,y)+L] + [(u,w)+L]$ on X/L corresponds to the addition $a_{(x,y)+L} + a_{(u,w)+L}$ and the scalar multiplication $b \cdot [(x,y)+L]$ on X/L corresponds to the multiplication $ba_{(x,y)+L}$ of reals. In other words, the mapping

$$e : X/L \to \mathbb{R},$$

$$(x,y)+L \mapsto a_{(x,y)+L},$$

is linear, and it is also a bijection. Hence, even without carrying out the exact calculations, it is "geometrically" expected that X/L is linearly isomorphic to $\mathbb{R}$. Hence, in this case we have that

$$\dim(\mathbb{R}^2) = \dim(L) + \dim(\mathbb{R}^2/L).$$

Definition 2.5.5. Let X be a linear space, $Y \preceq X$ and $x_1, \ldots, x_n \in X$. We say that $x_1, \ldots, x_n$ are *linearly dependent* $(\mathrm{mod}Y)$, if $x_1 + Y, \ldots, x_n + Y$ are linearly dependent in X/Y, and $x_1, \ldots, x_n$ are *linearly independent* $(\mathrm{mod}Y)$, if $x_1 + Y, \ldots, x_n + Y$ are linearly independent in X/Y.

One can show that $x_1, \ldots, x_n$ are linearly dependent $(\mathrm{mod}Y)$ if and only if there are $n \geq 1$ and $a_1, \ldots, a_n \in \mathbb{R}$ such that $a_i \neq 0$, for some $i \in \{1, \ldots, n\}$, and

$$\exists_{y \in Y}\left(\sum_{i=1}^{n} a_i x_i = y\right),$$

and hence, $x_1, \ldots, x_n$ are linearly independent (mod$Y$) if and only if for every $n \geq 1$ and every $a_1, \ldots, a_n \in \mathbb{R}$

$$\exists_{y \in Y} \left(\sum_{i=1}^{n} a_i x_i = y \right) \Rightarrow a_1 = \ldots = a_n = 0.$$

THEOREM 2.5.6. *Let X be a linear space and $Y \preceq X$. If B_Y is a basis of Y, and $B := B_Y \cup C$ is a basis of X, where $B_Y \cap C = \emptyset$, then*

$$C + Y := \{c + Y \mid c \in C\}$$

is a basis of X/Y, and $\dim(X) = \dim(Y) + \dim(X/Y)$.

PROOF. If $Z := \langle C \rangle$, then $X = Y + Z$. To see this let $x \in X$, and let $b_1, \ldots, b_m \in B_Y$, $c_1, \ldots, c_n \in C$, and $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n \in \mathbb{R}$ such that

$$x = \sum_{i=1}^{m} \lambda_i b_i + \sum_{j=1}^{n} \mu_j c_j = y + z,$$

where

$$y := \sum_{i=1}^{m} \lambda_i b_i \in Y \quad \& \quad z := \sum_{j=1}^{n} \mu_j c_j \in Z.$$

Clearly, we have that $x \in Y + Z$. Moreover, $Y \cap Z = \{\mathbf{0}\}$, since, if there was $x \in X$ such that $x \neq \mathbf{0}$ and $x \in Y \cap Z$ i.e.,

$$x = \sum_{i=1}^{m} \lambda_i b_i = \sum_{j=1}^{n} \mu_j c_j,$$

for some $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n \in \mathbb{R}$, then there would be some $\mu_j \neq 0$, and consequently the vector c_j could be written as a linear combination of the rest. Since $B_Y \cap C = \emptyset$, this is impossible. Hence, by the Proposition 2.2.19 we have that $X = Y \oplus Z$. Let the function

$$\phi : Z \to X/Y, \quad z \mapsto z + Y, \quad z \in Z,$$

Since ϕ is the restriction of the canonical projection π_Y to the subspace Z of X, it is also a linear map. First we show that ϕ is an injection. If $z_1, z_2 \in Z$, then

$$z_1 + Y = z_2 + Y \Leftrightarrow (z_1 - z_2) \in Y \Leftrightarrow (z_1 - z_2) \in Y \cap Z \Leftrightarrow (z_1 - z_2) = \mathbf{0} \Leftrightarrow z_1 = z_2.$$

Next we show that ϕ is a surjection. Let $x + Y \in X/Y$. Since there are $y \in Y$ and $z \in Z$ such that $x = y + z$, as we described above, we have that

$$\phi(z) := z + Y = (\mathbf{0} + Y) + (z + Y) = (y + Y) + (z + Y) := (y + z) + Y = x + Z.$$

Since C is a basis of Z, by the Lemma 2.4.6(vii) we have that $\phi(C) = C + Y$ is a basis of X/Y. Now the equality $\dim(X) = \dim(Y) + \dim(X/Y)$ follows[10]. $\square$

[10] Since $\dim(X) = |B| = |B_Y \cup C|$, we use here the fact that $|B_Y \cup C| = |B_Y| + |C|$, when $B_Y \cap C = \emptyset$. This fact about cardinalities of sets is trivial when the related sets are finite, and it can also be shown in the general case.

For the quotient space of the Example 2.5.4 we have that $L := \{(x, y) \in \mathbb{R}^2 \mid y = ax\}$, for some $a \in \mathbb{R}$, $B_L := \{(1, a)\}$ is a basis of L and if $C := \{(0, 1)\}$, then $B_L \cup C$ is a basis of $\mathbb{R}^2$, $B_L \cap C = \emptyset$, and $\{(0, 1) + L\}$ is a basis of $\mathbb{R}^2/L$.

If X is a finite dimensional space, and $f : X \to Y$ is linear, by the Theorem 2.4.8 we have that

$$\dim(X) = \dim(\mathtt{Ker}(f)) + \dim(\mathtt{Im}(f)),$$

and by the previous theorem we also have that

$$\dim(X) = \dim(\mathtt{Ker}(f)) + \dim(X/\mathtt{Ker}(f)).$$

Hence, if $m = \dim(\mathtt{Im}(f)) = \dim(X/\mathtt{Ker}(f))$, and since both these spaces are by the Corollary 2.4.15 isomorphic to R^m, we also have that $X/\mathtt{Ker}(f) \simeq \mathtt{Im}(f)$. Next we show this fact for any linear space X.

THEOREM 2.5.7. *If X, Y are linear spaces and $f \in \mathcal{L}(X, Y)$, then*

$$X/\mathtt{Ker}(f) \simeq \mathtt{Im}(f).$$

PROOF. Let the function $\phi : X/\mathtt{Ker}(f) \to \mathtt{Im}(f)$, defined by

$$\phi(x + \mathtt{Ker}(f) := f(x),$$

for every $x + \ker(f) \in X/\mathtt{Ker}(f)$. First we show that ϕ is indeed a function and an injection. If $x + \ker(f)$ and $x' + \ker(f)$ are in $X/\mathtt{Ker}(f)$, then

$$x + \ker(f) = x' + \ker(f) \Leftrightarrow (x - x') \in \ker(f) \Leftrightarrow f(x - x') = \mathbf{0} \Leftrightarrow f(x) = f(x').$$

Next we show that ϕ is linear:

$$\begin{aligned} \phi\big((x + \ker(f)) + (x' + \ker(f))\big) &:= \phi\big((x + x') + \ker(f)\big) \\ &:= f(x + x') \\ &= f(x) + f(x') \\ &:= \phi\big(x + \mathtt{Ker}(f)\big) + \phi\big(x' + \mathtt{Ker}(f)\big), \end{aligned}$$

$$\begin{aligned} \phi\big(a(x + \ker(f))\big) &:= \phi\big(ax + \ker(f)\big) \\ &:= f(ax) \\ &= af(x) \\ &:= a\phi\big(x + \mathtt{Ker}(f)\big). \end{aligned}$$

Since ϕ is, trivially, a surjection, we have that ϕ is a linear isomorphism. $\square$

From the previous theorem a linear map $f : X \to Y$ is written as the composition of an injection (ϕ) with a surjection ($\pi_{\mathtt{Ker}(f)}$)

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\pi_{\mathtt{Ker}(f)}"'] & Y \\ X/\mathtt{Ker}(f). \arrow[ur, "\phi"'] & \end{tikzcd}$$

PROPOSITION 2.5.8. *Let X be a linear space and $Y, Z \preceq X$.*

(i) $Y/(Y \cap Z) \simeq (Y + Z)/Z$.

(ii) If $X = Y \oplus Z$, *then* $Y \simeq X/Z$.

(iii) If $Z \preceq Y$, *then* $(X/Z)/(Y/Z) \simeq X/Y$.

PROOF. (i) Let $f : Y \to (Y + Z)/Z$ defined by $f(y) := y + Z$, for every $y \in Y$. Since f is the composition of linear maps

$$Y \xrightarrow{i_Y} Y + Z \xrightarrow{\pi_Z} (Y + Z)/Z,$$

where $i_Y(y) := y = y + \mathbf{0}$ is the linear injection of Y into $Y + Z$ (see also Proposition 2.4.10(i)), or since f is the restriction of the linear map π_Y to the subspace Y of $Y + Z$, we have that f is linear. If $(y + z) + Z \in (Y + Z)/Z$, then

$$f(y) := y + Z = (y + Z) + Z = (y + Z) + (z + Z) := (y + z) + Z$$

i.e., f is a surjection. Since

$$y \in \mathtt{Ker}(f) \Leftrightarrow f(y) = \mathbf{0} \Leftrightarrow y + Z = Z \Leftrightarrow y \in Z \Leftrightarrow y \in Y \cap Z,$$

we have that $\ker(f) = Y \cap Z$. By the Theorem 2.5.7 we get

$$Y/\mathtt{Ker}(f) \simeq \mathtt{Im}(f) \Leftrightarrow Y/(Y \cap Z) \simeq (Y + Z)/Z.$$

(ii) If $X = Y \oplus Z$, then by the Proposition 2.2.19 we have that $Y \cap Z = \{\mathbf{0}\}$, and by the case (i) we get

$$Y \simeq Y/\{\mathbf{0}\} \simeq (Y + Z)/Z = (Y \oplus Z)/Z = X/Z.$$

(iii) Exercise. □

DEFINITION 2.5.9. If X, Y are finite-dimensional linear spaces, and $f : X \to Y$ is linear, the *rank* of f is defined by

$$\mathtt{rank}(f) := \dim(\mathtt{Im}(f)).$$

By the Theorems 2.5.7 and 2.5.6 we have that

$$\mathtt{rank}(f) = \dim(X/\mathtt{Ker}(f)) = \dim(X) - \dim(\mathtt{Ker}(f)),$$

hence f is an injection if and only if $\mathtt{rank}(f) = \dim(X)$.

2.6. The free linear space

In this section we study the construction of a free object in the category **Lin**. The methods that we use and the results we prove here are found, in some disguise, in many different contexts in mathematics.

Definition 2.6.1. If X is a set, and $f : X \to \mathbb{R}$, the *zero set* $[f = 0]$ of f is the inverse image of $\{0\}$ under f, and the *cozero set* $[f \neq 0]$ of f is the complement[11] of $f = 0$ i.e.,

$$[f = 0] := \{x \in X \mid f(x) = 0\} = f^{-1}(\{0\}),$$
$$[f \neq 0] := \{x \in X \mid f(x) \neq 0\} = X \setminus [f = 0].$$

We denote by εX the set of all real-valued functions on X that are non-zero for *only* finitely many elements of X i.e.,

$$\varepsilon X := \big\{f \in \mathbb{F}(X) \mid [f \neq 0] \text{ is finite}\big\}.$$

If $f \in \varepsilon X$, we also say that f is *almost everywhere* 0.

If $f \in \varepsilon X$, there are $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in X$ such that

$$\begin{cases} f(x) \neq 0 & \text{, if } x \in \{x_1, \ldots, x_n\} \\ f(x) = 0 & \text{, otherwise.} \end{cases}$$

The constant zero function $\overline{0}^X$ on X is in εX, since $\big[\overline{0}^X \neq 0\big] = \emptyset$, and $\emptyset$ is trivially a finite subset of X. If $x \in X$, the function $f_x : X \to \mathbb{R}$, defined by

$$f_x(y) := \begin{cases} 1 & \text{, if } y = x \\ 0 & \text{, } y \neq x \end{cases}$$

for every $y \in X$, is also in εX.

Remark 2.6.2. Let $f, g \in \mathbb{F}(X)$ and $a \in \mathbb{R}$.
(i) $[(f + g) \neq 0] \subseteq [f \neq 0] \cup [g \neq 0]$.
(ii) $[(af) \neq 0] \subseteq [f \neq 0]$.

Proof. (i) Let $x \in X$ such that $f(x) + g(x) \neq 0$, and suppose that $f(x) = g(x) = 0$ i.e., $x \in [f = 0] \cap [g = 0]$. Then $f(x) + g(x) = 0$, which is a contradiction, hence $x \in X \setminus \big([f = 0] \cap [g = 0]\big)$, and consequently[12] we get $x \in [f \neq 0] \cup [g \neq 0]$.
(ii) If $a = 0$, then $[(af) \neq 0] = \emptyset \subseteq [f \neq 0]$. If $a \neq 0$, and $x \in X$ such that $af(x) \neq 0$, then $f(x) \neq 0$. $\square$

It is easy to find $f, g \in \mathbb{F}(X)$ such that $[(f + g) \neq 0] \subsetneq [f \neq 0] \cup [g \neq 0]$. Moreover, if $a = 0$, and $[f \neq 0] \neq \emptyset$, we have that $[(af) \neq 0] = \emptyset \subsetneq [f \neq 0]$, and if $a \neq 0$, then $[(af) \neq 0] = [f \neq 0]$.

[11] If $Y \subseteq X$, its complement $X \setminus Y$ is defined by $X \setminus Y := \{x \in X \mid x \notin Y\}$.

[12] If Y and Z are subsets of a set X, then $X \setminus (Y \cap Z) = (X \setminus Y) \cup (X \setminus Z)$. Interchanging $\cap$ and $\cup$, we get $X \setminus (Y \cup Z) = (X \setminus Y) \cap (X \setminus Z)$.

Corollary 2.6.3. *If X is a given set, then εX, equipped with the linear structure of $\mathbb{F}(X)$, is a linear subspace of $\mathbb{F}(X)$, which has as basis the set*

$$B_X := \{f_x \mid x \in X\},$$

and the map $i_X : X \to \varepsilon X$, defined by $x \mapsto f_x$, for every $x \in X$, is an injection.

Proof. If $X = \emptyset$, then $\varepsilon\emptyset$ is the trivial linear space, since

$$\varepsilon\emptyset := \{f : \emptyset \to \mathbb{R} \mid [f \neq 0] \text{ is finite}\} = \{\emptyset\},$$

and $B_\emptyset = \emptyset$. The injectivity of $i_\emptyset$ also follows trivially.

Let X be a non-empty set. If $f, g \in \varepsilon X$, then by the Remark 2.6.2(i) we have that $[(f + g) \neq 0]$ is a subset of the union of two finite sets. Since the union of two finite sets is also finite, and since any subset of a finite set is again finite, we conclude that $[(f+g) \neq 0]$ is a finite subset of X, and hence $f+g \in \varepsilon X$. Similarly, by the Remark 2.6.2(ii) we have that if $f \in \varepsilon X$ and $a \in \mathbb{R}$, then $af \in \varepsilon X$. Since $\overline{0}^X \in \varepsilon X$, we conclude that εX is a linear subspace of $\mathbb{F}(X)$.

Next we show that $\langle B_X \rangle = \varepsilon X$. Let $f \in \varepsilon X$. If $f = \overline{0}^X$, then

$$f = 0 f_{x_0} = f(x_0) f_{x_0},$$

where $x_0 \in X$. If $[f \neq 0] = \{x_1, \ldots, x_n\}$, for some $x_1, \ldots, x_n \in X$, then

$$f = \sum_{i=1}^{n} f(x_i) f_{x_i} \in \langle B_X \rangle.$$

To show that B_X is linearly independent subset of εX, let $f_{x_1}, \ldots, f_{x_n} \in B_X$ and $a_1, \ldots, a_n \in \mathbb{R}$ such that $a_1 f_{x_1} + \ldots + a_n f_{x_n} = \overline{0}^X$. Since for every $k \in \{1, \ldots, n\}$ we have that

$$\bigg(\sum_{i=1}^{n} a_i f_{x_i}\bigg)(x_k) = \overline{0}^X(x_k) \Leftrightarrow a_k = 0,$$

we get what we want. Finally, we suppose that $x, x' \in X$ such that $f_x = f_{x'}$. Since $1 = f_x(x) = f_{x'}(x)$, we conclude that $x' = x$. $\square$

"Identifying" X with B_X, we can view εX as a linear space with X as a basis. If $X := \mathbf{1} := \{0\}$, then

$$\varepsilon\mathbf{1} := \big\{f \in \mathbb{F}(\mathbf{1}) \mid [f \neq 0] \text{ is finite}\big\} := \mathbb{F}(\mathbf{1}),$$

and $\mathbb{F}(\mathbf{1})$ has as many elements as the set $\mathbb{R}$ of real numbers (why?). We see that εX can be much larger than X. It turns out that if X is already a linear space, then the linear space εX is very different from X.

Definition 2.6.4. If X is a set, the linear space εX is called the *free linear space* generated by X, or the free linear space over X.

The fundamental property of εX is that any *function* h from X to a linear space Y generates a *linear map* from εX to Y that "extends" h. In the proof of the following theorem we need to be careful, since εX is not necessarily a finite-dimensional linear space, and we have proved the Proposition 2.4.5(ii) only for finite-dimensional linear spaces.

THEOREM 2.6.5 (The universal property of the free linear space). *If X is a set, then for every linear space Y and every function $h : X \to Y$ there is a unique linear map $\varepsilon h : \varepsilon X \to Y$ such that "h factors through εX" i.e., the following diagram commutes*[13]

$$\begin{tikzcd} X \arrow[r, hook, "i_X"] \arrow[dr, "h"'] & \varepsilon X \arrow[d, dashed, "\varepsilon h"] \\ & Y. \end{tikzcd}$$

PROOF. Let $f \in \varepsilon X$, such that $[f \neq 0] = \{x_1, \ldots, x_n\}$, for some $n \geq 1$. As we have seen in the proof of the Corollary 2.6.3, f is written as follows:

$$f = \sum_{i=1}^{n} f(x_i) f_{x_i}.$$

This writing of f is unique. To show this, let $\{y_1, \ldots, y_m\} \subseteq X$ such that

$$\sum_{j=1}^{m} b_j f_{y_j} = f = \sum_{i=1}^{n} f(x_i) f_{x_i}.$$

Since, for every $l \in \{1, \ldots, m\}$, we have that

$$f(y_l) = \left(\sum_{j=1}^{m} b_j f_{y_j} \right)(y_l) = \sum_{j=1}^{m} b_j f_{y_j}(y_l) = b_l,$$

the previous equalities are written as follows:

$$\sum_{j=1}^{m} f(y_j) f_{y_j} = f = \sum_{i=1}^{n} f(x_i) f_{x_i}.$$

If $k \in \{1, \ldots, n\}$, then

$$\left(\sum_{j=1}^{m} f(y_j) f_{y_j} \right)(x_k) = f(x_k) \Leftrightarrow \sum_{j=1}^{m} f(y_j) f_{y_j}(x_k) = f(x_k),$$

[13] An injection $f : A \to B$ is also denoted by $f : A \hookrightarrow B$. We use such a "hook right arrow" in a diagram to indicate that i_X is an injection. A dashed arrow, like the arrow corresponding to εh, in a diagram is used to denote the uniqueness of that arrow.

which implies that $x_k = y_l$ for some $l \in \{1, \ldots, m\}$. Similarly we show that every y_j is equal to some x_i, hence $m = n$ and $x_i = y_i$, for every $i \in \{1, \ldots, n\}$. Let the function $h^* : B_X \to Y$ defined by

$$h^*(f_x) := h(x).$$

By the uniqueness of the writing of some f as above as a linear combination of the elements of B_X, the function h^* has the following unique linear extension εh on εX, where

$$(\varepsilon h)(f) := (\varepsilon h)\bigg(\sum_{i=1}^{n} f(x_i) f_{x_i}\bigg) := \sum_{i=1}^{n} f(x_i) h(x_i).$$

Notice that if $f = \overline{0}^X$, the writing of f as $f(x_0) f_{x_0}$ is not unique, since, for example, $f(x_0) f_{x_0} = \overline{0}^X = f(x_1) f_{x_1}$, for every $x_0, x_1 \in X$. Since[14]

$$(\varepsilon h)\big(\overline{0}^X\big) := (\varepsilon h)\big(f(x_0) f_{x_0}\big) = f(x_0) h(x_0) = 0 h(x_0) =$$
$$= \mathbf{0} = 0 h(x_1) = f(x_1) h(x_1) = (\varepsilon h)\big(f(x_1) f_{x_1}\big),$$

we have though, that the formula of εh applies also on $\overline{0}^X$ and gives the expected value $\mathbf{0}$. Since $(\varepsilon h)\big(i_X(x)\big) := (\varepsilon h)\big(f_x\big) := h(x)$, we get the required commutativity of the above diagram. □

Notice that the importance of εX lies exactly in its universal property. Although $\mathbb{F}(X)$ is a linear space that can also be considered as "a space generated by X", there is no interesting connection between X and $\mathbb{F}(X)$. In the case of εX instead, X "is" the basis of εX, a fact crucial to the previous proof of the universal property of εX. Next we show that εX is unique, up to isomorphism, in **Lin** i.e., if W is a linear space that satisfies the universal property of the free linear space generated by X, then W is linearly isomorphic to εX.

Theorem 2.6.6 (Uniqueness of the free linear space). *Let W be a linear space and $j_X : X \to W$ an injection, such that for every linear space Y and every function $h : X \to Y$ there is a unique linear map $h_W : W \to Y$ such that the following diagram commutes*

$$\begin{tikzcd} X \arrow[r, hook, "j_X"] \arrow[dr, "h"'] & W \arrow[d, dashed, "h_W"] \\ & Y. \end{tikzcd}$$

Then W is linearly isomorphic to εX.

Proof. If in the universal property for W we take $Y := W$ and $h := j_X$

[14]If we write $\overline{0}^X = \sum_{i=1}^{n} \overline{0}^X(x_i) f_{x_i}$, the formula of εh gives similarly that $(\varepsilon h)\big(\overline{0}^X\big) = \mathbf{0}$.

$$\begin{tikzcd} X \arrow[r, hook, "j_X"] \arrow[dr, "j_X"'] & W \arrow[d, dashed, "(j_X)_W"] \\ & W, \end{tikzcd}$$

then explain why $(j_X)_W = \mathrm{id}_W$. Next we proceed as in the proof of the up to isomorphism-uniqueness of the product (see the section 6.3 of the Appendix). $\square$

Proposition 2.6.7. *Let X, Y, Z be sets, and $h : X \to Y$, $g : Y \to Z$ functions.*
(i) There is a unique linear map $\varepsilon h : \varepsilon X \to \varepsilon Y$ such that the following diagram commutes

$$\begin{tikzcd} X \arrow[r, "h"] \arrow[d, hook, "i_X"'] & Y \arrow[d, hook, "i_Y"] \\ \varepsilon X \arrow[r, dashed, "\varepsilon h"'] & \varepsilon Y. \end{tikzcd}$$

(ii) The following lower outer diagram commutes

$$\begin{tikzcd} X \arrow[r, "h"'] \arrow[rr, bend left, "g \circ h"] \arrow[d, hook, "i_X"'] & Y \arrow[r, "g"'] \arrow[d, hook, "i_Y"] & Z \arrow[d, hook, "i_Z"] \\ \varepsilon X \arrow[r, dashed, "\varepsilon h"] \arrow[rr, bend right, "\varepsilon(g \circ h)"'] & \varepsilon Y \arrow[r, dashed, "\varepsilon g"] & \varepsilon Z \end{tikzcd}$$

i.e., $\varepsilon(g \circ h) = \varepsilon g \circ \varepsilon h$.
(iii) If X is a set, let $E_0(X) := \varepsilon X$, and, if $h : X \to Y$ is a function from the set X to the set Y, let $E_1(h) := \varepsilon h : \varepsilon X \to \varepsilon Y$ is the linear map determined in the case (i). Then the pair $E := (E_0, E_1)$ is a covariant functor from the category ***Set*** *to the category* ***Lin****.*

Proof. Exercise. $\square$

Proposition 2.6.8. *Let X, Y be linear spaces and $h : X \to Y$ a function.*
(i) There is a unique linear map $\pi_X : \varepsilon X \to X$ such that $\pi_X \circ i_X = \mathrm{id}_X$

$$\begin{tikzcd}
X \arrow[r, hook, "i_X"] \arrow[dr, "\mathrm{id}_X"'] & \varepsilon X \arrow[d, dashed, "\pi_X"] \\
 & X.
\end{tikzcd}$$

(ii) Let the following subset of εX

$$N(X) := \big\{f_{\lambda x+\mu y} - \lambda f_x - \mu f_y \mid x, y \in X \ \& \ \lambda, \mu \in \mathbb{R}\big\}.$$

Then $\langle N(X)\rangle \subseteq \mathtt{Ker}(\pi_X)$*.*

(iii) $\mathtt{Ker}(\pi_X) \subseteq \langle N(X)\rangle$*.*

(iv) The function h *is a linear map if and only if* $\pi_Y \circ \varepsilon h = h \circ \pi_X$

$$\begin{tikzcd}
X \arrow[r, "h"] \arrow[d, "i_X"] & Y \arrow[d, "i_Y"'] \\
\varepsilon X \arrow[r, dashed, "\varepsilon h"'] \arrow[u, bend left, "\pi_X"] & \varepsilon Y. \arrow[u, bend right, "\pi_Y"']
\end{tikzcd}$$

Proof. We describe only the steps to find the algorithm for the proof of the case (iii). Since $\overline{0}^X \in \mathtt{Ker}(\pi_X)$, we only want to show that for every $n \geq 1$, and for every $x_1, \ldots x_n \in X$,

$$f = \sum_{i=1}^{n} f(x_i) f_{x_i} \in \mathtt{Ker}(\pi_X) \Rightarrow f \in \langle N(X)\rangle.$$

(a) Show that $f_{\mathbf{0}} \in N(X)$, and show the case $n = 1$ i.e., if $f = f(x_1)f_{x_1} \in \mathtt{Ker}(\pi_X)$, then $f \in \langle N(X)\rangle$.

(b) Show the case $n = 2$ i.e., if $f = f(x_1)f_{x_1} + f(x_2)f_{x_2} \in \mathtt{Ker}(\pi_X)$, then $f \in \langle N(X)\rangle$.

(c) Show the case $n = 3$ i.e., if $f = f(x_1)f_{x_1} + f(x_2)f_{x_2} + f(x_3)f_{x_3} \in \mathtt{Ker}(\pi_X)$, then $f \in \langle N(X)\rangle$.

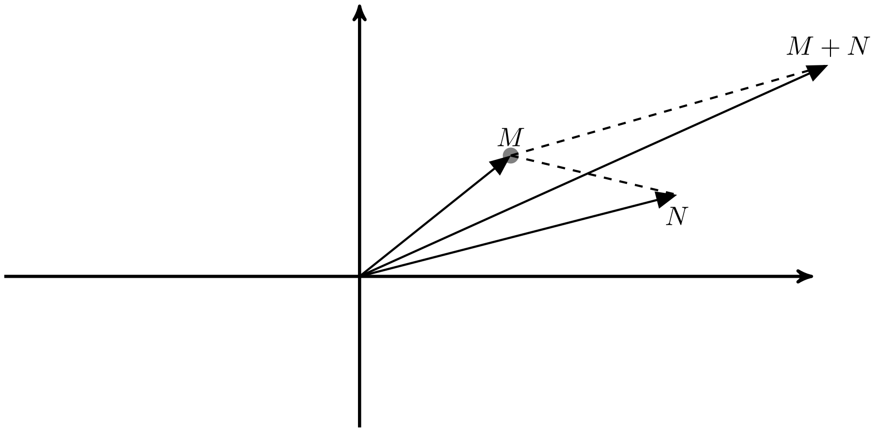
(d) Show the case $n = 4$ i.e., if $f = f(x_1)f_{x_1} + f(x_2)f_{x_2} + f(x_3)f_{x_3} + f(x_4)f_{x_4} \in \mathtt{Ker}(\pi_X)$, then $f \in \langle N(X)\rangle$.

The algorithm for the general case is either evident in the proof of this case, or, if not, it should be clear in the proof of the case $n = 5$. $\square$

2.7. Convex sets

In this section we study the notion of a convex set in a linear space, proving some very first results in the theory of convex sets. The area of "Convex Analysis" is very broad, and we refer to [**14**] for further reading.

