Linear Algebra I

Martin Otto

Winter Term 2006/07

Contents

Chapter 1

Introduction

1.1 Motivating Examples

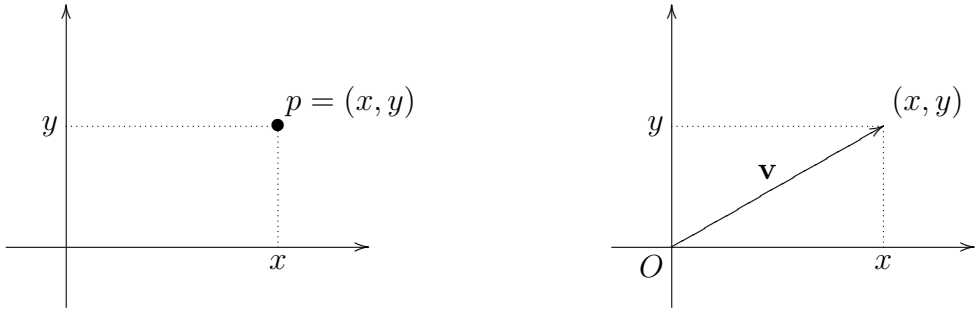
Linear algebra is concerned with the study of vector spaces. It investigates and isolates mathematical structure and methods encountered in phenomena to do with linearity. Instances of linearity arise in many different contexts of mathematics and its applications, and linear algebra provides a uniform framework for their treatment.

As is typical in mathematics, the extraction of common key features that are observed in various seemingly unrelated areas gives rise to an abstraction and simplification which allows to study these crucial features in isolation. The results of this investigation can then be carried back into all those areas where the underlying common feature arises, with the benefit of a unifying perspective. Linear algebra is a very good example of a branch of mathematics motivated by the observation of structural commonality across a wide range of mathematical experience.

In this rather informal introductory chapter, we consider a number of (partly very familiar) examples that may serve as a motivation for the systematic general study of "spaces with a linear structure" which is the core topic of linear algebra.

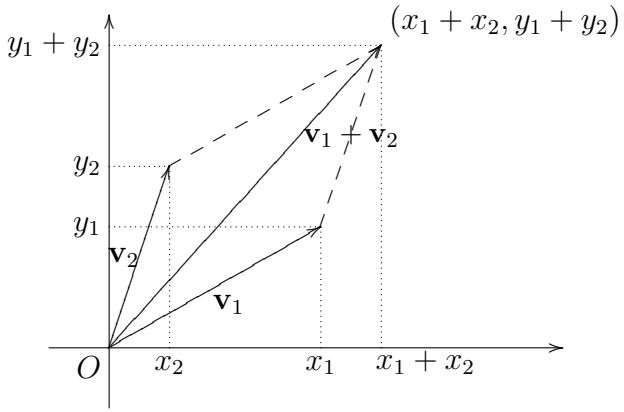
1.1.1 The two-dimensional real plane

The plane of basic planar geometry is modelled as $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, the set of *ordered pairs* [geordnete Paare] (x, y) of real numbers $x, y \in \mathbb{R}$.



One may also think of the directed arrow pointing from the origin $O = (0, 0)$ to the position $p = (x, y)$ as the vector [Vektor] $\mathbf{v} = (x, y)$. "Linear structure" in $\mathbb{R}^2$ has the following features.

Vector addition [Vektoraddition] There is a natural addition over $\mathbb{R}^2$, which may be introduced in two slightly different but equivalent ways.



Arithmetically we may just lift the addition operation $+^{\mathbb{R}}$ from $\mathbb{R}$ to $\mathbb{R}^2$, applying it *component-wise*:

$$(x, y) +^{\mathbb{R}^2} (x', y') := (x +^{\mathbb{R}} x', y +^{\mathbb{R}} y').$$

At first, we explicitly index the plus signs to distinguish their different interpretations: the new over $\mathbb{R}^2$ from the old over $\mathbb{R}$.

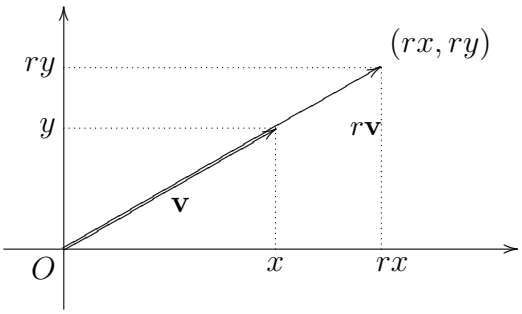
Geometrically we may think of vectors $\mathbf{v} = (x, y)$ and $\mathbf{v}' = (x', y')$ as acting as translations of the plane ; the vector $\mathbf{v} + \mathbf{v}'$ is then the vector which corresponds to the composition of the two translations, translation through $\mathbf{v}$ followed by translation through $\mathbf{v}'$. (Convince yourself that this leads to the same addition operation on $\mathbb{R}^2$.)

Scalar multiplication [Skalare Multiplikation] A real number $r \neq 0$ can also be used to re-scale vectors in the plane. This operation is called scalar multiplication.

$$r \cdot (x, y) := (r \cdot^{\mathbb{R}} x, r \cdot^{\mathbb{R}} y).$$

We include scalar multiplication by $r = 0$, even though it maps all vectors to the null vector $\mathbf{0} = (0, 0)$ and thus does not constitute a proper re-scaling.

Scalar multiplication is arithmetically induced by component-wise ordinary multiplication over $\mathbb{R}$, but is not an operation over $\mathbb{R}^2$ in the same sense that ordinary multiplication is an operation over $\mathbb{R}$. In scalar multiplication a number (a scalar) from the number domain $\mathbb{R}$ operates on a vector $\mathbf{v} \in \mathbb{R}^2$.



It is common practice to drop the $\cdot$ in multiplication notation; we shall later mostly write $r\mathbf{v} = (rx, ry)$ instead of $r \cdot \mathbf{v} = (r \cdot x, r \cdot y)$.

Remark There are several established conventions for vector notation; we here chose to write vectors in $\mathbb{R}^2$ just as pairs $\mathbf{v} = (x, y)$ of real numbers. One may equally well write the two components vertically, as in $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$.

In some contexts it may be useful to be able to switch between these two styles and explicitly refer to *row vectors* [Zeilenvektoren] versus *column vectors* [Spaltenvektoren]. Which style one adopts is largely a matter of convention – the linear algebra remains the same.

Basic laws We isolate some simple arithmetical properties of vector addition and scalar multiplication; these are in fact the crucial features of what "linear structure" means, and will later be our *axioms* [Axiome] for *vector spaces* [Vektorräume]. In the following we use $\mathbf{v}, \mathbf{v}_1, \ldots$ for arbitrary vectors (elements of $\mathbb{R}^2$ in our case) and r, s for arbitrary scalars (elements of the number domain $\mathbb{R}$ in this case).

V1 vector addition is *associative* [assoziativ]. For all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$:

$$(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3).$$

V2 vector addition has a *neutral element* [neutrales Element].
There is a *null vector* $\mathbf{0}$ such that for all $\mathbf{v}$:

$$\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v} = \mathbf{v}.$$

In $\mathbb{R}^2$, $\mathbf{0} = (0, 0) \in \mathbb{R}^2$ serves as the null vector.

V3 vector addition has *inverses* [inverse Elemente].
For every vector $\mathbf{v}$ there is a vector $-\mathbf{v}$ such that

$$\mathbf{v} + (-\mathbf{v}) = (-\mathbf{v}) + \mathbf{v} = \mathbf{0}.$$

For $\mathbf{v} = (x, y) \in \mathbb{R}^2$, $-\mathbf{v} := (-x, -y)$ is as desired.

V4 vector addition is *commutative* [kommutativ].
For all $\mathbf{v}_1, \mathbf{v}_2$:

$$\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1.$$

V5 scalar multiplication is *associative*.
For all vectors $\mathbf{v}$ and scalars r, s:

$$r \cdot \big(s \cdot \mathbf{v}\big) = (r \cdot s) \cdot \mathbf{v}.$$

V6 scalar multiplication has a *neutral element* 1.
For all vectors $\mathbf{v}$:

$$1 \cdot \mathbf{v} = \mathbf{v}.$$

V7 scalar multiplication is *distributive* [distributiv] w.r.t. the scalar.
For all vectors $\mathbf{v}$ and all scalars r, s:

$$(r + s) \cdot \mathbf{v} = r \cdot \mathbf{v} + s \cdot \mathbf{v}.$$

V8 scalar multiplication is *distributive* w.r.t. the vector.
For all vectors $\mathbf{v}_1, \mathbf{v}_2$ and all scalars r:

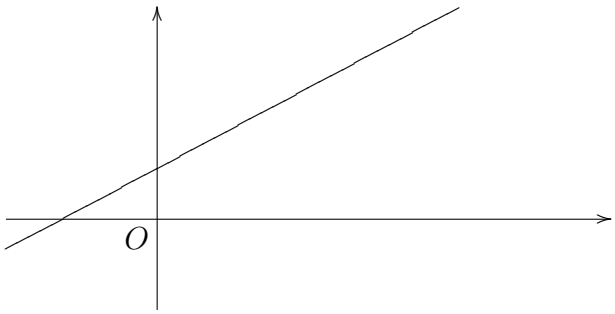
$$r \cdot \big(\mathbf{v}_1 + \mathbf{v}_2\big) = \big(r \cdot \mathbf{v}_1\big) + \big(r \cdot \mathbf{v}_2\big).$$

All the laws in the axioms (V1-4) are immediate consequences of the corresponding properties of ordinary addition over $\mathbb{R}$, since $+^{\mathbb{R}^2}$ is component-wise $+^{\mathbb{R}}$. Similarly (V5/6) are immediate from corresponding properties of multiplication over $\mathbb{R}$, because of the way in which scalar multiplication between $\mathbb{R}$ and $\mathbb{R}^2$ is component-wise ordinary multiplication. Similar comments apply to the distributive laws for scalar multiplication (V7/8), but here the relationship is slightly more interesting because of the asymmetric nature of scalar multiplication.

For associative operations like $+$, we freely write terms like $a + b + c + d$ without parentheses, as associativity guarantees that precedence does not matter; similarly for $r \cdot s \cdot \mathbf{v}$.

Exercise 1.1.1 Annotate all $+$ signs in the identities in (V1-8) to make clear whether they take place over $\mathbb{R}$ or over $\mathbb{R}^2$, and similarly mark those places where $\cdot$ stands for scalar multiplication and where it stands for ordinary multiplication over $\mathbb{R}$.

Linear equations over $\mathbb{R}^2$ and lines in the plane



Consider a line in the real plane. It can be thought of as the solution set of a *linear equation* [lineare Gleichung] of the form

$$E\colon ax + by = c.$$

In the equation, x and y are regarded as variables, a, b, c are fixed constants, called the coefficients of E. The *solution set* [Lösungsmenge] of the equation E is

$$S(E) = \{(x, y) \in \mathbb{R}^2 \colon ax + by = c\}.$$

Exercise 1.1.2 Determine coefficients a, b, c for a linear equation E so that its solution set is the line through points $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ for two distinct given points $p_1 \neq p_2$ in the plane.

Looking at arbitrary linear equations of the form E, we may distinguish two degenerate cases:

(i) $a = b = c = 0$: $S(E) = \mathbb{R}^2$ is not a line but the entire plane.

(ii) $a = b = 0$ and $c \neq 0$: $S(E) = \emptyset$ is empty (no solutions).

In all other cases, $S(E)$ really is a line. What does that mean arithmetically or algebraically, though? What can we say about the structure of the solution set in the remaining, non-degenerate cases?

It is useful to analyse the solution set of an arbitrary linear equation

$$E\colon ax + by = c$$

in terms of the associated homogeneous equation

$$E^*\colon ax + by = 0.$$

Generally a linear equation is called *homogeneous* if the right-hand side is 0.

Observation 1.1.1 *The solution set $S(E^*)$ of any homogeneous linear equation is non-empty and closed under scalar multiplication and vector addition over $\mathbb{R}^2$:*

(a) $\mathbf{0} = (0, 0) \in S(E^*)$.

(b) *if* $\mathbf{v} \in S(E^*)$, *then for any* $r \in \mathbb{R}$ *also* $r\mathbf{v} \in S(E^*)$.

(c) *if* $\mathbf{v}, \mathbf{v}' \in S(E^*)$, *then also* $\mathbf{v} + \mathbf{v}' \in S(E^*)$.

In other words, the solution set of a linear equation in the linear space $\mathbb{R}^2$ has itself the structure of a linear space; scalar multiplication and vector addition in the surrounding space naturally restrict to the solution space and obey the same laws (V1-8) in restriction to this subspace. We shall later consider such linear *subspaces* systematically.

Exercise 1.1.3 (i) Prove the claims of the observation.

(ii) Verify that the laws (V1-8) hold in restriction to $S(E^*)$.

We return to the arbitrary linear equation

$$E\colon ax + by = c$$

Observation 1.1.2 *E is homogeneous ($E^* = E$) if, and only if $\mathbf{0} \in S(E)$.*

Proof. (1) Suppose first that $c = 0$. $E\colon a \cdot x + b \cdot y = 0$. Then $(x, y) = \mathbf{0} = (0, 0)$ satisfies the equation, and thus $\mathbf{0} \in S(E)$.

Conversely, if $\mathbf{0} = (0, 0) \in S(E)$, then $(x, y) = (0, 0)$ satisfies the equation $E\colon a \cdot x + b \cdot y = c$. Therefore $a \cdot 0 + b \cdot 0 = c$ and thus $c = 0$ follows.

□

Suppose now that E has at least one solution, $S(E) \neq \emptyset$. So there is some $\mathbf{v}_0 \in S(E)$ [we already know that $\mathbf{v}_0 \neq \mathbf{0}$ if E is not homogeneous.] We claim that then the whole solution set has the form

$$S(E) = \big\{\mathbf{v}_0 + \mathbf{v}\colon \mathbf{v} \in S(E^*)\big\}.$$

In other words it is the result of translating the solution set of the associated homogeneous equation through $\mathbf{v}_0$ where $\mathbf{v}_0$ is any fixed but arbitrary solution of E.

Lemma 1.1.3 *Consider the linear equation $E\colon a \cdot x + b \cdot y = c$ over $\mathbb{R}^2$, with the associated homogeneous equation $E^*\colon a \cdot x + b \cdot y = 0$.*

(a) *$S(E) = \emptyset$ if and only if $c \neq 0$ and $a = b = 0$.*

(b) *Otherwise, if $\mathbf{v}_0 \in S(E)$ then*

$$S(E) = \big\{\mathbf{v}_0 + \mathbf{v}\colon \mathbf{v} \in S(E^*)\big\}.$$

Proof. (1) (a) "If": let $a = b = 0$ and $c \neq 0$. Then E is unsolvable as the left-hand side equals 0 for all x, y while the right-hand side is $c \neq 0$.

"Only if": let $S(E) = \emptyset$. Firstly, c cannot be 0, as otherwise $\mathbf{0} \in S(E) = S(E^*)$. Similarly, if we had $a \neq 0$, then $(x, y) = (c/a, 0) \in S(E)$ and if $b \neq 0$, then $(x, y) = (0, c/b) \in S(E)$.

(b) Let $\mathbf{v}_0 = (x_0, y_0) \in S(E)$, so that $a \cdot x_0 + b \cdot y_0 = c$.

We show the set equality $S(E) = \big\{\mathbf{v}_0 + \mathbf{v}\colon \mathbf{v} \in S(E^*)\big\}$ by showing two inclusions.

$S(E) \subseteq \big\{\mathbf{v}_0 + \mathbf{v}\colon \mathbf{v} \in S(E^*)\big\}$:

Let $\mathbf{v}' = (x', y') \in S(E)$. Then $a \cdot x' + b \cdot y' = c$. As also $a \cdot x_0 + b \cdot y_0 = c$, we have that $a \cdot x' + b \cdot y' - (a \cdot x_0 + b \cdot y_0) = a \cdot (x' - x_0) + b \cdot (y' - y_0) = 0$, whence $\mathbf{v} := (x' - x_0, y' - y_0)$ is a solution of E^*. Therefore $\mathbf{v}' = \mathbf{v}_0 + \mathbf{v}$ for this $\mathbf{v} \in S(E^*)$.

$\big\{\mathbf{v}_0 + \mathbf{v}\colon \mathbf{v} \in S(E^*)\big\} \subseteq S(E)$:

[1] Compare section 1.2.6 for basic proof patterns encountered in these simple examples.

Let $\mathbf{v}' = \mathbf{v}_0 + \mathbf{v}$ where $\mathbf{v} = (x, y) \in S(E^*)$. Note that $\mathbf{v}' = (x_0 + x, y_0 + y)$. As $a \cdot x + b \cdot y = 0$ and $a \cdot x_0 + b \cdot y_0 = c$, we have $a \cdot (x_0 + x) + b \cdot (y_0 + y) = c$ and therefore $\mathbf{v}' \in S(E)$.

□

Parametric representation Let $E\colon a \cdot x + b \cdot y = c$ be such that $S(E) \neq \emptyset$, and $S(E) \neq \mathbb{R}^2$ (these are the degenerate cases). From what we saw above E is non-degenerate if and only if $(a, b) \neq (0, 0)$. In this case we want to turn Lemma 1.1.3 (b) into an explicit *parametric form*. Put

$$\mathbf{w} := (-b, a). \qquad (^2)$$

We check that $\mathbf{w} \in S(E^*)$, and that – under the assumption that E is non-degenerate – $S(E^*) = \{\lambda \cdot \mathbf{w}\colon \lambda \in \mathbb{R}\}$. Combining this with Lemma 1.1.3 (b), we interpret $\mathbf{v}_0 \in S(E)$ as an arbitrary point on the line described by E, which in parametric form is therefore the set of points

$$\mathbf{v}_0 + \lambda \cdot \mathbf{w} \qquad (\lambda \in \mathbb{R}).$$

1.1.2 Three-dimensional real space

Essentially everything we did above in the two-dimensional case carries over to an analogous treatment of the n-dimensional case of $\mathbb{R}^n$. Because it is the second most intuitive case, and still easy to visualise, we now look at the three-dimensional case of $\mathbb{R}^3$.

$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z)\colon x, y, z \in \mathbb{R}\}$$

is the set of three-tuples (triples) of real numbers. Addition and scalar multiplication over $\mathbb{R}^3$ are defined (component-wise) according to

$$(x_1, y_1, z_1) + (x_2, y_2, z_2) := (x_1 + x_2, y_1 + y_2, z_1 + z_2)$$

and

$$r(x, y, z) := (rx, ry, rz),$$

for arbitrary $(x, y, z), (x_1, y_1, z_1), (x_2, y_2, z_2) \in \mathbb{R}^3$ and $r \in \mathbb{R}$.

The resulting structure on $\mathbb{R}^3$, with this addition and scalar multiplication, and null vector $\mathbf{0} = (0, 0, 0)$ satisfies the laws (axioms) (V1-8) from above. (Verify this, as an exercise!)

[2] Geometrically, the vector $(-b, a)$ is orthogonal to the vector (a, b) formed by the coefficients of E^*.

Linear equations over $\mathbb{R}^3$

A linear equation over $\mathbb{R}^3$ takes the form

$$E\colon ax + by + cz = d,$$

for coefficients $a, b, c, d \in \mathbb{R}$. Its solution set is

$$S(E) = \big\{(x, y, z) \in \mathbb{R}^3 \colon ax + by + cz = d\big\}.$$

Again, we consider the associated homogeneous equation

$$E^*\colon ax + by + cz = 0.$$

In complete analogy with Observation 1.1.1 above, we find firstly that $S(E^*)$ contains $\mathbf{0}$ and is closed under vector addition and scalar multiplication (and thus is a linear subspace). Further, in analogy with Lemma 1.1.3, either $S(E) = \emptyset$ or, whenever $S(E) \neq \emptyset$, then

$$S(E) = \big\{\mathbf{v}_0 + \mathbf{v} \colon \mathbf{v} \in S(E^*)\big\}$$

for any fixed but arbitrary solution $\mathbf{v}_0 \in S(E)$.

Exercise 1.1.4 Check the above claims and try to give rigorous proofs. Find out in exactly which cases E has no solution, and in exactly which cases $S(E) = \mathbb{R}^3$. Call these cases degenerate.
Convince yourself, firstly in an example, that in the non-degenerate case the solution set $\emptyset \neq S(E) \neq \mathbb{R}^3$ geometrically corresponds to a plane within $\mathbb{R}^3$. Furthermore, this plane contains the origin (null vector $\mathbf{0}$) iff E is homogeneous.[3] Can you provide a parametric representation of the set of points in such a plane?

1.1.3 Systems of linear equations over $\mathbb{R}^n$

A system of linear equations [lineares Gleichungssystem] consists of a tuple of linear equations that are to be solved simultaneously. The solution set is the *intersection* of the solution sets of the individual equations.

A single linear equation over $\mathbb{R}^n$ has the general form

$$E\colon a_1x_1 + \cdots + a_nx_n = b$$

[3] "iff" is shorthand for "if, and only if", logical equivalence or bi-implication.

with *coefficients* [Koeffizienten] $a_1, \ldots, a_n, b \in \mathbb{R}$ and *variables* [Variable, Unbestimmte] $x_1, \ldots, x_n$.

Considering a system of m linear equations $E_1, \ldots, E_m$ over $\mathbb{R}^n$, we index the coefficients doubly such that

$$E_i\colon a_{i1}x_1 + \cdots + a_{in}x_n = b_i$$

is the i-th equation with coefficients $a_{i1}, \ldots, a_{in}, b_i$. The entire system is

$$E\colon \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 & (E_1) \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 & (E_2) \\ a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n = b_3 & (E_3) \\ \qquad\qquad\qquad\qquad\quad \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m & (E_m) \end{cases}$$

with m *rows* [Zeilen] and n *columns* [Spalten] on the left-hand side and one on the right-hand side. Its *solution set* is

$$\begin{aligned} S(E) &= \big\{\mathbf{v} = (x_1, \ldots, x_n) \in \mathbb{R}^n\colon \mathbf{v} \text{ satisfies } E_i \text{ for } i = 1, \ldots, m\big\} \\ &= S(E_1) \cap S(E_2) \cap \ldots \cap S(E_m) \;=\; \textstyle\bigcap_{i=1,\ldots,m} S(E_i). \end{aligned}$$

The *associated homogeneous system* E^* is obtained by replacing the right-hand sides (the coefficients b_i) by 0.

With the same arguments as in Observation 1.1.1 and Lemma 1.1.3 (b) we find the following.

Lemma 1.1.4 *Let E be a system of linear equations over $\mathbb{R}^n$.*

(i) *The solution set of the associated homogeneous system E^* contains the null vector $\mathbf{0} \in \mathbb{R}^n$ and is closed under vector addition and scalar multiplication (and thus is a linear subspace).*

(ii) *If $S(E) \neq \emptyset$ and $\mathbf{v}_0 \in S(E)$ is any fixed but arbitrary solution, then*

$$S(E) = \big\{\mathbf{v}_0 + \mathbf{v}\colon \mathbf{v} \in S(E^*)\big\}.$$

An analogoue of Lemma 1.1.3 (a), which would tell us when the equations in E have any simultaneous solutions at all, is not so easily available at first. Consider for instance the different ways in which three planes in $\mathbb{R}^3$ may intersect or fail to intersect.

Remark 1.1.5 For a slightly different perspective, consider the vectors of coefficients formed by the columns of E, $\mathbf{a}_i = (a_{1i}, a_{2i}, \ldots, a_{mi}) \in \mathbb{R}^m$ for $i = 1, \ldots, n$ and $\mathbf{b} = (b_1, \ldots, b_m) \in \mathbb{R}^m$. Then E can be rewritten equivalently as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n = \mathbf{b}.$$

(Note, incidentally, that in order to align this view with the usual layout of the system E, one might prefer to think of the $\mathbf{a}_i$ and $\mathbf{b}$ as *column vectors*; for the mathematics of the equation, though, this makes no difference.)

We shall exploit this view further in later chapters. For now we stick with the focus on rows.

We now explore a well-known classical method for the effective solution of a system of linear equations. In the first step we consider individual transformations (of the schema of coefficients in E) that leave the solution set invariant. We then use these transformations systematically to find out whether E has any solutions, and if so, to find them.

Row transformations

If $E_i\colon a_{i1}x_1 + \ldots + a_{in}x_n = b_i$ and $E_j\colon a_{j1}x_1 + \ldots + a_{jn}x_n = b_j$ are rows of E and $r \in \mathbb{R}$ is a scalar, we let

(i) rE_i be the equation

$$rE_i\colon (ra_{i1})x_1 + \ldots + (ra_{in})x_n = rb_i.$$

(ii) $E_i + rE_j$ be the equation

$$E_i + rE_j\colon (a_{i1} + ra_{j1})x_1 + \ldots + (a_{in} + ra_{jn})x_n = b_i + rb_j.$$

Lemma 1.1.6 *The following transformations on a system of linear equations leave the solution set invariant, i.e., lead from E to a new system E' that is equivalent with E.*

(T1) exchanging two rows.

(T2) replacing some E_i by rE_i for a scalar $r \neq 0$.

(T3) replacing E_i by $E_i + rE_j$ for some scalar r and some $j \neq i$.

Proof. It is obvious that (T1) does not affect $S(E)$ as it just corresponds to a re-labelling of the equations.

For (T2), it is clear that $S(E_i) = S(rE_i)$ for $r \neq 0$: for any $(x_1, \ldots, x_n)$

$$a_{i1}x_1 + \ldots + a_{in}x_n = b_i \quad \text{iff} \quad ra_{i1}x_1 + \ldots + ra_{in}x_n = rb_i.$$

For (T3) we show for all $\mathbf{v} = (x_1, \ldots, x_n)$:

$$\mathbf{v} \in S(E_i) \cap S(E_j) \qquad \text{iff} \qquad \mathbf{v} \in S(E_i + rE_j) \cap S(E_j).$$

Assume first that $\mathbf{v} \in S(E_i) \cap S(E_j)$.

Then $a_{i1}x_1 + \ldots + a_{in}x_n = b_i$ and $a_{j1}x_1 + \ldots + a_{jn}x_n = b_j$ together imply that

$$\begin{array}{ll} & (a_{i1} + ra_{j1})x_1 + \ldots + (a_{in} + ra_{jn})x_n \\ = & (a_{i1}x_1 + \ldots + a_{in}x_n) + r(a_{j1}x_1 + \ldots + a_{jn}x_n) \\ = & b_i + rb_j. \end{array}$$

Therefore also $\mathbf{v} \in S(E_i + rE_j)$.

If, conversely, $\mathbf{v} \in S(E_i + rE_j) \cap S(E_j)$, we may appeal to the implication from left to right we just proved for arbitrary r, use it for $-r$ in place of r and get that $\mathbf{v} \in S((E_i + rE_j) + (-r)E_j)$. But $(E_i + rE_j) + (-r)E_j$ is E_i, whence $\mathbf{v} \in S(E_i)$ follows.

□

Gauß-Jordan algorithm

The basis of this algorithm for solving any system of linear equations is also referred to as *Gaussian elimination*, because it successively eliminates variables from some equations by means of the above equivalence transformations. The resulting system finally is of a form (*upper triangle* or *echelon* form [obere Dreiecksgestalt]) in which the solutions (if any) can be read off.

Key step Let

$$E\colon \left\{ \begin{array}{lcllcll} a_{11}x_1 & + & a_{12}x_2 & +\cdots+ & a_{1n}x_n & = & b_1 \qquad (E_1) \\ a_{21}x_1 & + & a_{22}x_2 & +\cdots+ & a_{2n}x_n & = & b_2 \qquad (E_2) \\ a_{31}x_1 & + & a_{32}x_2 & +\cdots+ & a_{3n}x_n & = & b_3 \qquad (E_3) \\ & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & +\cdots+ & a_{mn}x_n & = & b_m \qquad (E_m) \end{array} \right.$$

Assume first that $a_{11} \neq 0$. Then by repeated application of (T3), we may replace

$$\begin{array}{lll} E_2 & \text{by} & E_2 + (-a_{21}/a_{11})E_1 \\ E_3 & \text{by} & E_3 + (-a_{31}/a_{11})E_1 \\ \vdots & & \\ E_m & \text{by} & E_m + (-a_{m1}/a_{11})E_1 \end{array}$$

with the result that the only remaining non-zero coefficient in the first column is a_{11}:

$$E'\colon \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 & (E_1) \\ \qquad\quad a'_{22}x_2 + \cdots + a'_{2n}x_n = b'_2 & (E'_2) \\ \qquad\quad a'_{32}x_2 + \cdots + a'_{3n}x_n = b'_3 & (E'_3) \\ \qquad\qquad\qquad\quad \vdots & \\ \qquad\quad a'_{m2}x_2 + \cdots + a'_{mn}x_n = b'_m & (E'_m) \end{cases}$$

If $a_{11} = 0$ but some other $a_{j1} \neq 0$ we may apply the above steps after first exchanging E_1 with E_j, according to (T1).

In the remaining case that $a_{i1} = 0$ for all i, E itself already has the shape of E' above, even with $a_{11} = 0$.

Iterated application of the key step Starting with E with m rows, we apply the key step to eliminate all coefficients in the first column in rows $2, \ldots, m$;

We then keep the first row unchanged and apply the key step again to treat the first remaining non-zero column in

$$E''\colon \begin{cases} a'_{22}x_2 + \cdots + a'_{2n}x_n = b'_2 & (E'_2) \\ a'_{32}x_2 + \cdots + a'_{3n}x_n = b'_3 & (E'_3) \\ \qquad\qquad\quad \vdots & \\ a'_{m2}x_2 + \cdots + a'_{mn}x_n = b'_m & (E'_m) \end{cases}$$

In each round we reduce the number of rows and columns still to be transformed by at least one. After at most $\max(m, n)$ rounds therefore we

obtain a system

$$\hat{E}\colon \left\{\begin{array}{lllllllll} \hat{a}_{1j_1}x_{j_1} & +\ldots+ & \hat{a}_{1j_2}x_{j_2} & +\ldots+ & \hat{a}_{1j_r}x_{j_r} & +\ldots+ & \hat{a}_{1n}x_n & = & \hat{b}_1 \\ & & \hat{a}_{2j_2}x_{j_2} & +\ldots+ & \hat{a}_{2j_r}x_{j_r} & +\ldots+ & \hat{a}_{2n}x_n & = & \hat{b}_2 \\ & & & & \ldots & & \ldots & & \\ & & & & \hat{a}_{rj_r}x_{j_r} & +\ldots+ & \hat{a}_{rn}x_n & = & \hat{b}_r \\ & & & & & & 0 & = & \hat{b}_{r+1} \\ & & & & & & 0 & = & \hat{b}_{r+2} \\ & & & & & & & \vdots & \\ & & & & & & 0 & = & \hat{b}_m \end{array}\right.$$

in upper triangle (echelon) form:

- r is the number of rows whose left-hand sides have not been completely eliminated (note in particular that $r = m$ can occur).
- $\hat{a}_{ij_i} \neq 0$ is the first non-vanishing coefficient on the left-hand side in row i for $i = 1, \ldots, r$; these coefficients are called *pivot elements*; the corresponding variables $x_{j_i}$for $i = 1, \ldots, r$ are called *pivot variables*.
- the remaining rows $r + 1, \ldots, m$ are those whose left-hand sides have been eliminated completely.