If $M, N \in \mathbb{R}^n$, the *linear segment* $[M, M+N]$ between M and $M+N$



is the following set

$$[M, M+N] := \big\{M + tN \mid t \in [0,1]\big\}.$$

If $t := 0$, then $M + 0N = M$, and if $t := 1$, then $M + 1N = M + N$. If, for example, S is the middle point of the segment, then

$$S = \frac{M + (M+N)}{2} = M + \frac{N}{2} = M + \frac{1}{2}N.$$

Since $N = M + (N - M)$, we have that the linear segment $[M, N]$ between M and N is the following set:

$$\begin{aligned}
[M, N] &= [M, M + (N - M)] \\
&= \big\{M + t(N - M) \mid t \in [0,1]\big\} \\
&= \big\{tN + (1-t)M \mid t \in [0,1]\big\} \\
&= \big\{sM + (1-s)N \mid s \in [0,1]\big\} \\
&= \big\{t_1 M + t_2 N \mid t_1, t_2 \in [0,1] \ \& \ t_1 + t_2 = 1\big\} \\
&= \big\{t_1 M + t_2 N \mid t_1, t_2 \geq 0 \ \& \ t_1 + t_2 = 1\big\}.
\end{aligned}$$

The next definition is the generalisation of the description of $[M, N]$ in $\mathbb{R}^n$.

Definition 2.7.1. If X is a linear space and $x, y \in X$, the *linear segment* $[x, y]$ between x and y is defined by

$$[x, y] := \big\{tx + (1-t)y \mid t \in [0,1]\big\}.$$

Clearly, for every $x, y \in X$ we have that

$$[x, y] = [y, x] \quad \& \quad [x, x] = \{x\}.$$

Moreover, if $f : X \to Y$ is a linear map, then f preserves linear segments i.e.,

$$\begin{aligned} f([x_1, x_2]) &:= f\Big(\big\{tx_1 + (1-t)x_2 \mid t \in [0,1]\big\}\Big) \\ &:= \big\{f\big(tx_1 + (1-t)x_2\big) \mid t \in [0,1]\big\} \\ &= \big\{tf(x_1) + (1-t)f(x_2) \mid t \in [0,1]\big\} \\ &:= [f(x_1), f(x_2)]. \end{aligned}$$

If $y_1, y_2 \in Y$, the inverse image $f^{-1}([y_1, y_2])$ of the segment $[y_1, y_2]$ in Y under a linear map f need not be a linear segment (e.g., consider the constant $\mathbf{0}$ linear map and the segment $[\mathbf{0}, \mathbf{0}] = \{\mathbf{0}\}$ in Y). What we have in this case though, is that

$$x_1, x_2 \in f^{-1}([y_1, y_2]) \Rightarrow [x_1, x_2] \subseteq f^{-1}([y_1, y_2])$$

i.e., $f^{-1}([y_1, y_2])$ is "closed under initial segments", since $f(x_1), f(x_2) \in [y_1, y_2]$, and then $f([x_1, x_2]) = [f(x_1), f(x_2)] \subseteq [y_1, y_2]$.

Definition 2.7.2. If X is a linear space, $n \geq 1$, and $x_1, \ldots, x_n \in X$, the *convex hull*, or the *convex span* of $x_1, \ldots, x_n$, is defined by

$$\texttt{Conv}(x_1, \ldots, x_n) := \left\{ \sum_{i=1}^{n} t_i x_i \mid t_1, \ldots, t_n \geq 0 \ \& \ \sum_{i=1}^{n} t_i = 1 \right\} \subsetneq \langle x_1, \ldots, x_n \rangle.$$

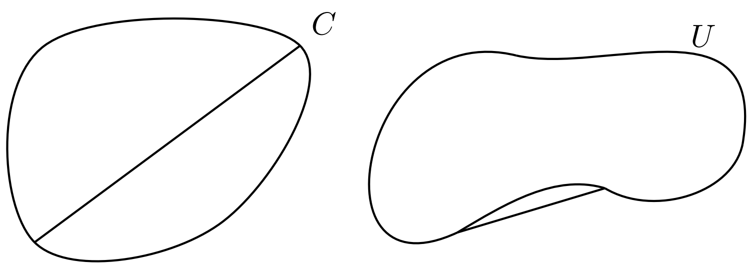
A linear combination

$$\sum_{i=1}^{n} t_i x_i, \quad \text{where } t_1, \ldots, t_n \geq 0 \ \& \ \sum_{i=1}^{n} t_i = 1$$

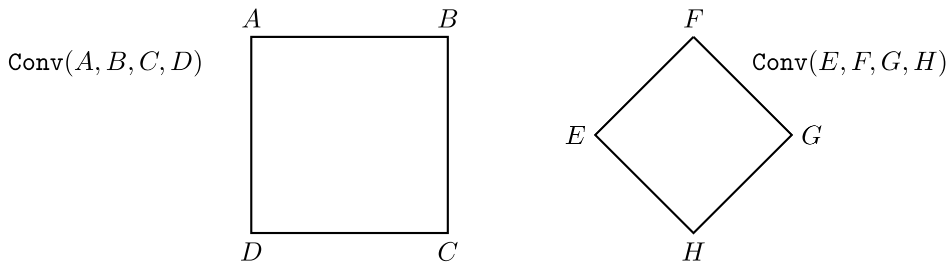
is called a *convex combination* of $x_1, \ldots, x_n$, and $t_1, \ldots, t_n$ are called the *barycentric coordinates* of $x = \sum_{i=1}^{n} t_i x_i$. A subset C of X is called *convex in* X, or, simply, *convex*, if the linear segment between any two elements of C is included in C, or, if C is closed under linear segment i.e.,

$$\forall_{c,d \in C}\big([c, d] \subseteq C\big).$$

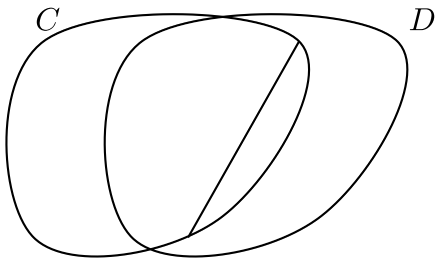
Next we draw a convex set C, a non-convex set U in $\mathbb{R}^2$,



and we also draw the convex hulls $\texttt{Conv}(A, B, C, D)$ and $\texttt{Conv}(E, F, G, H)$ in $\mathbb{R}^2$, which are the following figures $ABCD$ and $EFGH$ *with* their interior.



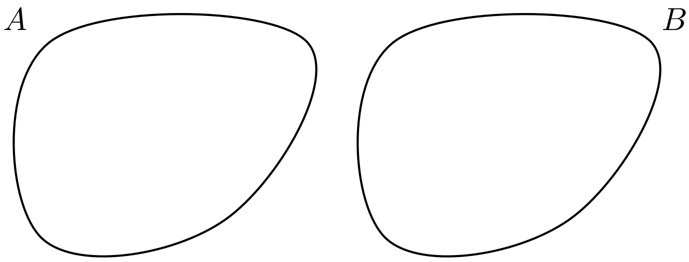
EXAMPLE 2.7.3. Any subspace Y of a linear space X is convex. Hence, X itself is convex. There are many convex subsets of X that are not subspaces of X. E.g., if X is a non-trivial linear space, and $x \neq \mathbf{0}$, then $\{x\}$ is convex in X. If $x, y \in X$, then $[x, y]$ is convex in X. If C, D are convex sets in X, their intersection $C \cap D$ is convex in X (why?).



Similarly, the intersection of any family $(C_i)_{i \in I}$ if convex sets in X is convex. The empty set $\emptyset$ is trivially convex in X, since the defining property of a convex set is trivially satisfied[15]

$$\forall_{c,d \in \emptyset}\big([c, d] \subseteq \emptyset\big) :\Leftrightarrow \forall_{c,d}\big(c \in \emptyset \ \& \ d \in \emptyset \Rightarrow [c, d] \subseteq \emptyset\big).$$

The union of two convex sets, like A and B in $\mathbb{R}^2$, is not always a convex set.



[15]Here we use the logical fact $(P \Rightarrow \bot) \Rightarrow (P \Rightarrow Q)$, where P, Q are formulas. For the derivation of this formula we use the logical rules Efq and Modus Ponens.

REMARK 2.7.4. Let X, Y be linear spaces, $C, C' \subseteq X$, and $D \subseteq Y$.
(i) If C is convex in X, and D is convex in Y, then $C \times D$ is convex in $X \times Y$.
(ii) If C, C' are convex, then $C + C' := \{c + c' \mid c \in C \ \&\ c' \in C'\}$ is convex.
(iii) If C is convex, and $a \in \mathbb{R}$, then $aC := \{a \cdot c \mid c \in C\}$ is convex.

PROOF. Exercise. $\square$

Notice that $\texttt{Conv}(x_1) = \{x_1\}$, and $\texttt{Conv}(x_1, x_2) = [x_1, x_2]$, while $\texttt{Conv}(x_1, x_2, x_3)$ is the *triangle*, with its interior, of x_1, x_2, x_3 in X. Generally, more than one convex combinations correspond to the same vector e.g.,

$$\frac{1}{2} = \frac{1}{2}\left(\frac{3}{8}\right) + \frac{1}{2}\left(\frac{5}{8}\right) = \frac{3}{4}\left(\frac{7}{16}\right) + \frac{1}{4}\left(\frac{11}{16}\right).$$

PROPOSITION 2.7.5. *If X is a linear space, $n \geq 1$, and $x_1, \ldots, x_n \in X$, their convex hull $\texttt{Conv}(x_1, \ldots, x_n)$ is the least convex set in X that contains $x_1, \ldots, x_n$.*

PROOF. First we show that $\texttt{Conv}(x_1, \ldots, x_n)$ is convex in X. Let

$$x = \sum_{i=1}^{n} t_i x_i, \quad t_1, \ldots, t_n \geq 0 \ \ \& \ \ \sum_{i=1}^{n} t_i = 1,$$

$$y = \sum_{i=1}^{n} s_i x_i, \quad s_1, \ldots, s_n \geq 0 \ \ \& \ \ \sum_{i=1}^{n} s_i = 1.$$

If $t \in [0, 1]$, then

$$\begin{aligned} tx + (1-t)y &= t\sum_{i=1}^{n} t_i x_i + (1-t)\sum_{i=1}^{n} s_i x_i \\ &= \sum_{i=1}^{n} t t_i x_i + \sum_{i=1}^{n}(1-t) s_i x_i \\ &= \sum_{i=1}^{n} [t t_i + (1-t)s_i] x_i. \end{aligned}$$

If $a_i := t t_i + (1-t)s_i$, for every $i \in \{1, \ldots, n\}$, and if $a, b \in \mathbb{R}$, let

$$a \wedge b := \min\{a, b\},$$

then, since $a \geq a \wedge b$ and $b \geq a \wedge b$, we have that

$$a_i \geq t(t_i \wedge s_i) + (1-t)(t_i \wedge s_i) = [t + (1-t)](t_i \wedge s_i) = t_i \wedge s_i \geq 0,$$

and

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} t t_i + \sum_{i=1}^{n}(1-t)s_i = t\sum_{i=1}^{n} t_i + (1-t)\sum_{i=1}^{n} s_i = t + (1-t) = 1.$$

Next, we show that if C is a convex set in X that contains $x_1, \ldots, x_n$, then $\texttt{Conv}(x_1, \ldots, x_n) \subseteq C$. For that we prove inductively the following formula:

$$\forall_{n \geq 1} \Big(\forall_{x_1, \ldots, x_n \in X} \big(x_1 \in C \ \& \ \ldots \ \& \ x_n \in C \Rightarrow \texttt{Conv}(x_1, \ldots, x_n) \subseteq C \big) \Big).$$

If $n = 1$, we show that $\forall_{x_1 \in X} \big(x_1 \in C \Rightarrow \texttt{Conv}(x_1) \subseteq C \big)$, which we get immediately from the equality $\texttt{Conv}(x_1) = \{x_1\}$. Next we suppose that

$$\forall_{x_1, \ldots, x_n \in X} \big(x_1 \in C \ \& \ \ldots \ \& \ x_n \in C \Rightarrow \texttt{Conv}(x_1, \ldots, x_n) \subseteq C \big),$$

and we show that

$$\forall_{x_1, \ldots, x_{n+1} \in X} \big(x_1 \in C \ \& \ \ldots \ \& \ x_{n+1} \in C \Rightarrow \texttt{Conv}(x_1, \ldots, x_{n+1}) \subseteq C \big).$$

Let $x_1, \ldots, x_n, x_{n+1} \in X$, such that $x_1 \in C \ \& \ \ldots \ \& \ x_n \in C \ \& \ x_{n+1} \in C$, and let

$$x = \sum_{i=1}^{n+1} t_i x_i, \quad t_1, \ldots, t_{n+1} \geq 0 \ \ \& \ \sum_{i=1}^{n+1} t_i = 1.$$

If there is some $i \in \{1, \ldots, n+1\}$, such that $t_i = 0$, then by the inductive hypothesis on the n-elements $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1}$ of X we get

$$x = \sum_{j=1, j \neq i}^{n+1} t_j x_j \in C.$$

If $t_i \neq 0$, for every $i \in \{1, \ldots, n+1\}$, then $t_{n+1} \neq 1$, and hence

$$x = (1 - t_{n+1})c + t_{n+1} x_{n+1},$$

$$c := \frac{t_1}{1 - t_{n+1}} x_1 + \ldots + \frac{t_n}{1 - t_{n+1}} x_n.$$

Since

$$\frac{t_i}{1 - t_{n+1}} \geq 0, \qquad i \in \{1, \ldots, n\},$$

and

$$\frac{t_1}{1 - t_{n+1}} + \ldots + \frac{t_n}{1 - t_{n+1}} = \frac{1}{1 - t_{n+1}} \big(t_1 + \ldots + t_n \big) = \frac{1 - t_{n+1}}{1 - t_{n+1}} = 1$$

we get $c \in \texttt{Conv}(x_1, \ldots, x_n)$. By the inductive hypothesis on $x_1, \ldots, x_n$ we have that $c \in C$, and hence, since C is convex, we get $x \in C$. $\square$

As in the case of the linear span, the notion of the convex hull is generalised to an arbitrary subset Y of a linear space X.

Definition 2.7.6. If X is a linear space, and $Y \subseteq X$, the *convex hull* $\texttt{Conv}_X(Y)$, or simpler $\texttt{Conv}(Y)$, of Y in X is defined by

$$\texttt{Conv}(Y) := \bigcap \{ C \subseteq X \mid Y \subseteq C \ \& \ C \text{ is convex} \}.$$

Clearly, $\texttt{Conv}(Y)$ is the least convex set including Y. Moreover, since $\emptyset$ is convex, $\texttt{Conv}(\emptyset) = \emptyset \subsetneq \{\mathbf{0}\} = \langle \emptyset \rangle$.

PROPOSITION 2.7.7. *Let X be a linear space, and $Y, Z \subseteq X$.*

(i) $Y \subseteq \mathtt{Conv}(Y)$.

(ii) If $Y \subseteq Z$, *then* $\mathtt{Conv}(Y) \subseteq \mathtt{Conv}(Z)$.

(iii) $\mathtt{Conv}(\mathtt{Conv}(Y)) = \mathtt{Conv}(Y)$.

(iv) $\mathtt{Conv}(Y \cap Z) \subseteq \mathtt{Conv}(Y) \cap \mathtt{Conv}(Z)$.

(v) $\mathtt{Conv}(Y) \cup \mathtt{Conv}(Z) \subseteq \mathtt{Conv}(Y \cup Z)$.

(vi) Y *is convex if and only if* $\mathtt{Conv}(Y) = Y$.

PROOF. Exercise. □

As in the case of linear span, the "top-down" definition of the convex hull of any subset Y has an equivalent "bottom-up" description.

PROPOSITION 2.7.8. *If X is a linear space, and $Y \subseteq X$, then the convex hull* $\mathtt{Conv}(Y)$ *of Y is equal to all convex combinations of the elements of Y.*

PROOF. What we need to show amounts to the equality $\mathtt{Conv}(Y) = \hat{Y}$, where

$$\hat{Y} := \bigg\{ \sum_{i=1}^{n} t_i y_i \mid n \geq 1 \ \& \ \forall_{i \in \{1,\ldots,n\}} \big(t_i \geq 0 \ \& \ y_i \in Y\big) \ \& \ \sum_{i=1}^{n} t_i = 1 \bigg\}.$$

By the Proposition 2.7.5, and since trivially $Y \subseteq \hat{Y}$, to show that $\mathtt{Conv}(Y) \subseteq \hat{Y}$, it suffices to show that $\hat{Y}$ is convex. Let

$$x = \sum_{i=1}^{n} t_i y_i, \quad y_1, \ldots, y_n \in Y \ \ \& \ \ t_1, \ldots, t_n \geq 0 \ \ \& \ \ \sum_{i=1}^{n} t_i = 1,$$

$$x' = \sum_{j=1}^{m} s_j {y_j}', \quad {y_1}', \ldots, {y_m}' \in Y \ \ \& \ \ s_1, \ldots, s_m \geq 0 \ \ \& \ \ \sum_{j=1}^{m} s_j = 1.$$

If $t \in [0,1]$, then

$$tx + (1-t)x' = \sum_{i=1}^{n} (tt_i) y_i + \sum_{j=1}^{m} (1-t) s_j {y_j}' \in \hat{Y},$$

since $tt_i \geq 0$, and $(1-t)s_j \geq 0$, and

$$\sum_{i=1}^{n} (tt_i) + \sum_{j=1}^{m} (1-t) s_j = t\bigg(\sum_{i=1}^{n} t_i\bigg) + (1-t)\bigg(\sum_{j=1}^{m} s_j\bigg) = t1 + (1-t)1 = 1.$$

By the definition of $\mathtt{Conv}(Y)$, to show the inclusion $\hat{Y} \subseteq \mathtt{Conv}(Y)$, it suffices to show that if C is convex in X, such that $Y \subseteq C$, then $\hat{Y} \subseteq C$. Let $x = \sum_{i=1}^{n} t_i y_i \in \hat{Y}$, for some $n \geq 1, t_1, \ldots, t_n \geq 0$, such that $\sum_{i=1}^{n} t_i = 1$, and $y_1, \ldots, y_n \in Y$. Hence, $x \in \mathtt{Conv}(y_1, \ldots, y_n)$. Since $\{y_1, \ldots, y_n\} \subseteq Y \subseteq C$, by Proposition 2.7.5 we get $x \in \mathtt{Conv}(y_1, \ldots, y_n) \subseteq C$. □

REMARK 2.7.9. Let X, Y be linear spaces, $f \in \mathcal{L}(X, Y)$, $C \subseteq X$, $D \subseteq Y$, and $Z \preceq X$.

(i) If D is convex in Y, then $f^{-1}(D) := \{x \in X \mid f(x) \in D\}$ is convex in X.

(ii) If C is convex in X, then $f(C) := \{f(x) \mid x \in C\}$ is convex in Y.

(iii) If C is convex in X, then $C + Z := \{c + Z \mid c \in C\}$ is convex in X/Z.

(iv) If $n \geq 1$, and $x_1, \ldots, x_n \in X$, then

$$f\big(\mathtt{Conv}(x_1, \ldots, x_n)\big) = \mathtt{Conv}(f(x_1), \ldots, f(x_n)).$$

PROOF. Exercise. $\square$

COROLLARY 2.7.10. *If X is a linear space, $f, g \in X^* := \mathcal{L}(X, \mathbb{R})$, and $a \in \mathbb{R}$, then the following sets*

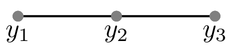
$$\begin{aligned}
[f > a] &:= \{x \in X \mid f(x) > a\},\\
[f \geq a] &:= \{x \in X \mid f(x) \geq a\},\\
[f = a] &:= \{x \in X \mid f(x) = a\},\\
[f < a] &:= \{x \in X \mid f(x) > a\},\\
[f \leq a] &:= \{x \in X \mid f(x) \geq a\},\\
[f > g] &:= \{x \in X \mid f(x) > g(x)\},\\
[f \geq g] &:= \{x \in X \mid f(x) \geq g(x)\},\\
[f = g] &:= \{x \in X \mid f(x) = g(x)\},
\end{aligned}$$

are convex in X.

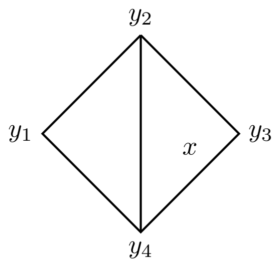
PROOF. Exercise. $\square$

2.8. Carathéodory's theorem

If $y_1, y_2, y_3 \in \mathbb{R}$ and $x \in \mathtt{Conv}(y_1, y_2, y_3)$, it is easy to see that x is in the convex hull of at most two elements of $\{y_1, y_2, y_3\}$. If $x \in \{y_1, y_2, y_3\}$, then it is in the convex hull of one of them, while if $x \notin \{y_1, y_2, y_3\}$, and supposing, without loss of generality, that $y_1 < y_2 < y_3$, then the hypothesis $x \in \mathtt{Conv}(y_1, y_2, y_3)$ implies that $x \in [y_1, y_3]$, and hence $x \in [y_1, y_2] = \mathtt{Conv}(y_1, y_2)$, or $x \in [y_2, y_3] = \mathtt{Conv}(y_2, y_3)$.



Similarly, if $y_1, y_2, y_3, y_4 \in \mathbb{R}^2$ and $x \in \mathtt{Conv}(y_1, y_2, y_3, y_4)$, it is easy to see that x is in the convex hull of at most three elements of $\{y_1, y_2, y_3, y_4\}$. If $x \in \{y_1, y_2, y_3, y_4\}$, then it is in the convex hull of one of them, while if x is in the line segment between two of them, it is in the convex hull of two of them. In any other case, we can find a triangle with vertices in $\{y_1, y_2, y_3, y_4\}$ in the interior of which x belongs to.



Carathéodory's theorem is a generalisation of these intuitive remarks. According to it, if $Y \subseteq \mathbb{R}^n$, and $x \in \mathtt{Conv}(Y)$, we can find at most $(n+1)$-elements in Y such that x is in their convex hull.

THEOREM 2.8.1 (Carathéodory). *If $Y \subseteq \mathbb{R}^n$, then for every $x \in \mathtt{Conv}(Y)$ there is $m \in \mathbb{N}$ such that $1 \leq m \leq n+1$, and there are $y_1, \ldots, y_m \in Y$, such that $x \in \mathtt{Conv}(y_1, \ldots, y_m)$.*

PROOF. If $x \in \mathtt{Conv}(Y)$, then by the Proposition 2.7.8 there is $k \geq 1$, and there are $y_1, \ldots, y_k \in Y$, such that

$$x = \sum_{i=1}^{k} t_i y_i, \quad t_1, \ldots, t_k \geq 0 \ \ \& \ \ \sum_{i=1}^{k} t_i = 1.$$

If $k \leq n+1$, we have nothing to prove. Suppose next that $k > n+1 \Leftrightarrow k-1 > n$, and suppose also that $t_1 > 0 \ \& \ \ldots \ \& \ t_k > 0$, since if $t_i = 0$, for some $i \in \{1, \ldots, k\}$, we can write immediately x as a convex combination of $k-1$ elements of Y. Actually, our strategy is to prove that there are $k-1$ elements of Y, such that x can be written as a convex combination of them. By repeating the same argument at most $k-(n+1)$ number of times, we reach the required conclusion. To find these $k-1$ elements of Y, we work as follows. Since by the Theorem 2.2.13 the $k-1 > n$ number of vectors in $\mathbb{R}^n$

$$(y_1 - y_k), \ldots, (y_{k-1} - y_k)$$

are linearly dependent, there are $a_1, \ldots, a_{k-1} \in \mathbb{R}$, such that

$$\sum_{j=1}^{k-1} a_j (y_j - y_k) = \mathbf{0},$$

and $a_j \neq 0$, for some $j \in \{1, \ldots, k-1\}$. If we define

$$a_k := -\sum_{j=1}^{k-1} a_j,$$

then

$$\sum_{i=1}^{k} a_i = 0 \quad \& \quad \sum_{i=1}^{k} a_i y_i = \mathbf{0},$$

since

$$\begin{aligned}
\sum_{i=1}^{k} a_i y_i &= \sum_{j=1}^{k-1} a_j y_j + a_k y_k \\
&= \sum_{j=1}^{k-1} a_j y_j + \bigg(- \sum_{j=1}^{k-1} a_j \bigg) y_k \\
&= \sum_{j=1}^{k-1} a_j y_j - \sum_{j=1}^{k-1} a_j y_k \\
&= \sum_{j=1}^{k-1} a_j (y_j - y_k) \\
&= \mathbf{0}.
\end{aligned}$$

We claim that there is some $i \in \{1, \ldots, k\}$ such that $a_i > 0$. If there is some $j \in \{1, \ldots, k-1\}$, such that $a_j > 0$, we are done. If $a_j \leq 0$, for every $j \in \{1, \ldots, k-1\}$, and since in this case there is some $j \in \{1, \ldots, k-1\}$ with $a_j < 0$, then $a_k > 0$. Hence, by our justified claim, and our hypothesis that each $t_i > 0$, there is some $l \in \{1, \ldots, k\}$ such that

$$\max\left\{\frac{a_1}{t_1}, \ldots, \frac{a_k}{t_k}\right\} = \frac{a_l}{t_l} := \frac{1}{M} > 0,$$

where $M := \frac{t_l}{a_l}$. We show that $x \in \texttt{Conv}(y_1, \ldots, y_{l-1}, y_{l+1}, \ldots, y_k)$. If

$$s_j := t_j - M a_j, \qquad j \in \{1, \ldots, l-1, l+1, \ldots, k\},$$

then, for every $j \in \{1, \ldots, l-1, l+1, \ldots, k\}$, we have that

$$\frac{a_j}{t_j} \leq \frac{1}{M} \Rightarrow M a_j \leq t_j \Rightarrow s_j := t_j - M a_j \geq 0,$$

and since $\sum_{i=1}^{k} a_i y_i = \mathbf{0}$, we have that $\sum_{j=1, j\neq l}^{k} a_j y_j = -(a_l y_l)$, and hence

$$\begin{aligned}
\sum_{j=1, j\neq l}^{k} s_j y_j &:= \sum_{j=1, j\neq l}^{k} (t_j - M a_j) y_j \\
&= \sum_{j=1, j\neq l}^{k} t_j y_j - M \sum_{j=1, j\neq l}^{k} a_j y_j \\
&= \sum_{j=1, j\neq l}^{k} t_j y_j - \frac{t_l}{a_l}\big[- (a_l y_l)\big] \\
&= \sum_{i=1}^{k} t_i y_i
\end{aligned}$$

$$= x.$$

Since $\sum_{i=1}^{k} a_i = 0$, we have that $\sum_{j=1, j\neq l}^{k} a_j = -a_l$, and hence

$$\begin{aligned}\sum_{j=1, j\neq l}^{k} s_j &:= \sum_{j=1, j\neq l}^{k} (t_j - Ma_j) \\ &= \sum_{j=1, j\neq l}^{k} t_j - M \sum_{j=1, j\neq l}^{k} a_j \\ &= \sum_{j=1, j\neq l}^{k} t_j - \frac{t_l}{a_l}[-a_l] \\ &= \sum_{i=1}^{k} t_i \\ &= 1\end{aligned}$$

i.e., x was written as a convex combination of $k-1$ elements of Y. □

By Carathéodory's theorem, an element x of $\mathtt{Conv}(Y)$, where $Y \subseteq \mathbb{R}^n$ can be written as

$$x = \sum_{i=1}^{n+1} t_i y_i, \quad t_1, \ldots t_{n+1} \geq 0, \quad y_1, \ldots, y_{n+1} \in Y,$$

since, even if the number m in the formulation of Carathéodory's theorem is smaller than $n+1$, one can consider in the above convex combination any elements of Y multiplied by 0's.

If Y is the union of a number of convex sets in $\mathbb{R}^n$, which is smaller than the dimension n of $\mathbb{R}^n$, then, according to the next corollary of Carathéodory's theorem, we can find a convex combination for the elements of $\mathtt{Conv}(Y)$ even smaller from the one determined by Carathéodory's theorem.

COROLLARY 2.8.2. *If $1 \leq k \leq n$, and $C_1, \ldots, C_k$ are convex subsets of $\mathbb{R}^n$, then for every*

$$x \in \mathtt{Conv}\left(\bigcup_{i=1}^{k} C_i\right)$$

there is $m \leq k$, and there are $y_1, \ldots, y_m \in \bigcup_{i=1}^{k} C_i$, such that

$$x \in \mathtt{Conv}(y_1, \ldots, y_m).$$

HINT: The case $k = 1$ is trivial, since by the convexity of C_1 we have that

$$x \in \mathtt{Conv}\left(\bigcup_{i=1}^{1} C_i\right) = \mathtt{Conv}(C_1) = C_1 \Rightarrow x \in \mathtt{Conv}(x) = \{x\}.$$

It is helpful to work next with two convex sets C_1, C_2 in $\mathbb{R}^2$, and to show that for every $x \in \texttt{Conv}(C_1 \cup C_2)$ there are at most two elements $y, y' \in C_1 \cup C_2$ such that $x \in \texttt{Conv}(y, y')$. The proof of the general case uses the trick of the case of $\mathbb{R}^2$ and the method of proof of Carathéodory's theorem. Namely, we suppose that the determined by Carathéodory's theorem convex combination of x with at most $m \leq n+1$ number of elements from $\bigcup_{i=1}^{k} C_i$ is larger than k, and we find a convex combination of x from $m-1$ number of vectors from $\bigcup_{i=1}^{k} C_i$. The rest is an exercise. $\square$

CHAPTER 3

Matrices

3.1. The linear space of matrices

DEFINITION 3.1.1. If $m, n \geq 1$, an array of real numbers

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

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

$$A_i := \begin{bmatrix} a_{i1} & \dots & a_{in} \end{bmatrix} := [a_{ij}]_i \quad (\in \mathbb{R}^n),$$

and we also write A as a column of its rows

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix}.$$

The *row-rank* of A is defined by

$$\texttt{rRank}(A) := \dim \big(\langle A_1, \dots, A_m \rangle\big)$$

and it is the maximum number of linearly independent rows of A. If $1 \leq j \leq n$, the j-th column of A is the array

$$A^j := \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} := [a_{ij}]^j \quad (\in \mathbb{R}^m),$$

and we also write A as a row of its columns

$$A = [A^1 \dots A^n].$$

The *column-rank* of A is defined by

$$\texttt{cRank}(A) := \dim \big(\langle A^1, \dots, A^n \rangle\big),$$

and it is the maximum number of linearly independent columns of A. 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 $\lambda \in \mathbb{R}$, we define

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

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

$$\lambda \cdot [a_{ij}] := [\lambda a_{ij}],$$

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

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

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

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

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

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

Remark 3.1.2. $M_{m,n}(\mathbb{R})$ 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 3.1.1. 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 Remark 2.4.14(ii) we get $\dim\big(M_{m,n}(\mathbb{R})\big) = \dim(\mathbb{R}^{mn}) = mn$. $\square$

DEFINITION 3.1.3. 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[1]

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

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 3.1.4. 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) $(a \cdot B)^t = a \cdot B^t$.

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

(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 3.1.5. 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

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

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

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 3.1.6. *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 3.1.7. *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) $I_n^t = I_n$.

(iii) If A is invertible, then $(A^{-1})^t = (A^t)^{-1}$.

PROOF. (i) $C = I_n C = (AB)C = (BA)C = B(AC) = BI_n = B$.
(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 3.1.6(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
$$(AB)(B^{-1}A^{-1}) = A[B(B^{-1}A^{-1})] = A[(BB^{-1})A^{-1}] = A[I_n A^{-1}] = AA^{-1} = I_n.$$

DEFINITION 3.1.8. If $A := [a_{ij}] \in M_n(\mathbb{R})$, the *trace* $\texttt{Tr}(A)$ is defined by
$$\texttt{Tr}(A) := \sum_{i=1}^{n} a_{ii}.$$

It is immediate to see that $\mathtt{Tr}(A^t) = \mathtt{Tr}(A)$, and $\mathtt{Tr}(I_n) = n$.

REMARK 3.1.9. Let $A, B, C \in M_n(\mathbb{R})$ and $\lambda \in \mathbb{R}$.
(i) $\mathtt{Tr}(A + B) = \mathtt{Tr}(A) + \mathtt{Tr}(B)$.
(ii) $\mathtt{Tr}(\lambda A) = \lambda \mathtt{Tr}(A)$.
(iii) $\mathtt{Tr}(AB) = \mathtt{Tr}(BA)$.
(iv) If B is invertible, then $\mathtt{Tr}(B^{-1}AB) = \mathtt{Tr}(A)$.
(v) $\mathtt{Tr}(A(B + C)) = \mathtt{Tr}(AB) + \mathtt{Tr}(AC)$.
(vi) $\mathtt{Tr}((\lambda A)B) = \lambda \mathtt{Tr}(AB)$.
(vii) There are no matrices $A, B \in M_n(\mathbb{R})$ such that

$$AB - BA = I_n.$$

(viii) If $A \in M_n(\mathbb{R})$ such that for every $B \in M_n(\mathbb{R})$, we have that $\mathtt{Tr}(AB) = 0$, then $A = \mathbf{0}_n$.

PROOF. Exercise. □

PROPOSITION 3.1.10. *(i) The set of symmetric matrices* $\mathtt{Sym}_n(\mathbb{R})$ *is a linear subspace of* $M_n(\mathbb{R})$.
(ii) If $A \in \mathtt{Sym}_n(\mathbb{R})$*, then* $\mathtt{Tr}(AA) \geq 0$.
(iii) If $A \in \mathtt{Sym}_n(\mathbb{R})$ *and* $A \neq \mathbf{0}_n$*, then* $\mathtt{Tr}(AA) > 0$.

PROOF. Exercise. □

As we shall explain later, because of the above properties, the function $(A, B) \mapsto \mathtt{Tr}(AB)$ is an inner product on $\mathtt{Sym}_n(\mathbb{R})$.

3.2. The linear map of a matrix

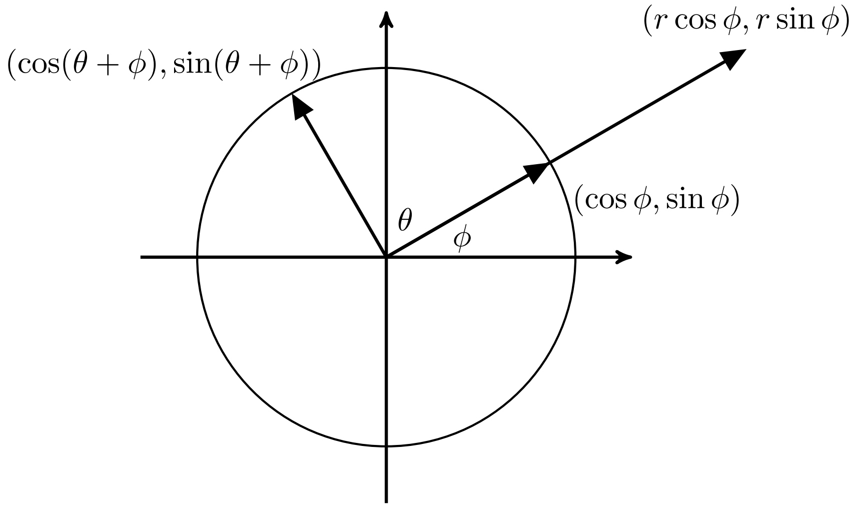
Matrices can be used to represent linear maps. Let's see the following characteristic example. If $\theta \in \mathbb{R}$, let the matrix

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

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

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

$$= r \begin{bmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi \\ \sin\theta\cos\phi + \cos\theta\cos\phi \end{bmatrix}$$
$$= r \begin{bmatrix} \cos(\theta+\phi) \\ \sin(\theta+\phi) \end{bmatrix},$$



where $r := \sqrt{x^2+y^2}$. Hence, R_θ is the anti-clockwise θ-rotation of the vector (x, y). From the geometric interpretation of R_θ we can infer that R_θ is a linear map. 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, a fact which is also expected from the geometric interpretation of R_θ. The 0-rotation R_0 is the identity map and $R(0) = I_2$, and if we consider the $(-\theta)$-rotation, then $R_\theta(R_{-\theta}(x)) = R_0(x) = x$. From the geometric interpretation of R_θ we expect that R_θ preserves the length of vectors in $\mathbb{R}^2$ i.e.,

$$|R_\theta\big((x,y)\big)| = |(x,y)|.$$

Definition 3.2.1. 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, and the output $T_A(x)$ is an $m \times 1$-matrix that represents a vector in $\mathbb{R}^m$.

If we unfold the above definition, we have that

$$\begin{aligned}
T_A(x) &= \begin{bmatrix} T_A(x)_1 \\ \vdots \\ T_A(x)_m \end{bmatrix} \\
&:= \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&:= \left(\sum_{j=1}^{n} a_{1j}x_j, \dots, \sum_{j=1}^{n} a_{ij}x_j, \dots, \sum_{j=1}^{n} a_{mj}x_j \right) \\
&= \big(x_1a_{11}, \dots, x_1a_{m1}\big) + \dots + \big(x_na_{1n}, \dots, x_na_{mn}\big) \\
&= x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \dots x_j \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \\
&= x_1A^1 + \dots x_jA^j + \dots + x_nA^n.
\end{aligned}$$

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

$$T_A(e_j) = 0A^1 + \dots + 0A^{j-1} + 1A^j + 0A^{j+1} + \dots + 0A^n = A^j,$$

and hence, for every $i \in \{1, \dots, m\}$, we have that

$$T_A(e_j)_i = a_{ij}.$$

From the linear structure of $M_n(\mathbb{R})$ we get $T_A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, since

$$T_A(X + Y) := A(X + Y) = AX + AY := T_AX + T_AY,$$

$$T_A(\lambda X) = A(\lambda X) = \lambda(AX) := \lambda T_AX.$$

Using the Proposition 3.1.6 we can show the following.

PROPOSITION 3.2.2. *If $A, B \in M_{m,n}(\mathbb{R})$, and $a \in \mathbb{R}$, the following hold:*

(i) If $T_A = \mathbf{0}$, then $A = \mathbf{0}_{mn}$.

(ii) $T_{A+B} = T_A + T_B$.

(iii) $T_{a \cdot A} = aT_A$.

(iv) If $T_A = T_B$, then $A = B$.

(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 $T_{AC} = T_A \circ T_C$

$$\begin{tikzcd}
\mathbb{R}^l \arrow[r, "T_C"] \arrow[rr, bend right, "T_{AC}"'] & \mathbb{R}^n \arrow[r, "T_A"] & \mathbb{R}^m
\end{tikzcd}$$

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

PROOF. Exercise. $\square$

PROPOSITION 3.2.3. *If $A \in M_n(\mathbb{R})$ and $A^1, \ldots, A^n$ are the columns of A, then A is invertible if and only if the vectors $A^1, \ldots, A^n$ are linearly independent in $\mathbb{R}^n$.*

PROOF. Exercise. [Hint: use the fact that if $x := (x_1, \ldots, x_n) \in \mathbb{R}^n$, then $T_A(x) = x_1 A^1 + \ldots + x_n A^n$.] $\square$

It is not a surprise now that the matrix

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

is not invertible. Matrices can be related to the theory of linear equations.

PROPOSITION 3.2.4. *If $A := [a_{ij}] \in M_{m,n}(\mathbb{R})$ and $(x_1, \ldots, x_n) \in \mathbb{R}^n$, the following are equivalent.*
(i) $(x_1, \ldots, x_n)$ is a solution of the following system of linear equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= 0 \\ \vdots \qquad\qquad \vdots \quad &\;\; \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= 0. \end{aligned}$$

(ii) $(x_1, \ldots, x_n)$ is a solution of the following equation in $\mathbb{R}^m$

$$x_1 A^1 + \ldots + x_n A^n = \mathbf{0}_m.$$

(iii) $(x_1, \ldots, x_n)$ is in $\mathtt{Ker}(T_A)$.

PROOF. The proof is immediate from the above unfolding of T_A. $\square$

The system of inhomogeneous equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ \vdots \qquad\qquad \vdots \quad &\;\; \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

does not always have a solution. Take e.g., the system

$$3x + 5y - z = 1,$$

$$3x + 5y - z = 2.$$

If there is one solution though, then all solutions are obtained from that solution and the solutions of the corresponding homogeneous system.

Proposition 3.2.5. *Let $A := [a_{ij}] \in M_{m,n}(\mathbb{R})$, $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$, $\mathtt{Sol}$ the set of solutions of the system of linear equations*

$$Ax = b,$$

and $\mathtt{Sol}_0$ the set of solutions of the homogeneous system of linear equations

$$Ax = \mathbf{0}_m,$$

where in both systems $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. If $x_0 \in \mathtt{Sol}$, then

$$\mathtt{Sol} = x_0 + \mathtt{Sol}_0.$$

Proof. Exercise. □

Corollary 3.2.6. *If $A \in M_{m,n}(\mathbb{R})$, then*

$$\mathtt{Rank}(T_A) = \mathtt{cRank}(A) \quad \& \quad \dim(\mathtt{Ker}(T_A)) = n - \mathtt{cRank}(A).$$

Proof. Exercise. □

3.3. The matrix of a linear map

So far we defined 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 3.3.1. *Let $n, m \geq 1$. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, there is a unique matrix $A_T \in M_{m,n}(\mathbb{R})$ such that $T = T_{A_T}$ i.e., for every $x \in \mathbb{R}^n$*

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

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 $j \in \{1, \ldots, n\}$ we write $T(e_j)$ as a linear combination of the standard basis of $\mathbb{R}^m$ i.e.,

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

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

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

By the Proposition 2.4.5, 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. As we have shown after the Definition 3.2.1, $T_A(e_j)$ is the j-column A^j of A. Hence, $T_{A_T}(e_j)$ is the j-column of A_T, which is exactly $T(e_j)$ by the definition of A_T. The uniqueness of A_T follows from the Proposition 3.2.2(iv); if $B \in M_{m,n}(\mathbb{R})$ such that $T_{A_T}(x) := A_T x = T(x) = Bx := T_B(x)$, we get $A_T = B$. $\square$

If $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map with $T(0,1) := (a,c)$ and $T(1,0) := (b,d)$ then

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

Proposition 3.3.2. *Let $\mathcal{A} : L(\mathbb{R}^n, \mathbb{R}^m) \to M_{m,n}(\mathbb{R})$, defined by $T \mapsto A_T$.*
(i) If $\mathcal{T}$ is the function defined in the Proposition 3.2.2(vii), then we have that $\mathcal{A} \circ \mathcal{T} = \mathrm{id}_{M_{m,n}(\mathbb{R})}$ and $\mathcal{T} \circ \mathcal{A} = \mathrm{id}_{L(\mathbb{R}^n,\mathbb{R}^m)}$

$$\begin{tikzcd} M_{m,n}(\mathbb{R}) \arrow[r, "\mathcal{T}"] \arrow[rr, bend right, "\mathrm{id}_{M_{m,n}(\mathbb{R})}"'] & L(\mathbb{R}^n, \mathbb{R}^m) \arrow[r, "\mathcal{A}"] \arrow[rr, bend left, "\mathrm{id}_{L(\mathbb{R}^n,\mathbb{R}^m}"] & M_{m,n}(\mathbb{R}) \arrow[r, "\mathcal{T}"] & L(\mathbb{R}^n, \mathbb{R}^m). \end{tikzcd}$$

(ii) If $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $S \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^l)$, then $A_{S \circ T} = A_S A_T$.
(iii) $A_{\mathrm{id}_{\mathbb{R}^n}} = I_n$ and $A_{0_n} = \mathbf{0}_n$.
(iv) If $T_1, T_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, then $A_{T_1 + T_2} = A_{T_1} + A_{T_2}$.
(v) If $\lambda \in \mathbb{R}$, then $A_{\lambda T} = \lambda A_T$.
(vi) If $U \in \mathcal{L}(\mathbb{R}^n)$ is invertible, then A_U is invertible and $A_U^{-1} = A_{U^{-1}}$.

Proof. Exercise. $\square$

If X, Y are finite-dimensional linear spaces and $f \in \mathcal{L}(X, Y)$, we can associate to f a matrix in a canonical way. This matrix is going to be the matrix of the linear map f. If $X := \mathbb{R}^n$ and $Y := \mathbb{R}^m$, the matrix of f is reduced to the matrix determined by the Theorem 3.3.1. The Proposition 3.3.2 is also generalised.

Definition 3.3.3. If $B_X := \{v_1, \ldots, v_n\}$ a basis of X, we denote by e_{B_X} the isomorphism $X \simeq \mathbb{R}^n$ with respect to B_X i.e.,

$$e_{B_X} : X \to \mathbb{R}^n \qquad e_{B_X}(x) := (x_1, \ldots, x_n), \qquad x = \sum_{j=1}^n x_j v_j,$$

$$e_{B_X}^{-1} : \mathbb{R}^n \to X \qquad e_{B_X}^{-1}\big((x_1, \ldots, x_n)\big) := \sum_{j=1}^n x_j v_j, \qquad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

If $x \in X$, we use the following notation for the column of the coefficients of x with respect to the basis B_X of X:

$$x_{B_X} := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad x = \sum_{j=1}^n x_j v_j,$$

and hence we have that

$$e_{B_X}(x) = x_{B_X}.$$

We also denote by $B_n := \{e_1, \ldots, e_n\}$ the standard basis of $\mathbb{R}^n$.

THEOREM 3.3.4. *Let X, Y be linear spaces, $n, m \geq 1$, $B_X := \{v_1, \ldots, v_n\}$ a basis of X, $B_Y := \{w_1, \ldots, w_m\}$ a basis of Y, and let $f : X \to Y$ be linear.*
(i) There is a unique linear map $f_{B_X B_Y} : \mathbb{R}^n \to \mathbb{R}^m$ such that the following diagram commutes

$$\begin{tikzcd} X \arrow[r, "f"] & Y \arrow[d, "e_{B_Y}"] \\ \mathbb{R}^n \arrow[u, "e_{B_X}^{-1}"] \arrow[r, dashed, "f_{B_X B_Y}"'] & \mathbb{R}^m. \end{tikzcd}$$

(ii) If f is an injection (surjection), then $f_{B_X B_Y}$ is an injection (surjection).
(iii) The matrix $A_{f_{B_X B_Y}} \in M_{m,n}(\mathbb{R})$ of $f_{B_X B_Y}$ has as columns the coefficients of $f(v_j)$ with respect to B_Y i.e.,

$$A_{f_{B_X B_Y}} = \begin{bmatrix} f(v_1)_{B_Y} & \ldots & f(v_n)_{B_Y} \end{bmatrix},$$

and for every $x \in X$ we have that

$$f(x)_{B_Y} = A_{f_{B_X B_Y}} x_{B_X}.$$

(iv) The matrix $A_{f_{B_X B_Y}}$ is the unique matrix in $M_{m,n}(\mathbb{R})$ that satisfies the second equation of the case (iii) i.e., if $A \in M_{m,n}(\mathbb{R})$ such that $f(x)_{B_Y} = Ax_{B_X}$, for every $x \in X$, then $A = A_{f_{B_X B_Y}}$.
(v) If $X := \mathbb{R}^n$, $Y := \mathbb{R}^m$, $B_X := B_n$, and $B_Y := B_m$, then $f_{B_n B_m} = f$ and $A_{f_{B_n B_n}} = A_f$

PROOF. (i) We define the function

$$f_{B_X B_Y} := e_{B_Y} \circ f \circ e_{B_X}^{-1},$$

which is linear, as a composition of linear maps, and makes the above diagram commutative. The uniqueness of $f_{B_X B_Y}$ is immediate.
(ii) If f is an injection, then $f_{B_X B_Y}$ is an injection, as a composition of injections. If $z \in \mathbb{R}^m$, there is $y \in Y$ such that $e_{B_Y}(y) = z$, and if f is a surjection, there is $x \in X$ such that $f(x) = y$, hence $f_{B_X B_Y}(e_{B_X}(x)) = e_{B_Y}(f(x)) = z$.
(iii) If $(x_1, \ldots, x_n) \in \mathbb{R}^n$, by definition we have that

$$f_{B_X B_Y}\big((x_1, \ldots, x_n)\big) := e_{B_Y}\bigg(f\bigg(\sum_{j=1}^n x_j v_j\bigg)\bigg) = e_{B_Y}\bigg(\sum_{j=1}^n x_j f(v_j)\bigg).$$

If $e_j := (0, \ldots, 0, 1, 0, \ldots, 0) \in B_n$, then we get

$$f_{B_X B_Y}(e_j) = e_{B_Y}(f(v_j)) := f(v_j)_{B_Y},$$

hence by the proof of the Theorem 3.3.1 we get

$$\begin{aligned}
A_{f_{B_X B_Y}} &:= \begin{bmatrix} f_{B_X B_Y}(e_1)_{B_m} & \ldots & f_{B_X B_Y}(e_n)_{B_m} \end{bmatrix} \\
&= \begin{bmatrix} f_{B_X B_Y}(e_1) & \ldots & f_{B_X B_Y}(e_n) \end{bmatrix} \\
&= \begin{bmatrix} f(v_1)_{B_Y} & \ldots & f(v_n)_{B_Y} \end{bmatrix}.
\end{aligned}$$

If $X \ni x = \sum_{j=1}^{n} x_j v_j$, then by the previous equalities we get

$$\begin{aligned}
f(x)_{B_Y} &:= e_{B_Y}(f(x)) \\
&:= e_{B_Y}\left(f\left(\sum_{j=1}^{n} x_j v_j\right)\right) \\
&:= f_{B_X B_Y}\big((x_1, \ldots, x_n)\big) \\
&= A_{f_{B_X B_Y}} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&:= A_{f_{B_X B_Y}} x_{B_X}.
\end{aligned}$$

(iv) If $x := u_j$, for some $j \in \{1, \ldots, n\}$, then

$$f(u_j)_{B_Y} = A {v_j}_{B_X} = A \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{mj} \end{bmatrix} = A^j,$$

hence $A = \begin{bmatrix} A^1 & \ldots & A^n \end{bmatrix} = \begin{bmatrix} f(v_1)_{B_Y} & \ldots & f(v_n)_{B_Y} \end{bmatrix} = A_{f_{B_X B_Y}}$.
(v) In this case $e_{B_n} = \mathrm{id}_{\mathbb{R}^n}$, $e_{B_m} = \mathrm{id}_{\mathbb{R}^m}$, and $f_{B_n B_m} := \mathrm{id}_{\mathbb{R}^m} \circ f \circ \mathrm{id}_{\mathbb{R}^n} = f$

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "f"] & \mathbb{R}^m \arrow[d, "\mathrm{id}_{\mathbb{R}^n}"] \\
\mathbb{R}^n \arrow[u, "\mathrm{id}_{\mathbb{R}^n}"] \arrow[r, dashed, "f_{B_n B_m}"'] & \mathbb{R}^m.
\end{tikzcd}$$

By the definition of A_T in the proof of the Theorem 3.3.1 we have that $A_{f_{B_n B_m}} = \begin{bmatrix} f(e_1)_{B_m} & \ldots & f(e_n)_{B_m} \end{bmatrix} = A_f$. □

By the above uniqueness of $A_{f_{B_X B_Y}}$ *the* matrix of $f \in \mathcal{L}(X, Y)$ is defined.

Definition 3.3.5. If X, Y are linear spaces, $n, m \geq 1$, $B_X := \{v_1, \ldots, v_n\}$ is a basis of X, $B_Y := \{w_1, \ldots, w_m\}$ is a basis of Y, and $f : X \to Y$ is linear, *the* *matrix of f with respect to B_X and B_Y* is the matrix $A_{f_{B_X B_Y}} \in M_{m,n}(\mathbb{R})$, where $f_{B_X B_Y} := e_{B_Y} \circ f \circ e_{B_X}^{-1}$. If $T : X \to X$ is linear, we use the notation $T_{B_X} := T_{B_X B_X}$

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, leftarrowtail, twoheadleftarrow, "e_{B_X}^{-1}"'] & X \arrow[d, rightarrowtail, twoheadrightarrow, "e_{B_X}"] \\ \mathbb{R}^n \arrow[r, dashed, "T_{B_X}"'] & \mathbb{R}^n. \end{tikzcd}$$

Corollary 3.3.6. *Let X be a linear space, $n \geq 1$, $B_X := \{v_1, \ldots, v_n\}$ and $C_X := \{w_1, \ldots, w_n\}$ are bases of X.*

(i) If $x \in X$, then $x_{C_X} = A_{(\mathrm{id}_X)_{B_X C_X}} x_{B_X}$ and

$$A_{(\mathrm{id}_X)_{B_X C_X}} = \begin{bmatrix} (v_1)_{C_X} & \ldots & (v_n)_{C_X} \end{bmatrix}$$

$$\begin{tikzcd} X \arrow[r, "\mathrm{id}_X"] \arrow[d, leftarrowtail, twoheadleftarrow, "e_{B_X}^{-1}"'] & X \arrow[d, rightarrowtail, twoheadrightarrow, "e_{C_X}"] \\ \mathbb{R}^n \arrow[r, dashed, "(\mathrm{id}_X)_{B_X C_X}"'] & \mathbb{R}^n. \end{tikzcd}$$

(ii) $(\mathrm{id}_X)_{B_X} = \mathrm{id}_{\mathbb{R}^n}$ *and* $A_{(\mathrm{id}_X)_{B_X}} = I_n$

$$\begin{tikzcd} X \arrow[r, "\mathrm{id}_X"] \arrow[d, leftarrowtail, twoheadleftarrow, "e_{B_X}^{-1}"'] & X \arrow[d, rightarrowtail, twoheadrightarrow, "e_{B_X}"] \\ \mathbb{R}^n \arrow[r, dashed, "(\mathrm{id}_X)_{B_X}"'] & \mathbb{R}^n. \end{tikzcd}$$

Proof. Exercise. □

If $\dim(X) = n$ and $\dim(Y) = m$, then by the Theorem 2.4.17 we have that $\dim(\mathcal{L}(X, Y)) = mn$ and by the Remark 3.1.2 we have that $\dim(M_{m,n}(\mathbb{R})) = mn$. Hence, by the Corollary 2.4.15 we get $\mathcal{L}(X, Y) \simeq \mathbb{R}^{mn} \simeq M_{mn}(\mathbb{R})$, hence $\mathcal{L}(X, Y) \simeq M_{m,n}(\mathbb{R})$. Next we provide a concrete isomorphism between these linear spaces.

Proposition 3.3.7. *Let X, Y be linear spaces, $n, m \geq 1$, $B_X := \{v_1, \ldots, v_n\}$ a basis of X, $B_Y := \{w_1, \ldots, w_m\}$ a basis of Y, and let $f, g : X \to Y$ be linear maps.*

(i) $A_{(f+g)_{B_X B_Y}} = A_{f_{B_X B_Y}} + A_{g_{B_X B_Y}}$.

(ii) If $\lambda \in \mathbb{R}$, then $A_{(\lambda f)_{B_X B_Y}} = \lambda A_{f_{B_X B_Y}}$.

(iii) The function $e_{B_X B_Y} : \mathcal{L}(X, Y) \to M_{m,n}(\mathbb{R})$, defined by $f \mapsto A_{f_{B_X B_Y}}$, for every $f \in \mathcal{L}(X, Y)$, is a linear isomorphism.

Proof. Exercise. □

The function $e_{B_X B_Y}$ also preserves the multiplication i.e., it sends the composition of linear maps to the product of the image-matrices.

Proposition 3.3.8. *Let X, Y and Z be linear spaces, $n, m, l \geq 1$, $B_X := \{v_1, \ldots, v_n\}$ a basis of X, $B_Y := \{w_1, \ldots, w_m\}$ a basis of Y, $B_Z := \{u_1, \ldots, u_l\}$ a basis of Z, and let $f : X \to Y$ and $g : Y \to Z$ be linear maps.*

(i) $(g \circ f)_{B_X B_Z} = g_{B_Y B_Z} \circ f_{B_X B_Y}$ *i.e., the following lower outer diagram commutes*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend left, "g \circ f"] & Y \arrow[r, "g"] \arrow[d, bend right, "e_{B_Y}"'] & Z \arrow[d, "e_{B_Z}"] \\ \mathbb{R}^n \arrow[u, "e_{B_X}^{-1}"] \arrow[r, dashed, "f_{B_X B_Y}"'] \arrow[rr, dashed, bend right, "(g \circ f)_{B_X B_Z}"'] & \mathbb{R}^m \arrow[u, bend right, "e_{B_Y}^{-1}"'] \arrow[r, dashed, "g_{B_Y B_Z}"'] & \mathbb{R}^l. \end{tikzcd}$$

(ii) $A_{(g \circ f)_{B_X B_Z}} = A_{g_{B_Y B_Z}} A_{f_{B_X B_Y}}$.