Applications of (T2) to the first r rows can further be used to make all pivot elements $\hat{a}_{ij_i} = 1$ if desired.

Most importantly, $S(\hat{E}) = S(E)$.

Reading off the solutions

Lemma 1.1.7 *For a system $\hat{E}$ in the above upper echelon form:*

(i) $S(\hat{E}) = \emptyset$ *unless* $\hat{b}_{r+1} = \hat{b}_{r+2} = \ldots = \hat{b}_m = 0$.

(ii) *If $\hat{b}_{r+1} = \hat{b}_{r+2} = \ldots = \hat{b}_m = 0$, then the values for all variables that are not pivot variables can be chosen arbitrarily, and matching values for the pivot variables computed, using the i-th equation to determine x_{ij_i}, and progressing in order of $i = r, r-1, \ldots, 1$.*

Moreover, all solutions are obtained in this way.

An obvious question that arises here, is whether the number of non-pivot variables that can be chosen freely in $S(E) = S(\hat{E})$ depends on the particular

sequence of steps in which E was tranformed into upper echelon form $\hat{E}$. We shall later see that this number is an invariant of the elimination procedure and related to the dimension of $S(E^*)$.

1.1.4 Linear spaces over $\mathbb{Z}_2$

We illustrate the point that scalar domains quite different from $\mathbb{R}$ give rise to analogous useful notions of linear structure. Linear algebra over $\mathbb{Z}_2$ has particular relevance in computer science – e.g., in relation to boolean functions, logic, cryptography and coding theory.

Arithmetic in $\mathbb{Z}_2$

[Compare section 1.3.2 for a more systematic account of $\mathbb{Z}_n$ for any n, and section 1.3.3 for $\mathbb{Z}_p$ where p is prime.]

Let $\mathbb{Z}_2 = \{0, 1\}$. One may think of 0 and 1 as integers or as boolean (bit) values here; both view points will be useful.

On $\mathbb{Z}_2$ we consider the following arithmetical operations of addition and multiplication: [4]

$+_2$	0	1
0	0	1
1	1	0

$\cdot_2$	0	1
0	0	0
1	0	1

In terms of integer arithmetic, 0 and 1 and the operations $+_2$ and $\cdot_2$ may be associated with the parity of integers as follows:

$$\begin{array}{ccl} 0 & — & \text{even integers;} \\ 1 & — & \text{odd integers.} \end{array}$$

Then $+_2$ and $\cdot_2$ describe the effect of ordinary addition and multiplication on parity. For instance, (odd) $\cdot$ (odd) = (odd) and (odd) + (odd) = (even).

In terms of boolean values and logic, $+_2$ is the "exclusive or" operation *xor* also denoted $\dot{\vee}$, while $\cdot_2$ is ordinary conjunction $\wedge$.

Exercise 1.1.5 Check the following arithmetical laws for $(\mathbb{Z}_2, +, \cdot)$ where $+$ is $+_2$ and $\cdot$ is $\cdot_2$, as declared above. We use $b, b_1, b_2, \ldots$ to denote arbitrary elements of $\mathbb{Z}_2$:

[4] We (at first) use subscripts in $+_2$ and $\cdot_2$ to distinguish these operations from their counterparts in ordinary arithmetic.

(i) $+$ and $\cdot$ are associative and commutative.
For all b_1, b_2, b_3: $(b_1 + b_2) + b_3 = b_1 + (b_2 + b_3)$;
$b_1 + b_2 = b_2 + b_1$.
Similarly for $\cdot$.

(ii) $\cdot$ is distributive over $+$.
For all b, b_2, b_3: $b \cdot (b_1 + b_2) = (b \cdot b_1) + (b \cdot b_2)$.

(iii) 0 is the neutral element for $+$.
For all b, $b + 0 = 0 + b = b$.
1 is the neutral element for $\cdot$.
For all b, $b \cdot 1 = 1 \cdot b = b$.

(iv) $+$ has inverses.
For all $b \in \mathbb{Z}_2$ there is a $-b \in \mathbb{Z}_2$ such that $b + (-b) = 0$.

(v) $\cdot$ has inverses for all $b \neq 0$.
$1 \cdot 1 = 1$ (as there is only this one instance).

We now look at the space

$$\mathbb{Z}_2^n := (\mathbb{Z}_2)^n = \underbrace{Z_2 \times \cdots \times \mathbb{Z}_2}_{n \text{ times}}$$

of n-tuples over $\mathbb{Z}_2$, or of length n bit-vectors $\mathbf{b} = (b_1, \ldots, b_n)$.

These can be added component-wise according to

$$(b_1, \ldots, b_n) +_2^n (b'_1, \ldots, b'_n) := (b_1 +_2 b'_1, \ldots, b_n +_2 b'_n). \qquad (5)$$

This addition operation provides the basis for a very simple example of a (symmetric) encryption scheme. Consider messages consisting of length n bit vectors, so that $\mathbb{Z}_2^n$ is our message space. Let the two parties who want to communicate messages over an insecure channel be called A for Alice and B for Bob (as is the custom in cryptography literature). Suppose Alice and Bob have agreed beforehand on some bit-vector $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{Z}_2^n$ to be their shared *key*, which they keep secret from the rest of the world.

If Alice wants to communicate message $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{Z}_2^n$ to Bob, she sends the encrypted message

$$\mathbf{m}' := \mathbf{m} + \mathbf{k}$$

[5] Again, the distinguishing markers for the different operations of addition will soon be dropped.

and Bob decrypts the bit-vector $\mathbf{m}'$ that he receives, according to

$$\mathbf{m}'' := \mathbf{m}' + \mathbf{k}$$

using the same key $\mathbf{k}$. Indeed, the arithmetic of $\mathbb{Z}_2$ and $\mathbb{Z}_2^n$ guarantees that always $\mathbf{m}'' = \mathbf{m}$. This is a simple consequence of the peculiar feature that

$$\begin{array}{rlll} & b + b = 0 & \text{for all } b \in \mathbb{Z}_2, & \text{(addition in } \mathbb{Z}_2) \\ \text{whence also} & \mathbf{b} + \mathbf{b} = \mathbf{0} & \text{for all } \mathbf{b} \in \mathbb{Z}_2^n. & \text{(addition in } \mathbb{Z}_2^n) \end{array}$$

Cracking this encryption is just as hard as to come into possession of the agreed key $\mathbf{k}$ – considered sufficiently unlikely in the short run if its length n is large and if there are no other regularities to go by (!). Note that the key is actually retrievable from any pair of plain and encrypted messages $\mathbf{k} = \mathbf{m} + \mathbf{m}'$.

Example, for $n = 8$ and with $k = (0, 0, 1, 0, 1, 1, 0, 1)$; remember that addition in $\mathbb{Z}_2^n$ is bit-wise $\dot\vee$ (exclusive or):

$$\begin{array}{r} \mathbf{m} = 10010011 \\ \mathbf{k} = 00101101 \\ \hline \mathbf{m}' = \mathbf{m} + \mathbf{k} = 10111110 \end{array} \qquad \text{and} \qquad \begin{array}{r} \mathbf{m}' = 10111110 \\ \mathbf{k} = 00101101 \\ \hline \mathbf{m} = \mathbf{m}' + \mathbf{k} = 10010011 \end{array}$$

We also define scalar multiplication over $\mathbb{Z}_2^n$, between $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{Z}_2^n$ and $\lambda \in Z_2$:

$$\lambda \cdot (b_1, \ldots, b_n) := (\lambda \cdot b_1, \ldots, \lambda \cdot b_n).$$

Exercise 1.1.6 Verify that $\mathbb{Z}_2^n$ with vector addition and scalar multiplication as introduced above satisfies all the laws (V1-8) (with null vector $\mathbf{0} = (0, \ldots, 0)$).

Parity check-bit

This is a basic example from coding theory. The underlying idea is used widely, for instance in supermarket bar-codes or in ISBN numbers. Consider bit-vectors of length n, i.e., elements of $\mathbb{Z}_2^n$. Instead of using all possible bit-vectors in $\mathbb{Z}_2^n$ as carriers of information, we restrict ourselves to some subspace $C \subseteq \mathbb{Z}_2^n$ of admissible codes. The possible advantage of this is that small errors (corruption of a few bits in one of the admissible vectors) may

become easily detectable, or even repairable – the fundamental idea in *error detecting* and *error correcting* codes. Linear algebra can be used to devise such codes and to derive efficient algorithms for dealing with them. This is particularly so for *linear codes* $C \subseteq \mathbb{Z}_2^n$, whose distinguishing feature is their closure under addition in $\mathbb{Z}_2^n$.

The following code with parity check-bit provides the most fundamental example, of a (weak) error-detecting code. Let $n > 2$ and consider the following linear equation over $\mathbb{Z}_2^n$:

$$E_+ \colon x_1 + \cdots + x_{n-1} + x_n = 0$$

with solution set

$$C_+ = S(E_+) = \big\{(b_1, \ldots, b_n) \in \mathbb{Z}_2^n \colon b_1 + \cdots + b_{n-1} + b_n = 0\big\}.$$

Note that the linear equation E_+ has coefficients over $\mathbb{Z}_2$, namely just 1s on the left hand side, and 0 on the right (hence homogeneous), and is based on addition in $\mathbb{Z}_2$. A bit-vector satisfies E_+ iff its parity sum is even, i.e., iff the number of 1s is even.

Exercise 1.1.7 How many bit-vectors are there in $\mathbb{Z}_2^n$? What is the proportion of bit-vectors in C_+?

Check that $C_+ \subseteq \mathbb{Z}_2^n$ contains the null vector and is closed under vector addition in $\mathbb{Z}_2^n$ (as well as under scalar multiplication). It thus provides an example of a linear subspace, and hence a so-called linear code.

Suppose that some information (like that on an identification tag for goods) is coded using not arbitrary bit-vectors in $\mathbb{Z}_2^n$ but just bit-vectors from the subspace C_+. Suppose further that some non-perfect data-transmission (e.g. through a scanner) results in some errors but that one can mostly rely on the fact that at most one bit gets corrupted. In this case a test whether the resulting bit-vector (as transmitted by the scanner say) still satisfies E_+ can reliably tell whether an error has occurred or not. This is because whenever $\mathbf{v} = (b_1, \ldots, b_n)$ and $\mathbf{v}' = (b'_1, \ldots, b'_n)$ differ in precisely one bit, then $\mathbf{v} \in C_+$ iff $\mathbf{v}' \notin C_+$.

From error-detecting to error-correcting

Better (and sparser) codes $C \subseteq \mathbb{Z}_2^n$ can be devised which allow not just to detect but even to repair corruptions that only affect a small number of bits.

If $C \subseteq \mathbb{Z}_2^n$ is such that any two distinct elements $\mathbf{v}, \mathbf{v}' \in C$ differ in at least $2t+1$ bits for some constant t, then C provides a *t-error-correcting code*. If $\hat{\mathbf{v}}$ is a possibly corrupted version of $\mathbf{v} \in C$ but differs from $\mathbf{v}$ in at most t places, then $\mathbf{v}$ is uniquely determined as the *unique* element of C which differs from $\hat{\mathbf{v}}$ in at most t places. In case of linear codes C, linear algebra provides techniques for efficient error-correction procedures in this setting.

Boolean functions in two variables

This is another example of spaces $\mathbb{Z}_2^n$ in the context of boolean algebra and propositional logic. Consider the set $\mathbb{B}^2$ of all boolean functions in two variables (which we here denote r and s)

$$\begin{array}{rcl} f\colon \mathbb{Z}_2 \times \mathbb{Z}_2 & \longrightarrow & \mathbb{Z}_2 \\ (r,s) & \longmapsto & f(r,s) \end{array}$$

Each such $f \in \mathbb{B}^2$ is fully represented by its table of values

r	s	$f(r,s)$
0	0	$f(0,0)$
0	1	$f(0,1)$
1	0	$f(1,0)$
1	1	$f(1,1)$

or more succinctly by just the 4-bit vector

$$\underline{f} := (f(0,0), f(0,1), f(1,0), f(1,1)) \in \mathbb{Z}_2^4.$$

We have for instance the following correspondences:

function $f \in \mathbb{B}^2$	arithmetical description of f	vector $\underline{f} \in \mathbb{Z}_2^4$
constant 0	$(r,s) \mapsto 0$	$(0,0,0,0)$
constant 1	$(r,s) \mapsto 1$	$(1,1,1,1)$
projection r	$(r,s) \mapsto r$	$(0,0,1,1)$
projection s	$(r,s) \mapsto s$	$(0,1,0,1)$
negation of r, $\neg r$	$(r,s) \mapsto 1-r$	$(1,1,0,0)$
negation of s, $\neg s$	$(r,s) \mapsto 1-s$	$(1,0,1,0)$
exclusive or, $\dot{\vee}$ (xor)	$(r,s) \mapsto r +_2 s$	$(0,1,1,0)$
conjunction, $\wedge$	$(r,s) \mapsto r \cdot_2 s$	$(0,0,0,1)$
Sheffer stroke, $\mid$ $(nand)$	$(r,s) \mapsto 1 - r \cdot_2 s$	$(1,1,1,0)$

The map

$$\begin{array}{rccc} \_ : & \mathbb{B}^2 & \longrightarrow & \mathbb{Z}_2^4 \\ & f & \longmapsto & \underline{f} \end{array}$$

is a bijection. In other words, it establishes a one-to-one and onto correspondence between the sets $\mathbb{B}^2$ and $\mathbb{Z}_2^4$. Compare section 1.2.2 on (one-to-one) functions and related notions. (In fact more structure is preserved in this case; more on this later.)

It is an interesting fact that all functions in $\mathbb{B}^2$ can be expressed in terms of compositions of just r, s and $|$.[6] For instance, $\neg r = r|r$, $1 = (r|r)|r$, and $r \wedge s = (r|s)|(r|s)$. Consider the following questions:

Is this is also true with $\dot\vee$ (exclusive or) instead of Sheffer's $|$?
If not, can all functions in $\mathbb{B}^2$ be expressed in terms of $0, 1, r, s, \dot\vee, \neg$?
If not all, which functions in $\mathbb{B}^2$ do we get?

Lemma 1.1.8 *Let $f_1, f_2 \in \mathbb{B}^2$ be represented by $\underline{f}_1, \underline{f}_2 \in \mathbb{Z}_2^4$. Then the function*

$$f_1 \dot\vee f_2 \colon (r, s) \longmapsto f_1(r, s) \dot\vee f_2(r, s)$$

is represented by

$$\underline{f_1 \dot\vee f_2} = \underline{f}_1 + \underline{f}_2. \qquad \text{(vector addition in } \mathbb{Z}_2^4)$$

Proof. By agreement of $\dot\vee$ with $+_2$. For all r, s:

$$(f_1 \dot\vee f_2)(r, s) = f_1(r, s) \dot\vee f_2(r, s) = f_1(r, s) +_2 f_2(r, s).$$

□

Corollary 1.1.9 *The boolean functions $f \in \mathbb{B}^2$ that are generated from $r, s, \dot\vee$ all satisfy*

$$\underline{f} \in C_+ = S(E_+) = \big\{\underline{f} \colon f(0,0) + f(0,1) + f(1,0) + f(1,1) = 0\big\}.$$

Here $E_+ \colon x_1 + x_2 + x_3 + x_4 = 0$ is the same homogeneous linear equation considered for the parity check bit above.

Conjunction $r \wedge s$, for instance, cannot be generated from $r, s, \dot\vee$.

[6] This set of functions is therefore said to be expressively complete; so are for instance also r, s with negation and conjunction.

Proof. We show that all functions that can be generated from $r, s, \dot\vee$ satisfy the condition by showing the following:

(i) the basic functions r and s satisfy the condition;

(ii) if some functions f_1 and f_2 satisfy the condition then so does $f_1 \dot\vee f_2$.

This implies that no functions generated from $r, s, \dot\vee$ can break the condition – what we want to show. [7]

Check (i): $\underline{r} = (0,0,1,1) \in C_+$; $\underline{s} = (0,1,0,1) \in C_+$.

(ii) follows from Lemma 1.1.8 as C_+ is closed under $+_2^4$, being the solution set of a homogeneous linear equation (compare Ecercise 1.1.7).

For the assertion about conjunction, observe that $\underline{\wedge} = (0,0,0,1) \notin C_+$. □

Exercise 1.1.8 Show that the functions generated by $r, s, \dot\vee$ do not even cover all of C_+ but only a smaller subspace of $\mathbb{Z}_2^4$. Find a second linear equation over $\mathbb{Z}_2^4$ that together with E_+ precisely characterises those $\underline{f}$ which are generated from $r, s, \dot\vee$.

Exercise 1.1.9 The subset of $\mathbb{B}^2$ generated from $r, s, 0, 1, \dot\vee, \neg$ (allowing the constant functions and negation as well) is still strictly contained in $\mathbb{B}^2$; conjunction is still not expressible. In fact the set of functions thus generated precisely corresponds to the subspace associated with $C_+ \subseteq \mathbb{Z}_2^4$.

1.2 Basics, Notation and Conventions

Note: This section is intended as a glossary of terms and basic concepts to turn to as they arise in context; and not so much to be read in one go.

1.2.1 Sets

Sets [Mengen] are unstructured collections of objects (the *elements* [Elemente] of the set) without repetitions. In the simplest case a set is denoted by an enumeration of its elements, inside *set brackets*. For instance, $\{0,1\}$ denotes the set whose only two elements are 0 and 1.

[7] This is a proof by a variant of the principle of proof by induction which works not just over $\mathbb{N}$. Compare section 1.2.6.

Some important standards sets:

$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$	the set of natural numbers [8]	[natürliche Zahlen]
$\mathbb{Z}$	the set of integers	[ganze Zahlen]
$\mathbb{Q}$	the set of rationals	[rationale Zahlen]
$\mathbb{R}$	the set of reals	[relle Zahlen]
$\mathbb{C}$	the set of complex numbers	[komplexe Zahlen]

Membership, set inclusion, set equality

For a set A: $a \in A$ (a is an element of A). $a \notin A$ is an abbreviation for "not $a \in A$", just as $a \neq b$ is shorthand for "not $a = b$".

$\emptyset$ denotes the *empty set* [leere Menge], so that $a \notin \emptyset$ for any a.

$B \subseteq A$ (B is a *subset* [Teilmenge] of A) if for all $a \in B$ we have $a \in A$. For instance, $\emptyset \subseteq \{0, 1\} \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$.

Two sets are *equal*, $A_1 = A_2$, if and only if they have precisely the same elements (for all a: $a \in A_1$ if and only if $a \in A_2$). It is often useful to test for equality via: $A_1 = A_2$ if and only if both $A_1 \subseteq A_2$ and $A_2 \subseteq A_1$.

The *strict subset* relation $A \subsetneq B$ says that $A \subseteq B$ and $A \neq B$, equivalently: $A \subseteq B$ and not $B \subseteq A$. ([9])

Set operations: intersection, union, difference, and products

The following are the most common (boolean operations) on sets.

Intersection [Durchschnitt] of sets, $A_1 \cap A_2$. The elements of $A_1 \cap A_2$ are precisely those that are elements of both A_1 and A_2.

Union [Vereinigung] of sets , $A_1 \cup A_2$. The elements of $A_1 \cup A_2$ are precisely those that are elements of at leat one of A_1 or A_2.

Set difference [Differenz], $A_1 \setminus A_2$, is defined to consist of precisely those $a \in A_1$ that are not elements of A_2.

[8] Note that we regard 0 as a natural number; there is a competing convention according to which it is not. It does not really matter but one has to be aware of the convention that is in effect.

[9] The subset symbol without the horizontal line below is often used in place of our $\subseteq$, but occasionally also to denote the strict subset relation. We here try to avoid it.

Cartesian products and tuples

Cartesian products provide sets of tuples. The simplest case of tuples is that of *(ordered) pairs* [(geordnete) Paare]. (a, b) is the ordered pair whose first component is a and whose second component is b. Two ordered pairs are equal iff they agree in both components:

$$(a, b) = (a', b') \quad \text{iff} \quad a = a' \text{ and } b = b'.$$

One similarly defines *n-tuples* [n-Tupel] $(a_1, \ldots, a_n)$ with n components for any $n \geqslant 2$. For some small n these have special names, namely pairs $(n = 2)$, triples $(n = 3)$, etc.

The *cartesian product* [Kreuzprodukt] of two sets, $A_1 \times A_2$, is the set of all ordered pairs (a_1, a_2) with $a_1 \in A_1$ and $a_2 \in A_2$.

Multiple cartesian products $A_1 \times A_2 \times \cdots \times A_n$ are similarly defined. The elements of $A_1 \times A_2 \times \cdots \times A_n$ are the n-tuples whose i-th components are elements of A_i, for $i = 1, \ldots, n$.

In the special case that the cartesian product is built from the same set A for all its components, one writes A^n for the set of n-tuples over A instead of $\underbrace{A \times A \times \cdots \times A}_{n \text{ times}}$.

Defined subsets New sets are often defined as subsets of a given sets. If p states a property that elements $a \in A$ may or may not have, then $B = \{a \in A \colon p(a)\}$ denotes the subset of A that consists of precisely those elements of A that do have property p. For instance, $\{n \in \mathbb{N} \colon 2 \text{ divides } n\}$ is the set of even natural numbers, or $\{(x, y) \in \mathbb{R}^2 \colon ax + by = c\}$ is the solution set of the linear equation $E \colon ax + by = c$.

1.2.2 Functions

Functions [Funktionen] (or maps [Abbildungen]) are the next most fundamental objects of mathematics after sets. Intuitively a function maps elements of one set (the *domain* [Definitionsbereich] of the function) to elements of another set (the *range* [Wertebereich] of the function). A function f thus prescribes for every element a of its domain precisely one element $f(a)$ in the range. The full specification of a function f therefore has three parts:

(i) the domain, a set $A = \mathrm{dom}(f)$,

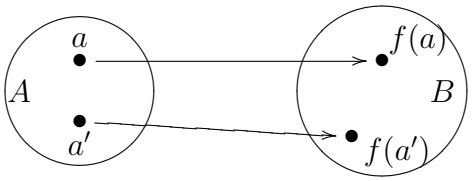
(ii) the range, a set $B = \mathrm{range}(f)$,

(iii) the association of precisely one $f(a) \in B$ with every $a \in A$.

Standard notation is as in

$$\begin{array}{rcl} f\colon A & \longrightarrow & B \\ a & \longmapsto & f(a) \end{array}$$

where the first line specifies sets A and B as domain and range, respectively, and the second line specifies the mapping prescription for all $a \in A$. For instance, $f(a)$ may be given as an arithmetical term. Any other description that *uniquely* determines an element $f(a) \in B$ for every $a \in A$ is admissible.



$f(a)$ is the *image of* a under f [Bild];
a is a *pre-image of* $b = f(a)$ [Urbild].

Examples

$$\begin{array}{rcl} \mathrm{id}_A\colon A & \longrightarrow & A \\ a & \longmapsto & a \end{array}$$
identity function on A

$$\begin{array}{rcl} \mathrm{succ}\colon \mathbb{N} & \longrightarrow & \mathbb{N} \\ n & \longmapsto & n+1 \end{array}$$
successor function on $\mathbb{N}$

$$\begin{array}{rcl} +\colon \mathbb{N}\times\mathbb{N} & \longrightarrow & \mathbb{N} \\ (n,m) & \longmapsto & n+m \end{array}$$
natural number addition ([10])

$$\begin{array}{rcl} \mathrm{prime}\colon \mathbb{N} & \longrightarrow & \{0,1\} \\ n & \longmapsto & \begin{cases} 1 & \text{if } n \text{ is prime} \\ 0 & \text{else} \end{cases} \end{array}$$
characteristic function of the set of primes

$$\begin{array}{rcl} f_{(a_1,\ldots,a_n)}\colon \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ (x_1,\ldots,x_n) & \longmapsto & a_1x_1+\cdots+a_nx_n \end{array}$$
a linear function

With a function $f\colon A \to B$ we associate its *image set* [Bildmenge]

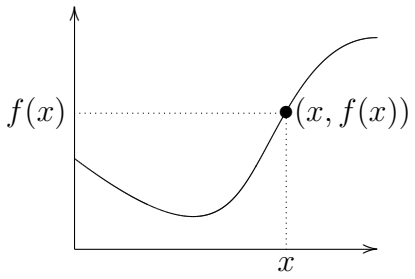
$$\mathrm{image}(f) = \big\{ f(a)\colon a \in A \big\} \subseteq B,$$

consisting of all the images of elements of A under f.

The actual association between $a \in \mathrm{dom}(f)$ and $f(a) \in \mathrm{range}(f)$ prescribed in f is often best visualised in term of the set of all pre-image/image pairs, called the *graph* [Graph] of f:

$$G_f = \big\{ (a, f(a))\colon a \in A \big\} \subseteq A \times B.$$

Exercise 1.2.1 Which properties must a subset $G \subseteq A \times B$ have in order to be the graph of *some* function $f\colon A \to B$?



Surjections, injections, bijections

Definition 1.2.1 A function $f\colon A \to B$ is said to be

(i) *surjective* (*onto*) if $\mathrm{image}(f) = B$,

(ii) *injective* (*one-to-one*) if for all $a, a' \in A$: $f(a) = f(a') \Rightarrow a = a'$,

(iii) *bijective* if it is injective and surjective.

Correspondingly, the function f is called a *surjection*, an *injection*, or a *bijection*, respectively.

Exercise 1.2.2 Classify the above example functions in these terms.

[10] Functions that are treated as binary operations like $+$, are sometimes more naturally written with the function symbol between the arguments, as in $n+m$ rather than $+(n, m)$, but the difference is purely cosmetic.

Note that injectivity of f means that every $b \in \mathrm{range}(f)$ has *at most one* pre-image; surjectivity says that it has *at least one*, and bijectivity says that it has *precisely one* pre-image.

Bijective functions $f\colon A \to B$ play a special role as they precisely translate one set into another at the element level. In particular, the existence of a bijection $f\colon A \to B$ means that A and B have the same size. [This is the basis of the set theoretic notion of *cardinality* of sets, applicable also for infinite sets.]

If $f\colon A \to B$ is bijective, then the following is a well-defined function

$$\begin{array}{rcl} f^{-1}\colon B & \longrightarrow & A \\ b & \longmapsto & \text{the } a \in A \text{ with } f(a) = b. \end{array}$$

f^{-1} is called the *inverse function* of f [Umkehrfunktion].

Composition

If $f\colon A \to B$ and $g\colon B \to C$ are functions, we may define their composition

$$\begin{array}{rcl} g \circ f\colon A & \longrightarrow & C \\ a & \longmapsto & g(f(a)) \end{array}$$

We read "$g \circ f$" as "g after f".

Note that in the case of the inverse f^{-1} of a bijection $f\colon A \to B$, we have

$$f^{-1} \circ f = \mathrm{id}_A \quad \text{and} \quad f \circ f^{-1} = \mathrm{id}_B.$$

Exercise 1.2.3 Find examples of functions $f\colon A \to B$ that have no inverse (because they are not bijective) but admit some $g\colon B \to A$ such that either $g \circ f = \mathrm{id}_A$ or $f \circ g = \mathrm{id}_B$.

How are these conditions related to injectivity/surjectivity of f?

Permutations

A *permutation* [Permutation] is a bijective function of the form $f\colon A \to A$. Their action on the set A may be viewed as a re-shuffling of the elements, hence the name permutation. Because they are bijective (and hence invertible) and lead from A back into A they have particularly nice composition properties.

In fact the set of all permutations of a fixed set A, together with the composition operation $\circ$ forms a group, which means that it satisfies the laws (G1-3) collected below (compare (V1-3) above and section 1.3.2).

For a set A, let $\mathrm{Sym}(A)$ be the set of permutations of A

$$\mathrm{Sym}(A) = \big\{ f\colon\ f \text{ a bijection from } A \text{ to } A \big\},$$

equipped with the composition operation

$$\begin{array}{rcl} \circ\colon \mathrm{Sym}(A) \times \mathrm{Sym}(A) & \longrightarrow & \mathrm{Sym}(A) \\ (f_1, f_2) & \longmapsto & f_1 \circ f_2. \end{array}$$

Exercise 1.2.4 Check that $(\mathrm{Sym}(A), \circ, \mathrm{id}_A)$ satisfies the following laws:

G1 (associativity)
For all $f, g, h \in \mathrm{Sym}(A)$:

$$f \circ \big(g \circ h\big) = \big(f \circ g\big) \circ h.$$

G2 (neutral element)
id_A is a neutral element w.r.t. $\circ$, i.e., for all $f \in \mathrm{Sym}(A)$:

$$f \circ \mathrm{id}_A = \mathrm{id}_A \circ f = f.$$

G3 (inverse elements)
For every $f \in \mathrm{Sym}(A)$ there is an $f' \in \mathrm{Sym}(A)$ such that

$$f \circ f' = f' \circ f = \mathrm{id}_A.$$

Give an example to show that $\circ$ is not commutative.

Definition 1.2.2 For $n \geqslant 1$, $S_n := (\mathrm{Sym}(\{1, \ldots, n\}), \circ, \mathrm{id}_{\{1,\ldots,n\}})$, the group of all permutations of an n element set, with composition, is called the *symmetric group* [symmetrische Gruppe] of n elements.

Common notation for $f \in S_n$ is $f = \begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix}$.

Exercise 1.2.5 What is the size of S_n, i.e., how many different permutations does a set of n elements have?
List all the elements of S_3 and compile the table of the operation $\circ$ over S_3.

1.2.3 Relations

An r-ary relation [Relation] R over a set A is a collection of r-tuples $(a_1, \ldots, a_r)$ over A, i.e., a subset $R \subseteq A^r$. For binary relations $R \subseteq A^2$ one often uses notation aRa' instead of $(a, a') \in R$.

For instance, the natural order relation $<$ is a binary relation over $\mathbb{N}$ consisting of all pairs (n, m) with $n < m$.

The graph of a function $f\colon A \to A$ is a binary relation.

One may similarly consider relations across different sets, as in $R \subseteq A \times B$; for instance the graph of a function $f\colon A \to B$ is a relation in this sense.

Equivalence relations

These are a particularly important class of binary relations. A binary relation $R \subseteq A^2$ is an equivalence relation [Äquivalenzrelation] over A iff it is

(i) *reflexive* [reflexiv]; for all $a \in A$, $(a, a) \in R$.

(ii) *symmetric* [symmetrisch]; for all $a, b \in A$: $(a, b) \in R$ iff $(b, a) \in R$.

(iii) *transitive* [transitiv]; for all $a, b, c \in A$: if $(a, b) \in R$ and $(b, c) \in R$, then $(a, c) \in R$.

Examples: equality (over any set); having the same parity (odd or even) over $\mathbb{Z}$; being divisible by exactly the same primes, over $\mathbb{N}$.

Non-examples: $\leqslant$ on $\mathbb{N}$; having absolute difference less than 5 over $\mathbb{N}$.

Exercise 1.2.6 Let $f\colon A \to B$ be a function. Show that the following is an equivalence relation:

$$R_f := \big\{(a, a') \in A^2 \colon f(a) = f(a')\big\}.$$

Exercise 1.2.7 Consider the following relationship between arbitrary sets A, B: $A \sim B$ if there exists some bijection $f\colon A \to B$. Show that this relationship has the properties of an equivalence relation: $\sim$ is reflexive, symmetric and transitive.

Any equivalence relation over a set A partitions A into *equivalence classes*. Let R be an equivalence relation over A. The R-equivalence class of $a \in A$ is the subset

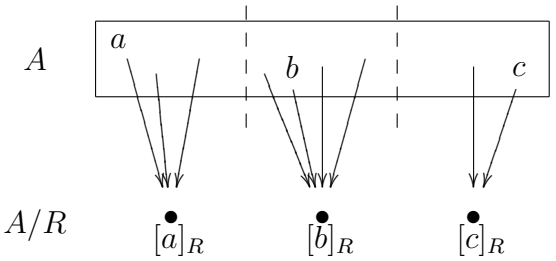
$$[a]_R = \big\{a' \colon (a, a') \in R\big\}.$$

Exercise 1.2.8 Show that for any two equivalence classes $[a]_R$ and $[a']_R$, either $[a]_R = [a']_R$ or $[a]_R \cap [a']_R = \emptyset$, and that A is the disjoint union of its equivalence classes w.r.t. R.

The set of equivalence classes is called the *quotient* of the unerlying set A w.r.t. the equivalence relation R, denoted A/R:

$$A/R = \big\{[a]_R \colon a \in A\big\}.$$

The function $\pi_R \colon A \to A/R$ that maps each element $a \in A$ to its equivalence class $[a]_R \in A/R$ is called the *natural projection*. Note that $(a, a') \in R$ iff $[a]_R = [a']_R$ iff $\pi_R(a) = \pi_R(a')$. Compare the diagram for an example of an equivalence relation with 3 classes that are represented, e.g., by the pairwise inequivalent elements $a, b, c \in A$:



For the following also compare section 1.3.2.

Exercise 1.2.9 Consider the equivalence relation $\equiv_n$ on $\mathbb{Z}$ defined as

$$a \equiv_n b \quad \text{iff} \quad a = kn + b \quad \text{for some } k \in \mathbb{Z}.$$

Integers a and b are equivalent in this sense if their difference is divisible by n, or if they leave the same remainder w.r.t. division by n.

Check that $\equiv_n$ is an equivalence relation over $\mathbb{Z}$, and that every equivalence class has a unique member in $\mathbb{Z}_n = \{0, \ldots, n-1\}$.

Show that addition and multiplication in $\mathbb{Z}$ operate class-wise, in the sense that for all $a \equiv_n a'$ and $b \equiv_n b'$ we also have $a + b \equiv_n a' + b'$ and $ab \equiv_n a'b'$.

1.2.4 Summations

As we deal with (vector and scalar) sums a lot, it is useful to adopt the usual concise summation notation. We write for instance

$$\sum_{1}^{n} a_i x_i \quad \text{for} \quad a_1 x_1 + \cdots + a_n x_n.$$

Relaxed variants of the general form $\sum_{i \in I} a_i$ where I is some index set that idexes a *family* $(a_i)_{i \in I}$ of terms to be summed up are useful. (In our usage of such notation, I has to be finite or at least only finitely many $a_i \neq 0$.) This convention implicitly appeals to associativity and commutativity of the underlying addition operation (why?).

Similar conventions apply to other associative and commutative operations, in particular set union and intersection (and here finiteness of the index set is not essential). For instance $\bigcup_{i \in I} A_i$ stands for the union of all the sets A_i for $i \in I$.

1.2.5 Propositional logic

We here think of propositions as assertions [Aussagen] about mathematical objects, and are mostly interested in (determining) their truth or falsity.

Typically propositions are structured, and composed from simpler propositions according to certain logical composition operators. Propositional logic [Aussagenlogik], and the standardised use of the propositional connectives comprising negation, conjunction and disjunction, plays an important role in mathematical arguments.

If A is an assertion then $\neg A$ (not A [nicht A]) stands for the *negation* [Negation] A and is true exactly when A itself is false and vice versa.

For assertions A and B, $A \wedge B$ (A and B [A und B]) stands for their *conjuction* [Konjunktion], which is true precisely when both A and B are true.

$A \vee B$ (A or B [A oder B]) stands for their *disjunction* [Disjunktion], which is true precisely when at least one of A and B is true.

The standardised semantics of these basic logical operators, and other derived ones, can be described in terms of truth tables. Using the boolean values 0 and 1 as truth values, 0 for *false* and 1 for *true*, the truth table for a

logical operator specifies the truth value of the resulting proposition in terms of the truth values for the component propositions. [11]

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$
0	0	1	0	0	1	1
0	1	1	0	1	1	0
1	0	0	0	1	0	0
1	1	0	1	1	1	1

The implication [Implikation] $A \Rightarrow B$ (A implies B [A impliziert B] is true unless A (the premise or assumption) is true and B (the conclusion) is false. The truth table therefore is the same as for $\neg A \vee B$.

The *equivalence* or *bi-implication* [Äquivalenz], $A \Leftrightarrow B$ (A is equivalent [äquivalent] with B, A if and and only if B [A genau dann wenn B]), is true precisely if A and B have the same truth value.

It is common usage to write and read a bi-implication as A *iff* B, where "iff" abbreviates "if and only if". [gdw: genau dann wenn]

We do not give any formal account of quantification, and only treat the symbolic quantifiers $\forall$ and $\exists$ as occasional shorthand notation for "for all" and "there exists", as in $\forall x, y \in \mathbb{R}\ (x + y = y + x)$.

1.2.6 Some common proof patterns

Implications It is important to keep in mind that an implication $A \Rightarrow B$ is proved if we establish that B must hold whenever A is true (think of A as the assumption, of B as the conclusion claimed *under this assumption*). No claim at all is made about settings in which (the assumption) A fails!

To prove an implication $A \Rightarrow B$, one can either assume A and establish B, or equivalently assume $\neg B$ (the negation of B) and work towards $\neg A$; the justification of the latter lies in the fact that $A \Rightarrow B$ is logically equivalent with $\neg B \Rightarrow \neg A$ (you might want to check the truth tables.)

[11] Note that these precise formal conventions capture some aspects of the natural everyday usage of "and" and "or", or "if ..., then ...", but not all. In particular, the truth or falsehood of a natural language composition may depend not just on the truth values of the component assertions but also on context. The standardised interpretation may therefore differ from your intuitive understanding at least in certain contexts.

Implications often occur as chains of the form

$$A \Rightarrow A_1,\ A_1 \Rightarrow A_2,\ A_2 \Rightarrow A_3, \ldots, A_n \Rightarrow B,$$

or $A \Rightarrow A_1 \Rightarrow A_2 \Rightarrow \ldots \Rightarrow B$ for short. The validity of (each step in) the chain then also implies the validity of $A \Rightarrow B$. Indeed, one often constructs a chain of intermediate steps in the course of a proof of $A \Rightarrow B$.

Indirect proof In order to prove A, it is sometimes easier to work indirectly, by showing that $\neg A$ leads to a contradiction. Then, as $\neg A$ is seen to be an impossibility, A must be true.

Bi-implications or equivalences To establish an equivalence $A \Leftrightarrow B$ one often shows separately $A \Rightarrow B$ and $B \Rightarrow A$. Chains of equivalences of the form

$$A \Leftrightarrow A_1,\ A_1 \Leftrightarrow A_2,\ A_2 \Leftrightarrow \ldots \Leftrightarrow B,$$

or $A \Leftrightarrow A_1 \Leftrightarrow A_2 \Leftrightarrow \ldots \Leftrightarrow B$ for short, may allow us to establish $A \Leftrightarrow B$ through intermediate steps. Another useful trick for establishing an equivalence between several assertions, say A, B and C for instance, is to prove a circular chain of one-sided implications, for instance

$$A \Rightarrow C \Rightarrow B \Rightarrow A$$

that involves all the assertions in some order that facilitates the proof.

Induction [vollständige Induktion] Proofs by induction are most often used for assertions $A(n)$ parametrised by the natural numbers $n \in \mathbb{N}$. In order to show that $A(n)$ is true for all $n \in \mathbb{N}$ one establishes

(i) the truth of $A(0)$ (the base case),

(ii) the validity of the implication $A(n) \Rightarrow A(n+1)$ in general, for all $n \in \mathbb{N}$ (the induction step).

As any individual natural number m is reached from the first natural number 0 in finitely many successor steps from n to $n+1$, $A(m)$ is established in this way via a chain of implications that takes us from the base case $A(0)$ to $A(m)$ via a number of applications of the induction step.

There are many variations of the technique. In particular, in order to prove $A(n+1)$ one may assume not just $A(n)$ but all the previous instances $A(0), \ldots, A(n)$ without violating the validity of the principle.

But also beyond the domain of natural numbers similar proof principles can be used, whenever the domain in question can similarly be generated from some basic instances via some basic construction steps ("inductive data types"). If A is true of the basic instances, and the truth of A is preserved in each construction step, then A must be true of all the objects that can be constructed in this fashion. We saw a simple example of this more general idea of induction in the proof of Corollary 1.1.9.

1.3 Algebraic Structures

In the most general case, an algebraic structure consist of a set (the domain of the structure) equipped with some operations, relations and distinguished elements (constants) over that set. A typical example is $(\mathbb{N}, +, <, 0)$ with domain $\mathbb{N}$, addition operation $+$, order relation $<$ and constant 0.

1.3.1 Binary operations on a set

A binary operation $*$ on a set A is a function

$$\begin{array}{rcl} *\colon A \times A & \longrightarrow & A \\ (a, a') & \longmapsto & a * a', \end{array}$$

where we write $a * a'$ rather than $*(a, a')$.

The operation $*$ is *associative* [assoziativ] iff for all $a, b, c \in A$:

$$a * (b * c) = (a * b) * c.$$

For associative operations, we may drop parentheses and write $a * b * c$ because precedence does not matter.

The operation $*$ is *commutative* [kommutativ] iff for all $a, b \in A$:

$$a * b = b * a.$$

$e \in A$ is a *neutral element* [neutrales Element] w.r.t. $*$ iff for all $a \in A$

$$a * e = e * a = a. \qquad ([12])$$

[12] An element just satisfying $a * e$ for all $a \in A$ is called a right-neutral element; similarly there are left-neutral elements; a neutral element as defined above is both, left- and right-neutral. Similarly for left and right inverses.

Examples of structures with an associative operation with a neutral element: $(\mathbb{N}, +, 0)$, $(\mathbb{N}, \cdot, 1)$, $(S_n, \circ, \mathrm{id})$; the first two operations are commutative, composition in S_n is not for $n \geqslant 3$.

Note that a neutral element, if any, is unique (why?).

If $*$ has neutral element e, then the element a' is called an *inverse* [inverses Element] of a w.r.t. $*$ iff

$$a * a' = a' * a = e. \qquad (^{12})$$

For instance, in $(\mathbb{N}, +, 0)$, 0 is the only element that has an inverse, while in $(\mathbb{Z}, +, 0)$, every element has an inverse.

Observation 1.3.1 *For an associative operation $*$ with neutral element e: if a has an inverse w.r.t. $*$, then this inverse is unique.*

Proof. Let a' and a'' be inverses of a: $a * a' = a' * a = a * a'' = a'' * a = e$. Then $a' = a' * e = a' * (a * a'') = (a' * a) * a'' = e * a'' = a''$. □

In additive notation one usually writes $-a$ for the inverse of a w.r.t. $+$, in multiplicative notation a^{-1} for the inverse w.r.t. $\cdot$.

1.3.2 Groups

An algebraic structure $(A, *, e)$ with binary operation $*$ and distinguished element e is a *group* [Gruppe] iff the following axioms (G1-3) are satisfied:

G1 (associativity): $*$ is associative.
For all $a, b, c \in A$: $a * (b * c) = (a * b) * c$.

G2 (neutral element): e is a neutral element w.r.t. $*$.
For all $a \in A$: $a * e = e * a = a$.

G3 (inverse elements): $*$ has inverses for all $a \in A$.
For every $a \in A$ there is an $a' \in A$: $a * a' = a' * a = e$.

A group with a commutative operation $*$ is called an *abelian* or *commutative group*. For these we have the additional axiom

G4 (commutativity): $*$ is commutative.
For all $a, b \in A$: $a * b = b * a$.

Observation 1.3.2 *Let $(A, *, e)$ be a group. For any $a, b, c \in A$:*
$a * c = b * c \;\Rightarrow\; a = b$.

Proof. Let $a * c = b * c$ and c' the inverse of c.
Then $a = a * e = a * (c * c') = (a * c) * c' = (b * c) * c' = b * (c * c') = b * e = b$. □

Examples of groups

Familiar examples of abelian groups are the additive groups over the common number domains $(\mathbb{Z}, +, 0)$, $(\mathbb{Q}, +, 0)$, $(\mathbb{R}, +, 0)$, or the additive group of vector addition over $\mathbb{R}^n$, $(\mathbb{R}^n, +, \mathbf{0})$. Further also the multiplicative groups over some of the common number domains without 0 (as it has no multiplicative inverse), as in $(\mathbb{Q}^*, \cdot, 1)$ and $(\mathbb{R}^*, \cdot, 1)$ where $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$ and $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$.

As examples of a non-abelian groups we have seen the symmetric groups S_n (non-abelian for $n \geqslant 3$), see 1.2.2.

Modular arithmetic and $\mathbb{Z}_n$

For $n \geqslant 2$ let $\mathbb{Z}_n = \{0, \ldots, n-1\}$ consist of the first n natural numbers.

Addition and multiplication of integers over $\mathbb{Z}$ induce operations of addition and multiplication over $\mathbb{Z}_n$, which we denote $+_n$ and $\cdot_n$ at first, via passage to remainders w.r.t. division by n. (Also compare Exercise 1.2.9.) For $a, b \in \mathbb{Z}_n$ put

$$\begin{aligned} a +_n b &:= \text{the remainder of } a + b \text{ w.r.t. division by } n. \\ a \cdot_n b &:= \text{the remainder of } ab \text{ w.r.t. division by } n. \end{aligned}$$

As an example we provide the tables for $+_4$ and $\cdot_4$ over $\mathbb{Z}_4$:

$+_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot_4$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Exercise 1.3.1 Check that $(\mathbb{Z}_4, +_4, 0)$ is an abelian group. Why does the operation $\cdot_4$ fail to form a group on $\mathbb{Z}_4$ or over $\mathbb{Z}_4 \setminus \{0\}$?

It is a fact from elementary number theory that for $a, b \in \mathbb{Z}$ the equation

$$ax + by = 1$$

has an integer solution for x and y if (and only if) a and b are relatively prime (their greatest common divisor is 1). If n is prime, therefore, the equation $ax + ny = 1$ has an integer solution for x and y for all $a \in \mathbb{Z}_n \setminus \{0\}$. But then x is an inverse w.r.t. $\cdot_n$ for a, since $ax + nk = 1$, for any integer k, means that ax leaves remainder 1 w.r.t. division by n.

It follows that for any prime p, $(\mathbb{Z}_p \setminus \{0\}, \cdot_p, 1)$ is an abelian group.

1.3.3 Rings and fields

Rings [Ringe] and fields [Körper] are structures of the format $(A, +, \cdot, 0, 1)$ with two binary operations $+$ and $\cdot$, and two distinguished elements 0 and 1.

Rings $(A, +, \cdot, 0, 1)$ is a *ring* if $(A, +, 0)$ is an abelian group, $\cdot$ is associative with neutral element 1 and the following *distributivity* laws [Distributivgesetze] are satisfied for all $a, b, c \in A$:

$$\begin{aligned} (a + b) \cdot c &= (a \cdot c) + (b \cdot c) \\ c \cdot (a + b) &= (c \cdot a) + (c \cdot b) \end{aligned}$$

A *commutative ring* is one with commutative multiplication operation $\cdot$.

One often adopts the convention that $\cdot$ takes precedence over $+$ in the absence of parentheses, so that $a \cdot c + b \cdot c$ stands for $(a \cdot c) + (b \cdot c)$.

Observation 1.3.3 *In any ring* $(A, +, \cdot, 0, 1)$, $0 \cdot a = a \cdot 0 = 0$ *for all* $a \in A$.

Proof. For instance, $a + (0 \cdot a) = 1 \cdot a + 0 \cdot a = (1 + 0) \cdot a = 1 \cdot a = a$. So $a + (0 \cdot a) = a + 0$, and $0 \cdot a = 0$ follows with Observation 1.3.2. □

Exercise 1.3.2 $(\mathbb{Z}_n, +, \cdot, 0, 1)$ is a commutative ring for any $n \geqslant 2$.

Fields A *field* [Körper] is a commutative ring $(A, +, \cdot, 0, 1)$ with $0 \neq 1$ and in which all $a \neq 0$ have inverses w.r.t. multiplication.

Familiar examples are the fields of rational numbers $(\mathbb{Q}, +, \cdot, 0, 1)$, the field of real numbers $(\mathbb{R}, +, \cdot, 0, 1)$, and the field of complex numbers (see below).

It follows from our considerations about modular arithmetic, $+_n$ and $\cdot_n$ over $\mathbb{Z}_n$, that for any prime number p, $(\mathbb{Z}_p, +_p, \cdot_p, 0, 1)$ is a field, usually denoted $\mathbb{F}_p$. Compare Exercise 1.1.5 for the case of $\mathbb{F}_2$.

The field of complex numbers, $\mathbb{C}$

We only give a very brief summary. As a set

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

where $i \notin \mathbb{R}$ is a "new" number, whose arithmetical role will become clear when we set $i^2 = i \cdot i = -1$ in complex multiplication.

We regard $\mathbb{R} \subseteq \mathbb{C}$ via the natural identification of $r \in \mathbb{R}$ with $r + 0i \in \mathbb{C}$. Similarly, i is identified with $0 + 1i$. The numbers λi for $\lambda \in \mathbb{R}$ are called *imaginary numbers* [imaginäre Zahlen], and a complex number $a + bi$ is said to have *real part* [Realteil] a and *imaginary part* [Imaginärteil] b.

The operation of addition over $\mathbb{C}$ corresponds to vector addition over $\mathbb{R}^2$ if we associate the complex number $a + bi \in \mathbb{C}$ with $(a, b) \in \mathbb{R}^2$:

$$(a_1 + b_1 i) + (a_2 + b_2 i) := (a_1 + a_2) + (b_1 + b_2)i.$$

Exercise 1.3.3 Check that $(\mathbb{C}, +, 0)$ is an abelian group.

The operation of multiplication over $\mathbb{C}$ is made to extend multiplication over $\mathbb{R}$, to satisfy distributivity and to make $i^2 = -1$. This leads to

$$(a_1 + b_1 i) \cdot (a_2 + b_2 i) := (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i.$$

Exercise 1.3.4 Check that $(\mathbb{C}, +, \cdot, 0, 1)$ is a field.

Over the field of complex numbers, any non-trivial polynomial has a zero [Fundamentalsatz der Algebra]. While the polynomial equation $x^2 + 1 = 0$ admits no solution over $\mathbb{R}$, $\mathbb{C}$ has been extended (in a minimal way) to provide solutions i and $-i$; but in adjoining this one extra number i, one in fact obtains a field over which any non-trivial polynomial equation is solvable (in technical terms: the field of complex numbers is algebraically closed).

1.3.4 Aside: isomorphisms of algebraic structures

An isomorphism [Isomorphismus] is a structure preserving bijection between (algebraic) structures. If there is an isomorphism between two structures, we say they are isomorphic. Isomorphic structures may be different (for instance have distinct domains) but they cannot be distinguished on structural grounds, and from a mathematical point of view one might as well ignore the difference.

For instance, two structures with a binary operation and a distinguished element (constant), $(A, *^A, e^A)$ and $(B, *^B, e^B)$ are *isomorphic* if there is an isomorphism between them, which in this case is a map

$$\begin{array}{lll} \varphi\colon A \longrightarrow B & \text{such that} & \\ & \varphi \text{ is bijective} & \\ & \varphi \text{ preserves } e\colon & \varphi(e^A) = e^B \\ & \varphi \text{ preserves } *\colon & \varphi(a *^A a') = \varphi(a) *^B \varphi(a') \text{ for all } a, a' \in A \end{array}$$

The diagram illustrates the way in which φ translates between $*^A$ and $*^B$, where $b = \varphi(a), b' = \varphi(a')$:

$$\begin{array}{ccc} (a, a') & \xrightarrow{\quad *^A \quad} & a *^A a' \\ \varphi \downarrow \; \downarrow \varphi & & \downarrow \varphi \\ (b, b') & \xrightarrow{\quad *^B \quad} & b *^B b' \end{array}$$

It may be instructive to verify that the existence of an isomorphism between $(A, *^A, e^A)$ and $(B, *^B, e^B)$ implies, for instance, that $(A, *^A, e^A)$ is a group if and only if $(B, *^B, e^B)$ is.

Exercise 1.3.5 Consider the additive group $(\mathbb{Z}_2^4, +, \mathbf{0})$ of vector addition in $\mathbb{Z}_2^4$ and the algebraic structure $(\mathbb{B}^2, \dot\vee, 0)$ where (as in section 1.1.4) $\mathbb{B}^2$ is the set of all $f\colon \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_2$, $0 \in \mathbb{B}^2$ is the constant function 0 and $\dot\vee$ operates on two functions in $\mathbb{B}^2$ by combining them with *xor*. Lemma 1.1.8 essentially says that our mapping $\varphi : f \longmapsto \underline{f}$ is an isomorphism between $(\mathbb{B}^2, \dot\vee, 0)$ and $(\mathbb{Z}_2^4, +, \mathbf{0})$. Fill in the details.

Exercise 1.3.6 Show that the following is an isomorphism between $(\mathbb{R}^2, +, \mathbf{0})$ and $(\mathbb{C}, +, 0)$:

$$\begin{array}{rcl} \varphi : \mathbb{R}^2 & \longrightarrow & \mathbb{C} \\ (a, b) & \longmapsto & a + bi. \end{array}$$

Exercise 1.3.7 Show that the symmetric group of two elements $(S_2, \circ, \mathrm{id})$ is isomorphic to $(\mathbb{Z}_2, +, 0)$.

Exercise 1.3.8 Show that there are two essentially different, namely non-isomorphic, groups with four elements. One is $(\mathbb{Z}_4, +_4, 0)$ (see section 1.3.2). So the task is to design a four-by-four table of an operation that satisfies the group axioms and behaves differently from that of $\mathbb{Z}_4$ modular arithmetic so that they cannot be isomorphic.

Isomorphisms within one and the same structure are called *automorphisms*; these correspond to permutations of the underlying domain that preserve the given structure and thus to *symmetries* of that structure.

Definition 1.3.4 Let $(A, \ldots)$ be an algebraic structure with domain A (with specified operations, relations, constants depending on the format). An *automorphism* of $(A, \ldots)$ is a permutation of A that, as a map $\varphi\colon A \to A$ is an isomorphism between $(A, \ldots)$ and $(A, \ldots)$.

The set of all automorphisms of a given structure $(A, \ldots)$ forms a group, the *automorphism group* of that structure, which is a subgroup of the full permutation group $\mathrm{Sym}(A)$.

Exercise 1.3.9 Check that the automorphisms of a fixed structure $(A, \ldots)$ with domain A form a group with composition and the identity id_A as the neutral element.

Chapter 2

Vector Spaces

Vector spaces are the key notion of linear algebra. Unlike the basic algebraic structures like groups, rings or fields considered in the last section, vector spaces are *two-sorted*, which means that we distinguish two kinds of objects with different status: vectors and scalars. The vectors are the elements of the actual vector space V, but on the side we always have a field $\mathbb{F}$ as the domain of scalars. Fixing the field of scalars $\mathbb{F}$, we consider the class of $\mathbb{F}$-vector spaces – or vector spaces over the field $\mathbb{F}$. So there are $\mathbb{R}$-vector spaces (real vector spaces), $\mathbb{C}$-vector spaces (complex vector spaces), $\mathbb{F}_p$-vector spaces, etc. Since there is a large body of common material that can be covered without specific reference to any particular field, it is most natural to consider $\mathbb{F}$-vector spaces for an arbitrary field $\mathbb{F}$ at first.

2.1 Vector spaces over arbitrary fields

We fix an arbitrary field $\mathbb{F}$. We shall use no properties of $\mathbb{F}$ apart from the general consequences of the field axioms, i.e., properties shared by all fields. Scalars 0 and 1 refer to the zero and one of the field $\mathbb{F}$. For a scalar $\lambda \in \mathbb{F}$ we write $-\lambda$ for the inverse w.r.t. addition; and if $\lambda \neq 0$, λ^{-1} for the inverse w.r.t. multiplication in $\mathbb{F}$.

An $\mathbb{F}$-vector space consists of a non-empty set V of vectors, together with a binary operation of vector addition

$$\begin{array}{rcl} +\colon V \times V & \longrightarrow & V \\ (v, w) & \longmapsto & v + w, \end{array}$$

an operation of scalar multiplication

$$\begin{array}{rcll} \cdot\colon \mathbb{F} \times V & \longrightarrow & V & \\ (\lambda, v) & \longmapsto & \lambda \cdot v & \text{or just } \lambda v, \end{array}$$

and a distinguished element $\mathbf{0} \in V$ called the null vector (not to be confused with the scalar $0 \in \mathbb{F}$).

2.1.1 The axioms

The axioms themselves are those familiar from section 1.1. Only, we have removed some of the more obvious redundancies from that preliminary version, in order to get a more economic set of rules that need to be checked.

(**V1**)	for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$:	$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
(**V2**)	for all $\mathbf{v} \in V$:	$\mathbf{v} + \mathbf{0} = \mathbf{v}$
(**V3**)	for all $\mathbf{v} \in V$:	$\mathbf{v} + ((-1) \cdot \mathbf{v}) = \mathbf{0}$
(**V4**)	for all $\mathbf{u}, \mathbf{v} \in V$:	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
(**V5**)	for all $\mathbf{v}$ and all $\lambda, \mu \in \mathbb{F}$:	$\lambda \cdot (\mu \cdot \mathbf{v}) = (\lambda\mu) \cdot \mathbf{v}$
(**V6**)	for all $\mathbf{v} \in V$:	$1 \cdot \mathbf{v} = \mathbf{v}$
(**V7**)	for all $\mathbf{v} \in V$ and all $\lambda, \mu \in \mathbb{F}$:	$(\lambda + \mu) \cdot \mathbf{v} = \lambda \cdot \mathbf{v} + \mu \cdot \mathbf{v}$
(**V8**)	for all $\mathbf{u}, \mathbf{v} \in V$ and all $\lambda \in \mathbb{F}$:	$\lambda \cdot (\mathbf{u} + \mathbf{v}) = (\lambda \cdot \mathbf{u}) + (\lambda \cdot \mathbf{v})$

Definition 2.1.1 Let $\mathbb{F}$ be a field. A non-empty set V together with operations $+\colon V \times V \to V$ and $\cdot\colon \mathbb{F} \times V \to V$ and distinguished element $\mathbf{0} \in V$ is an $\mathbb{F}$-*vector space* [$\mathbb{F}$-Vektorraum] if the above axioms V1-8 are satisfied.

Note that (V1-4) say that $(V, +, \mathbf{0})$ is an abelian group; (V5/6) says that scalar multiplication is associative with $1 \in \mathbb{F}$ acting as a neutral element; V7/8 assert (two kinds of) distributivity.

We generally adopt the following conventions when working in an $\mathbb{F}$-vector space V:

(i) $\cdot$ can be dropped. E.g., $\lambda\mathbf{v}$ stands for $\lambda \cdot \mathbf{v}$.

(ii) parentheses that would govern the order of precedence for multiple $+$ or $\cdot$ can be dropped (as justified by associativity). E.g., we may write $\mathbf{u} + \mathbf{v} + \mathbf{w}$.

(iii) between vector addition and scalar multiplication, scalar multiplication has the higher precedence. E.g., we write $\lambda\mathbf{u} + \mu\mathbf{v}$ for $(\lambda\mathbf{u}) + (\mu\mathbf{v})$.

(iv) we write $\mathbf{v} - \mathbf{w}$ for $\mathbf{v} + (-1)\mathbf{w}$, and $-\mathbf{v}$ for $(-1)\mathbf{v}$.

Exercise 2.1.1 Show that, in the presence of the other axioms, (V3) is equivalent to: for all $\mathbf{v} \in V$ there is some $\mathbf{w} \in V$ such that $\mathbf{v} + \mathbf{w} = \mathbf{0}$.

The following collects some important derived rules, which are direct consequences of the axioms.

Lemma 2.1.2 *Let V be an $\mathbb{F}$-vector space. Then, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $\lambda \in \mathbb{F}$:*

(i) $\mathbf{u} + \mathbf{v} = \mathbf{u} + \mathbf{w} \;\Rightarrow\; \mathbf{v} = \mathbf{w}$.

(ii) $0\mathbf{v} = \mathbf{0}$.

(iii) $\lambda\mathbf{0} = \mathbf{0}$.

(iv) $\lambda\mathbf{v} = \mathbf{0} \;\Rightarrow\; \mathbf{v} = \mathbf{0}$ *or* $\lambda = 0$.

Proof. Ad (i): add $-\mathbf{u}$ on both sides of the first equation.

Ad (ii): $0\mathbf{v} = (1-1)\mathbf{v} = \mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$. Note that the second equality uses (V7) and (V6).

Ad (iii): for any $\mathbf{u} \in V$: $\lambda\mathbf{0} = \lambda(\mathbf{u} + (-1)\mathbf{u}) = \lambda\mathbf{u} + (-\lambda\mathbf{u}) = \lambda\mathbf{u} + (-1)\lambda\mathbf{u} = \mathbf{0}$. This uses (V3), (V8), (V5) and (V3) again.

Ad (iv): suppose $\lambda\mathbf{v} = \mathbf{0}$ and $\lambda \neq 0$. Then $\lambda^{-1}\lambda\mathbf{v} = \mathbf{v} = \lambda^{-1}\mathbf{0} = \mathbf{0}$, where the last equality uses (iii) for λ^{-1}.

□

Isomorphisms of $\mathbb{F}$-vector spaces

As for algebraic structures, the notion of isomorphism of $\mathbb{F}$-vector spaces is to capture the situation where V and W are structurally the same, as vector spaces over the same field $\mathbb{F}$.

Definition 2.1.3 Consider two $\mathbb{F}$-vector spaces V and W. We say that a map $\varphi\colon V \to W$ is a *vector space isomorphism* between V and W iff

(i) $\varphi\colon V \to W$ is a bijection.

(ii) for all $\mathbf{u}, \mathbf{v} \in V$: $\quad \varphi(\mathbf{u} + \mathbf{v}) = \varphi(\mathbf{u}) + \varphi(\mathbf{v})$.