Proof. (i) Using the commutativity of the above inner diagrams we get

$$\begin{aligned} (g \circ f)_{B_X B_Z} &:= e_{B_Z} \circ (g \circ f) \circ e_{B_X}^{-1} \\ &= e_{B_Z} \circ [g \circ \mathrm{id}_Y \circ f] \circ e_{B_X}^{-1} \\ &= e_{B_Z} \circ [g \circ (e_{B_Y}^{-1} \circ e_{B_Y}) \circ f] \circ e_{B_X}^{-1} \\ &= (e_{B_Z} \circ g \circ e_{B_Y}^{-1}) \circ (e_{B_Y} \circ f \circ e_{B_X}^{-1}) \\ &= g_{B_Y B_Z} \circ f_{B_X B_Y}. \end{aligned}$$

We have that $A_{g_{B_Y B_Z}} \in M_{l,m}(\mathbb{R})$ and $A_{f_{B_X B_Y}} \in M_{m,n}(\mathbb{R})$, therefore their multiplication is well-defined and belongs in $M_{l,n}(\mathbb{R})$. Moreover, $A_{(g \circ f)_{B_X B_Z}} \in M_{l,n}(\mathbb{R})$. By the Theorem 3.3.4(iii) we have that

$$A_{(g \circ f)_{B_X B_Z}} x_{B_X} := [g(f(x)]_{B_Z} = A_{g_{B_Y B_Z}} f(x)_{B_Y} = A_{g_{B_Y B_Z}} A_{f_{B_X B_Y}} x_{B_X},$$

and by the uniqueness in the Theorem 3.3.4(iv) $A_{(g \circ f)_{B_X B_Z}} = A_{g_{B_Y B_Z}} A_{f_{B_X B_Y}}$. □

Corollary 3.3.9. *Let $n \geq 1$, let X be an n-dimensional linear space, and let B_X and C_X be bases of X.*

(i) $(\mathrm{id}_X)_{B_X} = (\mathrm{id}_X)_{C_X B_X} \circ (\mathrm{id}_X)_{B_X C_X}$

$$
\begin{tikzcd}
X \arrow[r, "\mathrm{id}_X"] \arrow[rr, bend left, "\mathrm{id}_X"] & X \arrow[r, "\mathrm{id}_X"] \arrow[d, bend right, "e_{C_X}"'] & X \arrow[d, "e_{B_X}"] \\
\mathbb{R}^n \arrow[u, "e_{B_X}^{-1}"] \arrow[r, dashed, "(\mathrm{id}_X)_{B_X C_X}"] \arrow[rr, dashed, bend right, "(\mathrm{id}_X)_{B_X}"'] & \mathbb{R}^n \arrow[r, dashed, "(\mathrm{id}_X)_{C_X B_X}"] \arrow[u, bend right, "e_{C_X}^{-1}"'] & \mathbb{R}^n.
\end{tikzcd}
$$

(ii) $A_{(\mathrm{id}_X)_{B_X}} = A_{(\mathrm{id}_X)_{C_X B_X}} A_{(\mathrm{id}_X)_{B_X C_X}}$*, and the matrices* $A_{(\mathrm{id}_X)_{C_X B_X}}$, $A_{(\mathrm{id}_X)_{B_X C_X}}$ *are inverse to each other.*

Proof. (i) It follows immediately from the Proposition 3.3.7(i).
(ii) The required equality follows immediately from the Proposition 3.3.7(ii). Interchanging the bases B_X and C_X we get $A_{(\mathrm{id}_X)_{C_X}} = A_{(\mathrm{id}_X)_{B_X C_X}} A_{(\mathrm{id}_X)_{C_X B_X}}$. Since by the Corollary 3.3.6(ii) we have that $A_{(\mathrm{id}_X)_{B_X}} = I_n = A_{(\mathrm{id}_X)_{C_X}}$, the matrices $A_{(\mathrm{id}_X)_{C_X B_X}}$ and $A_{(\mathrm{id}_X)_{B_X C_X}}$ are inverse to each other. $\square$

Corollary 3.3.10. *Let X, Y, Z and W be linear spaces, $n, m, l, s \geq 1$, $B_X := \{v_1, \ldots, v_n\}$ a basis of X, $B_Y := \{w_1, \ldots, w_m\}$ a basis of Y, $B_Z := \{u_1, \ldots, u_l\}$ a basis of Z, $B_W := \{\rho_1, \ldots, \rho_s\}$ a basis of W, and let $f : X \to Y$, $g : Y \to Z$, and $h : Z \to W$ be linear maps.*
(i) $(h \circ g \circ f)_{B_X B_W} = h_{B_Z B_W} \circ g_{B_Y B_Z} \circ f_{B_X B_Y}$

$$
\begin{tikzcd}
X \arrow[r, "f"] \arrow[rrr, bend left, "h \circ g \circ f"] & Y \arrow[r, "g"] \arrow[d, bend right, "e_{B_Y}"'] & Z \arrow[r, "h"] \arrow[d, bend right, "e_{B_Z}"'] & W \arrow[d, "e_{B_W}"] \\
\mathbb{R}^n \arrow[u, "e_{B_X}^{-1}"] \arrow[r, dashed, "f_{B_X B_Y}"] \arrow[rrr, dashed, bend right, "(h \circ g \circ f)_{B_X B_W}"'] & \mathbb{R}^m \arrow[r, dashed, "g_{B_Y B_Z}"'] \arrow[u, bend right, "e_{B_Y}^{-1}"'] & \mathbb{R}^l \arrow[r, dashed, "h_{B_Z B_W}"] \arrow[u, bend right, "e_{B_Z}^{-1}"'] & \mathbb{R}^s.
\end{tikzcd}
$$

(ii) $A_{(h \circ g \circ f)_{B_X B_W}} = A_{h_{B_Z B_W}} A_{g_{B_Y B_Z}} A_{f_{B_X B_Y}}$.

Proof. Exercise. $\square$

Theorem 3.3.11. *If $n \geq 1$, X is an n-dimensional linear space, B_X and C_X are bases of X, and $T : X \to X$ is a linear map, then there exists an invertible* *matrix* $B \in M_n(\mathbb{R})$ *such that*

$$A_{T_{C_X}} = B^{-1} A_{T_{B_X}} B.$$

Proof. By the Corollary 3.3.10 we have that

$$\begin{tikzcd}
X \arrow[r, "\mathrm{id}_X"] \arrow[rrr, bend left, "T"] & X \arrow[r, "T"] \arrow[d, bend right, "e_{B_X}"'] & X \arrow[r, "\mathrm{id}_X"] \arrow[d, bend right, "e_{B_X}"'] & X \arrow[d, "e_{C_X}"] \\
\mathbb{R}^n \arrow[u, "e_{C_X}^{-1}"] \arrow[r, dashed, "(\mathrm{id}_X)_{C_X B_X}"] \arrow[rrr, dashed, bend right, "T_{C_X}"'] & \mathbb{R}^n \arrow[r, dashed, "T_{B_X}"'] \arrow[u, bend right, "e_{B_X}^{-1}"'] & \mathbb{R}^n \arrow[r, dashed, "(\mathrm{id}_X)_{B_X C_X}"] \arrow[u, bend right, "e_{B_X}^{-1}"'] & \mathbb{R}^n
\end{tikzcd}$$

$T_{C_X} = (\mathrm{id}_X)_{B_X C_X} \circ T_{B_X} \circ (\mathrm{id}_X)_{C_X B_X}$ and $A_{T_{C_X}} = A_{(\mathrm{id}_X)_{B_X C_X}} A_{T_{B_X}} A_{(\mathrm{id}_X)_{C_X B_X}}$. By the Corollary 3.3.9(ii) $A_{(\mathrm{id}_X)_{C_X B_X}}$ and $A_{(\mathrm{id}_X)_{B_X C_X}}$ are inverse to each other. $\square$

Definition 3.3.12. If $n \geq 1$, X is an n-dimensional linear space, and $T : X \to X$ is a linear map, we say that a basis $B_X := \{v_1, \ldots, v_n\}$ of X *diagonalises* T, if $A_{T_{B_X}} = \big[T(v_1)_{B_X} \ldots T(v_n)_{B_X}\big] \in M_n(\mathbb{R})$ is a diagonal matrix. We say that T is *diagonalisable*, if there exists some basis C_X of X such that C_X diagonalises T. A matrix $A \in M_n(\mathbb{R})$ is called a *diagonalisable matrix*, if the linear map T_A of A is diagonalisable.

Theorem 3.3.13. *If* $n \geq 1$, X *is an* n*-dimensional linear space,* B_X *is a basis of* X, *and* $T : X \to X$ *is a linear map, then* T *is diagonalisable if and only if there exists some invertible matrix* $B \in M_n(\mathbb{R})$ *such that* $B^{-1} A_{T_{B_X}} B$ *is a diagonal matrix.*

Proof. Exercise. $\square$

Definition 3.3.14. If $A, A' \in M_n(\mathbb{R})$, we say that A and A' are *similar*, in symbols $A \sim A'$, if there is some invertible matrix $B \in M_n(\mathbb{R})$ such that $A' = B^{-1}AB$.

Remark 3.3.15. The relation $A \sim A'$ is an equivalence relation on $M_n(\mathbb{R})$.

Proof. Exercise. $\square$

CHAPTER 4

Inner product spaces

4.1. The Euclidean inner product space and the Minkowski spacetime

Definition 4.1.1. Let X be a linear space. A *bilinear map* on X is a function $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ such that for every $x, y, z \in X$ and $\lambda \in \mathbb{R}$ the following hold:
(i) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ (left additivity).
(ii) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ (right additivity).
(iii) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ (left homogeneity).
(iv) $\langle x, \lambda y \rangle = \lambda \langle x, y \rangle$ (right homogeneity).
An *inner product* on X is a symmetric bilinear map i.e., for every $x, y \in X$
(v) $\langle x, y \rangle = \langle y, x \rangle$ (symmetry),
and the pair $(X, \langle \cdot, \cdot \rangle)$ is called an *inner product space*. We call an inner product
(a) *non-degenerate*, if

$$\forall_{x \in X} \bigg(\forall_{z \in X} \big(\langle x, z \rangle = 0 \big) \Rightarrow x = \mathbf{0} \bigg),$$

(b) *positive*, if $\forall_{x \in X} \big(\langle x, x \rangle \geq 0 \big)$,
(c) *negative*, if $\forall_{x \in X} \big(\langle x, x \rangle \leq 0 \big)$,
(c) *positive definite*, if

$$\forall_{x \in X} \big(x \neq \mathbf{0} \Rightarrow \langle x, x \rangle > 0 \big),$$

(d) *negative definite*, if

$$\forall_{x \in X} \big(x \neq \mathbf{0} \Rightarrow \langle x, x \rangle < 0 \big),$$

(e) *indefinite*, if it is neither positive definite nor negative definite.
If $\langle \cdot, \cdot \rangle$ is an inner product on X the *quadratic form* $Q_{\langle \cdot, \cdot \rangle}$ associated with $\langle \cdot, \cdot \rangle$ is the function $Q_{\langle \cdot, \cdot \rangle} : X \to \mathbb{R}$, defined by

$$Q_{\langle \cdot, \cdot \rangle}(x) := \langle x, x \rangle,$$

for every $x \in X$. If the inner product $\langle \cdot, \cdot \rangle$ is clear from the context, we write $Q(x)$ instead of $Q_{\langle \cdot, \cdot \rangle}(x)$. An element x of X is called
(α) *lightlike*, or *null*, if $Q(x) := \langle x, x \rangle = 0$,

(β) *timelike*, if $Q(x) := \langle x, x\rangle < 0$,

(γ) *spacelike*, if $Q(x) := \langle x, x\rangle > 0$,

(δ) *unit*, if $Q(x) := \langle x, x\rangle \in \{1, -1\}$.

E.g., if $C([0,1])$ is the linear space of all continuous, real-valued functions on the compact interval $[0,1]$, the function

$$(f, g) \mapsto \langle f, g\rangle := \int_0^1 f(t)g(t)dt,$$

for every $f, g \in C([0,1])$ is a non-degenerate, positive, inner product on $C([0,1])$. As we have already seen in the Remark 3.1.9, the function

$$(A, B) \mapsto \mathtt{Tr}(AB)$$

is a non degenerate, indefinite inner product on $M_n(\mathbb{R})$, and a positive definite inner product on $\mathtt{Sym}_n(\mathbb{R})$.

Remark 4.1.2. Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space.

(i) Left additivity and left homogeneity imply right additivity and right homogeneity, and vice versa.

(ii) If $x \in X$, then $\langle \mathbf{0}, x\rangle = 0$.

(iii) If $\langle\cdot,\cdot\rangle$ is positive definite, then $\langle\cdot,\cdot\rangle$ is positive, and non-degenerate.

(iv) If $\langle\cdot,\cdot\rangle$ is positive, then $-\langle\cdot,\cdot\rangle$, where $-\langle\cdot,\cdot\rangle(x,y) := -\langle x, y\rangle$, for every $x, y \in X$, is a negative inner product on X.

(v) For every $x, y \in X$ we have that

$$\langle x, y\rangle = \frac{1}{4}\big(\langle x+y, x+y\rangle - \langle x-y, x-y\rangle\big).$$

(vi) If $\langle\langle\cdot,\cdot\rangle\rangle$ is an inner product on X such that $Q_{\langle\cdot,\cdot\rangle}(x) = Q_{\langle\langle\cdot,\cdot\rangle\rangle}(x)$, for every $x \in X$, then $\langle\langle x, y\rangle\rangle = \langle x, y\rangle$, for every $x, y \in X$.

Proof. Exercise. $\square$

Definition 4.1.3. 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

$$\langle x, y\rangle := \sum_{i=1}^n x_i y_i := x_1y_1 + \ldots + x_ny_n.$$

If $n = 1$, the Euclidean inner product on $\mathbb{R}$ is the product on $\mathbb{R}$. The quadratic form of the Euclidean inner product on $\mathbb{R}^n$ is given by

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

The Euclidean inner product is a positive definite inner product on $\mathbb{R}^n$, with $\mathbf{0}$ as the only lightlike vector, and with no timelike vectors.

DEFINITION 4.1.4. If $n \geq 2$, $x = (x_1, \ldots, x_{n-1}, s)$ and $y = (y_1, \ldots, y_{n-1}, t)$ are in $\mathbb{R}^n$, their *Minkowski inner product*, or their *Lorentz inner product*, is defined by

$$\langle x, y \rangle := \sum_{i=1}^{n-1} x_i y_i - st = x_1 y_1 + \ldots + x_{n-1} y_{n-1} - st.$$

The pair $\mathcal{M}_4 := (\mathbb{R}^4, \langle \cdot, \cdot \rangle)$, where $\langle \cdot, \cdot \rangle$ is the Minkowski inner product on $\mathbb{R}^4$, is called the *Minkowski spacetime*. We call the pair $\mathcal{M}_3 := (\mathbb{R}^3, \langle \cdot, \cdot \rangle)$, where $\langle \cdot, \cdot \rangle$ is the Minkowski inner product on $\mathbb{R}^3$, the *Minkowski planetime*, and the pair $\mathcal{M}_2 := (\mathbb{R}^2, \langle \cdot, \cdot \rangle)$, where $\langle \cdot, \cdot \rangle$ is the Minkowski inner product on $\mathbb{R}^2$, the *Minkowski linetime*.

The quadratic forms of the Minkowski inner product on $\mathbb{R}^n$ and $\mathbb{R}^4$ are

$$\begin{aligned} Q\big(x_1, \ldots, x_{n-1}, s)\big) &:= \langle (x_1, \ldots, x_{n-1}, s), (x_1, \ldots, x_{n-1}, s) \rangle \\ &:= \sum_{i=1}^{n-1} x_i^2 - s^2 = x_1^2 + \ldots + x_{n-1}^2 - s^2, \end{aligned}$$

$$Q\big((x_1, x_2, x_3, s)\big) := \langle (x_1, x_2, x_3, s), (x_1, x_2, x_3, s) \rangle := \sum_{i=1}^{3} x_i^2 - s^2 = x_1^2 + x_2^2 + x_3^2 - s^2.$$

The Minkowski spacetime is regarded in [**18**], p. 1, as "the appropriate arena within which to formulate those laws of physics that do not refer specifically to gravitational phenomena"[1]. A non-zero vector x in $\mathbb{R}^n$, equipped with the Minkowski inner product, where $x := (x_1, \ldots, x_{n-1}, 0)$ is spacelike,

$$Q\big((x_1, \ldots, x_{n-1}, 0)\big) := x_1^2 + \ldots + x_{n-1}^2 - 0 > 0,$$

while if $x := (0, \ldots, 0, s)$, then x is timelike, since

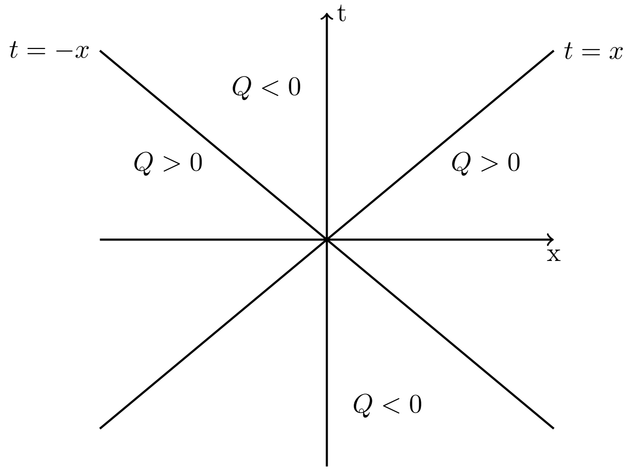
$$Q\big((0, \ldots, 0, s)\big) := 0^2 + \ldots + 0^2 - s^2 < 0.$$

Moreover, there are non-zero vectors in $\mathbb{R}^4$ that are lightlike e.g.,

$$Q\big((1, 1, 0, \sqrt{2})\big) := 1^2 + 1^2 + 0^2 - \sqrt{2}^2 = 0.$$

If we consider the Minkowski linetime, the lightlike vectors form the lines $t = x$ and $t = -x$, the timelike vectors are within the cone at the origin $(0, 0)$ formed by these two lines, and the spacelike vectors are outside the cone.

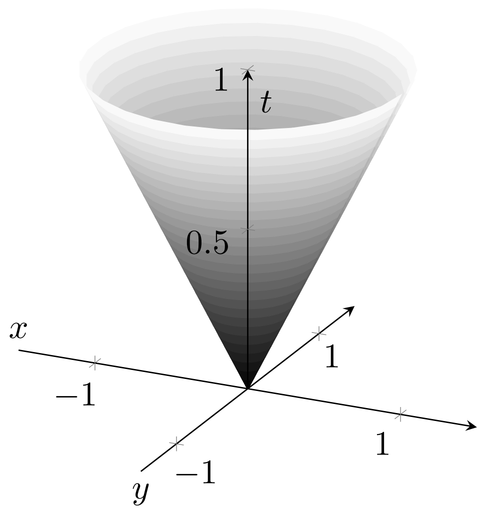
[1]See [**18**] for an elaborated study of the mathematical properties of the Minkowski spacetime.



If we consider the Minkowski planetime, then one can show that its lightlike vectors $(x, y, t) \in \mathbb{R}^3$ form a cone at the origin $(0, 0, 0)$ with equation

$$x^2 + y^2 - t^2 = 0,$$

its timelike vectors are inside this cone, and its spacelike vectors are outside this cone. Next we draw the upper lightcone at the origin $(0, 0, 0)$.



Proposition 4.1.5. *The Minkowski inner product is an indefinite and non-degenerate inner product on* $\mathbb{R}^n$.

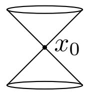
Proof. Exercise. $\square$

DEFINITION 4.1.6. A subset C of a linear space X is called a *cone*, if

$$\forall_{x \in C} \forall_{\lambda > 0} \big(\lambda x \in C\big).$$

If $\langle, \cdot, \cdot \rangle$ is an inner product on X, and $x_0 \in X$, the *light cone* $C(x_0)$ at x_0 is the set

$$C(x_0) := \{x \in X \mid Q(x - x_0) = 0\}.$$



The light cone $C(\mathbf{0})$ of all lightlike vectors in X is a cone in X; if $Q(x) = 0$, for some $x \in X$, then $Q(\lambda x) = \lambda^2 Q(x) = 0$. In the case of the Euclidean inner product $C(\mathbf{0}) = \{\mathbf{0}\}$, and the set of all timelike vectors "inside" $C(\mathbf{0})$ is empty. Generally, the light cone of a non-zero vector is not a cone in X; e.g., if $a \in \mathbb{R}$ and $a \neq 0$, then $C(a) = \{b \in \mathbb{R} \mid (b - a)^2 = 0\} = \{a\}$, which is not a cone in $\mathbb{R}$.

DEFINITION 4.1.7. If X is a linear 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*.

Notice that

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

If $x = \mathbf{0}$, then $||\mathbf{0}|| = 0$, 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 4.1.8. 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.

PROPOSITION 4.1.9 (Inequality of Cauchy). *If $x, y \in \mathbb{R}^n$, then*

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

The equality holds if and only if x, y are linearly dependent.

PROOF. By definition we need to show

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

which is equivalent to

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

This we get as follows:

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

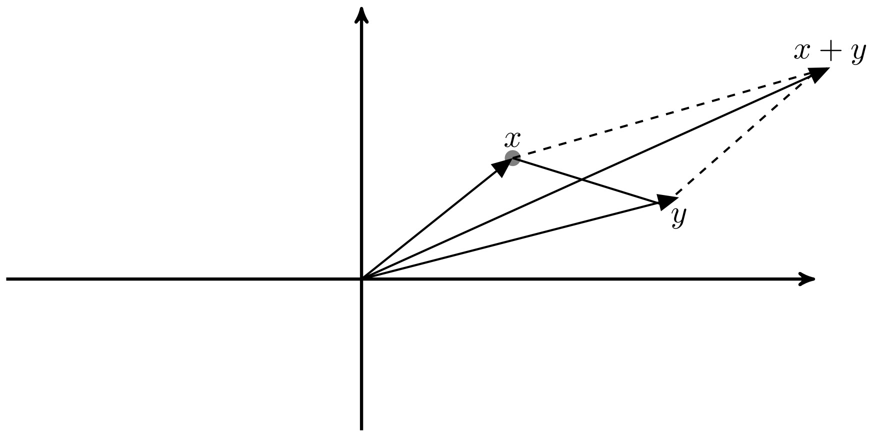
If $y = \mathbf{0}$, then $0x + 1y = \mathbf{0}$, hence $x, \mathbf{0}$ are linearly dependent, and the equality trivially holds. If $y_j \neq 0$, for some $j \in \{1, \ldots, n\}$, then $B = A \Leftrightarrow x = \lambda y$, where $\lambda := \frac{x_j}{y_j}$. Clearly, $x_j = \lambda y_j$. If $i \neq j$, then $x_i = \lambda y_i \Leftrightarrow x_i - \frac{x_j}{y_j} y_i = 0 \Leftrightarrow x_i y_j - x_j y_i = 0$. The last condition, for every $i, j \in \{1, \ldots, n\}$ such that $i \neq j$ is equivalent to $B - A = 0 \Leftrightarrow B = A$. $\square$

The inequality of Cauchy is necessary to the proof that $(\mathbb{R}^n, |.|)$ is a normed space; it is used in the proof of the triangle inequality for the Euclidean norm, while the rest of the proof is easy. 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[2] that

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

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

[2]The following figure also explains why $|x + y| \le |x| + |y|$.



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

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

PROOF. Since $x \perp y \Leftrightarrow \langle x, y\rangle = 0$, we have that

$$|x+y|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle = |x|^2 + |y|^2.$$

□

By the inequality of Cauchy we have that if $x, y \neq \mathbf{0}$, then

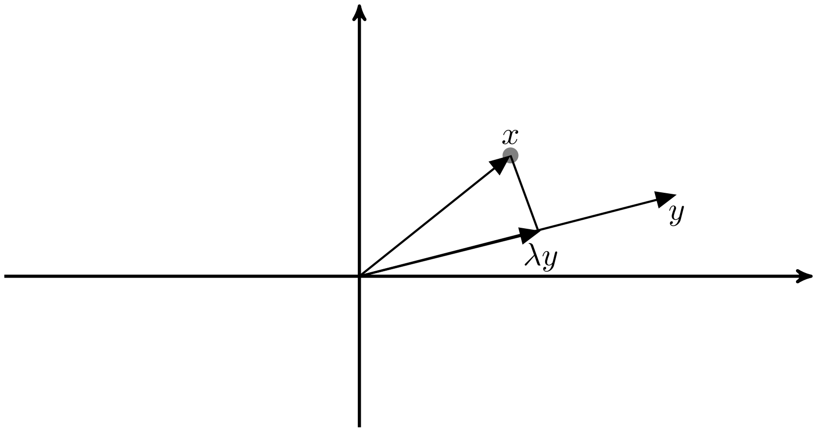
$$\left|\frac{|\langle x, y\rangle|}{|x||y|}\right| = \frac{|\langle x, y\rangle|}{|x||y|} \leq 1 \Leftrightarrow -1 \leq \frac{\langle x, y\rangle}{|x||y|} \leq 1.$$

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



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

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

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

$\square$

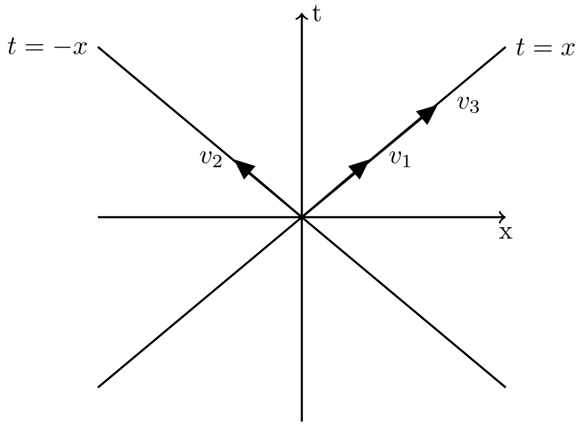
Because of the above characterisation of orthogonality in $\mathbb{R}^n$ through the Euclidean inner product, we give the following definition.

DEFINITION 4.1.12. Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be an inner product space. If $x, y \in X$, then x, y are $\langle\langle \cdot, \cdot \rangle\rangle$-*orthogonal*, or simpler, *orthogonal*, if $\langle\langle x, y\rangle\rangle = 0$, and we also write $x \perp y$. If $Y \subseteq X$, its *orthogonal complement* $Y^{\perp}$ is defined by

$$Y^{\perp} := \big\{x \in X \mid \forall_{y \in Y}\big(\langle\langle x, y\rangle\rangle = 0\big)\big\}.$$

If $x \in Y^{\perp}$, we also write $x \perp Y$. It is immediate to show that $Y^{\perp}$ is a linear subspace of X, which is called the *orthogonal space of* Y.

A lightlike vector v_1 in the Minkowski timeline that lies on the line $t = x$ is orthogonal to any lightlike vector v_2 that lies on the line $t = -x$ with respect to the Euclidean inner product. These vectors though, are not orthogonal with respect to the Minkowski product on $\mathbb{R}^2$. Actually, v_1 is orthogonal to any vector v_3 parallel to it, and the same for v_2. I.e., for lightlike vectors orthogonality means parallelism!



PROPOSITION 4.1.13. *Let $x, y \in \mathbb{R}^4$ such that x and y are lightlike with respect to the Minkowski product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^4$. Then x and y are $\langle \cdot, \cdot \rangle$-orthogonal if and only if x, y are linearly dependent.*

PROOF. Exercise. $\square$

It is not a coincidence that the inequality of Cauchy is not satisfied by the Minkowski inner product, as we can find (Exercise) timelike vectors u, w in the Minkowski spacetime satisfying

$$\langle u, w \rangle^2 > \langle u, u \rangle \langle w, w \rangle.$$

DEFINITION 4.1.14. Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be an inner product space. A basis $B := \{v_1, \ldots, v_n\}$ of X is called *orthogonal*, if

$$\forall_{i,j \in \{1,\ldots,n\}} \big(i \neq j \Rightarrow \langle\langle v_i, v_j \rangle\rangle = 0\big),$$

and it is called *orthonormal*, if it is orthogonal and $Q(v_i) \in \{-1, 0, 1\}$, for every $i \in \{1, \ldots, n\}$.

Clearly, the standard basis B_n of $\mathbb{R}^n$ is orthonormal.

DEFINITION 4.1.15. If $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ is an inner product space, a linear map $T : X \to X$ is called an *orthogonal transformation*, or an *inner product preserving* linear map, if for every $x, y \in X$ it satisfies

$$\langle\langle T(x), T(y) \rangle\rangle = \langle\langle x, y \rangle\rangle.$$

PROPOSITION 4.1.16. *Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a finite-dimensional, non-degenerate, inner product space and let $T : X \to X$ be an orthogonal transformation.*

(i) T preserves orthogonality i.e., $x \perp y \Leftrightarrow T(x) \perp T(y)$, for every $x, y \in X$.

(ii) T is an isomorphism.

(iii) For every $x \in X$, we have that $Q(T(x)) = Q(x)$.

(iv) If $S : X \to X$ is a linear map such that $Q(S(x)) = Q(x)$, for every $x \in X$, then S is an orthogonal transformation.

(v) If $B_X := \{v_1, \ldots, v_n\}$ is an orthonormal basis of X, then its image $T(B_X) := \{T(v_1), \ldots, T(v_n)\}$ under T is an orthonormal basis of X.

(vi) Suppose that X has an orthonormal basis, and let $S : X \to X$ be a linear map such that for every orthonormal basis $B_X := \{v_1, \ldots, v_n\}$ of X its image $S(B_X)$ under S is an orthonormal basis of X, and for every $i \in \{1, \ldots, n\}$, we have that $Q(S(v_i)) = Q(v_i)$. Then S is an orthogonal transformation.

Proof. Exercise. $\square$

The existence of orthogonal and orthonormal bases of an inner product space is studied next.

4.2. Existence of an orthogonal basis, the positive definite case

Throughout this section $\langle\langle \cdot, \cdot \rangle\rangle$ is a positive definite inner product on a linear space X. The properties of $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ generalise the properties of $\mathbb{R}^n$ equipped with the Euclidean norm, if we use the following definitions, motivated by their geometric interpretation in the case of $\mathbb{R}^n$. Namely, we define

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

$$\mathtt{pr}_y(x) := \lambda_y(x) y, \qquad \lambda_y(x) := \frac{\langle\langle x, y \rangle\rangle}{\langle\langle y, y \rangle\rangle},$$

where $y \neq \mathbf{0}$. Since $\langle\langle x, x \rangle\rangle \geq 0$, the *norm induced by the positive definite inner product* $\langle\langle \cdot, \cdot \rangle\rangle$ on X is given by

$$||x|| := \sqrt{\langle\langle x, x \rangle\rangle},$$

for every $x \in X$. We need to show the generalisation of the inequality of Cauchy, in order to prove that $||x||$ is a norm on X.

Proposition 4.2.1. *Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a positive definite, inner product space, $x, y, x_1, \ldots, x_n \in X$, and let $|| \cdot ||$ be the norm on X induced by $\langle\langle \cdot, \cdot \rangle\rangle$.*

(i) (Pythagoras) If $x \perp y$, then $||x + y||^2 = ||x||^2 + ||y||^2$.