(iii) for all $\lambda \in \mathbb{F}$, $\mathbf{v} \in V$: $\quad \varphi(\lambda\mathbf{v}) = \lambda\varphi(\mathbf{v})$.

Two $\mathbb{F}$-vector spaces are isomorphic iff there is an isomorphism between them.

In the above condition on an isomorphism, (ii) is compatibility with addition, (iii) compatibility with scalar multiplication. Check that these also imply compatibility with the null vectors, namely that $\varphi(\mathbf{0}^V) = \mathbf{0}^W$.

2.1.2 Examples old and new

Example 2.1.4 For $n \in \mathbb{N}$, let $\mathbb{F}^n$ be the set of n-tuples over $\mathbb{F}$ with component-wise addition

$$((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \longmapsto (a_1 + b_1, \ldots, a_n + b_n)$$

and scalar multiplication with $\lambda \in \mathbb{F}$ according to

$$(\lambda, (a_1, \ldots, a_n)) \longmapsto (\lambda a_1, \ldots, \lambda a_n)$$

and $\mathbf{0} = (0, \ldots, 0) \in \mathbb{F}^n$. This turns $\mathbb{F}^n$ into an $\mathbb{F}$-vector space. [The standard n-dimensional vector space over $\mathbb{F}^n$.]

We include the (degenerate) case of $n = 0$. The standard interpretation of A^0 is (irrespective of what A is) that $A^0 = \{\square\}$ has the empty tuple $\square$ as its only element. Letting $\lambda\square = \square$ and declaring $\square + \square = \square$ we find that $V = \mathbb{F}^0$ becomes a vector space whose only element $\square$ is also its null vector.

We saw the concrete examples of $\mathbb{R}^n$ and $\mathbb{Z}_2^n$ above.

With $\mathbb{Z}_p^n$ for arbitrary prime p we get more examples of vector spaces over finite fields. Can you determine the size of $\mathbb{Z}_p^n$? Since these are finite spaces their analysis is of a more combinatorial character than that of $\mathbb{R}^n$ or $\mathbb{C}^n$.

Example 2.1.5 Let A be a non-empty set, and let $\mathcal{F}(A, \mathbb{F})$ be the set of all functions $f\colon A \to \mathbb{F}$. We declare vector addition on $\mathcal{F}(A, \mathbb{F})$ by

$$f_1 + f_2 := f \quad \text{where} \quad \begin{array}{rccl} f\colon & A & \to & \mathbb{F} \\ & a & \longmapsto & f_1(a) + f_2(a) \end{array}$$

and scalar multiplication with $\lambda \in \mathbb{F}$ by

$$\lambda f := g \quad \text{where} \quad \begin{array}{rcl} g\colon A & \to & \mathbb{F} \\ a & \longmapsto & \lambda f(a). \end{array}$$

This turns $\mathcal{F}(A, \mathbb{F})$ into an $\mathbb{F}$-vector space. Its null vector is the constant function with value $0 \in \mathbb{F}$ for every $a \in A$.

This vector addition and scalar multiplication over $\mathcal{F}(A, \mathbb{F})$ is referred to as *point-wise* [punktweise] addition or multiplication.

Remark 2.1.6 Example 2.1.5 actually generalises Example 2.1.4 in the following sense. We may identify the set of n-tuples over $\mathbb{F}$ with the set of functions $\mathcal{F}(\{1, \ldots, n\}, \mathbb{F})$ via the association

$$(a_1, \ldots, a_n) \quad \leftrightharpoons \quad \left\{ \begin{array}{rcl} f\colon \{1, \ldots, n\} & \longrightarrow & \mathbb{F} \\ i & \longmapsto & a_i. \end{array} \right.$$

This yields a bijection between $\mathbb{F}^n$ and $\mathcal{F}(\{1, \ldots, n\}, \mathbb{F})$ that is a vector space isomorphism (compatible with the vector space operations, see Definition 2.1.3).

Two familiar concrete examples of spaces of the form $\mathcal{F}(A, \mathbb{R})$ are the following:

- $\mathcal{F}(\mathbb{N}, \mathbb{R})$, the $\mathbb{R}$-vector space of all real-valued sequences, where we identify a sequence $(a_i)_{i \in \mathbb{N}} = (a_0, a_1, a_2, \ldots)$ with the function $f\colon \mathbb{N} \to \mathbb{R}$ that maps $i \in \mathbb{N}$ to a_i.
- $\mathcal{F}(\mathbb{R}, \mathbb{R})$, the $\mathbb{R}$-vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$.

There are many other natural examples of vector spaces of functions, as for instance the following ones over $\mathbb{R}$:

- Pol($\mathbb{R}$), the $\mathbb{R}$-vector space of all polynomial functions over $\mathbb{R}$, i.e., of all functions

  $$\begin{array}{rcl} f\colon \mathbb{R} & \to & \mathbb{R} \\ x & \longmapsto & a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 = \sum_{i=0,\ldots,n} a_i x^i \end{array}$$

  for suitable $n \in \mathbb{N}$ and coefficients $a_i \in \mathbb{R}$.

- $\mathrm{Pol}_n(\mathbb{R})$, the $\mathbb{R}$-vector space of all polynomial functions over $\mathbb{R}$ of degree at most n, for some fixed n. So $\mathrm{Pol}_n(\mathbb{R})$ consists of all functions $f\colon x \mapsto \sum_{0=1,\ldots,n} a_i x^i$ for any choice of coefficients $a_0, a_1, \ldots, a_n$ in $\mathbb{R}$.

Exercise 2.1.2 Define vector addition and scalar multiplication in $\mathrm{Pol}(\mathbb{R})$ and $\mathrm{Pol}_n(\mathbb{R})$ in accordance with the stipulations in $\mathcal{F}(A, \mathbb{R})$. Check the vector space axioms.

Concentrating on the coefficients in the polynomials, can you pin down a natural correspondence between $\mathbb{R}^{n+1}$ and $\mathrm{Pol}_n(\mathbb{R})$ that is a vector space isomorphism (Definition 2.1.3)?

Exercise 2.1.3 Define the space $\mathrm{Pol}(\mathbb{F}_p)$ of polynomial functions over $\mathbb{F}_p$. A polynomial function is given by an arithmetical expression $\sum_{i=0,\ldots,n} a_i x^i$ with coefficients $a_i \in \mathbb{F}_p$, and viewed as a functions in $\mathcal{F}(\mathbb{F}_p, \mathbb{F}_p)$. Verify that we obtain an $\mathbb{F}_p$-vector space. What is the size of this space in the case of $\mathbb{F}_2$? Note that two distinct polynomials may define the same polynomial function!

Example 2.1.7 Let $\mathbb{F}^{(m,n)}$ be the set of all $m \times n$ *matrices* [Matrizen] with entries from $\mathbb{F}$. We write $A = (a_{ij})_{1\leqslant i\leqslant n; 1\leqslant j\leqslant m}$ for a matrix with m rows and n columns with entry $a_{ij} \in \mathbb{F}$ in row i and column j. On $\mathbb{F}^{(m,n)}$ we again declare addition and scalar multiplication in the natural component-wise fashion:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} :=$$

$$\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

and

$$\lambda \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} := \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

Let $\mathbf{0} \in \mathbb{F}^{(m,n)}$ be the matrix with entries $a_{ij} = 0$ throughout. Then $\mathbb{F}^{(m,n)}$ is an $\mathbb{F}$-vector space. Which standard vector space over $\mathbb{F}$ is this space isomorphic to?

2.2 Subspaces

A (linear) subspace $U \subseteq V$ of an $\mathbb{F}$-vector space V is a subset U that is itself an $\mathbb{F}$-vector space w.r.t. the *induced* linear structure, i.e., w.r.t. the addition and scalar multiplication inherited from V.

We have seen several examples of this subspace relationship above:

- the solution set $S(E^*) \subseteq \mathbb{R}^n$ of any homogeneous system of equations E^* over $\mathbb{R}^n$ forms a subspace; see Observation 1.1.1; this observation generalises to any n and any other field $\mathbb{F}$.

- the relationship between the $\mathbb{R}$-vector spaces $\mathrm{Pol}_n(\mathbb{R}) \subseteq \mathrm{Pol}(\mathbb{R})$ and $\mathrm{Pol}(\mathbb{R}) \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R})$.

Note, however, that the solution set $S(E)$ of a not necessarily homogeneous system of equations usually is not a subspace, as vector addition and scalar multiplication do not operate in restriction to this subset. (We shall return to this in section 2.2.2 below.)

2.2.1 Linear subspaces

Definition 2.2.1 Let V be an $\mathbb{F}$-vector space. A non-empty subset $U \subseteq V$ is a (linear) *subspace* [Untervektorraum] iff vector addition $+$ and scalar multiplication $\cdot$ of V restrict to the subset U in such a way that U with these induced operations is an $\mathbb{F}$-vector space.

That $+$ and $\cdot$ restrict to operations of the required format on $U \subseteq V$ means that

(i) for all $\mathbf{u}_1, \mathbf{u}_2 \in U$: $\mathbf{u}_1 + \mathbf{u}_2 \in U$.

(ii) for all $\mathbf{u} \in U$ and all $\lambda \in \mathbb{F}$: $\lambda\mathbf{u} \in U$.

These are referred to as *closure conditions* on U. And in fact, closure is all that is needed, as stated in the following.

Proposition 2.2.2 *Let $\emptyset \neq U \subseteq V$ where V is an $\mathbb{F}$-vector space. Then U is a subspace of V iff for all $\mathbf{u}_1, \mathbf{u}_2 \in U$ and all $\lambda_1, \lambda_2 \in \mathbb{F}$:*

$$\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 \in U.$$

Proof. It is clear that this closure condition is necessary for U to be a subspace, as it needs to be closed under both addition and scalar multiplication by definition.

Conversely, assume that $\emptyset \neq U \subseteq V$ and that the above closure condition is satisfied.

We firstly see that vector addition and scalar multiplication of V do restrict to U, in the sense that for $\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2 \in U$ and $\lambda \in \mathbb{F}$:

$$\begin{aligned} \mathbf{u}_1 + \mathbf{u}_2 &\in U \qquad (\text{put } \lambda_1 = \lambda_2 = 1) \\ \lambda \mathbf{u} &\in U \qquad (\text{put } \mathbf{u}_1 = \mathbf{u}_2 = \mathbf{u} \text{ and } \lambda_1 = \lambda,\ \lambda_2 = 0). \end{aligned}$$

Of the axioms (V1-8) that we need to verify in restriction to U, all but (V2) are trivial: any identity between terms that holds for all choices of vectors in V must in particular hold of all choices of vectors from $U \subseteq V$.

(V2) is different because it (implicitly) requires $\mathbf{0}$ to be in U; but this is no problem as our closure condition shows that $\mathbf{u} + (-1)\mathbf{u} = \mathbf{0} \in U$ (for $\mathbf{u}_1 = \mathbf{u}_2 = \mathbf{u}$ and $\lambda_1 = 1$, $\lambda_2 = -1$), as long as we have any $\mathbf{u} \in U$ to apply this to – and we do, as $U \neq \emptyset$ by assumption.

□

Exercise 2.2.1 Show that the closure condition expressed in the proposition is equivalent with the following, extended form, which is sometimes more handy in subspace testing:

(i) $\mathbf{0} \in U$.

(ii) for all $\mathbf{u}_1, \mathbf{u}_2 \in U$: $\mathbf{u}_1 + \mathbf{u}_2 \in U$.

(iii) for all $\mathbf{u} \in U$ and all $\lambda \in \mathbb{F}$: $\lambda \mathbf{u} \in U$.

Exercise 2.2.2 Verify that the following are subspace relationships:

(i) $\{(b_1, \ldots, b_m, 0, \ldots, 0) \in \mathbb{F}^n : (b_1, \ldots, b_m) \in \mathbb{F}^m\} \subseteq \mathbb{F}^n$. [For any $m \leqslant n$.]

(ii) $S(E^*) \subseteq \mathbb{F}^n$ where E^*: $a_1 x_1 + \ldots + a_n x_n = 0$ is a homogeneous linear equation over $\mathbb{F}^n$ with coefficients $a_i \in \mathbb{F}$.

(iii) $\mathrm{Pol}_n(\mathbb{R}) \subseteq \mathrm{Pol}(\mathbb{R})$.

(iv) $\mathrm{Pol}(\mathbb{R}) \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R})$.

Exercise 2.2.3 Check that the following are *not* subspace relationships:

(i) $S(E) \subseteq \mathbb{F}^n$ where $E\colon a_1x_i + \ldots + a_nx_n = 1$ is an inhomogeneous linear equation over $\mathbb{F}^n$ with coefficients $a_i \in \mathbb{F}$.

(ii) $\{f \in \mathrm{Pol}(\mathbb{R})\colon f(0) \leqslant 17\} \subseteq \mathrm{Pol}(\mathbb{R})$.

(iii) $\{f\colon f \text{ a bijection from } \mathbb{R} \text{ to } \mathbb{R}\} \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R})$.

(iv) $\mathcal{F}(\mathbb{N}, \mathbb{Z}) \subseteq \mathcal{F}(\mathbb{N}, \mathbb{R})$.

Proposition 2.2.3 *Any intersection of subspaces is a subspace. Let V be an $\mathbb{F}$-vector space, $U_i \subseteq V$ subspaces for all $i \in I$. Then $\bigcap_{i\in I} U_i \subseteq V$ is also a subspace.*

Proof. We use the criterion of Proposition 2.2.2 to show that $U := \bigcap_{i\in I} U_i$ is a subspace of V.

Note first that $U \neq \emptyset$ as $\mathbf{0} \in U_i$ for all i (U_i is a subspace); so $\mathbf{0} \in U$.

Let $\mathbf{u}_1, \mathbf{u}_2 \in U$, $\lambda_1, \lambda_2 \in \mathbb{F}$. We need to show that $\lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 \in U$.

$\mathbf{u}_1 \in U$ implies that $\mathbf{u}_1 \in U_i$ for each $i \in I$, similarly for $\mathbf{u}_2$. Therefore, as U_i is a subspace, $\lambda_1\mathbf{u}_1 + \lambda_2u_2 \in U_i$. As this holds for every individual $i \in I$, we have that $\lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 \in \bigcap_{i\in I} U_i = U$, as required.

□

This closure under intersection implies for instance that the solution set of any system of homogeneous linear equations is a subspace, just on the basis that the solution set of every single homogeneous linear equation is. And this applies even for infinite systems, like the following.

Example 2.2.4 Consider the $\mathbb{R}$-vector space $\mathcal{F}(\mathbb{N}, \mathbb{R})$ of all real-valued sequences. We write $(a_j)_{j\in\mathbb{N}}$ for a typical member. Let, for $i \in \mathbb{N}$, E_i be the following homogeneous linear equation

$$E_i\colon a_i + a_{i+1} - a_{i+2} = 0.$$

It is easily verified that $S(E_i) = \{(a_j)_{j\in\mathbb{N}}\colon a_i + a_{i+1} = a_{i+2}\}$ forms a subspace of $\mathcal{F}(\mathbb{N}, \mathbb{R})$. The intersection of all these subspaces, $\bigcap_{i\in\mathbb{N}} S(E_i)$, contains precisely those sequences $(a_j)_{j\in\mathbb{N}}$ for which

$$a_{j+2} = a_j + a_{j+1} \text{ for all } j \in \mathbb{N},$$

the *Fibonacci* sequences. [Of course one could also verify directly that the set of Fibonacci sequences forms a subspace of $\mathcal{F}(\mathbb{N}, \mathbb{R})$.]

2.2.2 Affine subspaces

We briefly consider subsets that are not linear subspaces but not far from such, like the solution sets $S(E)$ to inhomogeneous (systems of) linear equations.

Recall that, for instance, in $\mathbb{R}^2$ a single linear equation can have the following types of solution sets:

(i) $\emptyset$; this is of no further interest.

(ii) $\mathbb{R}^2$; the entire plane, which (although degenerate) is a linear subspace.

(iii) a line in $\mathbb{R}^2$, which may or may not contain $\mathbf{0}$; if it does, it is a linear subspace; if it does not, it fails to be closed under vector addition or scalar multiplication. However, it is still a simple translate of a linear subspace, as we saw in Lemma 1.1.3 (b).

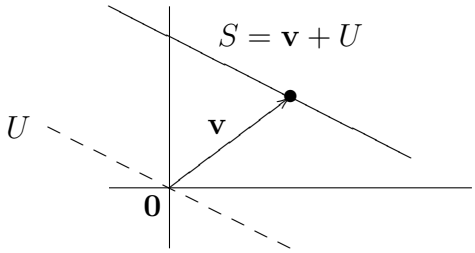
Definition 2.2.5 $S \subseteq V$ is an *affine subspace* [affiner Unterraum] of the vector space V if S is of the form

$$S = \{\mathbf{v} + \mathbf{u} \colon \mathbf{u} \in U\}$$

for some linear subspace $U \subseteq V$ and some $\mathbf{v} \in V$.

For convenience we introduce the notation $\mathbf{v} + U$ for

$$\mathbf{v} + U := \{\mathbf{v} + \mathbf{u} \colon \mathbf{u} \in U\}.$$



Exercise 2.2.4 Show that if $S = \mathbf{v}_0 + U \subseteq V$ is an affine subspace, then $S = \mathbf{v} + U$ for *every* $\mathbf{v} \in S$, and that the linear subspace $U \subseteq V$ is uniquely determined by S.

Exercise 2.2.5 Let $S \subseteq V$ be an affine subspace. Show that S is a linear subspace iff $\mathbf{0} \in S$.

Exactly along the lines of section 1.1.3 we find for instance the following for solution sets of systems of linear equations – and here the role of U is played by the solution set of the associated homogeneous system.

Exercise 2.2.6 Let E be a system of linear equations

$$E_j\colon \sum_{i=1,\ldots,n} a_{ji}x_i = b_j \qquad \text{for } j = 1,\ldots,m,$$

over $\mathbb{F}^n$, with coefficients $a_{ji}, b_j \in \mathbb{F}$. Let E^* be the associated homogeneous system with all b_j replaced by 0.

(a) Show that $S(E^*)$ is a linear subspace of $\mathbb{F}^n$, and that, if $\mathbf{v} \in S(E)$ is any solution of E, then $S(E) = \mathbf{v} + S(E^*)$ is an affine subspace of $\mathbb{F}^n$.

(b) Verify that the Gauß-Jordan elimination procedure works in this general case exactly as it does over $\mathbb{R}$.

(c) Show that $S(E^*)$ has non-trivial solutions $\mathbf{u} \neq \mathbf{0}$ for all systems E with more variables (columns on the left-hand side) than equations (rows), i.e., when $m < n$. [Hint: in this case, not all variables can be pivot.]

Proposition 2.2.6 *If E is any system of linear equations over $\mathbb{F}^n$, then its solution set is either empty or forms an affine subspace.*

Exercise 2.2.7 Show that the intersection of (any number of) affine subspaces of a vector space V is either empty or again an affine subspace. [Use Proposition 2.2.3 and Exercise 2.2.4.]

Exercise 2.2.8 Consider $\mathbb{Z}_2^3$ as an $\mathbb{F}_2$-vector space. How many different linear and affine subspaces, respectively, does this space have?

Exercise 2.2.9 In the game "SET", there are 81 cards which differ according to 4 properties for each of which there are 3 distinct states (3 colours: red, green, blue; 3 shapes: round, angular, wavy; 3 numbers: 1, 2, 3; and 3 faces: thin, medium, thick). A "SET", in terms of the rules of the game, is any set of 3 cards such that for each one of the four properties, either all 3 states of that property are represented or all 3 cards in the set have the same state. For instance,

(red, round,1, medium),
(red, angular,2, thin),
(red, wavy,3, thick)

form a "SET", while the following do not:

$$\begin{array}{c}(\text{red}, \text{round}, 1, \text{medium}),\\ (\text{blue}, \text{angular}, 2, \text{thin}),\\ (\text{green}, \text{wavy}, 1, \text{thick}).\end{array}$$

Modelling the set of all cards as $\mathbb{Z}_3^4$, verify that the game's notion of "SET" corresponds to affine subspaces of three elements (the lines in $\mathbb{Z}_3^4$).

Show that any two different cards uniquely determine a third one with which they form a "SET".
[Extra: what is the largest number of cards that can fail to contain a "SET"?]

Exercise 2.2.10 Consider the following subspace of the $\mathbb{R}$-vector space of all real-valued sequences, $\mathcal{F}(\mathbb{N}, \mathbb{R})$, defined in terms of a parameter $a \in \mathbb{R}$:

$$S_a := \big\{(a_i)_{i\in\mathbb{N}} \colon \lim_{i\to\infty} a_i = a\big\}.$$

Show that for any a, S_a is an affine subspace of $\mathcal{F}(\mathbb{N}, \mathbb{R})$.
For which a is it even a linear subspace?

2.3 Aside: affine and linear spaces

There is also a natural notion of *affine spaces*. These are point spaces with a linear structure embodied in an accompanying $\mathbb{F}$-vector space.

Definition 2.3.1 An *affine space* [affiner Raum] with associated $\mathbb{F}$-vector space V consists of a triple (A, V, ρ) where

(i) A is the set of *points* of the affine space (we write $P, Q, R, \ldots \in A$ for points).

(ii) V is an $\mathbb{F}$-vector space.

(iii) ρ describes an action of the group $(V, +, \mathbf{0})$ on A as a group of *translations*,

$$\begin{array}{rcl}\rho \colon V \times A & \longrightarrow & A\\ (\mathbf{v}, P) & \longmapsto & \rho(\mathbf{v}, P) =: P + \mathbf{v},\end{array}$$

such that:

(a) for all $\mathbf{u}, \mathbf{v} \in V$ and all $P \in A$: $(P + \mathbf{u}) + \mathbf{v} = P + (\mathbf{u} + \mathbf{v})$
(ρ is compatible with vector addition in V, a group action).

(b) for all $\mathbf{v} \in V$ and all $P \in A$: $P + \mathbf{v} = P$ iff $\mathbf{v} = \mathbf{0}$
(no non-trivial fixed points).

(c) for all $P, Q \in A$ there is some $\mathbf{v} \in V$ such that $P + \mathbf{v} = Q$
(all of A covered by translations of any given point).

Note that the vector $\mathbf{v}$ whose existence is postulated in (c) is uniquely determined by P and Q, by (a) and (b). We write $\overrightarrow{PQ}$ for this unique $\mathbf{v}$ such that $P + \mathbf{v} = Q$.

Any $\mathbb{F}$-vector space V gives rise to an affine space whose point space is V, and with $\rho\colon V \times V \to V$ just being vector addition.

The main reason why one wants to regard affine spaces as entities in their own right is geometric. The Euclidean plane, for instance, is in fact more closely modelled as an affine space with associated $\mathbb{R}$-vector space $\mathbb{R}^2$ than by the $\mathbb{R}$-vector space $\mathbb{R}^2$ itself. The reason is that the Euclidean plane is entirely translation invariant as a point space, with no distinguished origin, while any vector space always has a distinguished element, namely its null vector. And indeed, there is of course a distinguished translation corresponding to the null vector, but not a distinguished point.

Of the natural notion of affine subspaces (subspaces of affine spaces) we have seen above only its slightly deflated version in the form of affine subspaces of vector spaces (which we may regard as affine spaces by the above). But the point there was exactly the same: to forget about the distinguished role of the origin or null vector and to obtain a notion that is invariant under translation.

In the general setting of an affine space (A, V, ρ), a subset $A_0 \subseteq A$ is an *affine subspace* if it is of the form $P + U = \{P + \mathbf{u}\colon \mathbf{u} \in U\}$ for some point $P \in A$ and some linear subspace $U \subseteq V$. One checks that this implies that (A_0, U, ρ_0) is itself an affine space, where ρ_0 is the restriction of ρ to $U \times A_0$.

Exercise 2.3.1 Check that the following are equivalent for any subset $A_0 \subseteq A$ of an affine space (A, V, ρ):

(i) A_0 forms an affine subspace.

(ii) the set $\{\overrightarrow{PQ}\colon P, Q \in A_0\}$ forms a linear subspace of V.

(iii) for some fixed $P \in A_0$, $\{\overrightarrow{PQ}\colon Q \in A_0\}$ forms a linear subspace of V.

Remark: over fields $\mathbb{F}$, in which $1 + 1 \neq 0$ (characteristic $\neq 2$), these are also equivalent with the condition that for any two distinct points $P, Q \in A_0$, the line through P and Q, $\{P + \lambda\overrightarrow{PQ}\colon \lambda \in \mathbb{F}\}$, is contained in A_0. (How?)

2.4 Linear dependence and independence

2.4.1 Linear combinations and spans

A *linear combination* [Linearkombination] of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in an $\mathbb{F}$-vector space V is any vector of the form

$$\lambda_1\mathbf{v}_1 + \ldots + \lambda_k\mathbf{v}_k = \sum_{i=1,\ldots,k} \lambda_i\mathbf{v}_i$$

with $\lambda_i \in \mathbb{F}$ for $i = 1, \ldots, k$.

Exercise 2.4.1 Show that (in generalisation of the explicit closure condition in Proposition 2.2.2) any subspace $U \subseteq V$ is closed under arbitrary linear combinations: whenever $\mathbf{u}_1, \ldots, \mathbf{u}_k \in U$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$, then $\sum_{i=1,\ldots,k} \lambda_i\mathbf{u}_i \in U$. [Hint: by induction on $k \geqslant 1$].

Definition 2.4.1 Let $S \subseteq V$ be any subset of the $\mathbb{F}$-vector space V. A *linear combination* [Linearkombination] over S is a vector of the form

$$\sum_{i=1,\ldots,k} \lambda_i\mathbf{v}_i,$$

for any $k \in \mathbb{N}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_k \in S$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$. ([1]) The scalars λ_i are called the coefficients of the linear combination.

We denote the set of all linear combinations over S by

$$\operatorname{span}(S) := \Big\{ \sum_{i=1,\ldots,k} \lambda_i\mathbf{v}_i \colon \mathbf{v}_1, \ldots, \mathbf{v}_k \in S, \lambda_1, \ldots, \lambda_k \in \mathbb{F}; k \in \mathbb{N} \Big\} \subseteq V.$$

Note that $\mathbf{0} \in \operatorname{span}(S)$ for any S (even for $S = \emptyset$).

The span of S, $\operatorname{span}(S)$, is also called the *linear hull* of S.

Lemma 2.4.2 *Let $S \subseteq V$ be any subset of the $\mathbb{F}$-vector space V. Then $\operatorname{span}(S) \subseteq V$ is a subspace. Moreover it is the smallest (in the sense of $\subseteq$) subspace of V that contains S. As such it can also be characterised as*

$$\operatorname{span}(S) = \bigcap_{\substack{U \subseteq V \text{a subspace} \\ \text{with } S \subseteq U}} U.$$

[1] We explicitly want to allow the (degenerate) case of $k = 0$ and put the null vector $\mathbf{0} \in V$ to be the value of the empty sum!

Proof. We leave the verification that $\mathrm{span}(S)$ is a subspace as an exercise.

Clearly $S \subseteq \mathrm{span}(S)$, so $\mathrm{span}(S)$ is a subspace containing S. We next show that $\mathrm{span}(S)$ is the smallest such subspace of V by showing that $\mathrm{span}(S) \subseteq U$ for any subspace $U \subseteq V$ that contains S. Let U be some subspace with $S \subseteq U$. By closure of U (as a subspace) under linear combinations, we get that any linear combination over S remains inside U, whence $\mathrm{span}(S) \subseteq U$ as claimed.

Finally, $\mathrm{span}(S)$ is the intersection of all subspaces U that contain S. We just saw that $\mathrm{span}(S)$ is indeed a subset of every one of those subspaces, hence it is also contained in their intersection. As $\mathrm{span}(S)$ itself is a subspace of V that contains S, the intersection of all such must in particular be contained in $\mathrm{span}(S)$ – and the claimed equality follows.

□

Exercise 2.4.2 Show that $\mathrm{span}(S) \subseteq \mathrm{span}(S')$ whenever $S \subseteq S'$; and that $\mathrm{span}(\mathrm{span}(S)) = \mathrm{span}(S)$.

Show that in general *not* $\mathrm{span}(S) \cap \mathrm{span}(S') = \mathrm{span}(S \cap S')$ and *not* $\mathrm{span}(S) \cup \mathrm{span}(S') = \mathrm{span}(S \cup S')$. Which inclusions do hold in these cases?

Definition 2.4.3 If $U \subseteq V$ is a subspace, then $S \subseteq U$ is a *spanning set* [Erzeugendensystem] for U iff $U = \mathrm{span}(S)$.

In this case one also says that S *spans* U.

Note that $S = \emptyset$ is a spanning set for $U = \{\mathbf{0}\}$.

For the following compare the row transformations in Gauß-Jordan, section 1.1.3.

Lemma 2.4.4 *Let $S \subseteq V$, V an $\mathbb{F}$-vector space. Let S' be obtained from S by one of the following operations*

(T2) replacing some $\mathbf{u} \in S$ by $\lambda\mathbf{u}$ for some $0 \neq \lambda \in \mathbb{F}$.

(T3) replacing some $\mathbf{u} \in S$ by $\mathbf{u} + \lambda\mathbf{v}$ for some $\lambda \in \mathbb{F}$ and $\mathbf{u} \neq \mathbf{v} \in S$.

Then $\mathrm{span}(S') = \mathrm{span}(S)$.

Proof. We do the case for $(T3)$. Let $\mathbf{v} \neq \mathbf{u}$, $\mathbf{u}, \mathbf{v} \in S$, $\lambda \in \mathbb{F}$. Let $\mathbf{u}' := \mathbf{u} + \lambda\mathbf{v}$ and put $S' := (S \setminus \{\mathbf{u}\}) \cup \{\mathbf{u}'\}$.

We firstly show that $\mathrm{span}(S') \subseteq \mathrm{span}(S)$. Let $\mathbf{w} \in \mathrm{span}(S')$: $\mathbf{w} = \mathbf{w}_0 + \mu\mathbf{u}'$ where $\mathbf{w}_0 \in \mathrm{span}(S \setminus \{\mathbf{u}\})$. Therefore $\mathbf{w} = \mathbf{w}_0 + \mu(\mathbf{u} + \lambda\mathbf{v}) \in \mathrm{span}(S)$.

The opposite inclusion is shown similarly, using $\mathbf{u} = \mathbf{u}' - \lambda\mathbf{v}$.

□

Given a subspace $U \subseteq V$ one is often interested in finding a spanning set $S \subseteq U$. Particular importance will be attached to minimal spanning sets. In order to prepare their study, we look at ways in which sets of vectors can be redundant for the purpose of forming linear combinations.

2.4.2 Linear (in)dependence

Definition 2.4.5 Let V be an $\mathbb{F}$-vector space and $S \subseteq V$. The set of vectors S is *linearly independent* [linear unabhängig] iff for all $\mathbf{u} \in S$:

$$\mathbf{u} \notin \mathrm{span}(S \setminus \{\mathbf{u}\}).$$

Otherwise, i.e., if this condition fails to hold, S is said to be linearly dependent.

Note that the empty set is, by definition, linearly independent as well. Any set S with $\mathbf{0} \in S$ is linearly dependent.

A useful criterion for linear independence is established in the following.

Lemma 2.4.6 *Let $\emptyset \neq S \subseteq V$. The set of vectors S is linearly independent iff for all $k \in \mathbb{N}$ and pairwise distinct $\mathbf{u}_1, \ldots, \mathbf{u}_k \in S$ and all $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$:*

$$\lambda_1\mathbf{u}_1 + \ldots + \lambda_k\mathbf{u}_k = \mathbf{0} \quad \Longrightarrow \quad \lambda_1 = \lambda_2 = \ldots = \lambda_k = 0.$$

A linear combination $\lambda_1\mathbf{u}_1 + \ldots + \lambda_k\mathbf{u}_k = \mathbf{0}$ in which the $\mathbf{u}_i$ are pairwise distinct and not all coefficients λ are 0 is called a *non-trivial linear combination* of $\mathbf{0}$. The criterion says that S is linearly dependent iff it admits a non-trivial linear combination of the null vector.

Proof. We show first that the criterion is necessary for linear independence. Suppose $S \neq \emptyset$ violates the criterion: there is a non-trivial linear combination $\mathbf{0} = \sum_{i=1,\ldots,k} \lambda_i\mathbf{u}_i$ with pairwise distinct $\mathbf{u}_i \in S$ and, for instance $\lambda_1 \neq 0$. Then $\mathbf{u}_1 = -\lambda_1^{-1} \sum_{i=2,\ldots,k} \lambda_i\mathbf{u}_i$ shows that $\mathbf{u}_1 \in \mathrm{span}(S \setminus \{\mathbf{u}_1\})$. So S is linearly dependent.

Conversely, to establish the sufficiency of the criterion, assume that S is linearly dependent. Let $\mathbf{u} \in S$ and suppose that $\mathbf{u} \in \mathrm{span}(S \setminus \{\mathbf{u}\})$. So $\mathbf{u} = \sum_{i=1,\ldots,k} \lambda_i\mathbf{u}_i$ for suitable λ_i and $\mathbf{u}_i \in S \setminus \{\mathbf{u}\}$, which may be chosen pairwise distinct (why?). But then $\mathbf{0} = \mathbf{u} + \sum_{i=1,\ldots,k}(-\lambda_i)\mathbf{u}_i$ is a non-trivial linear combination of $\mathbf{0}$ over S.

□

Example 2.4.7 For a simple example consider sets $S = \{\mathbf{u}, \mathbf{v}\} \subseteq \mathbb{R}^2$ of two vectors in $\mathbb{R}^2$. Suppose S is linearly dependent. Then there must be $(\lambda, \mu) \neq (0, 0)$ such that $\lambda\mathbf{u} + \mu\mathbf{v} = \mathbf{0}$. If, for instance, $\lambda \neq 0$, this implies that $\mathbf{u} = (-\mu/\lambda)\mathbf{v}$ is a scalar multiple of $\mathbf{v}$.

Conversely, if $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ and $\mathbf{u}$ and $\mathbf{v}$ are not scalar multiples of each other, then $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent (why?).

Exercise 2.4.3 Show that $\{(1, 1, 2), (1, 2, 3), (1, 2, 4)\}$ is linearly independent in $\mathbb{R}^3$, and that $\{(1, 1, 2), (2, 1, 2), (1, 2, 4)\}$ are linearly dependent.

Exercise 2.4.4 Show that the three functions $x \mapsto x$, $x \mapsto x^2$ and $x \mapsto 1$ form a linearly independent set in $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

Lemma 2.4.8 *Let S be any set of $n > m$ many vectors in $\mathbb{F}^m$. Then S is linearly dependent.*

Proof. Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ where the $\mathbf{u}_i$ are pairwise distinct. We need to find coefficients λ_i such that

$$\lambda_1\mathbf{u}_1 + \ldots + \lambda_n\mathbf{u}_n = \mathbf{0}$$

but not all $\lambda_i = 0$.

Let $\mathbf{u}_i = (a_{1i}, \ldots, a_{mi})$ for $i = 1, \ldots, n$. Then the condition that $\lambda_1\mathbf{u}_1 + \ldots + \lambda_n\mathbf{u}_n = \mathbf{0}$ is expressed by the following homogeneous system of linear equations, in which $\lambda_1, \ldots, \lambda_n$ play the role of the variables:

$$E\colon \begin{cases} a_{11}\lambda_1 + a_{12}\lambda_2 + \cdots + a_{1n}\lambda_n = 0 \\ a_{21}\lambda_1 + a_{22}\lambda_2 + \cdots + a_{2n}\lambda_n = 0 \\ \qquad\qquad\qquad\qquad\qquad\quad \vdots \\ a_{m1}\lambda_1 + a_{m2}\lambda_2 + \cdots + a_{mn}\lambda_n = 0 \end{cases}$$

Recalling how Gauß-Jordan applies to finding all the solutions of a system like this, we know that – since there are more variables than equations – not all variables λ_i can turn out to be pivot variables. But values for non-pivot variables can be chosen freely, and then the tuple be completed to a solution. If we choose $\lambda_i = 1$ for some non-pivot variable λ_i, we are guaranteed to find a solution with $\lambda_i \neq 0$. This solution yields a non-trivial linear combination of $\mathbf{0}$ over S. Therefore S is linearly dependent.

□

Observation 2.4.9 *There are sets of n vectors in $\mathbb{F}^n$ that are linearly independent. For example, let, for $i = 1, \ldots, n$, $\mathbf{e}_i$ be the vector in $\mathbb{F}^n$ whose i-th component is 1 and all others 0. So $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $\mathbf{e}_n = (0, \ldots, 0, 1)$. Then $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is linearly independent. Moreover, S spans $\mathbb{F}^n$.*

Proof. For linear independence, consider any linear combination over S and note that

$$\sum_{i=1,\ldots,n} \lambda_i \mathbf{e}_i = (\lambda_1, \ldots, \lambda_n) \in \mathbb{F}^n.$$

Therefore, if $\sum_{i=1,\ldots,n} \lambda_i \mathbf{e}_i = \mathbf{0} = (0, \ldots, 0)$ it must be that $\lambda_i = 0$ for all i.

On the other hand, $(\lambda_1, \ldots, \lambda_n) = \sum_{i=1,\ldots,n} \lambda_i \mathbf{e}_i$ shows that any vector in $\mathbb{F}^n$ is a linear combination over S.

□

The following shows exactly how a linearly independent set can become linearly dependent as one extra vector is added.

Lemma 2.4.10 *Let $S \subseteq V$ be linearly independent. If $S \cup \{\mathbf{v}\}$ is linearly dependent, then $\mathbf{v} \in \mathrm{span}(S)$.*

Proof. If $S \cup \{\mathbf{v}\}$ is linearly dependent there must be a non-trivial linear combination of $\mathbf{0}$ over $S \cup \{\mathbf{v}\}$,

$$\mathbf{0} = \lambda \mathbf{v} + \sum_{i=1,\ldots,k} \lambda_i \mathbf{u}_i,$$

where the $\mathbf{u}_i \in S$ are pairwise distinct. Here $\lambda \neq 0$, as otherwise already S would be linearly dependent. Therefore

$$\mathbf{v} = -\lambda^{-1} \sum_{i=1,\ldots,k} \lambda_i \mathbf{u}_i \ \in \ \mathrm{span}(S).$$

□

2.5 Bases and dimension

2.5.1 Bases

Definition 2.5.1 Let V be a vector space. A *basis* [Basis] of V is a set $B \subseteq V$ of vectors such that

(i) B spans V: $V = \mathrm{span}(B)$.

(ii) B is linearly independent.

In the case of the trivial vector space $V = \{\mathbf{0}\}$ consisting of just the null vector, we admit $B = \emptyset$ as its basis (consistent with our stipulations regarding spanning sets and linear independence in this case.)

An equivalent formulation is that a basis is a minimal subset of V that spans V.

Exercise 2.5.1 Prove the equivalence of the following:

(i) $B \subseteq V$ is a basis.

(ii) $\mathrm{span}(B) = V$ and for every $\mathbf{b} \in B$, $\mathrm{span}(B \setminus \{\mathbf{b}\}) \subsetneq V$.

The following was already shown in Observation 2.4.9 above.

Example 2.5.2 The following is a basis for $\mathbb{F}^n$, the so-called *standard basis* of that space: $B = \{\mathbf{e}_i \colon 1 \leqslant i \leqslant n\}$ where for $i = 1, \ldots, n$:

$$\mathbf{e}_i = (b_{i1}, \ldots, b_{in}) \qquad \text{with} \quad b_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{else.} \end{cases}$$

Example 2.5.3 Before we consider bases in general and finite bases in particular, we look at an example of a vector space for which we can exhibit a basis, but which cannot have a finite basis. Consider the following subspace U of the $\mathbb{R}$-vector space $\mathcal{F}(\mathbb{N}, \mathbb{R})$ of all real-valued sequences $(a_i)_{i \in \mathbb{N}}$:

$$U := \mathrm{span}(\{\mathbf{u}_i \colon i \in \mathbb{N}\})$$

where $\mathbf{u}_i$ is the sequence which is 0 everywhere with the exception of the i-th value which is 1. $B := \{\mathbf{u}_i \colon i \in \mathbb{N}\} \subseteq \mathcal{F}(\mathbb{N}, \mathbb{R})$ is linearly independent [show this as an exercise]; so it forms a basis for U.

We note that U consists of precisely all those sequences in $\mathcal{F}(\mathbb{N}, \mathbb{R})$ that are zero in all but finitely many positions. Consider a sequence $\mathbf{a} = (a_i)_{i \in \mathbb{N}} \in$

U that is a linear combination over B; this linear combination involves only finitely many of the sequences $\mathbf{u}_i$ from B. Let m be the maximal index i such that $\mathbf{u}_i$ occurs with a non-zero coefficient in this representation of $\mathbf{a}$ as a linear combination over B. As all the sequences $\mathbf{u}_i$ with $i \leqslant m$ are zero in all places $i > m$, the same is true of $\mathbf{a}$. So $a_i = 0$ for all $i > m$ and any non-zero positions in $(a_i)_{i\in\mathbb{N}}$ must occur among $a_0, \ldots, a_m$. Conversely, if $\mathbf{a} = (a_i)_{i\in\mathbb{N}}$ and m are such that $a_i = 0$ for all $i > m$, then $\mathbf{a} = \sum_{i=0,\ldots,m} a_i\mathbf{u}_i \in \mathrm{span}(B) = U$.

Assume now that U also had a finite basis, say $B_0 = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$. By the above, each v_i is a sequence that is zero from some point m_i onwards. Let $m = \max(m_i)$. Then any sequence that is a linear combination over B_0 must be zero in all positions $i > m$. It follows that, for instance, $\mathbf{u}_{m+1} \notin \mathrm{span}(B_0)$, contradicting the assumption that B_0 spans U.

Exercise 2.5.2 Similarly provide infinite bases and show that there are no finite bases for:

(i) $\mathrm{Pol}(\mathbb{R})$, the $\mathbb{R}$-vector space of all polynomials over $\mathbb{R}$.

(ii) $\mathcal{F}(\mathbb{N}, \mathbb{Z}_2)$, the $\mathbb{F}_2$-vector space of all infinite bit-streams.

2.5.2 Finite-dimensional vector spaces

Definition 2.5.4 A vector space V is called *finite-dimensional* [endlich dimensional] iff V possesses a finite basis. Otherwise V is *infinite-dimensional* [unendlich dimensional].

Proposition 2.5.5 *Let $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $B' = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be (finite) bases of V with n and m elements, respectively. Then $n = m$.*

Proof. We give an indirect proof of the proposition. Assume $n > m$. We will show that B cannot in fact be linearly independent. The proof is strictly analogous to the one in Lemma 2.4.8.

We want to find coefficients λ_i in a non-trivial linear combination

$$\lambda_1\mathbf{u}_1 + \ldots + \lambda_n\mathbf{u}_n = \sum_{i=1,\ldots,n} \lambda_i\mathbf{u}_i = \mathbf{0}.$$

Let for, $i = 1, \ldots, n$,

$$\mathbf{u}_i = \sum_{j=1,\ldots,m} a_{ji}\mathbf{v}_j.$$

Such representations exist as B' spans V. Now the condition that $\lambda_1\mathbf{u}_1 + \ldots + \lambda_n\mathbf{u}_n = \mathbf{0}$ is equivalent with

$$\sum_{i=1,\ldots,n} \lambda_i \Big(\sum_{j=1,\ldots,m} a_{ji}\mathbf{v}_j \Big) = \mathbf{0}.$$

Regrouping these terms we get

$$\sum_{j=1,\ldots,m} \Big(\sum_{i=1,\ldots,n} a_{ji}\lambda_i \Big)\mathbf{v}_j = \mathbf{0}.$$

As the $\mathbf{v}_j$ form a basis (linear independence) this is equivalent with

$$\sum_{i=1,\ldots,n} a_{ji}\lambda_i = 0 \qquad \text{for } j = 1, \ldots, m.$$

Now this is the same homogeneous system of linear equations over $\mathbb{F}^n$ as considered in Lemma 2.4.8. We argued there that this system must have non-trivial solutions for the λ_i since this system has more variables than equations.

We therefore conclude that B is linearly dependent, contradicting the assumption that it was a basis.

□

The proposition justifies the following definition.

Definition 2.5.6 For finite-dimensional V we let its *dimension* [Dimension], $\dim(V)$, be the size of a (any) basis of V. So V is n-dimensional, $\dim(V) = n$, iff V has a basis consisting of precisely n vectors.

Lemma 2.5.7 *Let* $\dim(V) = n$ *and* $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ *a basis. Then the coefficients in any linear combination over* B *are uniquely determined in the following sense. For all* $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n \in \mathbb{F}$*:*

$$\sum_{i=1,\ldots,n} \lambda_i\mathbf{b}_i = \sum_{i=1,\ldots,n} \mu_i\mathbf{b}_i \qquad \Rightarrow \qquad \lambda_i = \mu_i \text{ for all } i.$$

Proof. The given equality implies $\sum_{i=1,\ldots,n}(\lambda_i - \mu_i)\mathbf{b}_i = \mathbf{0}$. As B is linearly independent, $\lambda_i - \mu_i = 0$ and hence $\lambda_i = \mu_i$, for $i = 1, \ldots, n$.

□

We sometimes want to be able to identify individual basis vectors in a basis, and for this introduce the notion of a *labelled basis* (also sometimes called *ordered basis* [geordnete Basis]), which is not a set of basis vectors but a labelled family or tuple of basis vectors.

Definition 2.5.8 A *labelled basis* of a vector space V is a family of vectors $(\mathbf{b}_i)_{i\in I}$ indexed by some index set I such that $\mathbf{b}_i \neq \mathbf{b}_j$ for $i \neq j \in I$ and $\{\mathbf{b}_i \colon i \in I\}$ forms a basis. In the finite-dimensional case, if $\dim(V) = n$, we may use the index set $I = \{1, \ldots, n\}$, and a labelled basis is just an n-tuple of pairwise distinct vectors that are linearly independent.

With a labelled basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of the n-dimensional $\mathbb{F}$-vector space V associate the following two maps:

$$\begin{array}{rcl} \xi_B \colon \mathbb{F}^n & \longrightarrow & V \\ (\lambda_1, \ldots, \lambda_n) & \longmapsto & \sum_{i=1,\ldots,n} \lambda_i \mathbf{b}_i \end{array}$$

$$\begin{array}{rcl} \text{and} \quad [\![\cdot]\!]_B \colon V & \longrightarrow & \mathbb{F}_n \\ \mathbf{v} & \longmapsto & (\lambda_1, \ldots, \lambda_n) \text{ if } \mathbf{v} = \sum_{i=1,\ldots,n} \lambda_i \mathbf{b}_i. \end{array}$$

That $[\![\mathbf{v}]\!]_B$ is well-defined follows from the previous lemma. Clearly both maps are linear, and inverses of each other. It follows that they constitute (an inverse pair of) vector space isomorphisms between V and $\mathbb{F}^n$. In particular we get the following.

Corollary 2.5.9 *Any $\mathbb{F}$-vector space of dimension n is isomorphic to $\mathbb{F}^n$.*

Exercise 2.5.3 Consider the $\mathbb{R}$-vector space Fib of all Fibonacci sequences (compare Example 2.2.4, where Fib was considered as a subspace of $\mathcal{F}(\mathbb{N}, \mathbb{R})$).

Recall that a sequence $(a_i)_{i\in\mathbb{N}} = (a_0, a_1, a_2, \ldots)$ is a Fibonacci sequence iff

$$a_{i+2} = a_i + a_{i+1} \text{ for all } i \in \mathbb{N}.$$

(a) Show that the space Fib has a basis consisting of two sequences; hence its dimension is 2.

(b) Show that Fib contains precisely two different geometric sequences, i.e., sequences of the form

$$a_i = \rho^i, \quad i = 0, 1, \ldots$$

for some real $\rho \neq 0$. [In fact the two reals involved are the so-called golden ratio and its negative reciprocal.]

(c) Show that the two sequences from part (b) are linearly independent, hence also from a basis.

(d) Express the standard Fibonacci sequence $\mathbf{f} = (0, 1, 1, 2, 3, \ldots)$ as a linear combination in the basis obtained in part (c). This yields a rather surprising closed term representation for the i-th member in $\mathbf{f}$.

We now analyse more closely the connection between bases, spanning sets and linearly independent sets in the finite dimensional case.

Lemma 2.5.10 *If V is spanned by a set S of m vectors, then V has a basis $B \subseteq S$. It follows that* $\dim(V) \leqslant m$.

Proof. Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and suppose $V = \mathrm{span}(S)$. We obtain a basis $B \subseteq S$ for V by succesively selecting members of S and discarding redundant members.

The iteration generates a sequence $\emptyset = S_0 \subseteq S_1 \subseteq S_2 \subseteq \cdots \subseteq S_m \subseteq S$ of linearly independent subsets $S_i \subseteq S$ such that $\mathrm{span}(S_i) = \mathrm{span}(\{\mathbf{u}_j : j \leqslant i\})$. Then $B := S_m$ will be as desired. It remains to define the S_i.

Let $S_0 := \emptyset$. Inductively let, for $0 \leqslant i < m$,

$$S_{i+1} := \begin{cases} S_i \cup \{\mathbf{u}_{i+1}\} & \text{if } \mathbf{u}_{i+1} \notin \mathrm{span}(S_i), \\ S_i & \text{if } \mathbf{u}_{i+1} \in \mathrm{span}(S_i). \end{cases}$$

The case distinction guarantees that all S_i are linearly independent (compare Lemma 2.4.10) and that $\mathbf{u}_i \in \mathrm{span}(S_i)$ for $i \leqslant m$.

□

Corollary 2.5.11 *Let* $\dim(V) = n$. *Then*

(a) *any set S of $m > n$ many vectors in V is linearly dependent.*

(b) *any set S of $m < n$ vectors in V fails to span V.*

Proof. (a) was shown in the proof of Proposition 2.5.5. (b) then follows from the previous lemma.

□

The argument in Lemma 2.5.10 can be extended as follows. Suppose we are given a finite spanning set S as above and a set A of linearly independent vectors, which we want to include in the basis we construct.

Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ with $\mathrm{span}(S) = V$, and let $A \subseteq V$ be linearly independent.

We can now use a variant of the above iteration to generate a basis comprising A, by starting from $S_0 := A$ rather than $S_0 = \emptyset$, and then proceeding as above. Then $B := S_m \supseteq A$ is a basis consisting of A and vectors from S. We have proved the following variant of Lemma 2.5.10.

Lemma 2.5.12 *Let V be a finite-dimensional vector space, spanned by $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$. Let $A \subseteq V$ be any linearly independent set of vectors.*

Then A can be extended to a basis B of V using vectors from S. We obtain a basis B with $A \subseteq B \subseteq A \cup S$.

In the special case where S is itself a basis, we obtain the so-called *Steinitzscher Austauschsatz*.

Corollary 2.5.13 *Let $\dim(V) = n$, $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis of V. Let $A \subseteq V$ be any linearly independent set.*

Then there is a basis $\hat{B}$ of V with $A \subseteq \hat{B} \subseteq A \cup B$.

Equivalently: there is a subset $B_0 \subseteq B$ such that $(B \setminus B_0) \cup A$ is a basis.

Proof. Use the previous lemma, with B in the place of S and call the resulting basis $\hat{B}$. For the second formulation let $B_0 := B \setminus \hat{B}$. □

The second formulation explains the name *exchange property* [Austauscheigenschaft] for this phenomenon. The members of B_0 are *exchanged* for the members of A in the new basis. We know from Proposition 2.5.5 that B and $\hat{B}$, and hence also B_0 and A must have the same size.

The following is another important corollary, known as the basis extension theorem [Basisergänzungssatz]. For its proof use any basis and apply the exchange property choosing the given linearly independent set for A.

Corollary 2.5.14 *Let V be finite-dimensional. Any linearly independent set of vectors in V can be extended to a basis of V.*

2.5.3 Dimensions of linear and affine subspaces

Let V be a finite-dimensional vector space.

If $U \subseteq V$ is a (linear) subspace of V, then $\dim(U)$ is just the dimension of U considered as a vector space in its own right. It follows from Corollary 2.5.11 that $\dim(U) \leqslant \dim(V)$, where $\dim(U) = \dim(V)$ iff $U = V$.

Affine subspaces (of a vector space V) have the form $S = \mathbf{v} + U$ for a linear subspace U of V, and this U is uniquely determined by S, compare Exercise 2.2.4.

Definition 2.5.15 For an affine subspace $S = \mathbf{v} + U$: $\dim(S) := \dim(U)$.

If $\dim(V) = n$, the following terminology is often used

dimension	1	2	$n-1$
affine subspaces	lines	planes	hyperplanes

Exercise 2.5.4 Let $\dim(V) = n$, $S \subseteq V$ a set of $m \leqslant n$ vectors. Show that there is a linear subspace $U \subseteq V$ with $S_0 \subseteq U$ and $\dim(U) = m$; if S is linearly independent then U is uniquely determined by this requirement. [Hint: put $U := \mathrm{span}(S)$ and apply reasoning as in Corollary 2.5.11].

Proposition 2.5.16 *Let $\dim(V) = n$, $S_0 \subseteq V$ a set of m vectors, where $1 \leqslant m \leqslant n+1$. Then there is an affine subspace $S \subseteq V$ with $S_0 \subseteq S$ and $\dim(S) = m-1$. Such S is uniquely determined if, for some/any $\mathbf{u}_0 \in S_0$ the set $\{\mathbf{u} - \mathbf{u}_0 \colon \mathbf{u} \in S_0 \setminus \{\mathbf{u}_0\}\}$ is linearly independent.*

Proof. Choose some $\mathbf{u}_0 \in S_0$, let $S_{\mathbf{u}_0} := \{\mathbf{u} - \mathbf{u}_0 \colon \mathbf{u} \in S_0 \setminus \{\mathbf{u}_0\}\}$ and put $U := \mathrm{span}(S_{\mathbf{u}_0})$. It is clear that $S = \mathbf{u}_0 + U$ is an affine subspace of V with $S_0 \subseteq S$ and $\dim(S) \leqslant m-1$.

For the uniqueness claim, observe that any affine subspace $S \supseteq S_0$ must contain $\mathbf{u}_0 + \mathrm{span}(S_{\mathbf{u}_0})$. If $S_{\mathbf{u}_0}$ is linearly independent then there is a unique subspace U of dimension $m-1$ that contains $\mathrm{span}(S_{\mathbf{u}_0})$, namely $\mathrm{span}(S_{\mathbf{u}_0})$. It follows that $S = \mathbf{u}_0 + \mathrm{span}(S_{\mathbf{u}_0})$ is also uniquely determined.

□

Exercise 2.5.5 Rephrase the above in term of point sets in an affine space, using the terminology of section 2.3.

2.5.4 Existence of bases

There is a general theorem that guarantees the existence of bases for any vector space – finite or infinite dimensional. Its proof is "non-constructive", i.e., it does not provide any explicit recipe for the construction of a basis in the general case (that would maybe be too much to expect). This general proof is based on an interesting principle of infinitary combinatorics known as *Zorn's Lemma* [Zornsches Lemma], which is logically equivalent (in the standard framework of set theory) with the *Axiom of Choice* [Auswahlaxiom].

We do not want to go into these details which are of a more foundational or logical nature. We mention these facts for the sake of completeness. The following is (a version of) Zorn's Lemma, which we then use without proof. For finite sets A, it can be proved from scratch.

Lemma 2.5.17 *Let $\mathcal{S} \neq \emptyset$ be a collection of subsets of the set A. Assume that $\mathcal{S}$ has the following property*

- *whenever $(S_i)_{i\in I}$ is a family of sets in $\mathcal{S}$ such that for any two members S_i and S_j of this family either $S_i \subseteq S_j$ or $S_j \subseteq S_i$, then $\bigcup_{i\in I} S_i \in \mathcal{S}$.*

Then $\mathcal{S}$ has maximal elements. *A maximal element is an $S \in \mathcal{S}$ such that $S \subsetneq S'$ for* no *$S' \in \mathcal{S}$.*

Theorem 2.5.18 *Every vector space has a basis. A basis may be obtained as a maximal element in $\mathcal{S} := \{S \subseteq V : S \text{ linearly independent }\}$.*

The proof on the basis of the lemma is not difficult. One firstly verifies that this $\mathcal{S}$ has the property required in the lemma. One then shows that *any* maximal element in this system $\mathcal{S}$ is a basis. Any element of $\mathcal{S}$ is linearly independent by definition. It remains to argue that a maximal element must be spanning for V.

Suppose S is a maximal element in $\mathcal{S}$. Assume $\mathrm{span}(S) \subsetneq V$. There must be some $\mathbf{v} \in V \setminus \mathrm{span}(S)$. But then $S \cup \{\mathbf{v}\} \supsetneq S$ is still linearly independent by Lemma 2.4.10. And this contradicts the maximality of S.

We concentrate again on the finite-dimensional case. The following is a "constructive" version of the above maximality argument. It re-proves Corollary 2.5.14. The analogue in the infinite-dimensional case can also be proved, with an application of Zorn's Lemma.

Lemma 2.5.19 *Let $S \subseteq V$ be linearly independent, V finite-dimensional. Then S can be extended to a basis $B \supseteq S$ of V.*

Proof. Construct an increasing chain of linearly independent subsets $S = S_0 \subsetneq S_1 \subsetneq \cdots \subsetneq S_m = S_{m+1} =: B$ until a basis is reached. Let $S_0 = S$. For $i = 0, 1, \ldots$ do

$$S_{i+1} := \begin{cases} S_i & \text{if } \operatorname{span}(S_i) = V \\ S_i \cup \{\mathbf{v}\} & \text{for some } \mathbf{v} \in V \setminus \operatorname{span}(S_i) \text{ else.} \end{cases}$$

The condition on the choice of $\mathbf{v}$ ensures that S_{i+1} will also be linearly independent. The sequence of the S_i becomes constant only when a basis is reached.

But this must happen before step $n+1$ if $\dim(V) = n$, because Corollary 2.5.11 tells us that we cannot have sets of more than n linearly independent vectors in V.

□

2.6 Products, sums and quotients of spaces

This section provides some of the standard methods to construct new vector spaces (or subspaces) from old – constructions that also occur naturally in applications when linear phenomena are being modelled by vector spaces. We accompany each construction principle with an account of the dimensions involved in the finite-dimensional case and of corresponding bases.

2.6.1 Direct products

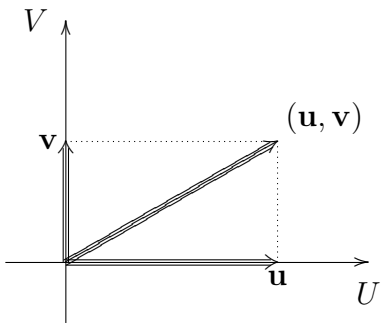
Definition 2.6.1 Let U and V be vector spaces over the same field $\mathbb{F}$. Their *direct product* [direktes Produkt] is the following $\mathbb{F}$-vector space W. The set of vectors of W is the cartesian product $W = U \times V = \{(\mathbf{u}, \mathbf{v}) \colon \mathbf{u} \in U, \mathbf{v} \in V\}$. Vector addition and scalar multiplication are defined component-wise according to

$$\begin{array}{rcl} +\colon (U \times V) \times (U \times V) & \longrightarrow & U \times V \\ ((\mathbf{u}, \mathbf{v}), (\mathbf{u}', \mathbf{v}')) & \longmapsto & (\mathbf{u} + \mathbf{u}', \mathbf{v} + \mathbf{v}') \end{array}$$

$$\begin{array}{rcl} \cdot\colon \mathbb{F} \times (U \times V) & \longrightarrow & U \times V \\ (\lambda, (\mathbf{u}, \mathbf{v})) & \longmapsto & (\lambda\mathbf{u}, \lambda\mathbf{v}) \end{array}$$

Exercise 2.6.1 Check that $U \times V$ with vector addition and scalar multiplication as given and with null vector $\mathbf{0} = (\mathbf{0}^U, \mathbf{0}^V) \in U \times V$ satisfies the vector space axioms.

The following represents the situation diagrammatically.



The direct product extends in a natural way to any number of factors. In particular, we write V^n for $\underbrace{V \times V \times \cdots \times V}_{n \text{ times}}$.

Noting that $\mathbb{F} = \mathbb{F}^1$ can be regarded as a 1-dimensional vector space over $\mathbb{F}$, we may re-interpret $\mathbb{F}^n$ – the standard n-dimensional $\mathbb{F}$-vector space – as the n-fold direct product of $\mathbb{F}$ with itself.

Proposition 2.6.2 *If $B^{(1)}$ is a basis of V_1 and $B^{(2)}$ a basis of V_2, then the following is a basis for the direct product $V = V_1 \times V_2$:*

$$B := \{(\mathbf{b}^{(1)}, \mathbf{0}^{(2)}) \colon \mathbf{b}^{(1)} \in B^{(1)}\} \cup \{(\mathbf{0}^{(1)}, \mathbf{b}^{(2)}) \colon \mathbf{b}^{(2)} \in B^{(2)}\},$$

where $\mathbf{0}^{(i)}$ stands for the null vector in V_i.

It follows that in the finite-dimensional case,

$$\dim(V_1 \times V_2) = \dim(V_1) + \dim(V_2).$$

Proof. We leave the verification of linear independence of B in $V_1 \times V_2$ as an exercise.