(ii) (Generalised Pythagoras) If $x_i \perp x_j$, for every $i, j \in \{1, \ldots, n\}$ such that $i \neq j$, then

$$||x_1 + \ldots + x_n||^2 = ||x_1||^2 + \ldots + ||x_n||^2.$$

(iii) (Parallelogram law) $||x + y||^2 + ||x - y||^2 = 2\big(||x||^2 + ||y||^2\big)$.

(iv) (Polarisation identity) $\langle\langle x, y \rangle\rangle := \frac{1}{4}\big(||x + y||^2 - ||x - y||^2\big)$.

(v) If $\lambda \in \mathbb{R}$, then $||\lambda x|| = \lambda ||x||$.

(vi) $||x|| = 0 \Leftrightarrow x = \mathbf{0}$.

(vii) If $x \neq \mathbf{0}$, then
$$\left|\left|\frac{x}{||x||}\right|\right| = 1.$$
(viii) (Inequality Cauchy-Schwarz) $|\langle\langle x, y\rangle\rangle| \leq ||x||\ ||y||$, and the equality holds if and only if x, y are linearly dependent.

(ix) (Triangle inequality) $||x + y|| \leq ||x|| + ||y||$. The equality holds if and only if either $x = \mathbf{0} \vee y = \mathbf{0}$ or there is some $\lambda > 0$ such that $x = \lambda y$.

(x) $||\cdot||$ is a norm on X.

(xi) $\big|||x|| - ||y||\big| \leq ||x - y||$.

(xii) $||x|| - ||y|| \leq \big|||x|| - ||y||\big| \leq ||x + y||$.

Proof. We show only the case (viii), and the rest is an exercise. If $y = \mathbf{0}$, then both terms of the inequality are equal to 0. If $y \neq \mathbf{0}$ and $||y|| = 1$, then, $(x - \lambda_y(x))\perp y$, hence
$$(x - \lambda_y(x)y)\perp\lambda_y(x)y.$$
By the theorem of Pythagoras we have that
$$\begin{aligned}||x||^2 &= ||(x - \lambda_y(x)y) + \lambda_y(x)y||^2\\ &= ||x - \lambda_y(x)y||^2 + ||\lambda_y(x)y||^2\\ &= ||x - \lambda_y(x)y||^2 + \big[|\lambda_y(x)|||y||\big]^2\\ &= ||x - \lambda_y(x)y||^2 + \lambda_y(x)^2,\end{aligned}$$
hence $\lambda_y(x)^2 \leq ||x||^2$, and consequently
$$|\lambda_y(x)| := \left|\frac{\langle\langle x, y\rangle\rangle}{\langle\langle y, y\rangle\rangle}\right| = |\langle\langle x, y\rangle\rangle| \leq ||x|| = ||x||1 = ||x||||y||.$$
If $y \neq \mathbf{0}$, then by the case (v) and the previous fact we have that
$$\left|\langle\langle x, \frac{y}{||y||}\rangle\rangle\right| \leq ||x|| \Leftrightarrow |\langle\langle x, y\rangle\rangle| \leq ||x||\ ||y||.$$
If x, y are linearly dependent, then the equality holds; if $y = \mathbf{0}$, it is trivial, and if $y \neq \mathbf{0}$, then $x = \lambda y$, for some $\lambda \in \mathbb{R}$, and
$$|\langle\langle\lambda y, y\rangle\rangle| = |\lambda||\langle\langle y, y\rangle\rangle| = |\lambda|||y||^2 = |\lambda|||y||||y|| = ||x||\ ||y||.$$
If the equality holds and one of x, y is $\mathbf{0}$, then x, y are linearly dependent. Suppose next that the equality holds and both $x \neq \mathbf{0}$ and $y \neq \mathbf{0}$. As it is expected geometrically, we have that
$$\begin{aligned}||x - \lambda_y(x)y||^2 &= \langle\langle x - \lambda_y(x)y, x - \lambda_y(x)y\rangle\rangle\\ &= \langle\langle x, x\rangle\rangle - 2\lambda_y(x)\langle\langle x, y\rangle\rangle + \lambda_y(x)^2\langle\langle y, y\rangle\rangle\\ &:= \langle\langle x, x\rangle\rangle - 2\frac{\langle\langle x, y\rangle\rangle}{\langle\langle y, y\rangle\rangle}\langle\langle x, y\rangle\rangle + \frac{\langle\langle x, y\rangle\rangle^2}{\langle\langle y, y\rangle\rangle}\end{aligned}$$

$$
\begin{aligned}
&= \langle\langle x, x\rangle\rangle - \frac{\langle\langle x, y\rangle\rangle^2}{\langle\langle y, y\rangle\rangle} \\
&= \frac{\langle\langle x, x\rangle\rangle\langle\langle y, y\rangle\rangle - \langle\langle x, y\rangle\rangle^2}{\langle\langle y, y\rangle\rangle} \\
&= 0,
\end{aligned}
$$

hence $x = \lambda_y(x)y$. Since $x \neq \mathbf{0}$ and $y \neq \mathbf{0}$, we get $\lambda_y(x) \neq 0$, and hence x, y are linearly dependent. □

THEOREM 4.2.2 (Jordan, von Neumann). *If $(X, ||.||)$ is a normed space, its norm $||.||$ is induced by some positive definite inner product $\langle\langle\cdot, \cdot\rangle\rangle$ on X if and only if $||.||$ satisfies the parallelogram law.*

PROOF. The satisfiability of the parallelogram law follows from the Proposition 4.2.1(iii). For the converse, we define, due to the polarization identity,

$$
\langle\langle x, y\rangle\rangle := \frac{1}{4}\Big(||x+y||^2 - ||x-y||^2\Big). \tag{4.1}
$$

It is immediate to show that $\langle\langle x, y\rangle\rangle$ is positive definite and symmetric form on X. It is also straightforward to see that

$$
\langle\langle -x, y\rangle\rangle = -\langle\langle x, y\rangle\rangle. \tag{4.2}
$$

In order to show left additivity, by the parallelogram law and the definition of $\langle\langle x, y\rangle\rangle$ we have that

$$
\begin{aligned}
4\langle\langle x+z, y\rangle\rangle &= ||x+z+y||^2 - ||x+z-y||^2 \\
&= \left|\left|\left(x+\frac{y}{2}\right) + \left(z+\frac{y}{2}\right)\right|\right|^2 - \left|\left|\left(x-\frac{y}{2}\right) + \left(z-\frac{y}{2}\right)\right|\right|^2 \\
&= 2\left|\left|x+\frac{y}{2}\right|\right|^2 + 2\left|\left|z+\frac{y}{2}\right|\right|^2 - ||x-z||^2 \\
&- \left(2\left|\left|x-\frac{y}{2}\right|\right|^2 + 2\left|\left|z-\frac{y}{2}\right|\right|^2 - ||x-z||^2\right) \\
&= 2\left(\left|\left|x+\frac{y}{2}\right|\right|^2 - \left|\left|x-\frac{y}{2}\right|\right|^2\right) + 2\left(\left|\left|z+\frac{y}{2}\right|\right|^2 - \left|\left|z-\frac{y}{2}\right|\right|^2\right) \\
&= 8\langle\langle x, \frac{y}{2}\rangle\rangle + 8\langle\langle z, \frac{y}{2}\rangle\rangle,
\end{aligned}
$$

where for the last equality we used the definition of $\langle\langle x, y\rangle\rangle$. Hence we get

$$
\langle\langle x+z, y\rangle\rangle = 2\Big(\langle\langle x, \frac{y}{2}\rangle\rangle + \langle\langle z, \frac{y}{2}\rangle\rangle\Big). \tag{4.3}
$$

If in (4.3) we set $z = 0$, we get for every $x, y \in X$

$$
\langle\langle x, y\rangle\rangle = 2\langle\langle x, \frac{y}{2}\rangle\rangle. \tag{4.4}
$$

Consequently, (4.3) becomes

$$\langle\langle x+z, y\rangle\rangle = 2\bigg(\langle\langle x, \frac{y}{2}\rangle\rangle + \langle\langle z, \frac{y}{2}\rangle\rangle\bigg) = \langle\langle x, y\rangle\rangle + \langle\langle z, y\rangle\rangle.$$

The rest of the proof is an exercise. □

In [**2**] there are around 350 characterizations of when a norm is induced by a positive definite inner product!

PROPOSITION 4.2.3. *Let $(X, \langle\langle\cdot,\cdot\rangle\rangle)$ be a positive definite, inner product space, and $||\cdot||$ the norm on X induced by $\langle\langle\cdot,\cdot\rangle\rangle$. Let $B := \{v_1, \dots, v_n\} \subseteq X$, such that (a) $v_i \perp v_j$, for every $i, j \in \{1, \dots, n\}$ with $i \neq j$, and (b) $v_i \neq \mathbf{0}$, for every $i \in \{1, \dots, n\}$. If $x \in X$, let*

$$\lambda_i(x) := \lambda_{v_i}(x) := \frac{\langle\langle x, v_i\rangle\rangle}{\langle\langle v_i, v_i\rangle\rangle} \quad \& \quad x_B := \sum_{i=1}^{n} \lambda_i(x) v_i.$$

(i) The vectors $v_1, \dots, v_n$ are linearly independent.
(ii) The vector $x - x_B$ is orthogonal to every element of B.
(iii) The vector x_B is the closest approximation to x from $\langle B\rangle$ i.e., for every $a_1, \dots, a_n \in \mathbb{R}$, we have that

$$||x - x_B|| := \bigg|\bigg|x - \sum_{i=1}^{n} \lambda_i(x) v_i\bigg|\bigg| \leq \bigg|\bigg|x - \sum_{i=1}^{n} a_i v_i\bigg|\bigg|.$$

(iv) Let $||v_i|| = 1$, for every $i \in \{1, \dots, n\}$.
(α) The following inequality, known as Bessel inequality, holds

$$\sum_{i=1}^{n} \lambda_i(x)^2 \leq ||x||^2.$$

(β) For every $x \in X$ we have that

$$||x||^2 = \sum_{i=1}^{n} \lambda_i(x)^2$$

if and only if B is a basis of X.

PROOF. (i) Let $a_1, \dots, a_n \in \mathbb{R}$ such that $\sum_{i=1}^{n} a_i v_i = \mathbf{0}$. If $k \in \{1, \dots, n\}$, then

$$0 = \bigg\langle\bigg\langle \sum_{i=1}^{n} a_i v_i, v_k \bigg\rangle\bigg\rangle = a_k \langle\langle v_k, v_k\rangle\rangle := a_k ||v_k||^2,$$

and by the hypothesis (b) we get $a_k = 0$.
(ii) If $k \in \{1, \dots, n\}$, then

$$\bigg\langle\bigg\langle x - \sum_{i=1}^{n} \lambda_i(x) v_i, v_k \bigg\rangle\bigg\rangle = \langle\langle x, v_k\rangle\rangle - \bigg\langle\bigg\langle \sum_{i=1}^{n} \lambda_i(x) v_i, v_k \bigg\rangle\bigg\rangle$$

$$
\begin{aligned}
&= \langle\langle x, v_k\rangle\rangle - \langle\langle \lambda_k(x)v_k, v_k\rangle\rangle \\
&= \langle\langle x, v_k\rangle\rangle - \lambda_k(x)\langle\langle v_k, v_k\rangle\rangle \\
&:= \langle\langle x, v_k\rangle\rangle - \frac{\langle\langle x, v_k\rangle\rangle}{\langle\langle v_k, v_k\rangle\rangle}\langle\langle v_k, v_k\rangle\rangle \\
&= 0.
\end{aligned}
$$

(iii) Since $x - x_B$ is orthogonal to B, it is also orthogonal to $\langle B\rangle$, hence

$$\left(x - \sum_{i=1}^{n} \lambda_i(x)v_i\right) \perp \sum_{i=1}^{n} \big(\lambda_i(x) - a_i\big)v_i.$$

By the theorem of Pythagoras we have that

$$
\begin{aligned}
\left|\left|x - \sum_{i=1}^{n} a_i v_i\right|\right|^2 &= \left|\left|x - \sum_{i=1}^{n} a_i v_i + \sum_{i=1}^{n} \lambda_i(x)v_i - \sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 \\
&= \left|\left|\left(x - \sum_{i=1}^{n} \lambda_i(x)v_i\right) + \sum_{i=1}^{n} \big(\lambda_i(x) - a_i\big)v_i\right|\right|^2 \\
&= \left|\left|x - \sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 + \left|\left|\sum_{i=1}^{n} \big(\lambda_i(x) - a_i\big)v_i\right|\right|^2,
\end{aligned}
$$

and the required inequality follows.
(iv)(α) Since $x - x_B$ is orthogonal to $\langle B\rangle$, we have that

$$\left(x - \sum_{i=1}^{n} \lambda_i(x)v_i\right) \perp \sum_{i=1}^{n} \lambda_i(x)v_i.$$

By the theorem of Pythagoras and its generalised version we have that

$$
\begin{aligned}
||x||^2 &= \left|\left|\left(x - \sum_{i=1}^{n} \lambda_i(x)v_i\right) + \sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 \\
&= \left|\left|x - \sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 + \left|\left|\sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 \\
&= \left|\left|x - \sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 + \sum_{i=1}^{n} ||\lambda_i(x)v_i||^2 \\
&= \left|\left|x - \sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 + \sum_{i=1}^{n} \lambda_i(x)^2||v_i||^2 \\
&= \left|\left|x - \sum_{i=1}^{n} \lambda_i(x)v_i\right|\right|^2 + \sum_{i=1}^{n} \lambda_i(x)^2,
\end{aligned}
$$

and the required inequality follows. The proof of (β) is an exercise. Notice that we do not need to suppose that X is finite-dimensional. $\square$

THEOREM 4.2.4. *If* $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ *is a positive definite, inner product space of dimension* $n \geq 1$, Y *is a proper subspace of* X, *and* $B_Y := \{y_1, \ldots, y_m\}$ *is an orthogonal basis of* Y, *there exist* $v_{m+1}, \ldots, v_n \in X$ *such that*

$$B_{\perp} := \{y_1, \ldots, y_m, v_{m+1}, \ldots, v_n\}$$

is an orthogonal basis of X.

PROOF. By the Corollary 2.2.16(iv) there exist $u_{m+1}, \ldots, u_n \in X$ such that $B := \{y_1, \ldots, y_m, u_{m+1}, \ldots, u_n\}$ is a basis of X. If $Y_{m+1} := \langle y_1, \ldots, y_m, u_{m+1} \rangle$, let

$$v_{m+1} := u_{m+1} - \sum_{j=1}^{m} \lambda_{y_j}(u_{m+1}) y_j := \sum_{j=1}^{m} \frac{\langle\langle u_{m+1}, y_j \rangle\rangle}{\langle\langle y_j, y_j \rangle\rangle} y_j.$$

By the Proposition 4.2.3(ii) we get $v_{m+1} \perp \{y_1, \ldots, y_m\}$. Since

$$u_{m+1} = v_{m+1} + \sum_{j=1}^{m} \lambda_{y_j}(u_{m+1}) y_j,$$

we get $u_{m+1} \in \langle y_1, \ldots, y_m, v_{m+1} \rangle$. Moreover, $v_{m+1} \neq \mathbf{0}$, since if $v_{m+1} = \mathbf{0}$, then $u_{m+1} = \sum_{j=1}^{m} \lambda_{y_j}(u_{m+1}) y_j$, and B is not linearly independent. Consequently, $\{y_1, \ldots, y_m, v_{m+1}\}$ is an orthogonal basis of Y_{m+1}. Repeating this step $m - n$ number of times (actually, an induction takes place here) we reach our conclusion. □

COROLLARY 4.2.5. *If* $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ *is a positive definite, inner product space of dimension* $n \geq 1$, *there exists an orthogonal (orthonormal) basis* $B_{\perp}$ $(B_{\perp}^1)$ *of* X.

PROOF. If $v_1 \in X$ such that $v_1 \neq \mathbf{0}$, then $\{v_1\}$ is an orthogonal basis of $\langle v_1 \rangle$, and by the Theorem 4.2.4 there is an orthogonal basis $B_{\perp}$ of X that extends $\{v_1\}$. Moreover, if $B_{\perp} := \{v_1, \ldots, v_n\}$ is an orthogonal basis of X, then

$$B_{\perp}^1 := \left\{ \frac{v_1}{||v_1||}, \ldots, \frac{v_n}{||v_n||} \right\}$$

is an orthonormal basis of X. □

Hence, if $B := \{u_1, \ldots, u_n\}$ is a basis of X, and

$$\begin{aligned}
v_1 &:= u_1 \\
v_2 &:= u_2 - \frac{\langle\langle u_2, v_1 \rangle\rangle}{\langle\langle v_1, v_1 \rangle\rangle} v_1 \\
v_3 &:= u_3 - \frac{\langle\langle u_3, v_2 \rangle\rangle}{\langle\langle v_2, v_2 \rangle\rangle} v_2 - \frac{\langle\langle u_3, v_1 \rangle\rangle}{\langle\langle v_1, v_1 \rangle\rangle} v_1 \\
&\vdots \qquad \vdots \\
v_n &:= u_n - \frac{\langle\langle u_n, v_{n-1} \rangle\rangle}{\langle\langle v_{n-1}, v_{n-1} \rangle\rangle} v_{n-1} - \ldots - \frac{\langle\langle u_n, v_1 \rangle\rangle}{\langle\langle v_1, v_1 \rangle\rangle} v_1,
\end{aligned}$$

then $B_\perp := \{v_1, \ldots, v_n\}$ is an orthogonal basis of X. The above algorithm for finding an orthogonal basis of X, given a finite basis of X, is called the *Gram-Schmidt orthogonalisation process.*

REMARK 4.2.6. Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a positive definite, inner product space of dimension $n \geq 1$, $B_X := \{v_1, \ldots, v_n\}$ a basis of X, and

$$x = \sum_{i=1}^{n} \lambda_i v_i \quad \& \quad y = \sum_{i=1}^{n} \mu_i v_i,$$

where $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n \in \mathbb{R}$. Let $a_{ij} := \langle\langle v_i, v_j \rangle\rangle$, for every $i, j \in \{1, \ldots, n\}$.
(i) $\langle\langle x, y \rangle\rangle = \sum_{i,j=1}^{n} \lambda_i \mu_j a_{ij}$.
(ii) If B_X is orthogonal, then $\langle\langle x, y \rangle\rangle = \sum_{i=1}^{n} \lambda_i \mu_i a_{ii}$.
(iii) If B_X is orthonormal, then $\langle\langle x, y \rangle\rangle = \sum_{i=1}^{n} \lambda_i \mu_i$.

PROOF. Exercise. $\square$

Since the standard basis B_n of $\mathbb{R}^n$ is orthonormal, by the Remark 4.2.6(iii) the Euclidean inner product is uniquely determined. Next we see why $Y^\perp$ is called the orthogonal complement of the subspace Y.

THEOREM 4.2.7. *If $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ is a positive definite, inner product space of dimension $n \geq 1$, and Y is a subspace of X of dimension r, where $0 \leq r \leq n$, then*

$$X = Y \oplus Y^\perp,$$

and hence $\dim(Y^\perp) = n - r$.

PROOF. Exercise. $\square$

The above decomposition of X does not hold, in general, if the inner product on X is not positive definite. E.g., we can find a subspace Y of the Minkowski linetime such that $\{\mathbf{0}\} \subsetneq Y^\perp \cap Y$ and $\mathbb{R}^2$ is not equal to $Y + Y^\perp$ (Exercise). Next we add one more equivalent condition to the Proposition 3.2.4.

PROPOSITION 4.2.8. *If $A := [a_{ij}] \in M_{m,n}(\mathbb{R})$ and $(x_1, \ldots, x_n) \in \mathbb{R}^n$, the following are equivalent.*
(i) $(x_1, \ldots, x_n)$ is a solution of the following system of linear equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= 0 \\ \vdots \qquad\qquad \vdots \quad &\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= 0. \end{aligned}$$

(ii) $(x_1, \ldots, x_n)$ is orthogonal to the row vectors $A_1, \ldots, A_m$ of A.

PROOF. Exercise. $\square$

Now we can complete the Corollary 3.2.6.

COROLLARY 4.2.9. *If $A \in M_{m,n}(\mathbb{R})$, then*

$$\mathtt{cRank}(A) = \mathtt{rRank}(A).$$

PROOF. Exercise. $\square$

4.3. Existence of an orthogonal basis, the general case

THEOREM 4.3.1. *If X is a non-trivial, finite-dimensional linear space and $\langle\langle \cdot, \cdot \rangle\rangle$ is an inner product on X, then X has an orthogonal basis.*

PROOF. We prove the following by induction on $n \geq 1$:

$$\forall_{n \geq 1} \forall_{X \in \mathbf{Lin}} \big(\dim(X) = n \Rightarrow X \text{ has an orthogonal basis} \big).$$

If $\dim(X) = 1$, then if $B_1 := \{v_1\}$ is a basis of X, then B_1 is orthogonal, since according to the Definition 4.1.12, the formula

$$\forall_{i,j \in \{1\}} \big(i \neq j \Rightarrow \langle\langle v_i, v_j \rangle\rangle = 0 \big)$$

holds trivially with the use of the rule Ex falso quodlibet (Efq). Suppose that the formula above is true for $n > 1$, and we prove it for $n + 1$. For this we suppose that X is a linear space with $\dim(X) = n + 1$. If for every $x \in X$ we have that $Q(x) = 0$, then by the polarisation identity (Remark 4.1.2(v)) the inner product $\langle\langle \cdot, \cdot \rangle\rangle$ is the constant function 0, and hence any basis of X is trivially orthogonal. Suppose next that there is $v_1 \in X$ such that $Q(v_1) \neq 0$. If $V_1 := \langle v_1 \rangle$, we show that

$$X = V_1 \oplus V_1^{\perp}.$$

First we show that $V_1 \cap V_1^{\perp} \subseteq \{\mathbf{0}\}$, hence, since both V_1 and $V_1^{\perp}$ are subspaces of X, we get $V_1 \cap V_1^{\perp} = \{\mathbf{0}\}$. If $x \in V_1 \cap V_1^{\perp}$, there is some $\lambda \in \mathbb{R}$ such that $x = \lambda v_1$, and $x \perp V_1$, hence

$$\langle\langle \lambda v_1, v_1 \rangle\rangle = \lambda \langle\langle v_1, v_1 \rangle\rangle = 0,$$

hence, since $Q(v_1) \neq 0$, we get $\lambda = 0$, and $x = \mathbf{0}$. Next we show that $X = V_1 + V_1^{\perp}$. If $x \in X$, then

$$x = \lambda_{v_1}(x) + (x - \lambda_{v_1}(x)),$$

and since $(x - \lambda_{v_1}(x)) \perp v_1$, we have that $x - \lambda_{v_1}(x) \in V_1^{\perp}$. Since $\dim(V_1^{\perp}) = n$, by the inductive hypothesis on $V_1^{\perp}$ there is some orthogonal basis $\{v_2, \ldots, v_{n+1}\}$ of $V_1^{\perp}$. Since $v_2, \ldots, v_{n+1} \perp v_1$, we have that

$$\{v_1, v_2, \ldots, v_{n+1}\}$$

is an orthogonal basis of X. $\square$

If $B_{\perp} = \{v_1, \ldots, v_n\}$ is an orthogonal basis of a finite-dimensional inner product space X, we assume without loss of generality, otherwise we re-enumerate $B_{\perp}$ accordingly, that $B_{\perp}$ is of the form

$$B_{\perp} = \big\{ v_1, \ldots, v_r, v_{r+1}, \ldots, v_s, v_{s+1}, \ldots, v_n \big\},$$

where

$$\begin{aligned} Q(v_j) &> 0, \quad j \in \{1, \ldots, r\}, \\ Q(v_k) &< 0, \quad k \in \{r+1, \ldots, s\}, \\ Q(v_l) &= 0, \quad l \in \{s+1, \ldots, n\}. \end{aligned}$$

If $x \in X$, there are $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that $x = \sum_{i=1}^n \lambda_i v_i$, and hence

$$\begin{aligned} Q(x) &= Q\bigg(\sum_{i=1}^n \lambda_i v_i\bigg) \\ &= \bigg\langle\bigg\langle \sum_{i=1}^n \lambda_i v_i, \sum_{i=1}^n \lambda_i v_i \bigg\rangle\bigg\rangle \\ &= \sum_{i=1}^n \lambda_i^2 Q(v_i) \\ &= \sum_{j=1}^r \lambda_j^2 Q(v_j) + \sum_{k=r+1}^s \lambda_k^2 Q(v_k) + \sum_{l=s+1}^n \lambda_l^2 Q(v_l) \\ &= \sum_{j=1}^r \lambda_j^2 Q(v_j) + \sum_{k=r+1}^s \lambda_k^2 Q(v_k). \end{aligned}$$

COROLLARY 4.3.2. *If X is a non-trivial, finite-dimensional linear space and $\langle\langle \cdot, \cdot \rangle\rangle$ is an inner product on X, then X has an orthonormal basis.*

PROOF. By the Theorem 4.3.1, and as we explained above, X has an orthogonal basis of the form

$$B_\perp = \big\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_s, v_{s+1}, \ldots, v_n\big\}.$$

Then

$$B_\perp^1 = \left\{\frac{v_1}{\sqrt{Q(v_1)}}, \ldots, \frac{v_r}{\sqrt{Q(v_r)}}, \frac{v_{r+1}}{\sqrt{-Q(v_{r+1})}}, \ldots, \frac{v_s}{\sqrt{-Q(v_s)}}, v_{s+1}, \ldots, v_n\right\}$$

is an orthonormal basis of X, since it is orthogonal, and

$$Q\bigg(\frac{v_j}{\sqrt{Q(v_j)}}\bigg) = \bigg\langle\bigg\langle \frac{v_j}{\sqrt{Q(v_j)}}, \frac{v_j}{\sqrt{Q(v_j)}} \bigg\rangle\bigg\rangle = 1,$$

$$Q\bigg(\frac{v_k}{\sqrt{-Q(v_k)}}\bigg) = \bigg\langle\bigg\langle \frac{v_k}{\sqrt{-Q(v_j)}}, \frac{v_k}{\sqrt{-Q(v_k)}} \bigg\rangle\bigg\rangle = -1,$$

for every $j \in \{1, \ldots, r\}$ and every $k \in \{r+1, \ldots, s\}$, respectively, and $Q(v_l) = 0$, for every $l \in \{s+1, \ldots, n\}$. □

In the next section we show that the numbers r and s in the above form of an orthogonal basis of X are the same in *any* orthogonal basis of X.

4.4. Sylvester's theorem

PROPOSITION 4.4.1. *Let X be a non-trivial, n-dimensional linear space, $\langle\langle \cdot, \cdot \rangle\rangle$ an inner product on X, and let*

$$X_0 := \big\{x \in X \mid \forall_{z \in X}\big(\langle\langle x, z\rangle\rangle = 0\big\}.$$

(i) X_0 is a linear subspace of X.

(ii) If $B_\perp := \{v_1, \ldots, v_n\}$ is an orthogonal basis of X, and if

$$i_0 := \Bigg|\Big\{i \in \{1, \ldots, n\} \mid Q(v_i) = 0\Big\}\Bigg|,$$

then $\dim(X_0) = i_0$.

PROOF. (i) This is immediate to show.
(ii) Let $B_\perp$ has the form

$$B_\perp = \big\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_s, v_{s+1}, \ldots, v_n\big\},$$

as explained in the previous section. We show that

$$X_0 = \langle v_{s+1}, \ldots, v_n\rangle,$$

and hence $\dim(X_0) = i_0 = n - s$. Let $x \in X_0$, and let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that

$$x = \sum_{j=1}^{r} \lambda_j v_j + \sum_{k=r+1}^{s} \lambda_k v_k + \sum_{l=s+1}^{n} \lambda_l v_l.$$

If $j \in \{1, \ldots, r\}$, and $k \in \{r+1, \ldots, s\}$, then $Q(v_j) > 0$ and $Q(v_k) < 0$, hence

$$\langle\langle x, v_j\rangle\rangle = \lambda_j Q(v_j) = 0 \Rightarrow \lambda_j = 0,$$

$$\langle\langle x, v_k\rangle\rangle = \lambda_k Q(v_k) = 0 \Rightarrow \lambda_k = 0.$$

Consequently,

$$x = \sum_{l=s+1}^{n} \lambda_l v_l \;\in\; \langle v_{s+1}, \ldots, v_n\rangle.$$

To show that $\langle v_{s+1}, \ldots, v_n\rangle \subseteq X_0$, let $x \in \langle v_{s+1}, \ldots, v_n\rangle$ and $\mu_{s+1}, \ldots, \mu_n \in \mathbb{R}$ such that $x = \sum_{l=s+1}^{n} \mu_l v_l$. If $z \in X$, let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that $x = \sum_{i=1}^{n} \lambda_i v_i$. Hence,

$$\begin{aligned}
\langle\langle x, z\rangle\rangle &= \Big\langle\Big\langle \sum_{i=1}^{n} \lambda_i v_i, \sum_{l=s+1}^{n} \mu_l v_l \Big\rangle\Big\rangle \\
&= \Big\langle\Big\langle \sum_{j=1}^{r} \lambda_j v_j + \sum_{k=r+1}^{s} \lambda_k v_k + \sum_{l=s+1}^{n} \lambda_l v_l, \sum_{l=s+1}^{n} \mu_l v_l \Big\rangle\Big\rangle \\
&= 0
\end{aligned}$$

i.e., $x \in X_0$. □

Since $i_0 = \dim(X_0)$, the number i_0 does not depend on the choice of $B_\perp$ i.e., it is the same for any orthogonal basis of X, and it is called the *index of nullity* of the inner product $\langle\langle\cdot,\cdot\rangle\rangle$. By the definition of X_0 we have that if $\langle\langle\cdot,\cdot\rangle\rangle$ is non-degenerate inner product on X, then $i_0 = 0$. As a consequence of the previous proposition, the number $s = n - i_0$ is the same in every orthogonal basis of X. Next we show that r is also the same in every orthogonal basis of X.