To see that $\operatorname{span}(B) = V_1 \times V_2$, note that if $\mathbf{v}^{(1)} = \sum_i \lambda_i^{(1)} \mathbf{b}_i^{(1)}$ is a representation of $\mathbf{v}^{(1)} \in V_1$ as a linear combination over $B^{(1)}$ and similarly $\mathbf{v}^{(2)} = \sum_j \mu_j^{(2)} \mathbf{b}_j^{(2)}$ in V_2 and over $B^{(2)}$, then in $V_1 \times V_2$:

$$(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}) = \sum_i \lambda_i^{(1)} (\mathbf{b}_i^{(1)}, \mathbf{0}) + \sum_j \mu_j^{(2)} (\mathbf{0}, \mathbf{b}_j^{(2)}).$$

□

2.6.2 Direct sums of subspaces

Definition 2.6.3 Let $U, W \subseteq V$ be subspaces of the $\mathbb{F}$-vector space V.

(i) The *sum* [Summe] of U and W, denoted $U+W$, is the subspace spanned by $U \cup W$.

(ii) If $U \cap W = \{\mathbf{0}\}$, the sum $U+W$ is called a *direct sum* [direkte Summe], written $U \oplus W$.

(iii) In the case that a direct sum spans all of V, $V = U \oplus W$, U and W are *linear complements* [Komplementärräume] of each other.

Observation 2.6.4 *For any two subspaces $U, W \subseteq V$:*

$$U + W = \{\mathbf{u} + \mathbf{w} \colon \mathbf{u} \in U, \mathbf{w} \in W\}.$$

If the sum is direct, then every $\mathbf{v} \in U \oplus W$ has a unique decomposition

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \quad \text{where} \quad \mathbf{u} \in U, \mathbf{w} \in W.$$

Proof. We only prove the uniqueness claim in $U \oplus W$. Let $\mathbf{v} = \mathbf{u} + \mathbf{w} = \mathbf{u}' + \mathbf{w}'$ where $\mathbf{u}, \mathbf{u}' \in U, \mathbf{w}, \mathbf{w}' \in W$.

It follows that $\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w} \in U \cap W$. As the sum is direct, $U \cap W = \{\mathbf{0}\}$ and thus $\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w} = \mathbf{0}$. So $\mathbf{u} = \mathbf{u}'$ and $\mathbf{w} = \mathbf{w}'$ follows.

□

This shows an interesting parallel between direct sums and products of two subspaces of the same vector space V.

Remark 2.6.5 Let subspaces $U_1, U_2 \subseteq V$ be such that $U_1 \cap U_2 = \{\mathbf{0}\}$. Then the direct product $U_1 \times U_2$ is isomorphic to the direct sum $U_1 \oplus U_2$.

Proof. In the light of the unique decomposition described in the observation above, we get an isomorphism of vector spaces (Definition 2.1.3) based on the mapping

$$\begin{array}{rccc} \varphi : & U_1 \times U_2 & \longrightarrow & U_1 \oplus U_2 \\ & (\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) & \longmapsto & \mathbf{u}^{(1)} + \mathbf{u}^{(2)}. \end{array}$$

□

Proposition 2.6.6 *Let $U_1, U_2 \subseteq V$ be finite-dimensional subspaces of V. Then*

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2).$$

Proof. Note first that $U_0 := U_1 \cap U_2$ is also a subspace of V (compare Lemma 2.2.3) and as it is contained in the given U_i, it must itself be finite-dimensional. Let $\dim(U_0) = n_0$ and let $B_0 = \{\mathbf{b}_1^{(0)}, \ldots, \mathbf{b}_{n_0}^{(0)}\}$ be a basis for U_0.

By Corollary 2.5.14, and as B_0 is a linearly independent subset both of U_1 and of U_2, we may extend B_0 in two different ways to obtain bases B_1 of U_1 and B_2 of U_2, respectively. Let $B_1 = B_0 \cup \{\mathbf{b}_1^{(1)}, \ldots, \mathbf{b}_{m_1}^{(1)}\}$ be the resulting basis of U_1, with m_1 pairwise distinct new basis vectors $\mathbf{b}_i^{(1)}$. Similarly let $B_2 = B_0 \cup \{\mathbf{b}_1^{(2)}, \ldots, \mathbf{b}_{m_2}^{(2)}\}$ be the basis of U_2, with m_2 pairwise distinct new basis vectors $\mathbf{b}_i^{(2)}$. So $\dim(U_1) = n_0 + m_1$ and $\dim(U_2) = n_0 + m_2$.

We claim that

$$B = B_0 \cup \{\mathbf{b}_1^{(1)}, \ldots, \mathbf{b}_{m_1}^{(1)}\} \cup \{\mathbf{b}_1^{(2)}, \ldots, \mathbf{b}_{m_2}^{(2)}\}$$

is a basis for $U = U_1 + U_2$. The dimension formula then follows as the above implies $\dim(U) = n_0 + m_1 + m_2 = (n_0 + m_1) + (n_0 + m_2) - n_0 = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2)$.

It remains to show that B is a basis. To establish linear independence, assume

$$\mathbf{0} = \underbrace{\sum_{i=1,\ldots,n_0} \lambda_i^{(0)} \mathbf{b}_i^{(0)}}_{=:\mathbf{u}_0} + \underbrace{\sum_{j=1,\ldots,m_1} \lambda_j^{(1)} \mathbf{b}_j^{(1)}}_{=:\mathbf{u}_1} + \underbrace{\sum_{k=1,\ldots,m_2} \lambda_k^{(2)} \mathbf{b}_k^{(2)}}_{=:\mathbf{u}_2}.$$

We need to show that this linear combination is trivial.

The equation implies that $\mathbf{u}_1 = -\mathbf{u}_0 - \mathbf{u}_2 \in U_1 \cap U_2 = U_0$ and similarly $\mathbf{u}_2 = -\mathbf{u}_0 - \mathbf{u}_1 \in U_0$.

By uniqueness of coefficients over bases (Lemma 2.5.7) we find that, for instance $\mathbf{u}_2 = \mathbf{0}$ as $\mathbf{u}_1 \in U_0$ means it can also be expressed without contributions $\mathbf{b}_k^{(2)}$. Similarly, we conclude that $\mathbf{u}_1 = \mathbf{0}$, whence also $\mathbf{u}_0 = -\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}$.

Now each $\mathbf{u}_i = \mathbf{0}$, for $1 = 0, 1, 2$, is expressed as a linear combination over a basis; this linear combination must therefore be trivial, i.e., all coefficients are 0, as we wanted to show.

It is obvious that B spans $U_1 + U_2$.

□

Example 2.6.7 Consider two 2-dimensional subspaces $U_1, U_2 \subseteq \mathbb{R}^3$. If (and only if) $U_1 = U_2$ can $\dim(U_1 \cap U_2)$ be 2.

Otherwise, $U_1 \cup U_2$ spans all of $\mathbb{R}^3$ (why?), i.e., $U_1 + U_2 = \mathbb{R}^3$. With the above dimension formula we find that

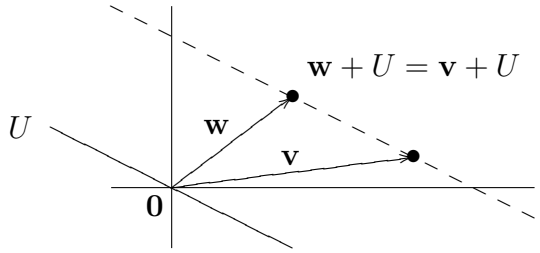
$$3 = \dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2) = 4 - \dim(U_1 \cap U_2),$$

whence $\dim(U_1 \cap U_2) = 1$ and we see that $U_1 \cap U_2$ is a line through $\mathbf{0}$.

Exercise 2.6.2 Transfer the reasoning of the example so the intersection of two 2-dimensional affine subspaces of $\mathbb{R}^3$.

2.6.3 Quotient spaces

Consider a subspace $U \subseteq V$ and imagine we want to look at vectors $\mathbf{v} \in V$ only up to components in U, i.e., we disregard any differences that lie within U. [Think of looking in the direction of U: what you see of $\mathbf{v}$ and $\mathbf{w}$ is the same; this is what we formalise as a quotient of V with respect to U in the following.]



We let $\equiv_U$ be the following equivalence relation on V:

$$\mathbf{v} \equiv_U \mathbf{w} \qquad \text{iff} \qquad \mathbf{w} - \mathbf{v} \in U.$$

This is the same as to say that $\mathbf{v} + U = \mathbf{w} + U$ (equality of affine subspaces).

Exercise 2.6.3 Verify that $\equiv_U$ is an equivalence relation on V (compare section 1.2.3).

The *quotient space* [Quotientenraum] $W := V/U$ has as its vectors the equivalence classes w.r.t. $\equiv_U$. Let us write

$$[\mathbf{v}]_U := \{\mathbf{w} \in V : \mathbf{v} \equiv_U \mathbf{w}\} = \mathbf{v} + U$$

for the equivalence class of $\mathbf{v}$.

We check that the following are well defined (independent of the representatives we pick from each equivalence class) as operations on these equivalence classes:

$$\begin{array}{rcl} [\mathbf{v}_1]_U + [\mathbf{v}_2]_U & := & [\mathbf{v}_1 + \mathbf{v}_2]_U, \\ \lambda \cdot [\mathbf{v}]_U & := & [\lambda \mathbf{v}]_U. \end{array}$$

For $+$, for instance, we need to check that, if $\mathbf{v}_1 \equiv_U \mathbf{v}_1'$ and $\mathbf{v}_2 \equiv_U \mathbf{v}_2'$ then also $(\mathbf{v}_1 + \mathbf{v}_2) \equiv_U (\mathbf{v}_1' + \mathbf{v}_2')$ (so that the resulting class $[\mathbf{v}_1 + \mathbf{v}_2]_U = [\mathbf{v}_1' + \mathbf{v}_2']_U$ is the same).

For this observe that $\mathbf{v}_1 \equiv_U \mathbf{v}_1'$ and $\mathbf{v}_2 \equiv_U \mathbf{v}_2'$ imply that $\mathbf{v}_1' = \mathbf{v}_1 + \mathbf{u}_1$ and $\mathbf{v}_2' = \mathbf{v}_2 + \mathbf{u}_2$ for suitable $\mathbf{u}_1, \mathbf{u}_2 \in U$. But then $(\mathbf{v}_1' + \mathbf{v}_2') = (\mathbf{v}_1 + \mathbf{v}_2) + (\mathbf{u}_1 + \mathbf{u}_2)$ and as $\mathbf{u}_1 + \mathbf{u}_2 \in U$, $(\mathbf{v}_1 + \mathbf{v}_2) \equiv_U (\mathbf{v}_1' + \mathbf{v}_2')$ follows.

The corresponding well-definedness for scalar multiplication is checked in a similar way.

Definition 2.6.8 Let $U \subseteq V$ be a subspace of the $\mathbb{F}$-vector space V. The *quotient space* [Quotientenraum] V/U is the $\mathbb{F}$-vector space whose vectors are the equivalence classes $[\mathbf{v}]_U = \mathbf{v} + U$ with respect to $\equiv_U$, with addition and scalar multiplication as defined above. Its null vector is the equivalence class $[\mathbf{0}^V]_U = \mathbf{0} + U = U$.

Exercise 2.6.4 Check the vector space axioms for V/U.

Quotients and direct sums Consider a direct sum $V = W \oplus U$, so that W is a complement of U in V.

We claim that every equivalence class (affine subspace) $[\mathbf{v}]_U = \mathbf{v} + U$ contains precisely one vector $\mathbf{w} \in W$.

Existence: as $V = W + U$, we know that $\mathbf{v} = \mathbf{w} + \mathbf{u}$ for some choice of $\mathbf{w} \in W$ and $\mathbf{u} \in U$. Then $[\mathbf{v}]_U = [\mathbf{w}]_U$ and $\mathbf{w} \in W \cap [\mathbf{v}]_U$.

Uniqueness: Let $\mathbf{w}, \mathbf{w}' \in [\mathbf{v}]_U$ for $\mathbf{w}, \mathbf{w}' \in W$. Then $\mathbf{w} - \mathbf{w}' \in U$. As the sum $V = W \oplus U$ is direct, and as $\mathbf{w} - \mathbf{w}' \in W \cap U = \{\mathbf{0}\}$, we find that $\mathbf{w} - \mathbf{w} = \mathbf{0}$ and $\mathbf{w} = \mathbf{w}'$ shows uniqueness as claimed.

This observation is the basis of the following.

Lemma 2.6.9 *$(W \oplus U)/U$ is isomorphic to W: a quotient space w.r.t. a subspace U is isomorphic to any complement of U in V.*

Proof. Let, for $[\mathbf{v}]_U \in V/U$, its image $\varphi(\mathbf{v}) \in W$ be the unique $\mathbf{w} \in W \cap [\mathbf{v}]_U$. We check that $\varphi\colon V/U \to W$ is a vector space isomorphism.

φ is bijective: injectivity was shown above;
surjectivity: $\varphi([\mathbf{w}]_U) = \mathbf{w}$ for every $\mathbf{w} \in W$.

φ is compatible with vector addition and scalar multiplication [check this as an exercise!].

□

We get a corollary about (finite) dimensions, if we observe that in the above situation $\dim(V) = \dim(W) + \dim(U)$ and $\dim(V/U) = \dim(W)$ due to isomorphy.

Corollary 2.6.10 *If V is finite-dimensional, then*

$$\dim(V/U) = \dim(V) - \dim(U)$$

for any subspace $U \subseteq V$.

Chapter 3

Linear Maps

3.1 Linear maps as homomorphisms

We consider maps from one $\mathbb{F}$-vector space to another.

Definition 3.1.1 A function $\varphi\colon V \to W$ between $\mathbb{F}$-vector spaces V and W (the same $\mathbb{F}$!) is called *linear* (a *linear map* or *linear function* [lineare Abbildung]) if for all $\mathbf{v}, \mathbf{v}' \in V$:

$$\varphi(\mathbf{v} + \mathbf{v}') = \varphi(\mathbf{v}) + \varphi(\mathbf{v}')$$

and for all $\mathbf{v} \in V$ and $\lambda \in \mathbb{F}$:

$$\varphi(\lambda\mathbf{v}) = \lambda\varphi(\mathbf{v}).$$

If φ is linear, it follows that it is compatible with arbitrary linear combinations. One shows by induction on n that for any $\mathbf{v}_i \in V$ and $\lambda_i \in \mathbb{F}$, $i = 1, \ldots, n$:

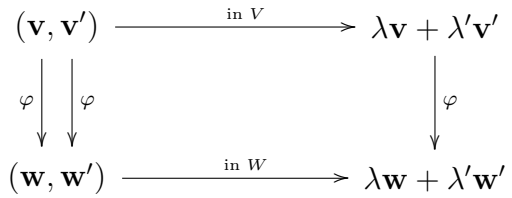
$$\varphi\big(\sum_{i=1,\ldots,n} \lambda_i \mathbf{v}_i\big) = \sum_{i=1,\ldots,n} \lambda_i \varphi(\mathbf{v}_i).$$

Observation 3.1.2 *If $\varphi\colon V \to W$ is linear, then $\varphi(\mathbf{0}^V) = \mathbf{0}^W$.*

Proof. By linearity, $\varphi(\mathbf{0}) = \varphi(\lambda\mathbf{0}) = \lambda\varphi(\mathbf{0})$ for all $\lambda \in \mathbb{F}$. Choosing $\lambda = 0 \in \mathbb{F}$, we see that $\varphi(\mathbf{0}) = 0\varphi(\mathbf{0}) = \mathbf{0}$. □

Linear maps are precisely the maps that preserve linear structure: they are compatible with vector addition and scalar multiplication and preserve null vectors. However, they need neither be injective nor surjective in general.

The diagram indicates this compatibility for a simple linear combination of two vectors with two arbitrary scalars.



We have seen such structure preserving maps already when we discussed isomorphisms between vector spaces (compare Definition 2.1.3). Vector space isomorphisms are particular instances of linear maps between vector spaces, namely bijective linear maps.

Remark 3.1.3 Structure preserving maps are called *homomorphisms* [Homomorphismen]. For vector spaces, they are called vector space homomorphisms. $\mathbb{F}$-vector space homomorphisms are precisely the linear maps between $\mathbb{F}$-vector spaces.

Homomorphisms may be classified according to their primary properties as maps (in particular injectivity and/or surjectivity) as follows. [1]

Zoology of (vector space) homomorphisms

Definition 3.1.4 Let V and W be $\mathbb{F}$-vector spaces, $\varphi\colon V \to W$ a linear map (i.e., an $\mathbb{F}$-vector space homomorphism). Then φ is

(i) an *epimorphism* [Epimorphismus] iff it is surjective.

(ii) a *monomorphism* [Monomorphismus] iff it is injective.

(iii) an *isomorphism* [Isomorphismus] iff it is bijective. [2]

[1] This classification of homomorphisms actually extends beyond the setting of vector spaces. Structure preserving maps in any other class of mathematical structures are similarly important and are classified according to the same terminology.

[2] Check that this is equivalent with Definition 2.1.3.

In the particular case where $W = V$, i.e., for a linear map $\varphi\colon V \to V$ from V to itself, one speaks of an *endomorphism* [Endomorphismus]. A bijective endomorphism – an isomorphism of V with itself – is called an *automorphism* [Automorphismus].

Vector space automorphisms play a special role: they are the *symmetries* of the linear structure of a vector space, namely permutations of the set of vectors that are compatible with vector addition, scalar multiplication and $\mathbf{0}$.

3.1.1 Images and kernels

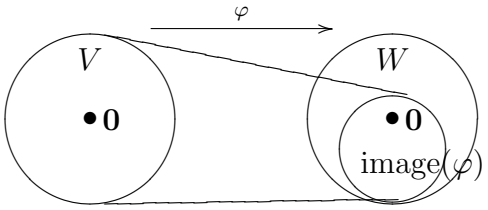
With any linear map $\varphi\colon V \to W$ between $\mathbb{F}$-vector spaces we associate two important sets (in fact subspaces): the image and the kernel of φ.

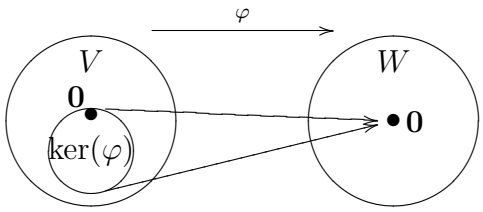
Definition 3.1.5 Let V and W be $\mathbb{F}$-vector spaces, $\varphi\colon V \to W$ linear. The *image* [Bild] of φ (as for any function) is defined to be the set

$$\text{image}(\varphi) = \{\varphi(\mathbf{v})\colon \mathbf{v} \in V\} \subseteq W.$$

The *kernel* [Kern] of φ is the set

$$\ker(\varphi) := \{\mathbf{v} \in V\colon \varphi(\mathbf{v}) = \mathbf{0}\} \subseteq V.$$





Lemma 3.1.6 $\text{image}(\varphi) \subseteq W$ *is a subspace of* W *and* $\ker(\varphi) \subseteq V$ *is a subspace of* V*.*

Proof. $\text{image}(\varphi) \subseteq W$ is a subspace. We check the closure conditions. Firstly, by linearity $\mathbf{0} = \varphi(\mathbf{0}) \in \text{image}(\varphi)$. Secondly, if $\mathbf{w}_1, \mathbf{w}_2 \in \text{image}(\varphi)$, then $\mathbf{w}_i = \varphi(\mathbf{v}_i)$ for suitable $\mathbf{v}_1, \mathbf{v}_2 \in V$. Therefore, as a consequence of

linearity, $\lambda_1\mathbf{w}_1 + \lambda_2\mathbf{w}_2 = \lambda_1\varphi(\mathbf{v}_1) + \lambda_2\varphi(\mathbf{v}_2) = \varphi(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2) \in \text{image}(\varphi)$, for any $\lambda_1, \lambda_2 \in \mathbb{F}$.

$\ker(\varphi) \subseteq V$ is a subspace. We check the closure conditions. Firstly, $\mathbf{0} \in \ker(\varphi)$ as $\varphi(\mathbf{0}) = \mathbf{0}$ is a consequence of linearity. Secondly, if $\mathbf{v}_1, \mathbf{v}_2 \in \ker(\varphi)$, then $\varphi(\mathbf{v}_i) = \mathbf{0}$. Linearity implies that $\varphi(\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2) = \lambda_1\varphi(\mathbf{v}_1)+\lambda_2\varphi(\mathbf{v}_2) = \mathbf{0}$ and hence $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 \in \ker(\varphi)$.

□

The following explains the great importance of kernels in the analysis of linear maps.

Lemma 3.1.7 *Let V and W be $\mathbb{F}$-vector spaces, $\varphi\colon V \to W$ linear. Then φ is injective (a monomorphism) iff* $\ker(\varphi) = \{\mathbf{0}\}$.

Proof. Assume first φ is injective. Then any $\mathbf{w} \in \text{image}(\varphi)$ has just one pre-image in V: there is precisely one $\mathbf{v} \in V$ such that $\varphi(\mathbf{v}) = \mathbf{w}$. For $\mathbf{w} = \mathbf{0}$ (in W) this implies that, $\mathbf{0}$ (in V) is the only vector in $\ker(\varphi)$.

Conversely, if φ is not injective, then there are $\mathbf{u} \neq \mathbf{v}$ in V such that $\varphi(\mathbf{u}) = \varphi(\mathbf{v})$. By linearity $\varphi(\mathbf{u} - \mathbf{v}) = \varphi(\mathbf{u}) - \varphi(\mathbf{v}) = \mathbf{0}$. As $\mathbf{u} \neq \mathbf{v}$ implies that $\mathbf{u} - \mathbf{v} \neq \mathbf{0}$, we find that $\ker(\varphi) \neq \{\mathbf{0}\}$.

□

3.1.2 Linear maps, bases and dimensions

Most of the basic assertions in this section extend to the infinite-dimensional case, with strictly analogous proofs. However, we restrict attention to the finite-dimensional case for the sake of simplicity.

Recall the notion of *labelled bases* from Definition 2.5.8.

The following are important existence and uniqueness properties for linear maps in terms of prescribed values on a basis.

Proposition 3.1.8 *Let V and W be $\mathbb{F}$-vector spaces,* $\dim(V) = n$ *and $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ a labelled basis of V. Let $f\colon \{\mathbf{b}_1, \ldots, \mathbf{b}_n\} \to W$ be an arbitrary function that maps every basis vector $\mathbf{b}_i$ in B to some vector $f(\mathbf{b}_i) \in W$. Then there is a unique linear map $\varphi\colon V \to W$ with $\varphi(\mathbf{b}_i) = f(\mathbf{b}_i)$ for all i. In other words, a linear map is fully determined by its values on any set of vectors that form a basis – and any choice of image vectors for these basis vectors can be realised by some linear map.*

Proof. Uniqueness. Let φ be as stipulated. As the basis spans V, any $\mathbf{v} \in V$ has a representation as $\mathbf{v} = \sum_i \lambda_i \mathbf{b}_i$ ([3]) for suitable $\lambda_i \in \mathbb{F}$. By linearity, $\varphi(\mathbf{v})$ is determined as

$$\varphi(\mathbf{v}) = \varphi(\textstyle\sum_i \lambda_i \mathbf{b}_i) = \displaystyle\sum_i \lambda_i \varphi(\mathbf{b}_i) = \sum_i \lambda_i f(\mathbf{b}_i).$$

This implies that φ is uniquely determined by f and the linearity requirement.

Existence. We know from Lemma 2.5.7 that the representation of $\mathbf{v} \in V$ as a linear combination $\mathbf{v} = \sum_i \lambda_i \mathbf{b}_i$ uniquely determines the coefficients $\lambda_i \in \mathbb{F}$. φ is therefore well-defined as a function $\varphi\colon V \to W$ by putting

$$\varphi(\mathbf{v}) := \sum_i \lambda_i f(\mathbf{b}_i) \quad \text{for} \quad \mathbf{v} = \sum_i \lambda_i \mathbf{b}_i.$$

Clearly this stipulation yields a function $\varphi\colon V \to W$, which agrees with f on the basis vectors $\mathbf{b}_i$. Moreover, φ is linear. For instance, if $\mathbf{v}' = \lambda\mathbf{v}$, then $\varphi(\mathbf{v}') = \varphi(\lambda \sum_i \lambda_i \mathbf{b}_i) = \varphi(\sum_i (\lambda\lambda_i)\mathbf{b}_i) = \sum_i (\lambda\lambda_i) f(\mathbf{b}_i) = \lambda \sum_i \lambda_i f(\mathbf{b}_i) = \lambda\varphi(\mathbf{v})$. Compatibility with vector addition is checked analogously. □

Lemma 3.1.9 *Let $B \subseteq V$ be a basis of V, $\varphi\colon V \to W$ a linear map. Let $\varphi(B) := \{\varphi(\mathbf{b})\colon \mathbf{b} \in B\}$. Then $\varphi(B)$ spans* $\mathrm{image}(\varphi) \subseteq W$.

Proof. Let $\mathbf{w} \in \mathrm{image}(\varphi)$, i.e., $\mathbf{w} = \varphi(\mathbf{v})$ for some $\mathbf{v} \in V$. If $\mathbf{v} = \sum_i \lambda_i \mathbf{b}_i$ then $\mathbf{w} = \varphi(\mathbf{v}) = \varphi(\sum_i \lambda_i \mathbf{b}_i) = \sum_i \lambda_i \varphi(\mathbf{b}_i) \in \mathrm{span}(\varphi(B))$. □

Lemma 3.1.10 *Let $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ be a labelled basis of V and $\varphi\colon V \to W$ an injective linear map (a monomorphism).*

Then $(\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_n))$ is a labelled basis of $\mathrm{image}(\varphi) \subseteq W$. *It follows that* $\dim(\mathrm{image}(\varphi)) = \dim(V)$ *and that*

$$\begin{array}{rcl} \varphi'\colon V & \longrightarrow & \mathrm{image}(\varphi) \subseteq W \\ \mathbf{v} & \longmapsto & \varphi(\mathbf{v}) \end{array}$$

is a vector space isomorphism.

[3] We sometimes drop the explicit bounds in summations if they are clear from context or do not matter. We write, for instance just $\mathbf{v} = \sum_i \lambda_i \mathbf{b}_i$ instead of $\mathbf{v} = \sum_{i=1}^n \lambda_i \mathbf{b}_i$ when working over a fixed labelled basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$.

Proof. Let φ be injective. As $\mathbf{b}_i \neq \mathbf{b}_j$ for $i \neq j$, injectivity implies that also $\varphi(\mathbf{b}_i) \neq \varphi(\mathbf{b}_j)$ for $i \neq j$. Hence the members of $(\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_n))$ are pairwise distinct.

By the previous lemma, $\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_n)$ span image(φ). It remains to show that the $\varphi(\mathbf{b}_i)$ are linearly independent in W. Let $\sum_i \lambda_i \varphi(\mathbf{b}_i) = \mathbf{0}$. Therefore $\varphi(\sum_i \lambda_i \mathbf{b}_i) = \sum_i \lambda_i \varphi(\mathbf{b}_i) = \mathbf{0}$. As $\ker(\varphi) = \{\mathbf{0}\}$ (Lemma 3.1.7), this implies that $\sum_i \lambda_i \mathbf{b}_i = \mathbf{0}$. As the $\mathbf{b}_i$ form a basis, $\lambda_i = 0$ for all i and the linear combination $\sum_i \lambda_i \varphi(\mathbf{b}_i) = \mathbf{0}$ is trivial. Hence the $\varphi(\mathbf{b}_i)$ are also linearly independent, and therefore form a labelled basis. Clearly φ' is surjective onto image(φ) and hence we obtain an isomorphism between V and image(φ).

□

In the special case of a vector space isomorphism $\varphi\colon V \to W$, image(φ) = W and hence $(\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_n))$ is a labelled basis of W. This implies the following.

Corollary 3.1.11 (a) *If $\varphi\colon V \to W$ is an isomorphism and $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ a labelled basis of V, then $(\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_n))$ is a labelled basis of W.*

(b) *Let $\dim(V) = \dim(W) = n$. Any choice of labelled bases $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of V and $(\mathbf{b}'_1, \ldots, \mathbf{b}'_n)$ of W gives rise to a unique vector space isomorphism φ between V and W such that $\varphi(\mathbf{b}_i) = \mathbf{b}'_i$.*

Proof. (b) is a direct application of Proposition 3.1.8.

(a) restates Lemma 3.1.10 for isomorphisms (surjective monomorphisms).

□

The assertions of the corollary are particularly interesting also for the case of automorphisms, i.e., isomorphisms $\varphi\colon V \to V$. Here we see that φ switches from one labelled bases of V to another.

The dimension formula We already saw that isomorphisms preserve dimensions, that monomorphisms have kernel dimension 0 and that the image under a monomorphism has the same dimension as the domain. These observation are generalised in the following important dimension formula.

Theorem 3.1.12 *Let V, W be finite-dimensional $\mathbb{F}$-vector spaces, $\varphi\colon V \to W$ a linear map (vector space homomorphism). Then*

$$\dim(V) = \dim(\ker(\varphi)) + \dim(\text{image}(\varphi)).$$

Proof. Let $\dim(V) = n$, $\dim(\ker(\varphi)) = k$.

We firstly choose a labelled basis $B_k = (\mathbf{b}_1, \ldots, \mathbf{b}_k)$ of the subspace $\ker(\varphi) \subseteq V$. Next we extend B_k to a basis of all of V, with a further $\ell = n - k$ many vectors $\mathbf{a}_1, \ldots, \mathbf{a}_\ell$ such that $B = (\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{a}_1, \ldots, \mathbf{a}_\ell)$ becomes a labelled basis of V. [NB: $\ell = n - k = 0$ is a possibility, namely if $\ker(\varphi) = V$.]

We claim that $(\varphi(\mathbf{a}_1), \ldots, \varphi(\mathbf{a}_\ell))$ is a labelled basis of $\text{image}(\varphi)$. That proves the dimension formula, as it implies that $\dim(\text{image}(\varphi)) = \ell = n - k$.

Clearly $\varphi(\mathbf{a}_1), \ldots, \varphi(\mathbf{a}_\ell) \in \text{image}(\varphi)$. These vectors are also pairwise distinct: suppose $\varphi(\mathbf{a}_i) = \varphi(\mathbf{a}_j)$ for some $1 \leqslant i < j \leqslant \ell$, then $\varphi(\mathbf{a}_i - \mathbf{a}_j) = \mathbf{0}$ and hence $\mathbf{a} := \mathbf{a}_i - \mathbf{a}_j$ a vector in $\ker(\varphi)$. So $\mathbf{a}$ is a linear combination $\mathbf{a} = \sum_{i=1}^{k} \lambda_i \mathbf{b}_i$ and therefore $\mathbf{a}_i - \mathbf{a}_j - \sum_{i=1}^{k} \lambda_i \mathbf{b}_i = \mathbf{0}$ is a non-trivial linear combination of $\mathbf{0}$ over a basis, which is a contradiction.

A similar argument shows that the $\varphi(\mathbf{a}_i)$ are linearly independent: if $\sum_{i=1}^{\ell} \mu_i \varphi(\mathbf{a}_i) = \mathbf{0}$, then $\varphi(\sum_{i=1}^{\ell} \mu_i \mathbf{a}_i) = \mathbf{0}$, whence $\mathbf{a} := \sum_{i=1}^{\ell} \mu_i \mathbf{a}_i \in \ker(\varphi)$. Then $\mathbf{a} = \sum_{i=1}^{k} \lambda_i \mathbf{b}_i$ for suitable λ_i and therefore $\sum_{i=1}^{k} \lambda_i \mathbf{b}_i - \sum_{i=1}^{\ell} \mu_i \mathbf{a}_i = \mathbf{0}$ would be a non-trivial linear combination of $\mathbf{0}$ over a basis, which is a contradiction.

Now $\text{image}(\varphi)$ is spanned by $\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_k), \varphi(\mathbf{a}_1), \ldots, \varphi(\mathbf{a}_\ell)$, and (as $\varphi(\mathbf{b}_i) = \mathbf{0}$) also by just $\varphi(\mathbf{a}_1), \ldots, \varphi(\mathbf{a}_\ell)$. So $(\varphi(\mathbf{a}_1), \ldots, \varphi(\mathbf{a}_\ell))$ is a labelled basis of $\text{image}(\varphi)$.

□

Exercise 3.1.1 Show that a homomorphism $\varphi\colon V \to W$ between finite-dimensional $\mathbb{F}$-vector spaces V and W, is surjective iff $\dim(\text{image}(\varphi)) = \dim(W)$.

Exercise 3.1.2 For a homomorphism $\varphi\colon V \to W$ let $U := \ker(\varphi) \subseteq V$, and consider the quotient space V/U with elements $\mathbf{v} + U$. Show that the following map is a well-defined monomorphism:

$$\begin{array}{rcccc} \varphi_U & : & V/U & \longrightarrow & W \\ [\mathbf{v}]_U = \mathbf{v} + U & & & \longmapsto & \varphi(\mathbf{v}). \end{array}$$

3.2 Vector spaces of homomorphisms

3.2.1 Linear structure on homomorphisms

Definition 3.2.1 For $\mathbb{F}$-vector spaces V, W, $\mathrm{Hom}(V, W)$ denotes the set of all linear maps $\varphi\colon V \to W$.

$\mathrm{Hom}(V, W)$ carries natural linear structure. Namely, there are natural addition operation (point-wise addition of functions) and scalar multiplication (point-wise multiplication by a scalar for functions). Namely, for $\varphi, \psi \in \mathrm{Hom}(V, W)$ and $\lambda \in \mathbb{F}$ the following maps are also members of $\mathrm{Hom}(V, W)$:

$$\begin{array}{rcll} (\varphi+\psi)\colon V & \longrightarrow & W & \\ \mathbf{v} & \longmapsto & \varphi(\mathbf{v})+\psi(\mathbf{v}). & \text{(point-wise addition)} \end{array}$$

$$\begin{array}{rcll} (\lambda\varphi)\colon V & \longrightarrow & W & \\ \mathbf{v} & \longmapsto & \lambda\varphi(\mathbf{v}). & \text{(point-wise scalar multiplication)} \end{array}$$

With these two operations, $\mathrm{Hom}(V, W)$ is itself an $\mathbb{F}$-vector space.

Proposition 3.2.2 $\mathrm{Hom}(V, W)$ *with point-wise addition and scalar multiplication and with the constant map* $\mathbf{0}\colon \mathbf{v} \mapsto \mathbf{0} \in W$ *for all* $\mathbf{v} \in V$*, is an* $\mathbb{F}$*-vector space.*

Proof. $\big(\mathrm{Hom}(V, W), +, \mathbf{0}\big)$ is an abelian group. Associativity and commutativity of (point-wise!) addition follow from associativity and commutativity of vector addition in W; addition of the constant-zero map $\mathbf{0}$ operates trivially, whence this constant-zero map is the neutral element w.r.t. addition; for $\varphi \in \mathrm{Hom}(V, W)$, the map $-\varphi\colon V \to W$ with $\mathbf{v} \mapsto -\varphi(\mathbf{v})$ acts as an inverse w.r.t. addition.

Point-wise scalar multiplication inherits associativity from associativity of scalar multiplication in W; the neutral element is $1 \in \mathbb{F}$.

Both distributivity laws follow from the corresponding laws in W. For instance, let $\lambda \in F$, $\varphi, \psi \in \mathrm{Hom}(V, W)$. Then, for any $\mathbf{v} \in V$:

$$\begin{aligned} [\lambda(\varphi+\psi)](\mathbf{v}) = \lambda\big([\varphi+\psi](\mathbf{v})\big) &= \lambda\big(\varphi(\mathbf{v})+\psi(\mathbf{v})\big) \\ &= \lambda\varphi(\mathbf{v})+\lambda\psi(\mathbf{v}) \;=\; [\lambda\varphi](\mathbf{v})+[\lambda\psi](\mathbf{v}). \end{aligned}$$

Hence $\lambda(\varphi+\psi) = \lambda\varphi + \lambda\psi$ in $\mathrm{Hom}(V, W)$.

□

Exercise 3.2.1 Let $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $(\mathbf{b}'_1, \ldots, \mathbf{b}'_m)$ be labelled bases for V and W respectively. Show that the following maps φ_{ij} are linear, pairwise distinct and linearly independent, and that they span $\mathrm{Hom}(V, W)$. For $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$, let

$$\begin{array}{rcll} \varphi_{ij}\colon V & \longrightarrow & W & \\ \mathbf{v} & \longmapsto & \lambda_j \mathbf{b}'_i & \text{where } \mathbf{v} = \sum_{j=1,\ldots,n} \lambda_j \mathbf{b}_j. \end{array}$$

So these maps form a labelled basis for $\mathrm{Hom}(V, W)$ and it follws that $\dim(\mathrm{Hom}(V, W)) = nm = \dim(V)\dim(W)$.

We shall later see again, in a different manner, that the dimension of $\mathrm{Hom}(V, W)$ as an $\mathbb{F}$-vector space, for finite-dimensional V, W, is the product of the dimensions of V and W (compare Theorem 3.3.2).

3.2.2 The dual space

NB: some of the assertions of this section really rely on finite dimension. We only consider finite-dimensional $\mathbb{F}$-vector spaces for the whole section.

The dual space is a special case of a space $\mathrm{Hom}(V, W)$, namely with $W = \mathbb{F} = \mathbb{F}^1$ the standard one-dimensional $\mathbb{F}$-vector space.

Definition 3.2.3 For any $\mathbb{F}$-vector space V, the $\mathbb{F}$-vector space $\mathrm{Hom}(V, \mathbb{F})$, with point-wise addition and scalar multiplication and constant-zero function as null-vector, is called the *dual space* [Dualraum] of V. It is denoted $V^* := \mathrm{Hom}(V, \mathbb{F})$.

Exercise 3.2.2 Check that the vector space structure on $\mathrm{Hom}(V, \mathbb{F})$ (as a special case of $\mathrm{Hom}(V, W)$) is the same as if we consider $\mathrm{Hom}(V, \mathbb{F})$ as a subspace of $\mathcal{F}(V, \mathbb{F})$ (the $\mathbb{F}$-vector space of all $\mathbb{F}$-valued functions on domain V). For this, firstly establish that $\mathrm{Hom}(V, \mathbb{F}) \subseteq \mathcal{F}(V, \mathbb{F})$ is a subspace, and secondly that the operations it thus inherits from $\mathcal{F}(V, \mathbb{F})$ are the same as those introduced in the context of $\mathrm{Hom}(V, W)$ for $W = \mathbb{F}$.

Example 3.2.4 Consider the left-hand side of a linear equation $E\colon a_1 x_1 + \cdots + a_n x_n = b$ over $\mathbb{F}^n$ as a map from $V = \mathbb{F}^n$ to $\mathbb{F}$:

$$\begin{array}{rcl} \varphi_E\colon \mathbb{F}^n & \longrightarrow & \mathbb{F} \\ (x_1, \ldots, x_n) & \longmapsto & a_1 x_1 + \cdots + a_n x_n. \end{array}$$

This is a linear map, and hence a member of $(\mathbb{F}^n)^*$. Note that the solution set of the associated homogeneous linear equation $E^*\colon a_1x_1 + \cdots + a_nx_n = 0$ is the kernel of this map: $S(E^*) = \ker(\varphi_E)$.

Example 3.2.5 Consider $V = \mathbb{F}^n$ with standard basis $B = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$.

For $i = 1, \ldots, n$, consider the linear map

$$\begin{array}{rcl} \eta_i\colon \mathbb{F}^n & \longrightarrow & \mathbb{F} \\ (\lambda_1, \ldots, \lambda_n) & \longmapsto & \lambda_i. \end{array}$$

η_i picks out the i-th component of every vector in $\mathbb{F}^n$.

Clearly $\eta_i \in (\mathbb{F}^n)^*$, and in fact $B^* := (\eta_1, \ldots, \eta_n)$ forms a basis of $(\mathbb{F}^n)^*$. B^* is called the *dual basis* of $V^* = (\mathbb{F}^n)^*$ associated with the given basis $B = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ of $V = \mathbb{F}^n$.

For linear independence, let $\eta = \sum_i \lambda_i\eta_i = \mathbf{0}$ be a linear combination of the constant-zero function, which has value $0 \in \mathbb{F}$ on all of $\mathbb{F}^n$. Apply η to the basis vector $\mathbf{e}_j$ to see that

$$0 = \eta(\mathbf{e}_j) = \sum_i \lambda_i\eta_i(\mathbf{e}_j) = \lambda_j.$$

Therefore $\lambda_j = 0$ for all j follows, whence B^* consists of linearly independent vectors.

To see that $(\mathbb{F}^n)^* = \mathrm{span}(B^*)$, let $\varphi\colon \mathbb{F}^n \to \mathbb{F}$ be linear. By Proposition 3.1.8 we know that φ is uniquely determined by its values on the basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$. In order to represent φ as a linear combination of the η_i, we just need to find coefficients λ_i such that the linear combination $\sum_i \lambda_i\eta_i$ returns the same values as φ on each $\mathbf{e}_j$ for $j = 1, \ldots, n$. Then $\varphi = \sum_i \lambda_i\eta_i$ follows by linearity alone.

As $(\sum_i \lambda_i\eta_i)(\mathbf{e}_j) = \lambda_j$ we merely need to put $\lambda_j := \varphi(\mathbf{e}_j)$ for $j = 1, \ldots, n$. So, for any $\varphi \in (\mathbb{F}^n)^*$:

$$\varphi = \sum_i \varphi_i(\mathbf{e}_i)\ \eta_i.$$

These considerations extend to arbitrary labelled bases $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of any n-dimensional space V. The corresponding linear maps, that form the *dual basis* $B^* = (\mathbf{b}_1^*, \ldots, \mathbf{b}_n^*)$ of V^* associated with the given basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of V, are the maps

$$\begin{array}{rcll} \mathbf{b}_i^*\colon V & \longrightarrow & \mathbb{F} & \\ \mathbf{v} & \longmapsto & \lambda_i & \text{where } \mathbf{v} = \sum_{i=1,\ldots,n} \lambda_i\mathbf{b}_i. \end{array}$$

Observation 3.2.6 *Let V be finite-dimensional. Then every labelled basis of V induces a dual basis of V^*. For each basis vector $\mathbf{b}$ in the basis of V the dual basis has a basis vector $\mathbf{b}^*$ which maps $\mathbf{b}$ to $1 \in \mathbb{F}$ and all other basis vectors to $0 \in \mathbb{F}$. As these bases have the same number of vectors,* $\dim(V) = \dim(V^*)$.

3.3 Linear maps and matrices

3.3.1 Matrix representation of linear maps

In this section we investigate the connection between $m \times n$ matrices over $\mathbb{F}$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{F}^{(m,n)}$$

and linear maps from some n-dimensional $\mathbb{F}$-vector space V to some m-dimensional $\mathbb{F}$-vector space W.

Standard spaces with standard bases

Think of the standard spaces $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$ of respective dimensions. We may read the above matrix as consisting of a sequence of n many column vectors

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{F}^m,$$

inducing the linear map

$$\begin{array}{rcl} \varphi_A \colon \mathbb{F}^n & \longrightarrow & \mathbb{F}^m \\ (x_1, \ldots, x_n) & \longmapsto & x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \sum_{j=1}^n x_j\mathbf{a}_j. \end{array}$$

In other words, φ_A is the vector space homomorphism from $\mathbb{F}^n$ to $\mathbb{F}^m$ that is determined by the stipulation that

$$\varphi_A(\mathbf{e}_j) := \mathbf{a}_j,$$

where

$$\mathbf{e}_j = (0, \ldots, 0, \underset{j}{1}, 0, \ldots, 0) \in \mathbb{F}^n$$

is the j-th basis vector of the standard labelled basis for $\mathbb{F}^n$, with entries 0 in slots $i \neq j$ and entry 1 in position j. That this stipulation precisely determines a unique linear map φ_A was shown in Proposition 3.1.8.

Looking at the computation of image vectors under this map, it is convenient now to think of the entries of $\mathbf{v} \in \mathbb{F}^n$ and its image $\varphi(\mathbf{v}) \in \mathbb{F}^m$ as *column vectors* [Spaltenvektoren].

If $\mathbf{v} = (x_1, \ldots, x_n)$, we obtain the entries y_i in $\varphi(\mathbf{v}) = (y_1, \ldots, y_m)$ according to [compare multiplication, Definition 3.3.3]

$$\begin{pmatrix} y_1 \\ \vdots \\ \boxed{y_i} \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{ij}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix} = \left(\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \hline a_{i1} & a_{i2} & \rightarrow & a_{in} \\ \hline \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}\right) \begin{pmatrix} x_1 \\ x_2 \\ \\ \downarrow \\ \\ x_n \end{pmatrix}$$

Example 3.3.1 Consider $V = W = \mathbb{R}^2$ and a 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{R}^{(2,2)}.$$

In terms of the standard basis $(\mathbf{e}_1, \mathbf{e}_2) = \big((1,0),(0,1)\big)$, A describes the map

$$\begin{array}{rcl} \varphi\colon \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ (x,y) & \longmapsto & x\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + y\begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = \begin{pmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \end{array}$$

For instance, we have the following familiar linear transformations of the real plane, represented by the following matrices in $\mathbb{R}^{(2,2)}$ in terms of the standard basis $(\mathbf{e}_1, \mathbf{e}_2)$:

(i) the identity $\mathrm{id}_{\mathbb{R}^2}$, represented by the *unit matrix*: $E_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

(ii) scalar transformation with factor λ: $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \lambda E_2$.

(iii) reflection in the line through $\mathbf{0}$ and $(1,1)$: $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

(iv) rotation through angle α: $\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$.

(v) a projection onto the line through $\mathbf{0}$ and $(1,1)$: $\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$.

(vi) a shear transformation along the x-axis: $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

It is clear that any linear map $\varphi\colon \mathbb{F}^n \mapsto \mathbb{F}^m$, $\varphi \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)$, is representable (with respect to the standard bases of these spaces) by a matrix $A_\varphi \in \mathbb{F}^{(m,n)}$ in this fashion. We merely put, as the j-th column vector of A_φ, the image of the j-th basis vector $\mathbf{e}_j \in V$ under φ:

$$\varphi(\mathbf{e}_j) =: \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{F}^m.$$

The association between $\mathbb{F}^{(m,n)}$ and $\mathrm{Hom}(V,W)$ generalises to arbitrary n- and m-dimensional $\mathbb{F}$-vector spaces V and W and arbitrary choices of labelled bases for them. This is discussed in the following section.

The matrix of a homomorphism w.r.t. to arbitrary bases

Let V and W be finite-dimensional $\mathbb{F}$-vector spaces. Fix labelled bases $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ for V, and $\hat{B} = (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_m)$ for W.

Consider $\varphi \in \mathrm{Hom}(V,W)$. The representation of φ with respect to the chosen bases, as an $m \times n$ matrix $A = [\![\varphi]\!]^B_{\hat{B}}$ is obtained as follows.

Let $\mathbf{w}_j := \varphi(\mathbf{b}_j)$ and express $\mathbf{w}_j$ in terms of the basis $(\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_m)$ as $\mathbf{w}_j = \sum_{i=1}^m a_{ij}\hat{\mathbf{b}}_i$. These coefficients $a_{1j}, \ldots, a_{mj}$ are therefore determined according to $(a_{1j}, \ldots, a_{mj}) = [\![\mathbf{w}_j]\!]_{\hat{B}}$. We now take these coefficients $a_{1j}, \ldots, a_{mj}$ to form the j-th column vector in the matrix $A = [\![\varphi]\!]^B_{\hat{B}} = (a_{ij})_{1 \leqslant i \leqslant m; 1 \leqslant j \leqslant n} \in \mathbb{F}^{(m,n)}$.

$$[\![\varphi(\mathbf{b}_j)]\!]_{\hat{B}} = (a_{ij})_{1 \leqslant i \leqslant m} \qquad \longrightarrow \qquad \left(\begin{array}{ccc|c|ccc} a_{11} & \cdots & & a_{1j} & & \cdots & a_{1n} \\ \vdots & & & \downarrow & & & \vdots \\ a_{m1} & \cdots & & a_{mj} & & \cdots & a_{mn} \end{array}\right)$$

Then, for any $\mathbf{v} \in V$ with representation $(\lambda_1, \dots \lambda_n)$ in terms of the basis $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$, i.e, with $(\lambda_1, \dots \lambda_n) = [\![\mathbf{v}]\!]_B$,

$$\mathbf{v} = \sum_{j=1}^{n} \lambda_j \mathbf{b}_j$$

implies that its image under φ is

$$\varphi(\mathbf{v}) = \sum_{j=1}^{n} \lambda_j \varphi(\mathbf{b}_j) = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} \lambda_j \Big)\, \hat{\mathbf{b}}_i.$$

This means that the representation $(\mu_1, \dots, \mu_m)$ of $\varphi(\mathbf{v})$ in terms of the basis $\hat{B} = (\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_m)$ is

$$[\![\varphi(\mathbf{v})]\!]_{\hat{B}} = (\mu_1, \dots, \mu_m) \quad \text{where} \quad \mu_i = \sum_{j=1}^{n} a_{ij} \lambda_j.$$

In terms of just the coefficients in the respective bases B and $\hat{B}$, we get the same relationship as if we interpreted everything over $\mathbb{F}^n$ and $\mathbb{F}^m$ with the standard bases. For $\mathbf{v} = \sum_{j=1}^{n} \lambda_j \mathbf{b}_j$ as above, we obtain the coefficients μ_i of $\varphi(\mathbf{v})$ in $\varphi(\mathbf{v}) = \sum_{i=1}^{m} \mu_i \hat{\mathbf{b}}_i$ according to

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \boxed{\mu_i} \\ \vdots \\ \mu_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}\lambda_j \\ \vdots \\ \sum_{j=1}^{n} a_{ij}\lambda_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}\lambda_j \end{pmatrix} = \left(\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \hline a_{i1} & a_{i2} & \rightarrow & a_{in} \\ \hline \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array} \right) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \\ \downarrow \\ \\ \lambda_n \end{pmatrix}$$

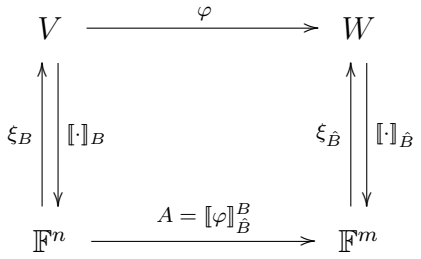
Diagrammatically, the following clarifies the view of φ involved in this representation. Recall that, for instance for V, since $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ is a basis, the maps

$$\begin{array}{rcl} [\![\cdot]\!]_B \colon V & \longrightarrow & \mathbb{F}^n \\ \mathbf{v} & \longmapsto & [\![\mathbf{v}]\!]_B = (\lambda_1, \dots, \lambda_n) \text{ if } \mathbf{v} = \sum_{j=1}^{n} \lambda_j \mathbf{b}_j \end{array}$$

and its inverse

$$\begin{array}{rcl} \xi_B \colon \mathbb{F}^n & \longrightarrow & V \\ (\lambda_1, \dots, \lambda_n) & \longmapsto & \sum_{j=1}^{n} \lambda_j \mathbf{b}_j \end{array}$$

are vector space isomorphisms. Similarly, for the basis $\hat{B} = (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_m)$ of W we have a corresponding isomorphism $[\![\cdot]\!]_{\hat{B}} \colon W \to \mathbb{F}^m$ and its inverse $\xi_{\hat{B}} \colon \mathbb{F}^m \to W$. Then the relationship between the map described by $A = [\![\varphi]\!]^B_{\hat{B}}$ on the coefficients and the map φ itself is indicated in the following commuting diagram:



Note that the representation of φ by $A = [\![\varphi]\!]^B_{\hat{B}} \in \mathbb{F}^{(m,n)}$ crucially depends on the chosen labelled bases in both the domain V and the range W of φ. In general the same homomorphism φ is represented by entirely different matrices, if different bases are used.

See section 3.3.3 in particular. Some of the key ideas in linear algebra revolve around the following:

- to separate essential features of the underlying homomorphisms from the purely incidental features of representational matrices.
- to isolate those properties of the representational matrices that capture properties of the underlying homomorphisms rather than being incidental to the choice of bases.

Exercise 3.3.1 Let $\dim(V) = n$.

(i) Show that the identity transformation $\mathrm{id}_V \colon V \to V$ is represented by the unit matrix E_n whenever the same labelled basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ is chosen for V both as domain and as range. In other words, $[\![\mathrm{id}_V]\!]^B_B = E_n$ for every B.

(ii) Precisely which homomorphisms $\varphi \colon V \to W$ can have the unit matrix E_n as their representation for suitable choices of bases $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of V and $(\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_n)$ of W? In other words, what does $[\![\varphi]\!]^B_{\hat{B}} = E_n$ for *some* choice of B and $\hat{B}$ imply about φ?

Converse to the above passage from homomorphisms to matrices, we may interpret any matrix $A \in \mathbb{F}^{(m,n)}$ as a representation of a homomorphism $\varphi \in \mathrm{Hom}(V,W)$, w.r.t. any choice of labelled bases in V and W.

Let $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ a labelled basis for V, $\hat{B} = (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_m)$ a labelled basis for W. Starting now from an arbitrary $m \times n$ matrix over $\mathbb{F}$, $A \in \mathbb{F}^{(m,n)}$, we find that there is a unique homomorphism $\varphi\colon V \to W$ for which $[\![\varphi]\!]^B_{\hat{B}} = A$:

$$\begin{array}{rcccl} \varphi = \varphi_A^{B\hat{B}} & : & V & \longrightarrow & W \\ & & \sum_{j=1}^n \lambda_j \mathbf{b}_j & \longmapsto & \sum_{i=1}^m \sum_{j=1}^n a_{ij}\lambda_j \hat{\mathbf{b}}_i. \end{array}$$

Note that φ is uniquely determined by linearity and the stipulation that it maps $\mathbf{b}_j$ to $\mathbf{w}_j := \sum_{i=1}^m a_{ij}\hat{\mathbf{b}}_i$ for $j = 1, \ldots, n$. Note also that again the j-th column in matrix A is taken to consist of the coefficients of the image $\varphi(\mathbf{b}_j)$ of the j-th basis vector of V, expressed as a linear combination over the chosen basis of W.

A similar diagram describes the relationship; but we now think of the transformation matrix A as given, and read it as representing a homomorphism $\varphi = \varphi_A^{B\hat{B}}$ for which $[\![\varphi]\!]^B_{\hat{B}} = A$:

$$\begin{array}{ccc} \mathbb{F}^n & \xrightarrow{\quad A \quad} & \mathbb{F}^m \\ {\scriptstyle [\![\cdot]\!]_B}\uparrow \;\downarrow{\scriptstyle \xi_B} & & {\scriptstyle [\![\cdot]\!]_{\hat{B}}}\uparrow \;\downarrow{\scriptstyle \xi_{\hat{B}}} \\ V & \xrightarrow{\quad \varphi = \varphi_A^{B\hat{B}} \quad} & W \end{array}$$

It is clear that these associations between $\mathrm{Hom}(V,W)$ and $\mathbb{F}^{(m,n)}$,

$$\begin{array}{rcl} \mathbb{F}^{(m,n)} & \longrightarrow & \mathrm{Hom}(V,W) \\ A & \longmapsto & \varphi_A^{B\hat{B}} \end{array}$$

and

$$\begin{array}{rcl} \mathrm{Hom}(V,W) & \longrightarrow & \mathbb{F}^{(m,n)} \\ \varphi & \longmapsto & [\![\varphi]\!]^B_{\hat{B}} \end{array}$$

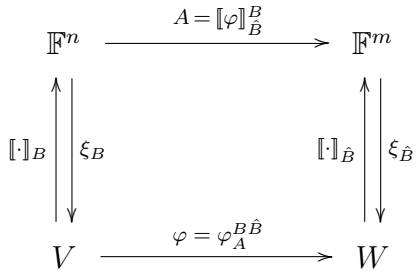
are inverses of each other. They are also compatible with the linear structure of $\mathrm{Hom}(V,W)$ and $\mathbb{F}^{(m,n)}$ as $\mathbb{F}$-vector spaces, and hence constitute an isomorphism between these two vector spaces.

In both directions we work with a fixed choice of labelled bases B and $\hat{B}$ for V and W. Different choices of bases would yield different isomorphisms.

Theorem 3.3.2 *Let V, W be finite-dimensional $\mathbb{F}$-vector spaces of dimension n and m, respectively. Fix labelled bases $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $\hat{B} = (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_m)$ of V and W, respectively. Then the $\mathbb{F}$-vector spaces*

- *$\mathbb{F}^{(m,n)}$, the $\mathbb{F}$-vector space of $m \times n$ matrices over $\mathbb{F}$, and*
- *$\mathrm{Hom}(V, W)$, the $\mathbb{F}$-vector space of homomorphisms from V to W,*

are isomorphic via the association between $\varphi \in \mathrm{Hom}(V, W)$ with its representation $[\![\varphi]\!]^B_{\hat{B}} \in \mathbb{F}^{(m,n)}$ and (conversely) between $A \in \mathbb{F}^{(m,n)}$ and $\varphi^{B\hat{B}}_A \in \mathrm{Hom}(V, W)$. It follows that $\dim(\mathrm{Hom}(V, W)) = \dim(\mathbb{F}^{(m,n)}) = mn$.



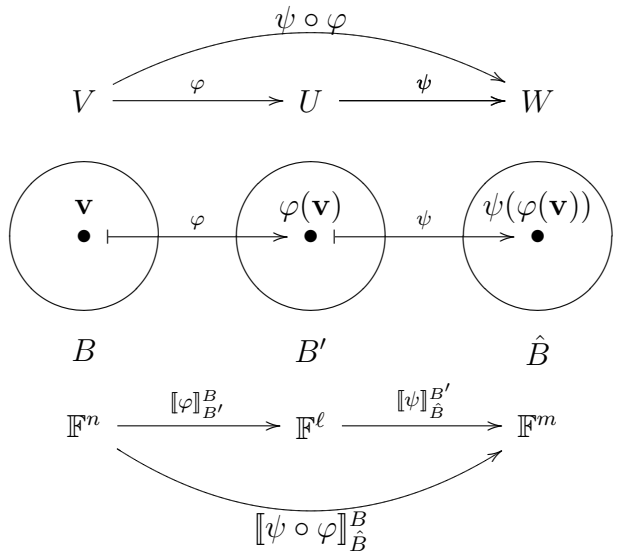
Exercise 3.3.2 Find representations of some of the sample endomorphisms $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ from Example 3.3.1 above w.r.t. other labelled bases. In particular, for (iii) and (v), a basis which uses a basis vector in the direction of the distinguished line of fixed points of these maps would be of interest.

Composition and matrix multiplication

The association between homomorphisms and matrices – based on distinguished labelled bases in the constituent vector spaces – tells us that any operation on homomorphisms has a natural counterpart in an operation on matrices.

The most important operation on maps is that of composition; translated into the language of the associated matrices it gives rise to the operation of matrix multiplication.

Consider three finite-dimensional $\mathbb{F}$-vector spaces, U, V, W, $\dim(V) = n$, $\dim(U) = \ell$, $\dim(W) = m$, and homomorphisms $\varphi \in \mathrm{Hom}(V, U)$ and $\psi \in \mathrm{Hom}(U, W)$.



It is easily checked that the composition $\psi \circ \varphi \colon V \to W$ is linear, and hence $\psi \circ \varphi \in \mathrm{Hom}(V, W)$.

Fix labelled bases
$$\begin{cases} B &= (\mathbf{b}_1, \ldots, \mathbf{b}_n) & \text{for } V; \\ B' &= (\mathbf{b}'_1, \ldots, \mathbf{b}'_\ell) & \text{for } U; \\ \hat{B} &= (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_m) & \text{for } W. \end{cases}$$

Let, w.r.t. these bases
$$\begin{cases} G &:= [\![\varphi]\!]^B_{B'} & \in \mathbb{F}^{(\ell,n)} \\ H &:= [\![\psi]\!]^{B'}_{\hat{B}} & \in \mathbb{F}^{(m,\ell)} \\ C &:= [\![\psi \circ \varphi]\!]^B_{\hat{B}} & \in \mathbb{F}^{(m,n)} \end{cases}$$

Then the coefficients of C are uniquely determined by those of G and H, according to:

$$\begin{aligned} (\psi \circ \varphi)(\mathbf{b}_j) &= \textstyle\sum_{i=1}^m c_{ij} \hat{\mathbf{b}}_i \\ &= \psi\big(\varphi(\mathbf{b}_j)\big) \\ &= \psi\Big(\textstyle\sum_{k=1}^\ell g_{kj} \mathbf{b}'_k\Big) && [G = [\![\varphi]\!]^B_{B'}] \\ &= \textstyle\sum_{k=1}^\ell g_{kj} \psi(\mathbf{b}'_k) \\ &= \textstyle\sum_{k=1}^\ell g_{kj} \sum_{i=1}^m h_{ik} \hat{\mathbf{b}}_i && [H = [\![\psi]\!]^{B'}_{\hat{B}}] \\ &= \textstyle\sum_{i=1}^m \Big(\sum_{k=1}^\ell h_{ik} g_{kj}\Big) \hat{\mathbf{b}}_i. \end{aligned}$$

This means that for $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$:

$$c_{ij} = \sum_{k=1}^{\ell} h_{ik} g_{kj}.$$

The summation that produces the coefficient in the i-th row and j-th column of C extends over the consecutive products of coefficients in the i-th row of H and the j-th column of G, as indicated in the diagram:

$$\begin{pmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1n} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & \boxed{\mathbf{c_{ij}}} & \cdots & c_{in} \\ \vdots & & \vdots & & \vdots \\ c_{m1} & \cdots & c_{mj} & \cdots & c_{mn} \end{pmatrix}$$

$$\Uparrow$$

$$\left(\begin{array}{cccc} h_{11} & h_{12} & \cdots & h_{1\ell} \\ \vdots & \vdots & & \vdots \\ \hline \mathbf{h_{i1}} & \mathbf{h_{i2}} & \rightarrow & \mathbf{h_{i\ell}} \\ \hline \vdots & \vdots & & \vdots \\ h_{m1} & h_{m2} & \cdots & h_{m\ell} \end{array}\right) \quad \left(\begin{array}{cc|c|cc} g_{11} & \cdots & \mathbf{g_{1j}} & \cdots & g_{1n} \\ g_{21} & \cdots & \mathbf{g_{2j}} & \cdots & g_{2n} \\ & & & & \\ \vdots & & \downarrow & & \vdots \\ & & & & \\ g_{\ell 1} & \cdots & \mathbf{g_{\ell j}} & \cdots & g_{\ell n} \end{array}\right)$$

Definition 3.3.3 The *product of matrices* [Matrizenprodukt] is defined as follows. Let $n, m, \ell \geqslant 1$.