THEOREM 4.4.2 (Sylvester's theorem). *Let X be a non-trivial, n-dimensional linear space, and let $\langle\langle\cdot,\cdot\rangle\rangle$ be an inner product on X. There is a natural number $r \geq 0$ such that for every orthogonal basis $B_\perp = \{v_1, \ldots, v_n\}$ of X there are exactly r many integers i in $\{1, \ldots, n\}$ such that $Q(v_i) > 0$.*

PROOF. Let $B_\perp$ and $C_\perp$ be orthogonal bases of X, where

$$B_\perp = \big\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_s, v_{s+1}, \ldots, v_n\big\},$$
$$C_\perp = \big\{w_1, \ldots, w_{r'}, w_{r'+1}, \ldots, w_s, w_{s+1}, \ldots, w_n\big\},$$

for some $r, r' \geq 0$. We show that the vectors

$$v_1, \ldots, v_r, w_{r'+1}, \ldots, w_s, w_{s+1}, \ldots, w_n$$

are linearly independent in X. Let $a_1, \ldots, a_r, b_{r'+1}, \ldots, b_n \in \mathbb{R}$ such that

$$a_1 v_1 + \ldots + a_r v_r + b_{r'+1} w_{r'+1} + \ldots + b_n w_n = \mathbf{0},$$

hence

$$a_1 v_1 + \ldots + a_r v_r = -\big(b_{r'+1} w_{r'+1} + \ldots + b_n w_n\big).$$

By the last equality we get

$$\begin{aligned} Q\big(a_1 v_1 + \ldots + a_r v_r\big) &= Q\big(-\big(b_{r'+1} w_{r'+1} + \ldots + b_n w_n\big)\big) \\ &= Q\big(-\big(b_{r'+1} w_{r'+1} + \ldots + b_s w_s\big)\big), \end{aligned}$$

where

$$Q\big(a_1 v_1 + \ldots + a_r v_r\big) = a_1^2 Q(v_1) + \ldots + a_r^2 Q(v_r)$$

and $Q(v_1), \ldots, Q(v_r) > 0$ and

$$Q\big(-\big(b_{r'+1} w_{r'+1} + \ldots + b_s w_s\big)\big) = b_{r'+1}^2 Q(w_{r'+1}) + \ldots + b_s^2 Q(w_s),$$

with $Q(w_{r'+1}), \ldots, Q(w_s) < 0$. Since then

$$0 \leq a_1^2 Q(v_1) + \ldots + a_r^2 Q(v_r) = b_{r'+1}^2 Q(w_{r'+1}) + \ldots + b_s^2 Q(w_s) \leq 0,$$

we get

$$a_1^2 Q(v_1) + \ldots + a_r^2 Q(v_r) = 0 = b_{r'+1}^2 Q(w_{r'+1}) + \ldots + b_s^2 Q(w_s),$$

hence $a_1 = \ldots = a_r = b_{r'+1} = \ldots = b_s = 0$. hence, the supposed equality

$$a_1 v_1 + \ldots + a_r v_r + b_{r'+1} w_{r'+1} + \ldots + b_n w_n = \mathbf{0},$$

is reduced to the equality

$$b_{s+1} w_{s+1} + \ldots + b_n w_n = \mathbf{0},$$

hence $b_{s+1} = \ldots = b_n = 0$, since $w_{s+1}, \ldots, w_n$ are linearly independent as elements of $C_\perp$. Since $v_1, \ldots, v_r, w_{r'+1}, \ldots, w_n$ are linearly independent, we get

$$r + (n - r') \leq n \Leftrightarrow r \leq r'.$$

Working similarly, we show that the vectors

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

are linearly independent in X, hence

$$r' + (n - r) \leq n \Leftrightarrow r' \leq r.$$

Since $r \leq r'$ and $r' \leq r$, we get the required equality $r = r'$. $\square$

The integer r in Sylvester's theorem is called the *index of positivity* of $\langle\langle \cdot, \cdot \rangle\rangle$, and the integer $s - r$, which is also independent from the choice of the orthogonal basis, is called the *index of negativity* of $\langle\langle \cdot, \cdot \rangle\rangle$. If $\langle\langle \cdot, \cdot \rangle\rangle$ is the constant 0 inner product, then $r = 0$, and if $\langle\langle \cdot, \cdot \rangle\rangle$ is the Minkowski inner product on $\mathbb{R}^4$, then $r = 3$. Of course, the index of negativity of the Euclidean inner product is 0.

Corollary 4.4.3. *Let X be a non-trivial, n-dimensional linear space, and let $\langle\langle \cdot, \cdot \rangle\rangle$ be an inner product on X. There is a direct sum decomposition*

$$X = X^+ \oplus X^- \oplus X_0$$

of X, where X_0 is defined in the Proposition 4.4.1 and X^+ and X^- satisfy the following conditions:

$$\forall_{x \in X^+}\big(x \neq \mathbf{0} \Rightarrow Q(x) > 0\big),$$
$$\forall_{x \in X^-}\big(x \neq \mathbf{0} \Rightarrow Q(x) < 0\big).$$

Proof. Exercise. $\square$

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

Definition 4.5.1. If $x \in \mathbb{R}^n$, the *open ball* $\mathcal{B}(x, \epsilon)$ with center x and radius $\epsilon > 0$ is defined by

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

The *closed ball* $\mathcal{B}(x, \epsilon]$ with center x and radius ϵ is defined by

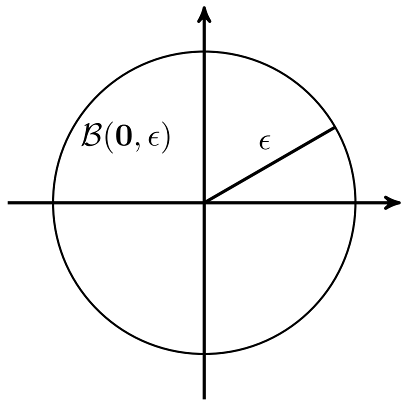
$$\mathcal{B}(x, \epsilon] := \{y \in \mathbb{R}^n \mid d(x, y) \leq \epsilon\}.$$

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.

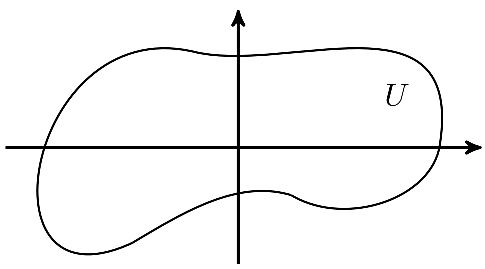
The open ϵ-ball $\mathcal{B}(\mathbf{0}, \epsilon)$ at the origin $(0,0)$ is the open ϵ-disc around $(0,0)$



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. The open ϵ-ball $\mathcal{B}(x, \epsilon)$ is an open set, since if we take any point y in the ball $\mathcal{B}(x, \epsilon)$, we can find an $\epsilon' > 0$ such that $\mathcal{B}(y, \epsilon') \subseteq \mathcal{B}(x, \epsilon)$. If $\epsilon' := \epsilon - |y - x_0|$, then $\epsilon' > 0$, for $y \neq x_0$. If $y = x_0$, we can take $\epsilon' = \epsilon$, and what we want follows immediately. Let $z \in \mathcal{B}(y, \epsilon')$ i.e., $|z - y| < \epsilon'$. Then we have that

$$\begin{aligned}
|z - x_0| &= |(z - y) + (y - x_0)| \\
&\leq |z - y| + |y - x_0| \\
&< \epsilon' + |y - x_0| \\
&= \epsilon - |y - x_0| + |y - x_0| \\
&= \epsilon.
\end{aligned}$$

The closed ϵ-ball $\mathcal{B}(\mathbf{0}, \epsilon]$ is *not* open, since a ball 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. Similarly, 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 4.5.2. *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 open, and $\mathbb{R}^n$ is 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) Exercise. $\square$

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 $\mathbb{R}^3$.

4.6. Differentiable functions on open sets in $\mathbb{R}^n$

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

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

and the following concept is well-defined.

Definition 4.6.1. Let U be an open subset of $\mathbb{R}^n$, $x=(x_1,\dots,x_n)\in U$, and $f:U\to\mathbb{R}$. If the following limit exists

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

we let

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

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 $B_n := \{e_1,\dots,e_i,\dots,e_n\}$ 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

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

then

$$D_1 f(x) := \frac{\partial f}{\partial x}(x)$$

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

where the term $2x$ 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,

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

where the term $3y^2$ 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

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

$$D_i(\lambda f)(x) = \lambda D_i f(x), \qquad \lambda \in \mathbb{R}.$$

DEFINITION 4.6.2. 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

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

E.g., if $f : \mathbb{R}^2 \to \mathbb{R}$ is defined as above by $f(x, y) := x^2y^3$, 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 $\lambda \in \mathbb{R}$. Before defining when a function $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$, is differentiable at some point $x_0 \in U$, we notice the following fact.

Remark 4.6.3. 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. (i) $\Rightarrow$ (ii) 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

$$g(h) := \begin{cases} \phi(h) & , \text{if } h \geq 0 \\ -\phi(h) & , \text{if } h < 0, \end{cases}$$

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,

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

(ii) $\Rightarrow$ (i) 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 4.6.4. 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 4.6.5. *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 [**16**], p. 322. □

4.7. The dual space of an inner product space

In the Definition 2.4.1 we defined the dual space of a linear space as the space of *functionals* on X i.e.,

$$X^* := \{f : X \to \mathbb{R} \mid f \text{ is linear}\}.$$

There are many examples of important functionals from the physical point of view. E.g., the *Dirac functional* $\delta_{x_0} : C([0,1]) \to \mathbb{R}$, where $x_0 \in [0,1]$, is defined by

$$\delta_{x_0}(f) := f(x_0),$$

for every $f \in C([0,1])$. If $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ is an n-dimensional inner product space we can describe in an explicit way the isomorphism between X and X^* (Corollary 2.4.16).

Theorem 4.7.1. *Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a finite-dimensional inner product space, and let $x_0 \in X$.*

(i) The function $L_{x_0} : X \to \mathbb{R}$, defined by

$$L_{x_0}(x) := \langle\langle x, x_0 \rangle\rangle,$$

for every $x \in X$, is in X^*.

(ii) The function $L : X \to X^*$, defined by

$$L(x_0) := L_{x_0},$$

for every $x_0 \in X$, is a linear map.

(iii) If $\langle\langle \cdot, \cdot \rangle\rangle$ is non-degenerate, then L is an isomorphism between X and X^*.

Proof. (i) and (ii) are immediate to show. Since $\dim(X) = \dim(X^*)$, to show (iii) it suffices to show that L is an injection i.e., $\mathtt{Ker}(L) = \{\mathbf{0}\}$. If $L(x) = \mathbf{0}$, then $\forall_{y \in X}\big(L_x(y) = \langle\langle y, x \rangle\rangle = 0\big)$, hence, since $\langle\langle \cdot, \cdot \rangle\rangle$ is non-degenerate, $x = \mathbf{0}$. $\square$

Consequently, if $\langle\langle \cdot, \cdot \rangle\rangle$ is a non-degenerate inner product on X, then for every functional $f \in X^*$ there is a *unique* $x_0 \in X$ such that

$$f = L_{x_0},$$

and the functional f is *represented* by the vector x_0. The condition (b) in the Definition 4.6.4 takes the form

$$f(x_0 + h) - f(x_0) = L_{(\mathrm{grad} f)(x_0)}(h) + |h| g(h),$$

where $L_{(\mathrm{grad} f)(x_0)} : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$L_{(\mathrm{grad} f)(x_0)}(y) := \langle (\mathrm{grad} f)(x_0), y \rangle,$$

for every $y \in \mathbb{R}^n$. I.e., *the differentiability of a function f on an open subset of $\mathbb{R}^n$ means that "locally" f is approximated by some linear functional on $\mathbb{R}^n$*. The functional $L_{(\mathrm{grad} f)(x_0)}$ is called the *derivative* of f at x_0.

Recall that if $B_X := \{v_1, \ldots, v_n\}$ is a basis of a linear space X, then

$$B_{X^*} := \{v_1^*, \ldots, v_n^*\}$$

is a basis of X^*, where for every $i, j \in \{1, \ldots, n\}$ we have that

$$v_i^*(v_j) := \delta_{ij}.$$

Corollary 4.7.2. *Let X be an n-dimensional linear space, and let Y be a subspace of X. Let*

$$Y^{\mathrm{perp}} := \big\{\phi \in X^* \mid \forall_{y \in Y}\big(\phi(y) = 0\big)\big\}.$$

(i) $\dim(Y) + \dim(Y^{\mathrm{perp}}) = n$.

(ii) If $\langle\langle \cdot, \cdot \rangle\rangle$ is a non-degenerate inner product on X, and if $L : X \to X^*$ is the isomorphism between X and X^* defined in the Theorem 4.7.1, then the restriction $L_{|Y^{\perp}} : Y^{\perp} \to Y^{\mathrm{perp}}$, defined by

$$x \mapsto L_x,$$

for every $x \in Y^{\perp}$, is an isomorphism.

Proof. Exercise. □

CHAPTER 5

Operators

5.1. Symmetric and unitary operators

In the Definition 2.4.1 we defined the set of *operators* $\mathcal{L}(X)$ on X as

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

Note that if $\langle\langle \cdot, \cdot \rangle\rangle$ is a non-degenerate inner product on X, then by the definition of non-degeneracy we have that

$$\forall_{z,z' \in X} \Big(\forall_{x \in X} \big(\langle\langle x, z \rangle\rangle = \langle\langle x, z' \rangle\rangle \big) \Rightarrow z = z' \Big). \tag{5.1}$$

LEMMA 5.1.1. *Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a finite-dimensional, non-degenerate inner product space, and $T \in \mathcal{L}(X)$.*
(i) If $y \in X$, the function $\phi_y : X \to \mathbb{R}$, defined, for every $x \in X$, by

$$\phi_y(x) ::= \langle\langle T(x), y \rangle\rangle,$$

is in X^*.
(ii) There is a unique $S \in \mathcal{L}(X)$ such that, for every $x, y \in X$,

$$\langle\langle T(x), y \rangle\rangle = \langle\langle x, S(y) \rangle\rangle, \tag{5.2}$$

PROOF. (i) We can use the fact that the composition of linear maps is linear.
(ii) We fix some $y \in X$. From the case (i) $\phi_y \in X^*$, hence by the Theorem 4.7.1(iii) there is a unique $z \in X$ such that $\phi_y = L_z$, hence for every $x \in X$ we have that

$$\langle\langle T(x), y \rangle\rangle = \phi_y(x) = L_z(x) = \langle\langle x, z \rangle\rangle.$$

Let $S : X \to X$, defined by $y \mapsto z$, for every $z \in X$. Since then the last equality becomes the required equality 5.2, it suffices to show that S is the unique operator satisfying (5.2). First we show that S is a linear map. Let $y_1, y_2 \in X$. Due to (5.1), to show that $S(y_1 + y_2) = S(y_1) + S(y_2)$, it suffices to show that

$$\forall_{x \in X} \Big(\langle\langle x, S(y_1 + y_2) \rangle\rangle = \langle\langle x, S(y_1) + S(y_2) \rangle\rangle \Big).$$

If $x \in X$, we have that

$$\langle\langle x, S(y_1 + y_2) \rangle\rangle = \langle\langle T(x), y_1 + y_2 \rangle\rangle$$

$$
\begin{aligned}
&= \langle\langle T(x), y_1\rangle\rangle + \langle\langle T(x), y_2\rangle\rangle \\
&= \langle\langle x, S(y_1)\rangle\rangle + \langle\langle x, S(y_2)\rangle\rangle \\
&= \langle\langle x, S(y_1) + S(y_2)\rangle\rangle.
\end{aligned}
$$

Similarly, to show that $S(\lambda y) = \lambda S(y)$, where $\lambda \in \mathbb{R}$, it suffices to show that

$$\forall_{x \in X}\Big(\langle\langle x, S(\lambda y)\rangle\rangle = \langle\langle x, \lambda S(y)\rangle\rangle\Big).$$

If $x \in X$, we have that

$$
\begin{aligned}
\langle\langle x, S(\lambda y)\rangle\rangle &= \langle\langle T(x), \lambda y\rangle\rangle \\
&= \lambda\langle\langle T(x), y\rangle\rangle \\
&= \lambda\langle\langle x, S(y)\rangle\rangle \\
&= \langle\langle x, \lambda S(y)\rangle\rangle.
\end{aligned}
$$

Let $S' \in \mathcal{L}(X)$ such that S' satisfies (5.2). If $y \in X$, then

$$\langle\langle x, S'(y)\rangle\rangle = \langle\langle T(x), y\rangle\rangle = \langle\langle x, S(y)\rangle\rangle,$$

hence by (5.1) we get $S'(y) = S(y)$. Since $y \in X$ is arbitrary, we get $S' = S$. $\square$

Definition 5.1.2. Let $(X, \langle\langle \cdot, \cdot\rangle\rangle)$ be a finite-dimensional, non-degenerate inner product space, and $T \in \mathcal{L}(X)$. The unique $S \in \mathcal{L}(X)$ satisfying (5.2) is called the *transpose operator* of T, and it is denoted by T^t. If $T^t = T$, we call T a *symmetric* operator.

If $T \in \mathcal{L}(X)$, where X is as above, then, for every $x, y \in X$, (5.2) becomes

$$\langle\langle T(x), y\rangle\rangle = \langle\langle x, T^t(y)\rangle\rangle, \tag{5.3}$$

and if T is symmetric, then, for every $x, y \in X$, (5.2) becomes

$$\langle\langle T(x), y\rangle\rangle = \langle\langle x, T(y)\rangle\rangle. \tag{5.4}$$

If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, then

$$
\begin{aligned}
\langle x, y\rangle &:= \sum_{i=1}^{n} x_i y_i \\
&= \begin{bmatrix} x_1 \ldots x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}^t \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\
&= x^t y,
\end{aligned}
$$

if we view x, y as $(n \times 1)$-matrices. If $T \in \mathcal{L}(\mathbb{R}^n)$, then

$$\begin{aligned}
\langle x, T^t(y)\rangle &= \langle T(x), y\rangle \\
&= \langle A_T x, y\rangle \\
&= \big(A_T x\big)^t y \\
&= \big(x^t A_T^t\big) y \\
&= x^t \big(A_T^t y\big) \\
&= \langle x, A_T^t y\rangle,
\end{aligned}$$

Since x, y are arbitrary, we get $T^t(y) = A_T^t y$, and since $T^t(y) = A_{T^t} y$, we get

$$A_{T^t} = A_T^t,$$

a fact that explains the use of the same notation and terminology for the transpose of an operator as in the case of a transpose of a matrix.

Proposition 5.1.3. *Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a finite-dimensional, non-degenerate inner product space, $S, T \in \mathcal{L}(X)$ and $\lambda \in \mathbb{R}$.*

(i) $(S + T)^t = S^t + T^t$.

(ii) $(\lambda S)^t = \lambda S^t$.

(iii) $(S \circ T)^t = T^t \circ S^t$.

(iv) $(S^t)^t = S$.

Proof. We work as in the proof of the Lemma 5.1.1 (exercise). □

In the Definition **??** an orthogonal transformation was defined as an operator that was preserving the corresponding inner product. If the inner product is positive definite, we use the following special term.

Definition 5.1.4. Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a finite-dimensional, positive definite inner product space, and let $U \in \mathcal{L}(X)$. We call U *unitary*, if it is inner product-preserving i.e., if for every $x, y \in X$ we have that

$$\langle\langle U(x), U(y)\rangle\rangle = \langle\langle x, y\rangle\rangle.$$

Proposition 5.1.5. *Let $(X, \langle\langle \cdot, \cdot \rangle\rangle)$ be a finite-dimensional, positive definite inner product space, and let $U \in \mathcal{L}(X)$. The following are equivalent.*

(i) U is unitary.

(ii) U is norm-preserving i.e., for every $x \in X$ we have that

$$||U(x)|| = ||x||,$$

where $||.||$ is the norm on X induced by $\langle\langle \cdot, \cdot \rangle\rangle$.

(iii) U sends a unit vector to a unit vector i.e.,

$$\forall_{x \in X}\big(||x|| = 1 \Rightarrow ||U(x)|| = 1\big).$$

Proof. (i) $\Rightarrow$ (ii) If $x \in X$, then we use the following equality

$$||U(x)||^2 = \langle\langle U(x), U(x)\rangle\rangle = \langle\langle x, x\rangle\rangle = ||x||^2.$$

(ii) $\Rightarrow$ (iii) If $x \in X$ such that $||x|| = 1$, then $||U(x)|| = ||x|| = 1$.

(iii) $\Rightarrow$ (ii) If $x = \mathbf{0}$, then $U(\mathbf{0}) = \mathbf{0}$ and the required equality is immediate from the fact that $||.||$ is a function. If $x \neq \mathbf{0}$, then

$$\left|\left|\frac{x}{||x||}\right|\right| = 1,$$

and by (iii) we have that

$$1 = \left|\left|\frac{x}{||x||}\right|\right| = \left|\left|U\left(\frac{x}{||x||}\right)\right|\right| = \left|\left|\frac{1}{||x||}U(x)\right|\right| = \left|\frac{1}{||x||}\right| ||T(x)|| = \frac{1}{||x||}||U(x)||,$$

hence $||U(x)|| = ||x||$.

(ii) $\Rightarrow$ (i) If $x, y \in X$, using the polarisation identity twice we get

$$\begin{aligned}
\langle\langle x, y\rangle\rangle &= \frac{1}{4}\big(\langle\langle x+y, x+y\rangle\rangle - \langle\langle x-y, x-y\rangle\rangle\big)\\
&= \frac{1}{4}\big(||x+y||^2 - ||x-y||^2\big)\\
&= \frac{1}{4}\big(||U(x+y)||^2 - ||U(x-y)||^2\big)\\
&= \frac{1}{4}\big(||U(x)+U(y)||^2 - ||U(x)-U(y)||^2\big)\\
&= \frac{1}{4}\big(\langle\langle U(x)+U(y), U(x)+U(y)\rangle\rangle - \langle\langle U(x)-U(y), U(x)-U(y)\rangle\rangle\big)\\
&= \langle\langle U(x), U(y)\rangle\rangle.
\end{aligned}$$

$\square$

Hence, a unitary operator is characterised from the fact that it sends unit vectors to unit vectors, a property that explains the use of the term "unitary". If

$$\mathbb{S}_1(X) := \{x \in X \mid ||x|| = 1\}$$

is the *unit sphere* of $(X, ||.||)$, then a unitary operator sends $\mathbb{S}_1(X)$ to itself i.e.,

$$U_{|\mathbb{S}_1(X)} : \mathbb{S}_1(X) \to \mathbb{S}_1(X).$$

By definition, a unitary operator preserves orthogonality since

$$\langle\langle x, y\rangle\rangle = 0 \Rightarrow \langle\langle U(x), U(y)\rangle\rangle = 0.$$

The converse is not always true i.e., there are operators that preserve orthogonality but are not unitary (exercise).

Theorem 5.1.6. *If $(X, \langle\langle \cdot, \cdot\rangle\rangle)$ is a finite-dimensional, positive definite inner product space, and $U \in \mathcal{L}(X)$, then U is unitary if and only if*

$$U^t \circ U = \mathrm{id}_X.$$

PROOF. Exercise. □

It is easy to see that a unitary operator is invertible and with some more effort that $U^{-1} = U^t$ (exercise). Moreover, it can be shown (exercise) that if U is unitary, then U^t is unitary. It can also be shown that the only unitary operators of $\mathbb{R}^2$ are the linear maps whose matrices are of the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ or } \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

Note that the determinant of the first is 1 and the determinant of the second is -1.

5.2. Eigenvalues and eigenvectors

DEFINITION 5.2.1. Let X be a linear space and $T \in \mathcal{L}(X)$. An element x of X is called an *eigenvector* of T, if

$$\exists_{\lambda\in\mathbb{R}}\big(T(x) = \lambda x\big). \tag{5.5}$$

If $x \neq \mathbf{0}$, and x is an eigenvector of T, then there is a unique[1] $\lambda \in \mathbb{R}$ satisfying (5.5), which is called an *eigenvalue* of T that *belongs to* x. If $A \in M_n(\mathbb{R})$, an *eigenvector* of A is an eigenvector of $T_A \in \mathcal{L}(\mathbb{R}^n)$.

If we consider the linear space

$$C^\infty(\mathbb{R}) := \big\{f \in \mathbb{F}(\mathbb{R}) \mid f \text{ is infinitely differentiable}\big\},$$

where $f : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable, if there exist all finite derivatives

$$f^{(1)} := f',\ f^{(2)} := f'',\ \ldots,\ f^{(n)} := \big(f^{(n-1)}\big)',\ \ldots,$$

the *derivative operator* is the linear operator

$$' : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$$
$$f \mapsto f',$$

for every $f \in C^\infty(\mathbb{R})$. If $\lambda \in \mathbb{R}$, the function $f \in C^\infty(\mathbb{R})$, defined by

$$f(t) := e^{\lambda t},$$

for every $t \in \mathbb{R}$, is an eigenvector of the derivative and λ belongs to f, since $f'(t) = \lambda f(t)$, for every $t \in \mathbb{R}$. Note that if λ is an eigenvalue of $T \in \mathcal{L}(X)$ that belongs to $x \in X$, then for every $a \in \mathbb{R}$, we have that ax is an eigenvector of T that also belongs to λ;

$$T(ax) = aT(x) = a(\lambda x) = \lambda(ax).$$

Clearly, the set $X_\lambda(T)$ of all eigenvectors of x with $\lambda \in \mathbb{R}$ as an eigenvalue

$$X_\lambda(T) := \big\{x \in X \mid T(x) = \lambda x\big\}$$

[1] If $T(x) = \lambda x = \mu x$, then $(\lambda - \mu)x = \mathbf{0}$, and since $x \neq \mathbf{0}$, we get $\lambda = \mu$.

is a subspace of X, and actually

$$X_\lambda(T) = \mathtt{Ker}(T - \lambda \mathrm{id}_X).$$

We call $X_\lambda(T)$ the *eigenspace of X belonging to λ*.

THEOREM 5.2.2. *If X is a linear space, $T \in \mathcal{L}(X)$, and $v_1, \ldots, v_m$ are eigenvectors of T with eigenvalues $\lambda_1, \ldots, \lambda_m$, respectively, such that $\lambda_i \neq \lambda_j$ for every $i, j \in \{1, \ldots, m\}$ with $i \neq j$, then $v_1, \ldots, v_m$ are linearly independent.*

PROOF. We prove by induction on $n \geq 1$ the following formula

$$\forall_{m \geq 1}\Big(\forall_{v_1,\ldots,v_m \in X}\forall_{\lambda_1,\ldots,\lambda_m \in \mathbb{R}}\big(\mathrm{PWU}(\lambda_1, \ldots, \lambda_m)\ \&$$

$$\&\ \mathrm{EIGEN}_T(v_1, \ldots, v_m; \lambda_1, \ldots, \lambda_m) \Rightarrow \mathrm{LIND}(v_1, \ldots, v_m)\big)\Big),$$

where

$$\mathrm{PWU}(\lambda_1, \ldots, \lambda_m) :\Leftrightarrow \forall_{i,j \in \{1,\ldots,m}\big(i \neq j \Rightarrow \lambda_i \neq \lambda_j\big),$$
$$\mathrm{EIGEN}_T(v_1, \ldots, v_m; \lambda_1, \ldots, \lambda_m) :\Leftrightarrow T(v_1) = \lambda_1 v_1\ \&\ \ldots\ \&\ T(v_m) = \lambda_m v_m,$$
$$\mathrm{LIND}(v_1, \ldots, v_m) :\Leftrightarrow v_1, \ldots, v_m \text{ are linearly independent.}$$

If $m = 1$, and if we fix some $v_1 \in X$ and $\lambda_1 \in \mathbb{R}$, then $\mathrm{PWU}(\lambda_1)$ holds trivially, and if $\mathrm{EIGEN}_T(v_1; \lambda_1)$, then necessarily $v_1 \neq \mathbf{0}$, hence $\mathrm{LIND}(v_1)$. We suppose that the formula for $m > 1$ and we show it for $m + 1$. Let $v_1, \ldots, v_{m+1} \in X$, and $\lambda_1, \ldots, \lambda_{m+1}$ such that $\mathrm{PWU}(\lambda_1, \ldots, \lambda_{m+1})$ and

$$\mathrm{EIGEN}_T(v_1, \ldots, v_{m+1}; \lambda_1, \ldots, \lambda_{m+1}),$$

and we show that $\mathrm{LIND}(v_1, \ldots, v_{m+1})$. For that, let $c_1, \ldots, c_{m+1} \in \mathbb{R}$ such that

$$c_1 v_1 + c_2 v_2 + \ldots + c_{m+1} v_{m+1} = \mathbf{0}. \tag{5.6}$$

Multiplying (5.6) by λ_1 we get

$$c_1 \lambda_1 v_1 + c_2 \lambda_1 v_2 + \ldots + c_{m+1} \lambda_1 v_{m+1} = \mathbf{0}. \tag{5.7}$$

Since

$$\begin{aligned} \mathbf{0} &= T(\mathbf{0}) \\ &= T\big(c_1 v_1 + c_2 v_2 + \ldots + c_{m+1} v_{m+1}\big) \\ &= c_1 T(v_1) + c_2 T(v_2) + \ldots + c_{m+1} T(v_{m+1}) \\ &= c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \ldots + c_{m+1} \lambda_{m+1} v_{m+1}, \end{aligned}$$

subtracting (5.7) from the last equality we get

$$c_2(\lambda_2 - \lambda_1) v_2 + \ldots + c_{m+1}(\lambda_{m+1} - \lambda_1) v_{m+1} = \mathbf{0}. \tag{5.8}$$

Since

$$\mathrm{PWU}(\lambda_1, \ldots, \lambda_{m+1}) \Rightarrow \mathrm{PWU}(\lambda_2, \ldots, \lambda_{m+1}),$$

and

$$\mathrm{EIGEN}_T(v_1, \ldots, v_{m+1}; \lambda_1, \ldots, \lambda_{m+1}) \Rightarrow$$

$$\mathrm{EIGEN}_T((\lambda_2 - \lambda_1)v_2, \ldots, (\lambda_{m+1} - \lambda_1)v_{m+1}; \lambda_2, \ldots, \lambda_{m+1}),$$

by the inductive hypothesis on $v_2, \ldots, v_{m+1} \in X$ and $\lambda_2 - \lambda_1, \ldots, \lambda_{m+1} - \lambda_1 \in \mathbb{R}$ we get

$$\mathrm{LIND}(v_2, \ldots, v_{m+1}),$$

and by (5.8) we get

$$c_2(\lambda_2 - \lambda_1) = \ldots = c_{m+1}(\lambda_{m+1} - \lambda_1) = 0.$$

since by $\mathrm{PWU}(\lambda_1, \ldots, \lambda_{m+1})$ we have that $(\lambda_2 - \lambda_1) \neq 0$ and $\ldots$ and $(\lambda_{m+1} - \lambda_1) \neq 0$, we conclude that

$$c_2 = \ldots = c_{m+1} = 0.$$

Since then (5.7) becomes $c_1 v_1 = \mathbf{0}$, we get $c_1 = 0$. Since $c_1, \ldots, c_{m+1}$ are arbitrary, we get $\mathrm{LIND}(v_1, \ldots, v_{m+1})$. $\square$

By the Theorem 5.2.2 we get another proof that the functions

$$f_1(t) = e^{\lambda_1 t}, \ldots, f_m(t) = e^{\lambda_m t}$$

where $\mathrm{PWU}(\lambda_1, \ldots, \lambda_m)$, are linearly independent in $C^\infty(\mathbb{R})$, since

$$\mathrm{EIGEN}(f_1, \ldots, f_m; \lambda_1, \ldots, \lambda_m).$$

COROLLARY 5.2.3. *If X is a linear space with* $\dim(X) = n \geq 1$*, $T \in \mathcal{L}(X)$, and $v_1, \ldots, v_n$ are eigenvectors of T with eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively, such that $\lambda_i \neq \lambda_j$ for every $i, j \in \{1, \ldots, n\}$ with $i \neq j$, then $\{v_1, \ldots, v_n\}$ is a basis of X.*

PROOF. Exercise. $\square$

PROPOSITION 5.2.4. *Let X be a non-trivial finite-dimensional linear space, $B_X = \{v_1, \ldots, v_n\}$ a basis of X, and $T \in \mathcal{L}(X)$.*

(i) B_X diagonalises T if and only if $v_1, \ldots, v_n$ are eigenvectors of T.

(ii) T is diagonalisable if and only if there is a basis of X consisting of eigenvectors of T.

PROOF. Exercise (use the Definition 3.3.12). $\square$

Next we describe a space having a basis consisting of eigenvectors.

5.3. The simplest ode, but one of the most important

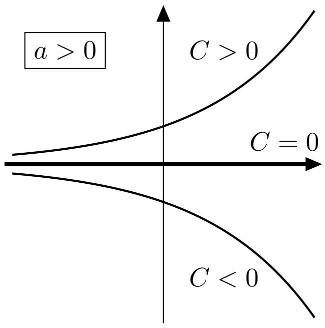
If $a \in \mathbb{R}$ and $x : J \to \mathbb{R}$ is differentiable, where J is an interval of $\mathbb{R}$, one can show (exercise) that the *ordinary differential equation* (ode)

$$x'(t) = ax(t) \tag{5.9}$$

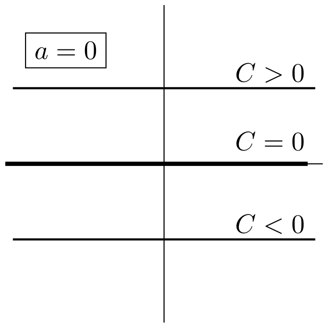
has as set of solutions the set

$$\mathtt{Sol}(5.9) = \big\{s : J \to \mathbb{R} \mid \exists_{C \in \mathbb{R}} \forall_{t \in J}\big(s(t) = Ce^{at}\big)\big\}.$$

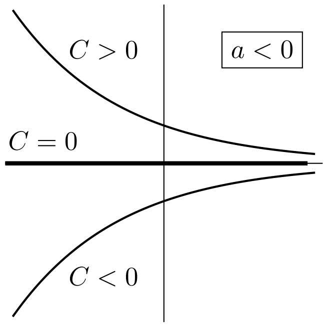
The equation (5.9) is the simplest ode. If $s \in \texttt{Sol}(5.9)$, then $s(0) = C$. Conversely, there is a unique function $s \in \texttt{Sol}(5.9)$ such that $s(0) = C$. This is a special case of the existence of a unique $s \in \texttt{Sol}(5.9)$ satisfying the initial condition $s(t_0) = s_0$, where $t_0 \in J$. The *parameter* a in (5.9) influences dramatically the way the solution curve s looks like. If $a > 0$, then we have the following three cases:



If $C > 0$, then $\lim_{t \longrightarrow +\infty} Ce^{at} = +\infty$, and if $C < 0$, then $\lim_{t \longrightarrow +\infty} Ce^{at} = -\infty$. If $a = 0$, the solution curves are constant functions



If $a < 0$, we have the following three cases:



In this case, if $C \neq 0$, then

$$\lim_{t \longrightarrow +\infty} Ce^{at} = C \lim_{t \longrightarrow +\infty} e^{-|a|t} = C \lim_{t \longrightarrow +\infty} \frac{1}{e^{|a|t}} = 0.$$

The above graphs reflect the *qualitative behavior* of the solution curves. If $a \neq 0$, equation (5.9) is *stable* in the following sense: If a is replaced by some a' sufficiently close to a, the qualitative behavior of the solution curves does not change. E.g., we have that

$$|a' - a| < |a| \Rightarrow \mathtt{sign}(a') = \mathtt{sign}(a),$$

since, if $a > 0$, then $|a' - a| < a \Leftrightarrow -a < a' - a < a \Rightarrow 0 < a' < 2a$, while, if $a < 0$, then $|a' - a| < -a \Leftrightarrow a < a' - a < -a \Rightarrow 2a < a' < 0$. If $a = 0$, equation (5.9) is *unstable*, since the slightest change in the value of a implies a big change in the qualitative behavior of the solution curves. For this reason we say that $a = 0$ is a *bifurcation point* in the one-parameter family of equations

$$\Big(x'(t) = ax(t)\Big)_{a \in \mathbb{R}}.$$

Let the following *system* of two odes in two unknown functions:

$$\begin{aligned} {x_1}'(t) &= a_1 x_1(t), \\ {x_2}'(t) &= a_2 x_2(t). \end{aligned} \tag{5.10}$$

Since there is no relation between $x_1(t)$ and $x_2(t)$, we have that

$$\mathtt{Sol}(5.10) = \bigg\{ s : J \to \mathbb{R}^2 \mid \exists_{C_1, C_2 \in \mathbb{R}} \forall_{t \in J} \Big(s(t) = \big(C_1 e^{a_1 t}, C_2 e^{a_2 t}\big)\Big)\bigg\}.$$

If $s_1(t) = C_1 e^{a_1 t}$ and $s_2(t) = C_2 e^{a_2 t}$, we get $C_1 = s_1(0)$ and $C_2 = s_2(0)$. The equation (5.10) can be written as

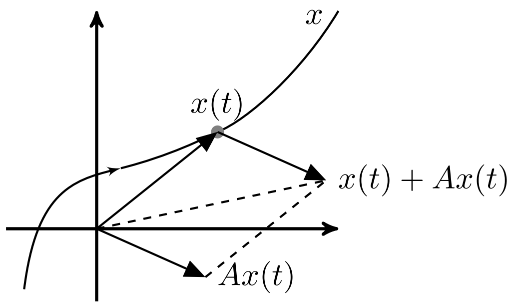
$$x'(t) = Ax(t), \tag{5.11}$$

where

$$A : \mathbb{R}^2 \to \mathbb{R}^2,$$

$$A(x_1, x_2) := (a_1 x_1, a_2, x_2)$$

is a *vector field* on $\mathbb{R}^2$, which geometrically we understand as a function that assigns to each vector $x \in \mathbb{R}^2$ the directed line segment from x to $x + Ax$.



We can write the equation (5.10) using matrices as follows

$$\begin{bmatrix} x_1{}'(t) \\ x_2{}'(t) \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \tag{5.12}$$

If J is an interval of $\mathbb{R}$, $x_1, \ldots, x_n : J \to \mathbb{R}$ are differentiable functions, and $a_{ij} \in \mathbb{R}$, for every $i, j \in \{1, \ldots, n\}$, let the following system of odes

$$\begin{aligned} x_1{}'(t) &= a_{11}x_1(t) + \ldots + a_{1n}x_n(t), \\ &\vdots \\ x_i{}'(t) &= a_{i1}x_1(t) + \ldots + a_{in}x_n(t), \\ &\vdots \\ x_n{}'(t) &= a_{n1}x_1(t) + \ldots + a_{nn}x_n(t). \end{aligned} \tag{5.13}$$

We can write equation (5.13) using matrices as follows

$$\begin{bmatrix} x_1{}'(t) \\ \vdots \\ x_i{}'(t) \\ \vdots \\ x_n{}'(t) \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ \vdots \\ x_i(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \tag{5.14}$$

or, generalizing the simplest ode, we can write it in the form

$$x'(t) = Ax(t), \tag{5.15}$$

where

$$A := \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} =: [a_{ij}]. \tag{5.16}$$

PROPOSITION 5.3.1. *Let the system of odes (5.15), and let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that*

$$A = \mathtt{Diag}(\lambda_1, \ldots, \lambda_n).$$

If $\mathtt{Sol}$ *is the set of solutions of the system (5.15), then* $\mathtt{Sol}$ *is a linear space and*

$$\mathtt{Sol} = \big\langle \big[e^{\lambda_1 t}\big], \ldots, \big[e^{\lambda_n t}\big] \big\rangle,$$

where

$$\big[e^{\lambda_1 t}\big] := \big(e^{\lambda_1 t}, \mathbf{0}, \ldots, \mathbf{0}\big), \ \ldots, \ \big[e^{\lambda_n t}\big] := \big(\mathbf{0}, \mathbf{0}, \ldots, e^{\lambda_n t}\big).$$

PROOF. Exercise. $\square$

5.4. Determinants

DEFINITION 5.4.1. If
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
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 5.4.2. 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.
$$\begin{aligned} \mathtt{Det}(A^1 + C^1, B^2) &:= \begin{vmatrix} a_1 + c_1 & b_1 \\ a_2 + c_2 & b_2 \end{vmatrix} \\ &:= (a_1 + c_1)b_2 - b_1(a_2 + c_2) \\ &= (a_1 b_2 - b_1 a_2) + (c_1 b_2 - b_1 c_2) \\ &:= \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} + \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} \\ &:= \mathtt{Det}(A^1, B^2) + \mathtt{Det}(C^1, B^2). \end{aligned}$$
$\square$

Although one can use the definition of $\texttt{Det}(A)$ to show the following corollary, its proof is simpler, if we use the fundamental properties of the Remark 5.4.2.

Corollary 5.4.3. *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* $\texttt{Det}(A^1 + \lambda B^2, B^2) = \texttt{Det}(A^1, B^2)$.

(ii) If $\lambda \in \mathbb{R}$*, then* $\texttt{Det}(A^1, B^2 + \lambda A^1) = \texttt{Det}(A^1, B^2)$.

(iii) $\texttt{Det}(A^1, B^2) = -\texttt{Det}(B^2, A^1)$.

Proof. Exercise. □

The determinant of a matrix A provides non-trivial information on vectors related to A. We have seen that $\texttt{Det}(I_2) = 1 \neq 0$, 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 5.4.4. *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*

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

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

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

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

($\Leftarrow$) Suppose that $\texttt{Det}(A, B) \neq 0$, 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 5.4.2 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. $\square$

Proposition 5.4.5. *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 then 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 5.4.6. 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 for the determinant of a matrix in $M_n(\mathbb{R})$. To define $\mathtt{Det}(A)$, where $A \in M_n(\mathbb{R})$ and $n > 2$, we suppose that we have defined $\mathtt{Det}(B)$, for every $B \in M_{n-1}(\mathbb{R})$, and we let

$$\mathtt{Det}(A) := (-1)^{i+1} a_{i1} \mathtt{Det}(A_{i1}) + \ldots + (-1)^{i+n} a_{in} \mathtt{Det}(A_{in}),$$

where the matrix $A_{ij} \in M_{n-1}(\mathbb{R})$ is obtained from A by crossing out the i-th row and the j-th column. E.g., if $n = 3$ and $i = 1$, we get

$$\begin{aligned} \mathtt{Det}(A) &:= (-1)^{1+1} a_{11} \mathtt{Det}(A_{11}) + (-1)^{1+2} a_{12} \mathtt{Det}(A_{12}) + (-1)^{1+3} a_{13} \mathtt{Det}(A_{13}) \\ &= a_{11} \mathtt{Det}(A_{11}) - a_{12} \mathtt{Det}(A_{12}) + a_{13} \mathtt{Det}(A_{13}). \end{aligned}$$

It can be shown that $\mathtt{Det} : M_n(\mathbb{R}) \to \mathbb{R}$ is the unique function from $M_n(\mathbb{R})$ to $\mathbb{R}$ that satisfies Remark 5.4.2(i)-(iii) and sends I_n to 1 (see [**17**] p. 149 and p. 171), and it is the unique function from $M_n(\mathbb{R})$ to $\mathbb{R}$ that satisfies the following conditions:

(D_1) $\mathtt{Det}(AB) = \mathtt{Det}(A)\mathtt{Det}(B)$,

(D_2) $\mathtt{Det}(I_n) = 1$,

(D_3) $\mathtt{Det}(A) \neq 0$ if and only if A is invertible.

If B is invertible, it is easy to see that

(D_4) $\mathtt{Det}(B^{-1}) = \mathtt{Det}(B)^{-1}$,

(D_5) $\mathtt{Det}(BAB^{-1}) = \mathtt{Det}(A)$.

From (D_3) on shows that if $A \in M_n(\mathbb{R})$, then

$$\mathtt{Det}(A) \neq 0 \Leftrightarrow \mathtt{Ker}(T_A) = \{0\},$$

and equivalently we have that

$$\mathtt{Det}(A) = 0 \Leftrightarrow \{0\} \subsetneq \mathtt{Ker}(T_A) \Leftrightarrow A \text{ is not invertible.}$$

5.5. The characteristic polynomial

Proposition 5.5.1. *If X is a finite-dimensional linear space, $T \in \mathcal{L}(X)$ and $\lambda \in \mathbb{R}$, then λ is an eigenvalue of T if and only if $T - \lambda \mathrm{id}_X$ is not invertible.*

Proof. We have that

$$\begin{aligned} T - \lambda \mathrm{id}_X \text{ is not invertible } &\Leftrightarrow T \text{ is not an injection} \\ &\Leftrightarrow \{\mathbf{0}\} \subsetneq \mathtt{Ker}(T - \lambda \mathrm{id}_X) \\ &\Leftrightarrow \exists_{x \in X}\big(x \neq \mathbf{0} \ \& \ T(x) = \lambda x\big) \\ &\Leftrightarrow \lambda \text{ is an eigenvalue of } T. \end{aligned}$$

□

Definition 5.5.2. If $A \in M_n(\mathbb{R})$, the *characteristic polynomial of* A is

$$p_A(t) := \mathtt{Det}(tI_n - A) = \begin{vmatrix} t - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & t - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ -a_{n1} & -a_{n2} & \dots & t - a_{nn} \end{vmatrix}.$$

If $T \in \mathcal{L}(\mathbb{R}^n)$, the *characteristic polynomial* p_T of T is p_{A_T}. If X is an n-dimensional linear space and $T \in \mathcal{L}(X)$, the *characteristic polynomial* p_T of T is $p_{A_{T_{B_X}}}$, where B_X is a basis of X.

By the definition of $\mathtt{Det}(A)$ we can show inductively that

$$p_A(t) = t^n + \text{ a polynomial of a degree lower than } n.$$

If $T \in \mathcal{L}(X)$, p_T is independent from the choice of the basis B_X. For this we use the Theorem 3.3.11 and the fact that similar matrices have equal characteristic polynomials; if $C = B^{-1}AB$, and if $t \in \mathbb{R}$, then by the property (D_5) of `Det` we get

$$\begin{aligned}\mathtt{Det}(tI_n - B^{-1}AB) &= \mathtt{Det}(tB^{-1}B - B^{-1}AB)\\ &= \mathtt{Det}(B^{-1}tB - B^{-1}AB)\\ &= \mathtt{Det}(B^{-1}(tB - AB))\\ &= \mathtt{Det}(B^{-1}(tI_nB - AB))\\ &= \mathtt{Det}(B^{-1}(tI_n - A)B)\\ &\overset{(D_5)}{=} \mathtt{Det}(tI_n - A).\end{aligned}$$

E.g., if

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

$$p_A(t) = \begin{vmatrix} t-1 & 1 & -3\\ 2 & t-1 & -1\\ 0 & -1 & t+1\end{vmatrix} = t^3 - t^2 - 2t + 4.$$

Proposition 5.5.3. *If $A \in M_n(\mathbb{R})$ and $\lambda \in \mathbb{R}$, then λ is an eigenvalue of A if and only if λ is a root of the characteristic polynomial of A.*

Proof. By the Definition 5.2.1 and the Proposition 5.5.1 we have that

$$\begin{aligned}\lambda \text{ is an eigenvalue of } A :&\Leftrightarrow \lambda \text{ is an eigenvalue of } T_A\\ &\Leftrightarrow T_A - \lambda \mathrm{id}_X \text{ is not invertible}\\ &\Leftrightarrow \lambda \mathrm{id}_X - T_A \text{ is not invertible}\\ &\Leftrightarrow \mathtt{Det}(A_{[\lambda \mathrm{id}_X - T_A]}) = 0\\ &\Leftrightarrow \mathtt{Det}(\lambda I_n - A) = 0\\ &\Leftrightarrow p_A(\lambda) = 0.\end{aligned}$$

□

E.g., if

$$A = \begin{bmatrix} 1 & 4\\ 2 & 3\end{bmatrix},$$

$$p_A(t) = \begin{vmatrix} t-1 & -4\\ -2 & t-3\end{vmatrix} = t^2 - 4t - 5 = (t-5)(t+1),$$

and the eigenvalues of A are 5 and -1. If (x, y) is an eigenvector of A, the system

$$\begin{bmatrix} 1 & 4\\ 2 & 3\end{bmatrix}\begin{bmatrix} x\\ y\end{bmatrix} = \begin{bmatrix} \lambda x\\ \lambda y\end{bmatrix}$$

takes the form

$$x + 4y = \lambda x,$$
$$2x + 3y = \lambda y,$$

or equivalently

$$(1-\lambda)x + 4y = 0,$$
$$2x + (3-\lambda)y = 0.$$

If $x = 1$, the pair $\left(1, \frac{2}{\lambda-3}\right)$ is an eigenvector of A. If $\lambda = 5$, the eigenspace X_5 has as basis the set $\{(1,1)\}$, and if $\lambda = -1$, then X_{-1} has as basis the set $\{\big(1, -\frac{1}{2}\big)\}$.

PROPOSITION 5.5.4. *If X is a linear space, $T \in \mathcal{L}(X)$, $n \geq 1$, and $B_X := \{v_1, \ldots, v_n\}$ is a basis of X consisting of eigenvectors of T with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively, an eigenvector of T is a scalar multiple of some $v_i \in B_X$.*

PROOF. Exercise. □

5.6. The spectral theorem

DEFINITION 5.6.1. If $A \in M_n(\mathbb{R})$ is symmetric, the *quadratic form q_A associated with A* is the function $q_A : \mathbb{R}^n \to \mathbb{R}$, defined, for every $x \in \mathbb{R}^n$, by

$$q_A(x) := \langle x, Ax \rangle.$$

If $x \in \mathbb{R}^n$, by our remark after the Definition 5.1.2 we have that

$$\begin{aligned}
q_A(x) &= x^t A x \\
&= \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} x_1 a_{11} + \ldots + x_n a_{n1} & \ldots & x_1 a_{1n} + \ldots x_n a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n x_i a_{i1} & \ldots & \sum_{i=1}^n x_i a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1 \sum_{i=1}^n x_i a_{i1} + \ldots + x_n \sum_{i=1}^n x_i a_{in} \\
&= \sum_{j=1}^n \left(\sum_{i=1}^n x_i a_{ij} \right)
\end{aligned}$$

$$= \sum_{j=1,i=1}^{n} x_j x_i a_{ij}$$
$$= \sum_{i,j=1}^{n} a_{ij} x_i x_j,$$

hence, since $x \in \mathbb{R}^n$ is arbitrary, we get

$$q_A = \sum_{i,j=1}^{n} a_{ij} \mathtt{pr}_i \mathtt{pr}_j \tag{5.17}$$

DEFINITION 5.6.2. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called *continuous*, if for every $V \subseteq \mathbb{R}^m$ open its inverse image

$$f^{-1}(V) := \{x \in \mathbb{R}^n \mid f(x) \in V\}$$

under f is open in $\mathbb{R}^n$. We denote by $C(\mathbb{R}^n, \mathbb{R}^m)$ the set of continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, and we also use the notation $C(\mathbb{R}^n) := C(\mathbb{R}^n, \mathbb{R})$.

It is easy to see that $C(\mathbb{R}^n)$ is a linear space and it is also closed under multiplication. The composition of continuous functions is also continuous. If G is open in $\mathbb{R}$, we have that

$$\mathtt{pr}_i^{-1}(G) = \mathbb{R} \times \ldots \times \mathbb{R} \times G \times \mathbb{R} \ldots \times \mathbb{R},$$

hence $\mathtt{pr}_i$ is a continuous function, and by (5.17) $q_A \in C(\mathbb{R}^n)$.

DEFINITION 5.6.3. A subset B of $\mathbb{R}^n$ is called *bounded*, if there are $r > 0$ and $x_0 \in \mathbb{R}^n$ such that $B \subseteq \mathcal{B}(x_0, r)$. A subset K of $\mathbb{R}^n$ is called *compact*, if K is closed and bounded.

Clearly, the unit sphere $S_1(\mathbb{R}^n)$ of $\mathbb{R}^n$, where

$$S_1(\mathbb{R}^n) := \{x \in \mathbb{R}^n \mid |x| = 1\},$$

is closed and bounded, therefore a compact subset of $\mathbb{R}^n$. One can show that if $f \in C(\mathbb{R}^n)$, its restriction $f_{|K} : K \to \mathbb{R}$ to a compact subset K of $\mathbb{R}^n$ has a maximum on K i.e., there is some $u \in K$ such that

$$\forall_{x \in K} \big(f(u) \geq f(x)\big).$$

This fundamental result is a generalisation of the fact that a continuous function $f : [a, b] \to \mathbb{R}$ has a maximum value. For a proof of this see e.g., [**11**], p. 31.

THEOREM 5.6.4. *Let $A \in M_n(\mathbb{R})$ be symmetric, and $q_A : \mathbb{R}^n \to \mathbb{R}$ the quadratic form associated with A. If $P \in S_1(\mathbb{R}^n)$ such that $q_A(P)$ is the maximum value of q_A on $S_1(\mathbb{R}^n)$, then P is an eigenvector of A.*

PROOF. If we define $Y := \langle P \rangle^{\perp}$, then $\dim(Y) = n - 1$. Let also

$$Y_1 := Y \cap S_1(\mathbb{R}^n) = \{y \in Y \mid |y| = 1\}.$$

If $Q \in Y_1$, let the curve[2] $\boldsymbol{x}_Q : \mathbb{R} \to \mathbb{R}^n$, defined by

$$\boldsymbol{x}_Q(t) := (\cos t)P + (\sin t)Q,$$

for every $t \in \mathbb{R}$. Clearly, $\boldsymbol{x}_Q(0) = P$. If $t \in \mathbb{R}$, and since $Q \perp P$, we have that

$$\begin{aligned}
|\boldsymbol{x}_Q(t)| &= \sqrt{\langle \boldsymbol{x}_Q(t), \boldsymbol{x}_Q(t)\rangle} \\
&= \sqrt{\langle (\cos t)P + (\sin t)Q, (\cos t)P + (\sin t)Q\rangle} \\
&= \sqrt{\langle (\cos t)P, (\cos t)P\rangle + \langle (\sin t)Q, (\sin t)Q\rangle} \\
&= \sqrt{(\cos^2 t)\langle P, P\rangle + (\sin^2 t)\langle Q, Q\rangle} \\
&= \sqrt{(\cos^2 t) + (\sin^2 t)} \\
&= 1
\end{aligned}$$

i.e., $\boldsymbol{x}_Q : \mathbb{R} \to S_1(\mathbb{R}^n)$. Moreover, if $t \in \mathbb{R}$, then

$$\boldsymbol{x}_Q{}'(t) := (-\sin t)P + (\cos t)Q,$$

hence $\boldsymbol{x}_Q{}'(0) = Q$. If $g := q_A \circ \boldsymbol{x}_Q : \mathbb{R} \to \mathbb{R}$

$$\begin{tikzcd}
\mathbb{R} \arrow[r, "\boldsymbol{x}_Q"] \arrow[dr, "g"'] & \mathbb{R}^n \arrow[d, "q_A"] \\
 & \mathbb{R},
\end{tikzcd}$$

then for every $t \in \mathbb{R}$ we have that

$$g(t) := q_A(\boldsymbol{x}_Q(t)) := \langle \boldsymbol{x}_Q(t), A\boldsymbol{x}_Q(t)\rangle,$$

hence using the Leibniz rule for the derivative of the product of two real-valued functions[3] and by the symmetry of A we get

$$\begin{aligned}
g'(t) &= \langle \boldsymbol{x}_Q{}'(t), A\boldsymbol{x}_Q(t)\rangle + \langle \boldsymbol{x}_Q(t), A\boldsymbol{x}_Q{}'(t)\rangle \\
&= \langle \boldsymbol{x}_Q{}'(t), A\boldsymbol{x}_Q(t)\rangle + \langle A\boldsymbol{x}_Q(t), \boldsymbol{x}_Q{}'(t)\rangle \\
&= 2\langle \boldsymbol{x}_Q{}'(t), A\boldsymbol{x}_Q(t)\rangle
\end{aligned}$$

---

[2]A curve in $\mathbb{R}^n$ is a function $\boldsymbol{x} : I \to \mathbb{R}^n$, where I is an interval of $\mathbb{R}$. If $\boldsymbol{x}(t) = \big(x_1(t), \ldots, x_n(t)\big)$, for every $t \in I$, we say that $\boldsymbol{x}$ is *differentiable* on I if the coordinate functions $x_1(t), \ldots, x_n(t)$ of $\boldsymbol{x}$ are differentiable on I. 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 $\boldsymbol{x}'(t_0) := \frac{d\boldsymbol{x}}{dt}(t_0) := \big(x_1{}'(t_0), \ldots, x_n{}'(t_0)\big) := \big(\frac{dx_1}{dt}(t_0), \ldots, \frac{dx_n}{dt}(t_0)\big)$. We call $\boldsymbol{x}'(t_0)$ the *velocity vector* of $\boldsymbol{x}(t)$ at time t_0.

[3]If I is an interval of $\mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{y} : I \to \mathbb{R}^n$ are differentiable curves, then it is easy to show that the function $\langle \boldsymbol{x}, \boldsymbol{y}\rangle : I \to \mathbb{R}$ satisfies the Leibniz rule: $\langle \boldsymbol{x}, \boldsymbol{y}\rangle'(t) = \langle \boldsymbol{x}'(t), \boldsymbol{y}(t)\rangle + \langle \boldsymbol{x}(t), \boldsymbol{y}'(t)\rangle$.

for every $t \in \mathbb{R}$. Since $g(0) = q_A(\boldsymbol{x}_Q(0)) = q_A(P)$ is the maximum value of q_A on $S_1(\mathbb{R}^n)$, we have that

$$\begin{aligned} 0 = g'(0) &\Leftrightarrow 2\langle \boldsymbol{x}_Q{}'(0), A\boldsymbol{x}_Q(0)\rangle = 0 \\ &\Leftrightarrow \langle Q, AP\rangle = 0 \\ &\Leftrightarrow AP \perp Q. \end{aligned}$$

Since Q is an arbitrary element of Y_1, we get $AP \perp Y_1$ i.e., $AP \in Y_1{}^{\perp}$, hence $AP \in \langle P\rangle$ i.e., there is some $\lambda \in \mathbb{R}$ such that $AP = \lambda P$. □

THEOREM 5.6.5. *Let $(X, \langle\langle , \rangle\rangle)$ be a finite-dimensional, positive definite inner product space. If $T \in \mathcal{L}(X)$ is symmetric, then T has a non-zero eigenvector.*

PROOF. We use the Theorem 5.6.4 and the fact that A_T is symmetric. □

DEFINITION 5.6.6. Let $(X, \langle\langle \cdot, \cdot\rangle\rangle)$ be a linear space, Y is a subspace of X and $T \in \mathcal{L}(X)$. We say that Y is *invariant under* T, or T-*invariant*, if $T(Y) \subseteq Y$ i.e.,

$$\forall_{y \in Y}\big(T(y) \in Y\big),$$

and we write $\mathrm{INV}(Y; T)$.

PROPOSITION 5.6.7. *Let X be a finite-dimensional, positive definite inner product space, Y a subspace of X, $T \in \mathcal{L}(X)$ symmetric with respect to $\langle\langle \cdot, \cdot\rangle\rangle$, and $x_0 \in X$ a non-zero eigenvector of T.*
(i) If $z \in X$ such that $z \perp x_0$, then $T(z) \perp x_0$.
(ii) If $\mathrm{INV}(Y; T)$, then $\mathrm{INV}(Y^{\perp}; T)$.

PROOF. (i) If $\lambda \in \mathbb{R}$ such that $T(x_0) = \lambda x_0$, then by the symmetry of T we get

$$0 = \lambda\langle\langle x_0, z\rangle\rangle = \langle\langle \lambda x_0, z\rangle\rangle = \langle\langle T(x_0), z\rangle\rangle = \langle\langle x_0, T(z)\rangle\rangle.$$

(ii) If $z \in Y^{\perp}$, we show that $T(z) \in Y^{\perp}$. If $y \in Y$, by the symmetry of T we get

$$\langle\langle T(z), y\rangle\rangle = \langle\langle y, T(z)\rangle\rangle = \langle\langle T(y), z\rangle\rangle = 0,$$

since $T(y) \in Y$ and $z \in Y^{\perp}$. □

DEFINITION 5.6.8. A basis B_X of a linear space consisting of eigenvectors of some $T \in \mathcal{L}(X)$ is called a *spectral* basis for T.

Clearly, a spectral basis for T diagonalises T.

THEOREM 5.6.9 (Spectral theorem). *Let $(X, \langle\langle \cdot, \cdot\rangle\rangle)$ be a non-trivial, finite-dimensional, positive definite inner product space. If $T \in \mathcal{L}(X)$ is symmetric with respect to $\langle\langle \cdot, \cdot\rangle\rangle$, then X has an orthogonal spectral basis for T.*

PROOF. By the Theorem 5.6.5 T has a non-zero eigenvector v_1 with eigenvalue $\lambda_1 \in \mathbb{R}$. Let $Y_1 := \langle y_1\rangle$. If $x = av_1 \in Y_1$, for some $a \in \mathbb{R}$, then

$$T(x) = T(av_1) = aT(v_1) = a\lambda_1 v_1 = (a\lambda_1)v_1 \in Y_1$$

i.e., $\mathrm{INV}(Y_1;T)$. By the Proposition 5.6.7(ii) we get $\mathrm{INV}(Y_1^{\perp};T)$. Since $X = Y_1 \oplus Y_1^{\perp}$, we get $\dim(Y_1^{\perp}) = n-1 \geq 0$. If $n-1=0$, then what we want follows immediately. If $n-1>0$, we repeat the previous argument for the symmetric operator

$$T_{|Y_1^{\perp}} : Y_1^{\perp} \to Y_1^{\perp}.$$

Note that a non-zero eigenvector of $T_{|Y_1^{\perp}}$ is a non-zero eigenvector of T. After n number of steps an orthogonal basis of X is formed consisting of eigenvectors of T. □

If B_X is an orthogonal spectral basis for T, then multiplying each element of B_X with the inverse of its norm we get an orthonormal spectral basis for T.

Corollary 5.6.10. *Let $(X, \langle\langle \cdot,\cdot \rangle\rangle)$ be a non-trivial, finite-dimensional, positive definite inner product space, $T \in \mathcal{L}(X)$ symmetric with respect to $\langle\langle \cdot,\cdot \rangle\rangle$, and v_1, v_2 eigenvectors of T with eigenvalues λ_1, λ_2, respectively. If $\lambda_1 \neq \lambda_2$, then $v_1 \perp v_2$.*

Proof. Exercise. □

CHAPTER 6

Appendix

In the appendix we include some material from various areas that is necessary to the understanding of the main topics of the lecture course.

6.1. Some logic

If P is any formula[1], we define its negation $\neg P$ (not P) by

$$\neg P := P \Rightarrow \bot,$$

where $\bot$ is a formula representing an absurdity.
When we prove a formula we use some logical rules, like the following:

$$[P \;\&\; (P \Rightarrow Q)] \Rightarrow Q,$$

which is called *Modus Ponens* (MP), or like the following rules for the *conjunction* "*P and Q*" ($P \;\&\; Q$) of the formulas P, Q:

$$(P \;\&\; Q) \Rightarrow P,$$

$$(P \;\&\; Q) \Rightarrow Q.$$

Conversely, if we have a proof of P and a proof of Q, we get a proof of $P \;\&\; Q$. A logical rule used in classical mathematics[2] is the *double negation shift* (DNS)

$$(\neg\neg P) \Rightarrow P.$$

The *principle of the excluded middle* (PEM) is the rule

$$P \vee (\neg P),$$

[1] A comprehensive introduction to first-order logic can be found in [**20**].

[2] Mathematics can be done also without the use of DNS and of PEM, in a way that does not contradict classical mathematics i.e., mathematics done with *classical logic*. This more general kind of mathematics is called constructive mathematics (see [**5**] and [**4**]). The mathematics used in physics is mainly classical mathematics. There are though, applications of constructive mathematics in physics (see [**13**] and [**7**]). The logical system generated by excluding DNS, or PEM, is called *intuitionistic logic* and its father is the great Dutch topologist L. E. J. Brouwer.

where $P \vee Q$ is the *disjunction*, "*P or Q*", of the formulas P, Q. The *Ex falso quodlibet* (Efq) - "from falsity anything follows" - is the rule, where P is any formula,

$$\bot \Rightarrow P,$$

LEMMA 6.1.1. *If P, Q are formulas, then* $\big[\neg(P \ \& \ Q)\big] \Leftrightarrow \big(Q \Rightarrow \neg P\big)$.

PROOF. First we show that $\big[\neg(P \ \& \ Q)\big] \Rightarrow \big(Q \Rightarrow \neg P\big)$. For that we suppose

$$\neg(P \ \& \ Q) := (P \ \& \ Q) \Rightarrow \bot,$$

and we prove

$$Q \Rightarrow \neg P := Q \Rightarrow (P \Rightarrow \bot).$$

For that we suppose Q and we show $P \Rightarrow \bot$. Suppose P. Since we already have supposed a proof of Q, we get a proof of $P \ \& \ Q$, and by MP we get $\bot$.
Next we show the converse implication $\big(Q \Rightarrow \neg P\big) \Rightarrow \big[\neg(P \ \& \ Q)\big]$. For that we suppose

$$Q \Rightarrow \neg P := Q \Rightarrow (P \Rightarrow \bot),$$

and we prove

$$\neg(P \ \& \ Q) := (P \ \& \ Q) \Rightarrow \bot.$$

For that we suppose $P \ \& \ Q$ and we show $\bot$. If we have $P \ \& \ Q$, we get Q. Hence by MP with our hypothesis we get $P \Rightarrow \bot$. Similarly, if we have $P \ \& \ Q$, we get P, hence by MP with the hypothesis $P \Rightarrow \bot$ we get $\bot$. $\square$

If $\phi(x)$ is a formula, we use the following abbreviations:

$$\exists_{x \in X} \phi(x) := \exists_x \big(x \in X \ \& \ \phi(x)\big),$$
$$\forall_{x \in X} \phi(x) := \forall_x \big(x \in X \Rightarrow \phi(x)\big).$$

PROPOSITION 6.1.2. *(i)* $\neg \exists_{x \in X} \phi(x) \Leftrightarrow \forall_{x \in X} \big(\neg \phi(x)\big)$.
(ii) $\neg \forall_{x \in X} \phi(x) \Leftrightarrow \exists_x \big(\neg \phi(x)\big)$.

PROOF. (i) First we show that $\neg \exists_{x \in X} \phi(x) \Rightarrow \forall_{x \in X} \big(\neg \phi(x)\big)$. For that we suppose

$$\neg \exists_{x \in X} \phi(x) := \big[\exists_{x \in X} \phi(x)\big] \Rightarrow \bot := \big[\exists_x \big(x \in X \ \& \ \phi(x)\big)\big] \Rightarrow \bot,$$

and we prove

$$\forall_{x \in X} \big(\neg \phi(x)\big) := \forall_x \big(x \in X \Rightarrow \neg \phi(x)\big) := \forall_x \big(x \in X \Rightarrow (\phi(x) \Rightarrow \bot)\big).$$

Let x such that $x \in X$. We show that $\phi(x) \Rightarrow \bot$. For that we suppose $\phi(x)$. But then we have that $\exists_x \big(x \in X \ \& \ \phi(x)\big)$, and by MP we get $\bot$.
Next we show that $\forall_{x \in X} \big(\neg \phi(x)\big) \Rightarrow \neg \exists_{x \in X} \phi(x)$. For that we suppose

$$\forall_{x \in X} \big(\neg \phi(x)\big) :=:= \forall_x \big(x \in X \Rightarrow \neg \phi(x)\big) := \forall_x \big(x \in X \Rightarrow (\phi(x) \Rightarrow \bot)\big),$$

and we show that

$$\neg \exists_{x \in X} \phi(x) := \big[\exists_{x \in X} \phi(x)\big] \Rightarrow \bot := \big[\exists_x \big(x \in X \ \& \ \phi(x)\big)\big] \Rightarrow \bot.$$

Suppose $\exists_x\big(x \in X \ \& \ \phi(x)\big)$. By our hypothesis instantiated to this $x \in X$ we get $\phi(x) \Rightarrow \bot$. Since from $x \in X \ \& \ \phi(x)$ we get $\phi(x)$, by MP we get $\bot$.
(ii) Exercise. $\square$

If $\phi(x, y)$ is a formula, we use the following abbreviations:

$$\exists_{x,y \in X}\phi(x, y) := \exists_{x \in X}\exists_{y \in X}\phi(x, y),$$

$$\forall_{x,y \in X}\phi(x, y) := \forall_{x \in X}\forall_{y \in X}\phi(x, y),$$

and, generally, if $n \geq 1$, and $\phi(x_1, \ldots, x_n)$ is a formula, we use the following abbreviations:

$$\exists_{x_1,\ldots,x_n \in X}\phi(x_1, \ldots, x_n) := \exists_{x_1 \in X} \ldots \exists_{x_n \in X}\phi(x_1, \ldots, x_x),$$

$$\forall_{x_1,\ldots,x_n \in X}\phi(x_1, \ldots, x_n) := \forall_{x_1 \in X} \ldots \forall_{x_n \in X}\phi(x_1, \ldots, x_x).$$

Corollary 6.1.3.

$$(i) \qquad \neg\bigg(\exists_{x_1,\ldots,x_n \in X}\phi(x_1, \ldots, x_n)\bigg) \Leftrightarrow \forall_{x_1,\ldots,x_n \in X}\big(\neg\phi(x_1, \ldots, x_n)\big).$$

$$(ii) \qquad \neg\bigg(\forall_{x_1,\ldots,x_n \in X}\phi(x_1, \ldots, x_n)\bigg) \Leftrightarrow \exists_{x_1,\ldots,x_n \in X}\big(\neg\phi(x_1, \ldots, x_n)\big).$$

Proof. We prove only (i), and for $n = 2$. The proof of the inductive step is the same. For (ii) we work similarly. By the repeated use of the Proposition 6.1.2(i) we get

$$\begin{aligned}
\neg\exists_{x,y \in X}\phi(x, y) &:= \neg\exists_{x \in X}\exists_{y \in X}\phi(x, y) \\
&:= \neg\exists_{x \in X}\big[\exists_{y \in X}\phi(x, y)\big] \\
&\Leftrightarrow \forall_{x \in X}\neg\big[\exists_{y \in X}\phi(x, y)\big] \\
&\Leftrightarrow \forall_{x \in X}\forall_{y \in X}\big(\neg\phi(x, y)\big) \\
&:= \forall_{x,y \in X}\big(\neg\phi(x, y)\big)
\end{aligned}$$

$\square$

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

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

6.2. Some set theory

The theory of sets is a very recent enterprise in the history of mathematics, which was created by Cantor. At the beginning, Cantor used the *Full Comprehension Scheme* (FCS):

$$\exists_u(u = \{x \mid \phi(x)\}),$$

which guarantees the existence of a set generated by *any* formula ϕ that is formed by the symbol of elementhood $\in$. Russell, and independently Zermelo, found that Cantor's principle was contradictory: if $\phi(x) := x \notin x$, then by FCS we have that

$$R = \{x \mid x \notin x\}$$

is a set. The contradiction

$$R \in R \Leftrightarrow R \notin R$$

is known as *Russell's paradox.* Zermelo's *Restricted Comprehension Scheme* (RCS), also known as *Separation Scheme*,

$$\exists_u(u = \{x \in v \mid \phi(x)\})$$

replaced the problematic principle FCS, and Russell's paradox was avoided. If V is the universe of all sets i.e.,

$$V := \{x \mid x = x\},$$

then RCS implies that $V \notin V$: if $V \in V$, then by RCS we have that $u = \{x \in V \mid x \notin x\} \in V$ and then $u \in u \leftrightarrow u \notin u$. If FCS was not contradictory, we wouldn't need so many axioms to describe our intuition about sets. E.g., the union of two sets would be defined as $u \cup v = \{x \mid x \in u \vee x \in v\}$. The first-order non-logical axioms of the *Zermelo-Fraenkel set theory* ZF in the first-order language of the symbol $\in$ are the following[3]:

Extensionality: $\forall_{x,y}\big(\forall_z(z \in x \Leftrightarrow z \in y) \Rightarrow x = y\big)$.

Empty set: $\exists_x\forall_y(y \notin x)$.

Unordered pair: $\forall_{x,y}\exists_z\forall_w(w \in z \Leftrightarrow w = x \vee w = y)$.

Union: $\forall_x\exists_y\forall_z\big(z \in y \Leftrightarrow \exists_w(w \in x \ \&\ z \in w)\big)$.

Replacement Scheme: If $\phi(x, y, \vec{w})$ is a function formula, then

$$\forall_x\exists_v\forall_y\big(y \in v \Leftrightarrow \exists_z(z \in x \ \&\ \phi(z, y, \vec{w}))\big).$$

Power-set: $\forall_x\exists_y\forall_z\big(z \in y \Leftrightarrow \forall_w(w \in z \Rightarrow w \in x)\big)$.

Foundation: $\forall_x\Big(x \neq \emptyset \Rightarrow \exists_z\big(z \in x \ \&\ \neg\exists_w(w \in z \wedge w \in x)\big)\Big)$.

Infinity: $\exists_x\big(\emptyset \in x \ \&\ \forall_y(y \in x \Rightarrow y \cup \{y\} \in x)\big)$.

---

[3] An introduction to the axiomatic set theory can be found in [**10**]. A more advanced book on the subject is [**15**].

Unlike the axioms for linear spaces (first came the examples, or *models*, of linear spaces, and then came the axioms for a linear space) the axioms for sets were given first and then their models were studied! The axioms of ZF are generally "accepted" by the classical mathematicians, and ZF is considered the standard foundation of classical mathematics. A function $f : x \to y$, where x, y are sets in ZF, is defined as an appropriate subset of $x \times y$, where the notion of an ordered pair can be defined through the notion of an unordered pair as follows

$$(a, b) := \big\{\{a\}, \{a, b\}\big\}.$$

Clearly,

$$\big\{\{a\}, \{a, b\}\big\} = \big\{\{c\}, \{c, d\}\big\} \Leftrightarrow a = c \ \& \ b = d.$$

If we add to ZF the axiom of choice, we get the system ZFC, where the axiom of choice can be formulated[4] as follows.

Axiom of choice (AC): If $(A_j)_{j \in J}$ is a family of non-empty sets indexed by some set J, there is a function

$$f : J \to \bigcup_{j \in J} A_j,$$

such that $f(j) \in A_j$, for every $j \in J$.

The axiom of choice might be considered trivial, in case I is finite, but as we show in the next theorem, in conjunction with the classical notion of sets, the AC, even for finite I, has non-trivial consequences. If one considers a constructive notion of set[5], then the AC for I finite is not problematic, but the acceptance of the AC when J is *countable* (i.e., when J has as many elements as $\mathbb{N}$) is debatable. The extrapolation though, to the case of an arbitrary set J has non-trivial, and sometimes unexpected, consequences even from the point of view of classical mathematics[6].

Theorem 6.2.1 (Diaconescu 1975). *The* AC *together with some very small part of Zermelo-Fraenkel set theory implies* (*constructively*[7]) *the principle of the excluded middle* PEM.

Proof. Let P be any well-formed formula, $\mathbf{2} := \{0, 1\}$, and let the sets

$$A_0 := \{x \in \mathbf{2} \mid x = 0 \vee P\},$$

[4] Another formulation of AC is the following: $\forall_v \exists_{f : v \to V} \forall_{x \in v} \big(x \neq \emptyset \Rightarrow f(x) \in x\big)$.

[5] There is a a constructive theory of sets CZF that uses intuitionistic logic and it is equivalent to ZF, if PEM is added to it. See [**1**] for a comprehensive study of CZF.

[6] A consequence of the AC is the Banach-Tarski "paradox", according to which, given a solid ball in 3-dimensional space, there exists a decomposition of the ball into a finite number of disjoint subsets, which can then be put back together in a different way to yield two identical copies of the original ball. Indeed, the reassembly process involves only moving the pieces around and rotating them without changing their shape. However, the pieces themselves are not solids in the usual sense, but infinite scatterings of points. The reconstruction can work with as few as five pieces.

[7] I.e., without the use of PEM.

$$A_1 := \{x \in \mathbf{2} \mid x = 1 \vee P\}.$$

Since $0 \in A_0$ and $1 \in A_1$, the sets A_0 and A_1 are non-empty. By AC there is

$$f : \mathbf{2} \to \bigcup_{j \in \mathbf{2}} A_j = A_0 \cup A_1 \subseteq \mathbf{2}$$

such that

$$f(0) \in A_0 \Leftrightarrow f(0) = 0 \vee P \quad \text{and}$$

$$f(1) \in A_1 \Leftrightarrow f(1) = 1 \vee P.$$

Since f takes values in $\mathbf{2}$, we consider the following cases. If $f(0) = 1$, then, since $f(0) \in A_0$, we get P. If $f(0) = 0$, we consider the two possible cases for $f(1)$. If $f(1) = 0$, then, since $f(1) \in A_1$, we get P. If $f(1) = 1$, we show $\neg P$ i.e., we reach a contradiction by supposing P. Suppose P. In this case $A_0 = A_1 = \mathbf{2}$. Let the set $\{A_0, A_1\}$ and the function $f^* : \{A_0, A_1\} \to \mathbf{2}$, defined by $f^*(A_j) := f(j)$, for every $j \in \mathbf{2}$. Since $A_0 = A_1 \Rightarrow f^*(A_0) = f^*(A_1) \Leftrightarrow f(0) = f(1) \Leftrightarrow 0 = 1$, we get the required contradiction. Hence, we showed $P \vee \neg P$. □

In the previous proof we used the (full) Separation Scheme, and the axioms for the unordered pair, the empty set, to define 0 and 1, and the extensionality axiom.

Zorn's lemma (ZL) [1935]: If I is a non-empty poset, such that every chain in I is bounded, then I has a maximal element.

THEOREM 6.2.2. AC $\Rightarrow$ ZL.

PROOF. **(Sketch)** Let $(I, \preceq)$ be a non-empty poset, such that every chain in I is bounded. Let C be a fixed chain in I, where by hypothesis

$$B(C) := \{i \in I \mid i \text{ is a bound of } C\} \neq \emptyset.$$

With the use of PEM, either C contains a maximal element of I, or not. In the first case, the existence of a maximal element follows immediately. Hence, we suppose that C does not contain a maximal element of I (Hyp_1). In this case, we show that

$$B^*(C) := \{i \in B(C) \mid i \notin C\} \neq \emptyset.$$

Suppose that $B^*(C) = \emptyset$ i.e.,

$$\forall_{i \in B(C)}(i \in C) \qquad (\mathrm{Hyp}_2).$$

Let $i_0 \in B(C)$. By Hyp_2 we get $i_0 \in C$. If $i \in I$, such that $i_0 \preceq i$, then $i \in B(C)$ too, hence by Hyp_2 we have that $i \in C$. Since $i_0 \in B(C)$, we get $i \preceq i_0$, hence by the transitivity of $\preceq$ we conclude that $i_0 = i$. Since i is an arbitrary element of I, we have that i_0 is a maximal element in I that it is also in C, which contradicts the hypothesis Hyp_1. Let

$$\mathcal{C} := \big\{C \subseteq I \mid C \text{ is a chain in } I \text{ that contains no maximal element of } I\big\}.$$

By the previous remark the family $\big(B^*(C)\big)_{C\in\mathcal{C}}$ is a family of non-empty sets indexed by $\mathcal{C}$, hence by AC there is a function

$$f:\mathcal{C}\to\bigcup_{C\in\mathcal{C}}B^*(C)$$

such that

$$f(C)\in B^*(C),\quad \text{for every } C\in\mathcal{C}.$$

The idea of the rest of the proof is the following. Let $i_0\in I$. By PEM, either i_0 is maximal in I, and we are done, or it is not. In the latter case, $\{i_0\}$ is a chain in I that contains no maximal element of I i.e., $\{i_0\}\in\mathcal{C}$. Hence,

$$f\big(\{i_0\}\big)\in B^*\big(\{i_0\}\big):=\big\{i\in B\big(\{i_0\}\big)\mid i\notin\{i_0\}\big\}$$

i.e., $f\big(\{i_0\}\big):=i_1$, such that $i_0\prec i_1:\Leftrightarrow i_0\preceq i_1\ \&\ i_0\neq i_1$. Repeating the same argument, either i_1 is maximal in I, or not. In the latter case, $\{i_0,i_1\}$ is a chain in I that contains no maximal element of I i.e., $\{i_0,i_1\}\in\mathcal{C}$. Consequently,

$$i_0\prec i_1\prec f\big(\{i_0,i_1\}\big).$$

Proceeding similarly, and repeating these steps at most as many times as the cardinality of I, the procedure will terminate, something which is equivalent to the existence of a maximal element in I. $\square$

6.3. Some more category theory

Definition 6.3.1. If $\boldsymbol{C}$ is a category, and $f:A\to B$ is an arrow in $\boldsymbol{C}$, f is called an *iso*, if there is an arrow $g:B\to A$ such that $g\circ f=\mathbf{1}_A$ and $f\circ g=\mathbf{1}_B$.

In the Proposition 2.4.9 we showed that the product linear space $X\times Y$ of X and Y satisfies the universal property of the products. Next we show that $X\times Y$ is unique up to isomorphism i.e., if there is some linear space W and linear maps $\varpi_X:W\to X$ and $\varpi_Y:W\to Y$ such that W also satisfies the universal property of the products

$$\begin{tikzcd} & Z \arrow[dl, "f"'] \arrow[d, dashed, "h"'] \arrow[dr, "g"] & \\ X & W \arrow[l, "\varpi_X"] \arrow[r, "\varpi_Y"'] & Y, \end{tikzcd}$$

then W is linearly isomorphic to $X\times Y$. Our proof holds in any category i.e., if A,B and C are objects in $\boldsymbol{C}$, and if $A\times B$ and C satisfy the universal property of the products in $\boldsymbol{C}$, then there is an iso from $A\times B$ to C.

First we remark that if we consider $Z:=W, f:=\varpi_X$ and $g:=\varpi_Y$

$$\begin{tikzcd} & W \arrow[dl, "\varpi_X"'] \arrow[d, dashed, "\mathrm{id}_W"] \arrow[dr, "\varpi_Y"] & \\ X & W \arrow[l, "\varpi_X"] \arrow[r, "\varpi_Y"'] & Y, \end{tikzcd}$$

then, since id_W makes the above inner diagrams commutative, and since the linear map h is unique, we have that $h = \mathrm{id}_W$. Since $X \times Y$ and W both satisfy the universal property of the products we get from the previous remark that

$$\begin{tikzcd} & X \times Y \arrow[dl, "\mathtt{pr}_X"'] \arrow[d, dashed, "h"'] \arrow[dr, "\mathtt{pr}_Y"] & \\ X & W \arrow[l, "\varpi_X"] \arrow[r, "\varpi_Y"'] \arrow[d, dashed, "g"'] & Y \\ & X \times Y \arrow[ul, "\mathtt{pr}_X"] \arrow[ur, "\mathtt{pr}_Y"'] & \end{tikzcd}$$

$g \circ h = \mathrm{id}_{X \times Y}$. Similarly from the commutative diagrams

$$\begin{tikzcd} & W \arrow[dl, "\varpi_X"'] \arrow[d, dashed, "g"'] \arrow[dr, "\varpi_Y"] & \\ X & X \times Y \arrow[l, "\mathtt{pr}_X"] \arrow[r, "\mathtt{pr}_Y"'] \arrow[d, dashed, "h"'] & Y \\ & W \arrow[ul, "\varpi_X"] \arrow[ur, "\varpi_Y"'] & \end{tikzcd}$$

we get that $h \circ g = \mathrm{id}_W$, hence $X \times Y$ and W are (linearly) isomorphic.

DEFINITION 6.3.2. In a commutative diagram of sets and functions, the arrows

$$X \overset{f}{\rightarrowtail} Y$$

$$X \overset{f}{\twoheadrightarrow} Y$$

$$X \overset{f}{\rightarrowtail\!\!\!\!\twoheadrightarrow} Y$$

denote an injection, a surjection, and a bijection, respectively.

Bibliography

[1] P. Aczel, M. Rathjen: *Constructive Set Theory*, book draft, 2010 (available in the internet).
[2] D. Amir: *Characterizations of inner product spaces*, Springer Basel, 1986.
[3] S. Awodey: Category Theory, Oxford University Press, 2010.
[4] A. Bauer: Five stages of accepting constructive mathematics, Bulletin of the American Mathematical Society, Volume 54, Number 3, 2017, 481-498.
[5] E. Bishop and D. S. Bridges: *Constructive Analysis*, Grundlehren der math. Wissenschaften 279, Springer-Verlag, Heidelberg-Berlin-New York, 1985.
[6] A. Blass: Existence of bases implies the axiom of choice, Contemporary Mathematics, Volume 31, 1984, 31-33.
[7] D. S. Bridges: Can constructive mathematics be applied in physics?, Journal of Philosophical Logic 28 (5), 1999, 439-453.
[8] B. Coecke: Kindergarten quantum mechanics-lecture notes, In: *Quantum Theory: Reconsiderations of the Foundations* III. A. Khrennikov, Ed., American Institute of Physics Press. arXiv:quant-ph/0510032, 2005.
[9] B. Coecke: Introducing categories to the practicing physics, arXiv:0808.1032, 2008.
[10] H. B. Enderton: *Elements of Set Theory*, Academic Press, 1977.
[11] O. Forster: *Analysis 2*, Vieweg+Teubner, 2008.
[12] W. Greub: *Linear Algebra*, Springer-Verlag, 1975.
[13] G. Hellman: Constructive mathematics and quantum mechanics: unbounded operators and the spectral theorem, Journal of Philosophical Logic 22, 1993, 221-248.
[14] J-B. Hiriart-Urruty, C. Lemaréchal: *Fundamentals of Convex Analysis*, Springer-Verlag, 2001.
[15] K. Kunen: *Set Theory*, Individual author and College Publications, 2013.
[16] S. Lang: *Undergraduate Analysis*, Springer, 1983.
[17] S. Lang: *Linear Algebra*, Springer-Verlag, 1987.
[18] G. L. Naber: *The Geometry of Minkowski Spacetime, An Introduction to the Mathematics of the Special Theory of Relativity*, Springer, 2012.
[19] H. Rubin, J. E. Rubin: *Equivalents of the axiom of choice II*, North-Holland, 1985.
[20] H. Schwichtenberg, S. Wainer: *Proofs and Computations*, Cambridge University Press, 2012.
[21] Higher category theory and physics, in https://ncatlab.org/nlab/show/higher+category+theory+and+physics