Let $A \in \mathbb{F}^{(m,\ell)}$, $A = (a_{ik})_{1 \leqslant i \leqslant m; 1 \leqslant k \leqslant \ell}$, with m rows and ℓ columns;

Let $B \in \mathbb{F}^{(\ell,n)}$, $B = (b_{kj})_{1 \leqslant k \leqslant \ell; 1 \leqslant j \leqslant n}$, with ℓ rows and n columns.

Then the matrix product AB is the matrix $C \in \mathbb{F}^{(m,n)}$, $C = (c_{ij})_{1 \leqslant i \leqslant m; 1 \leqslant j \leqslant n}$, with m rows and n columns, whose entries are

$$c_{ij} = \sum_{k=1}^{\ell} a_{ik} b_{kj}.$$

Note that the number of columns of A must be the same as the number of rows in B. These conditions precisely match the requirements that A and B represent homomorphisms φ_A and φ_B that can be composed in the order

$\varphi_A \circ \varphi_B$. The resulting product matrix AB has the format to represent the homomorphism $\varphi_A \circ \varphi_B$, with as many rows as A and as many columns as B.

We also note that this definition of matrix multiplication covers the way in which we obtained coefficients of image vectors $\varphi(\mathbf{v})$ from coefficients of $\mathbf{v}$ and the representation of φ by a matrix $A = [\![\varphi]\!]^B_{\hat{B}}$ w.r.t. bases B and $\hat{B}$:

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \boxed{\mu_i} \\ \vdots \\ \mu_m \end{pmatrix} = \left(\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \hline a_{i1} & a_{i2} & \rightarrow & a_{in} \\ \hline \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}\right) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \\ \downarrow \\ \\ \lambda_n \end{pmatrix}$$

is just matrix multiplication between the $m \times n$ matrix A and the $n \times 1$ matrix (i.e., column vector from $\mathbb{F}^n$) which consists of the coefficients $[\![\mathbf{v}]\!]_B = (\lambda_1, \ldots, \lambda_n)$ and produces the coefficients $[\![\varphi(\mathbf{v})]\!]_{\hat{B}} = (\mu_1, \ldots, \mu_m)$ as an $m \times 1$ matrix (i.e., column vector from $\mathbb{F}^m$).

In the above analysis of the composition of homomorphisms we have shown the following.

Proposition 3.3.4 *If homomorphisms $\varphi\colon V \to U$ and $\psi\colon U \to W$ are represented w.r.t. chosen labelled bases B, B', $\hat{B}$ of each of these spaces by matrices $A_\varphi = [\![\varphi]\!]^B_{B'}$ and $A_\psi = [\![\psi]\!]^{B'}_{\hat{B}}$, then the composition $\psi \circ \varphi\colon V \to W$ is represented, w.r.t. bases B and $\hat{B}$ by the product matrix $A_{\psi\circ\varphi} = A_\psi A_\varphi$, i.e.,*

$$[\![\psi \circ \varphi]\!]^B_{\hat{B}} = [\![\psi]\!]^{B'}_{\hat{B}} \cdot [\![\varphi]\!]^B_{B'}.$$

Definition 3.3.5 The n-dimensional *unit matrix* [Einheitsmatrix] over $\mathbb{F}$ is the matrix $E_n \in \mathbb{F}^{(n,n)}$ with entries $a_{ii} = 1$ and $a_{ij} = 0$ for $1 \leqslant i \neq j \leqslant n$.

We collect some properties of matrix multiplication, also in relation to the vector space operations on $\mathbb{F}^{(m,n)}$ (or $\mathrm{Hom}(V, W)$). Recall that the $m \times n$ null matrix (the null vector in $\mathbb{F}^{(m,n)}$) is the matrix with entries $0 \in \mathbb{F}$ throughout.

Lemma 3.3.6 *Whenever the matrices involved are such that the products are defined (we put superscripts to indicate the space $\mathbb{F}^{(m,n)}$ they come from), then*

(i) *matrix multiplication is associative:*

$$A^{(m,n)}(B^{(n,k)}C^{(k,\ell)}) = (A^{(m,n)}B^{(n,k)})C^{(k,\ell)}.$$

(ii) *matrix multiplication is distributive w.r.t. to addition:*

$$A^{(m,n)}(B^{(n,k)} + C^{(n,k)}) = A^{(m,n)}B^{(n,k)} + A^{(m,n)}C^{(n,k)}.$$

(iii) *the null matrix* $\mathbf{0}^{(n,m)}$ *annihilates any matrix:* $\mathbf{0}^{(n,m)}A^{(m,k)} = \mathbf{0}^{(n,k)}$ *and* $B^{(k,n)}\mathbf{0}^{(n,m)} = \mathbf{0}^{(k,m)}$.

(iv) *The* $n \times n$ *unit matrix* E_n *acts as a neutral element for matrix multiplication:* $A^{(m,n)}E_n = A^{(m,n)}$ *and* $E_nB^{(n,k)} = B^{(n,k)}$.

Proof. All assertions can be checked arithmetically. However, they also follow directly from corresponding properties of the composition of homomorphisms, with the use of Proposition 3.3.4.

E.g., (i) follows from associativity of the composition of maps.

(ii) reflects a corresponding distributivity of composition over addition of homomorphisms. One simply checks that

$$\psi \circ (\varphi_1 + \varphi_2) = (\psi \circ \varphi_1) + (\psi \circ \varphi_2),$$

for any $\psi \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)$ and $\varphi_i \in \mathrm{Hom}(\mathbb{F}^k, \mathbb{F}^n)$ say.

(iii) uses the fact that the null map executed before or after any other homomorphism, results in the null map.

(iv) finally relies on the observation that the unit matrix E_n represents the identity endomorphism $\mathrm{id}_{\mathbb{F}^n}$.

□

Example 3.3.7 Matrix multiplication is not commutative. In terms of homomorphisms of $\mathbb{R}^2$, for instance, the composition of a rotation through angle $\alpha = \pi/2$ does not commute with the reflection in the line through $(1, 1)$. The two products of the corresponding matrices therefore must be different. In fact,

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

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

Exercise 3.3.3 Compute the matrix products of the matrices representing a rotation through angle α and a reflection in the line through $\mathbf{0}$ at an angle β in terms of the standard basis $(\mathbf{e}_1, \mathbf{e}_2)$ of $\mathbb{R}^2$, and determine for which values of α and β these two matrices do commute.

Exercise 3.3.4 Over any field $\mathbb{F}$, show that:

(a) the matrices $A \in \mathbb{F}^{(2,2)}$ for which $AB = BA$ for all $B \in \mathbb{F}^{(2,2)}$ are precisely the matrices λE_2, for $\lambda \in \mathbb{F}$.

(b) matrix multiplication is commutative on the subset of *diagonal* matrices in $\mathbb{F}^{(n,n)}$. These are the matrices $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ with $a_{ij} = 0$ for all $i \neq j$.

$\mathbf{Hom}(V, V)$ and $\mathbb{F}^{(n,n)}$ as rings

Let V be an n-dimensional $\mathbb{F}$-vector space V and fix a labelled basis B of V. Then the matrices $A = [\![\varphi]\!]_B^B$ associated with endomorphisms $\varphi \in \mathrm{Hom}(V, V)$, $\varphi \colon V \to V$, correspond to the space of all $n \times n$ square matrices $A \in \mathbb{F}^{(n,n)}$.

With matrix addition and multiplication, $\mathbb{F}^{(n,n)}$ forms a ring

$$\big(\mathbb{F}^{(n,n)}, +, \cdot, \mathbf{0}, E_n\big).$$

Via the bijective correspondence with $\mathrm{Hom}(V, V)$, $A \mapsto \varphi_A^{BB}$, this structure is isomorphic to

$$\big(\mathrm{Hom}(V, V), +, \circ, \mathbf{0}, \mathrm{id}_V\big),$$

the *ring of endomorphisms of* V [Endomorphismenring]. [4]

Theorem 3.3.8 *For any $n \geqslant 1$ and for any n-dimensional $\mathbb{F}$-vector space V, the rings*

- $\big(\mathbb{F}^{(n,n)}, +, \cdot, \mathbf{0}, E_n\big)$
- $\big(\mathrm{Hom}(V, V), +, \circ, \mathbf{0}, \mathrm{id}_V\big)$

are isomorphic. For every choice of a labelled basis B for V, the association between endomorphisms and their matrix representations w.r.t. B induces an isomorphism between these two rings.

[4] Note that $\mathbf{0} \in \mathbb{F}^{(n,n)}$ is the null matrix (all entries equal $0 \in \mathbb{F}$) while $\mathbf{0} \in \mathrm{Hom}(V, V)$ is the null endomorphism, i.e., the constant map that sends all $\mathbf{v} \in V$ to the null vector $\mathbf{0} \in V$.

3.3.2 Invertible homomorphisms and regular matrices

Let $\varphi \in \mathrm{Hom}(V, W)$, V, W finite-dimensional $\mathbb{F}$-vector spaces.

We know from our general considerations in section 3.1.2 that the homomorphism $\varphi\colon V \to W$ is injective iff $\ker(\varphi) = \{\mathbf{0}\}$. Moreover, φ is surjective iff $\mathrm{image}(\varphi) = W$ iff $\dim(\mathrm{image}(\varphi)) = \dim(W)$.

From the dimension formula, φ is bijective (an isomorphism) iff $\ker(\varphi) = \{\mathbf{0}\}$ and $\dim(V) = \dim(W)$.

Invertible homomorphisms

A homomorphism $\varphi \in \mathrm{Hom}(V, W)$, $\varphi\colon V \to W$, is *invertible* [invertierbar] iff it has an inverse $\varphi^{-1}\colon W \to V$ such that

$$\varphi \circ \varphi^{-1} = \mathrm{id}_W \text{ and } \varphi^{-1} \circ \varphi = \mathrm{id}_V.$$

It is easy to show that the inverse φ^{-1} of a linear map φ is necessarily linear. So $\varphi^{-1} \in \mathrm{Hom}(W, V)$.

As the existence of an invertible homomorphism in $\mathrm{Hom}(V, V)$ implies that V and W are isomorphic, one may essentially restrict attention to the case in which $W = V$ and study invertible endomorphisms (i.e., automorphisms) of V.

Definition 3.3.9 Let V be an $\mathbb{F}$-vector space. $\mathrm{Aut}(V)$ stands for the subset of $\mathrm{Hom}(V, V)$ consisting of the automorphisms of V.

Note that the null endomorphism, which is the null element of the endomorphism ring $\mathrm{Hom}(V, V)$ is not present in $\mathrm{Aut}(V)$. $\mathrm{Aut}(V)$ is not a ring. On the other hand, invertibility means that the multiplicative operation $\circ$ (composition) has inverses, and $\mathrm{Aut}(V)$ forms a group with $\circ$ and neutral element id_V. ([5]) The following is easily proved by checking the axioms.

Theorem 3.3.10 $\big(\mathrm{Aut}(V), \circ, \mathrm{id}_V\big)$ *is a group.*

[5] Far more generally, the automorphisms (i.e., self-isomorphisms or symmetries) of any mathematical object form a group, with composition as the group operation and with the identity map as the trivial symmetry acting as the neutral element.

Regular matrices and $\mathrm{GL}_n(\mathbb{F})$

Via the association between $\mathrm{Hom}(V,V)$ with $\mathbb{F}^{(n,n)}$ for $n = \dim(V)$ – based on any choice of a labelled basis for V – we obtain a subset of $\mathbb{F}^{(n,n)}$ corresponding to $\mathrm{Aut}(V)$. This subset consists of those $n \times n$ matrices over $\mathbb{F}$ that possess an inverse w.r.t. matrix multiplication. With Theorem 3.3.10 and via the isomorphism asserted in Theorem 3.3.8 we already know that this subset forms a group.

Definition 3.3.11 A matrix $A \in \mathbb{F}^{(n,n)}$ is called *regular* [regulär] iff it possesses an inverse A^{-1} with respect to matrix multiplication, i.e., if there is an $A^{-1} \in F^{(n,n)}$ such that

$$AA^{-1} = A^{-1}A = E_n.$$

$\mathrm{GL}_n(\mathbb{F}) \subseteq \mathbb{F}^{(n,n)}$ stands for the set of regular $n \times n$ matrices over $\mathbb{F}$.

Definition 3.3.12 The *general linear group* $\big(\mathrm{GL}_n(\mathbb{F}), \cdot, E_n\big)$ is the group consisting of the regular $n \times n$ matrices over $\mathbb{F}$, with matrix multiplication as the group operation and with the n-dimensional unit matrix E_n as its neutral element.

Recall that for $n = 0$ we let $\mathbb{F}^0 = \{\mathbf{0}\}$ be the $\mathbb{F}$-vector space consisting of just the null vector. Matching that, we let $\mathrm{Aut}(\mathbb{F}^0)$ consist of just the identity (mapping $\mathbf{0}$ to itself) and $\mathrm{GL}_0(\mathbb{F})$ the group consisting of just its neutral element.

Exercise 3.3.5 Check that the following are equivalent for $A \in \mathbb{F}^{(n,n)}$:

(i) there is a $B \in \mathbb{F}^{(n,n)}$ such that $AB = E_n$.

(ii) there is a $B \in \mathbb{F}^{(n,n)}$ such that $BA = E_n$.

Hint: work in $\mathrm{Aut}(\mathbb{F}^n)$.

Exercise 3.3.6 Let $A, B \in \mathrm{GL}_n(\mathbb{F})$. Show that $(AB)^{-1} = B^{-1}A^{-1}$. [This rule follows from the group axioms, but here you may also work in $\mathrm{Aut}(\mathbb{F}^n)$.]

Exercise 3.3.7 Show that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{F}^{(2,2)}$ is invertible iff $ad - bc \neq 0$. [Hint: argue via endomorphisms and consider linear independence of (a, c) and (b, d).] Can you give explicit formulae for the entries in A^{-1}?

Exercise 3.3.8 What is $\mathrm{GL}_1(\mathbb{F})$?

We shall return to the problem of finding the inverse of a matrix in the next chapter.

3.3.3 Change of basis transformations

A particular point of interest in the choice of bases for the representation of vectors and homomorphisms is the effect of *basis transformations* [Basistransformationen]: the effect of switching from one basis of V to another on these representations.

Change of basis: coefficients of vectors

Let $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $B' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_n)$ be labelled bases of V. Any vector $\mathbf{v} \in V$ has a unique representation with respect to either basis.

$$\begin{aligned}&\mathbf{v} = \textstyle\sum_i^n \lambda_i \mathbf{b}_i = \sum_i \lambda'_i \mathbf{b}'_i \quad \text{where}\\ &(\lambda_1, \ldots, \lambda_n) = [\![\mathbf{v}]\!]_B \text{ and } (\lambda'_1, \ldots, \lambda'_n) = [\![\mathbf{v}]\!]_{B'}.\end{aligned}$$

What is the relationship between $[\![\mathbf{v}]\!]_B$ and $[\![\mathbf{v}]\!]_{B'}$? We want to capture the transformation $[\![\mathbf{v}]\!]_B \mapsto [\![\mathbf{v}]\!]_{B'}$ as a transformation in $\mathbb{F}^n$. We know that the maps

$$\begin{array}{rcl} [\![\cdot]\!]_B \colon V & \longrightarrow & \mathbb{F}^n \\ \mathbf{v} & \longmapsto & [\![\mathbf{v}]\!]_B \end{array} \quad \text{with inverse} \quad \begin{array}{rcl} \xi_B \colon \mathbb{F}^n & \longrightarrow & V \\ (\lambda_1, \ldots, \lambda_n) & \longmapsto & \sum_i \lambda_i \mathbf{b}_i \end{array}$$

and

$$\begin{array}{rcl} [\![\cdot]\!]_{B'} \colon V & \longrightarrow & \mathbb{F}^n \\ \mathbf{v} & \longmapsto & [\![\mathbf{v}]\!]_{B'} \end{array} \quad \text{with inverse} \quad \begin{array}{rcl} \xi_{B'} \colon \mathbb{F}^n & \longrightarrow & V \\ (\lambda_1, \ldots, \lambda_n) & \longmapsto & \sum_i \lambda_i \mathbf{b}'_i, \end{array}$$

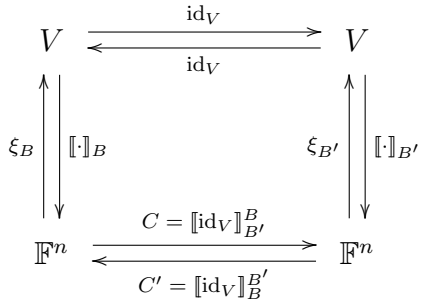
are vector space isomorphisms between V and $\mathbb{F}^n$. The desired transformation from coefficients λ_i to coefficients λ'_i is governed by the requirement that

$$(\lambda'_1, \ldots, \lambda'_n) = [\![\xi_B(\lambda_1, \ldots, \lambda_n)]\!]_{B'}.$$

Similarly for the converse transformation from from coefficients λ'_i to coefficients λ_i, the requirement is $(\lambda_1, \ldots, \lambda_n) = [\![\xi_{B'}(\lambda'_1, \ldots, \lambda'_n)]\!]_B$.

The transformations therefore are obtained as compositions $[\![\cdot]\!]_{B'} \circ \xi_B$ (from B to B') and $[\![\cdot]\!]_B \circ \xi_{B'}$ (from B' to B), respectively. Their matrix representations, C (for the transformation from B to B') and $C' = C^{-1}$ (for the transformation from B' to B) are best read off from the following diagram,

as representations of the identity map on V in corresponding (mixed) bases:



The transformation that maps $(\lambda_1, \ldots, \lambda_n)$ to $(\lambda'_1, \ldots, \lambda'_n)$ is thus

$$\lambda'_i = \sum_j c_{ij}\lambda_j,$$

or

$$\begin{pmatrix} \lambda'_1 \\ \vdots \\ \lambda'_n \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix},$$

where the j-th column of the matrix C consists just of the coefficients of $\mathbf{b}_j$ w.r.t. basis B', i.e.,

$$(c_{1j}, \ldots, c_{nj}) = [\![\mathbf{b}_j]\!]_{B'}.$$

Clearly, $CC' = C'C = E_n$, and $C' = C^{-1}$ is the inverse of C w.r.t. matrix multiplication. This can also be inferred from compositionality according to $[\![\mathrm{id}_V]\!]_B^{B'}[\![\mathrm{id}_V]\!]_{B'}^B = [\![\mathrm{id}_V \circ \mathrm{id}_V]\!]_B^B = [\![\mathrm{id}_V]\!]_B^B = E_n$.

Exercise 3.3.9 Determine the coefficients c_{ij} of the transformation matrix $C = (c_{ij}) \in \mathbb{F}^{(n,n)}$ directly from the requirement that

$$\mathbf{v} = \sum_j \lambda_j \mathbf{b}_j = \sum_j \sum_i \lambda_j c_{ij} \mathbf{b}'_i.$$

Change of basis: matrices of endomorphisms

What is the effect of a change of basis for V on the representation of a given endomorphism $\varphi \in \mathrm{Hom}(V, V)$?

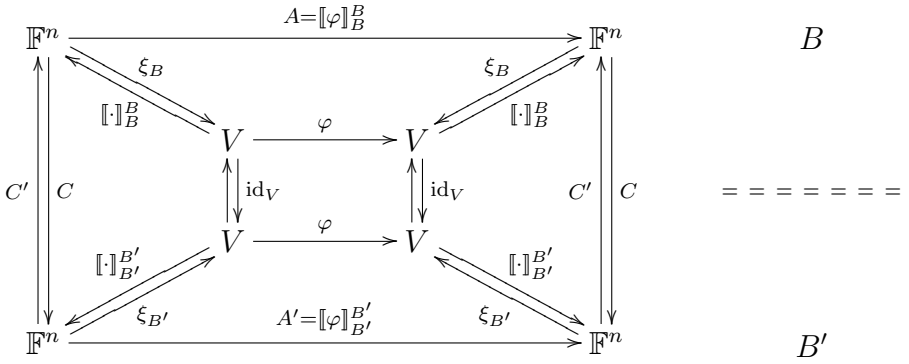
We saw above how the change of basis from $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ to $B' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_n)$ is expressed by the matrix $C = [\![\mathrm{id}_V]\!]_{B'}^B$. This matrix C and its inverse $C^{-1} = C' = [\![\mathrm{id}_V]\!]_B^{B'}$ are used here again.

Let $\varphi \in \mathrm{Hom}(V,V)$ and consider its representation by matrices $A = [\![\varphi]\!]^B_B$ w.r.t. B and $A' = [\![\varphi]\!]^{B'}_{B'}$ w.r.t. B'.

From compositionality we find that

$$[\![\mathrm{id}_V]\!]^B_{B'}[\![\varphi]\!]^B_B[\![\mathrm{id}_V]\!]^{B'}_B = [\![\mathrm{id}_V \circ \varphi \circ \mathrm{id}_V]\!]^{B'}_{B'} = [\![\varphi]\!]^{B'}_{B'},$$

i.e., $A' = [\![\mathrm{id}_V]\!]^B_{B'} A [\![\mathrm{id}_V]\!]^{B'}_B = CAC' = CAC^{-1}$.



Checking the same relationship by hand, consider $\mathbf{v} = \sum_j \lambda_j \mathbf{b}_j = \sum_j \lambda'_j \mathbf{b}'_j$ and its image under φ, $\varphi(\mathbf{v}) = \sum_i \mu_i \mathbf{b}_i = \sum_i \mu'_i \mathbf{b}'_i$.

Note that both λ_j and λ'_j, as well as the μ_i and μ'_i are related in the above manner by C and $C' = C^{-1}$. On the other hand, the μ_i are related to the λ_j via A, while the μ'_i are related to the λ'_j via A'. Together these relationships fully determine A' in terms of A and vice versa. We find that

$$\begin{array}{rcll}
\sum_j a'_{ij}\lambda'_j & = & \mu'_i & [A' = [\![\varphi]\!]^{B'}_{B'}] \\
& = & \sum_k c_{ik}\mu_k & [\mu \stackrel{C}{\longmapsto} \mu'] \\
& = & \sum_k c_{ik}\Big(\sum_\ell a_{k\ell}\lambda_\ell\Big) & [A = [\![\varphi]\!]^B_B] \\
& = & \sum_k c_{ik}\Big(\sum_\ell a_{k\ell}\big(\sum_j c'_{\ell j}\lambda'_j\big)\Big). & [\lambda' \stackrel{C'}{\longmapsto} \lambda] \\
& = & \sum_j \big(CAC^{-1}\big)_{ij}\lambda'_j. &
\end{array}$$

Hence we find again that

$$A' = CAC^{-1}$$

This relationship between different matrix representations of the same endomorphism is a central notion of matrix calculus.

Definition 3.3.13 Two matrices $A, B \in \mathbb{F}^{(n,n)}$ are *similar* [ähnlich], $A \sim B$, iff

$$A = CBC^{-1} \quad \text{for some regular } C \text{ with inverse } C^{-1}.$$

Lemma 3.3.14 *Similarity is an equivalence relation on the set $\mathbb{F}^{(n,n)}$ of all $n \times n$ matrices over $\mathbb{F}$.*

Proof. We need to show that $\sim$ is reflexive, symmetric and transitive:

(i) reflexivity: $A \sim A$ as $A = E_n A E_n$ (and $E_n = (E_n)^{-1}$).

(ii) symmetry: $A \sim B$ implies $B \sim A$, as $A = CBC^{-1}$ iff $B = C^{-1}AC$ (and $C = (C^{-1})^{-1}$).

(iii) transitivity: $A \sim B$ and $B \sim D$ implies $A \sim D$. Let $A = C_1 B (C_1)^{-1}$ and $B = C_2 D (C_2)^{-1}$; then $A = C_1 C_2 D (C_2)^{-1}(C_1)^{-1} = CDC^{-1}$ for $C := C_1 C_2$ (note that $(C_1 C_2)^{-1} = (C_2)^{-1}(C_1)^{-1}$).

□

Lemma 3.3.15 *Two matrices $A, A' \in \mathbb{F}^{(n,n)}$ are similar iff they are the representations of the same endomorphism of $\mathbb{F}^n$ with respect to some choice of two labelled bases of $\mathbb{F}^n$. In other words, the equivalence classes w.r.t. similarity are in a bijective correspondence with* $\mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^n)$.

Proof. Let $A = CA'C^{-1}$, i.e., $A' = C^{-1}AC$. Let $B = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be the standard basis of $\mathbb{F}^n$ and choose $\varphi \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^n)$ such that $[\![\varphi]\!]_B^B = A$ ($\varphi = \varphi_A$ is just multiplication with A).

Let $C = (c_{ij})_{1 \leqslant i,j \leqslant n}$. We observe that $C = [\![\mathrm{id}]\!]_B^{B'}$ if we let $B' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_n)$ consist of the column vectors of C:

$$\mathbf{b}'_j := \sum_j c_{ij}\mathbf{e}_i = \begin{pmatrix} c_{1j} \\ \vdots \\ c_{nj} \end{pmatrix}$$

Then $B' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_n)$ is another labelled basis of $\mathbb{F}^n$ (the columns of the regular matrix C always are) and A' is the matrix representation of φ w.r.t. this new basis B':

$$[\![\varphi]\!]_{B'}^{B'} = [\![\mathrm{id}]\!]_{B'}^{B} [\![\varphi]\!]_B^B [\![\mathrm{id}]\!]_B^{B'} = C^{-1}AC.$$

Conversely, we saw above that whenever A and A' represent the same endomorphism w.r.t. to any two bases B and B' then A and A' are similar.

□

3.3.4 Ranks

We consider the matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{F}^{(m,n)}$$

with *row vectors* $\mathbf{r}_i = \big(a_{i1}, \ldots, a_{in}\big) \in \mathbb{F}^n$ for $i = 1, \ldots, m$

and *column vectors* $\mathbf{c}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{F}^m$ for $j = 1, \ldots, n$.

With A we associate the following ranks:

- *row-rank*, the dimension of the span of the row vectors of A, $\text{r-rank}(A) := \dim\big(\text{span}(\mathbf{r}_1, \ldots, \mathbf{r}_m)\big)$.
- *column-rank*, the dimension of the span of the column vectors of A, $\text{c-rank}(A) := \dim\big(\text{span}(\mathbf{c}_1, \ldots, \mathbf{c}_n)\big)$.

We shall later see that always $\text{r-rank}(A) = \text{c-rank}(A)$ – and this number will correspondingly be called the rank of A, $\text{rank}(A)$.

A connection between column-rank and homomorphisms represented by A is apparent. If we take A to represent $\varphi_A \colon \mathbb{F}^n \to \mathbb{F}^m$ w.r.t. the standard bases in these spaces:

$$\begin{array}{rcl} \varphi_A \colon \mathbb{F}^n & \longrightarrow & \mathbb{F}^m \\ (\lambda_1, \ldots, \lambda_n) & \longmapsto & A \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \lambda_1 \mathbf{c}_1 + \cdots + \lambda_n \mathbf{c}_n = \sum_{j=1}^n \lambda_j \mathbf{c}_j, \end{array}$$

then $\text{image}(\varphi) = \text{span}(\mathbf{c}_1, \ldots, \mathbf{c}_n)$ and therefore

$$\text{c-rank}(A) = \dim(\text{image}(\varphi_A)).$$

Definition 3.3.16 Let $\varphi \in \mathrm{Hom}(V, W)$, V and W finite-dimensional $\mathbb{F}$-vector spaces. The *rank* of φ is defined to be the dimension of the image of φ [which is a subspace of W]: $\mathrm{rank}(\varphi) := \dim(\mathrm{image}(\varphi))$.

The above therefore says the following:

Lemma 3.3.17 $\text{c-rank}(A) = \mathrm{rank}(\varphi_A)$.

Lemma 3.3.18 *For $A \in \mathbb{F}^{(n,n)}$:* $\text{c-rank}(A) = n$ *iff A is regular.*

Proof. Let $\text{c-rank}(A) = n$. Then $\varphi_A \colon \mathbb{F}^n \to \mathbb{F}^n$ is surjective, and hence (by the dimension formula) also injective, and thus an automorphism. Therefore, A is regular.

Conversely, if A is regular, then $\varphi_A \in \mathrm{Aut}(\mathbb{F}^n)$ has image $\mathbb{F}^n$, whence $\text{c-rank}(A) = \dim(\mathrm{image}(\varphi_A)) = n$.

□

Lemma 3.3.19 *If $C \in \mathbb{F}^{(m,m)}$ is regular then*

(i) $\text{c-rank}(CA) = \text{c-rank}(A)$ *for any $A \in \mathbb{F}^{(m,n)}$.*

(ii) $\text{c-rank}(AC) = \text{c-rank}(A)$ *for any $A \in \mathbb{F}^{(n,m)}$.*

Proof. As φ_C, the endomorphism of $\mathbb{F}^m$ represented by C w.r.t. the standard basis, is an automorphism,

(i) $\dim(\mathrm{image}(\varphi_A)) = \dim(\mathrm{image}(\varphi_C \circ \varphi_A))$.

(ii) $\dim(\mathrm{image}(\varphi_A)) = \dim(\mathrm{image}(\varphi_A \circ \varphi_C))$.

The claims about the column-ranks follow with Lemma 3.3.17.

□

Lemma 3.3.20 *Any application of the row transformations as considered in the Gauß-Jordan procedure leaves the row-rank invariant. Let A' be obtained from A by one of the following*

(T1) exchanging two rows.

(T2) replacing some row $\mathbf{r}_i$ by $\lambda\mathbf{r}_i$ for a scalar $\lambda \neq 0$.

(T3) replacing some row $\mathbf{r}_i$ by $\mathbf{r}_i + \lambda\mathbf{r}_j$ for some scalar λ and some $j \neq i$.

Then $\text{r-rank}(A') = \text{r-rank}(A)$.

Proof. For (T1) the claim is obvious.

For (T2) and (T3) compare Lemma 2.4.4.

□

The following will therefore allow us to determine the row-rank of a matrix quite easily, via Gauß-Jordan. Recall that a sequence of applications of the above operations will transform any matrix into a matrix in echelon (upper triangle) form.

Lemma 3.3.21 *Let $A \in \mathbb{F}^{(m,n)}$ be in echelon form, with r rows that have non-zero entries, i.e., $\mathbf{r}_1, \ldots, \mathbf{r}_r \neq \mathbf{0}$ and $\mathbf{r}_{r+1} = \cdots = \mathbf{r}_m = \mathbf{0}$. Then r-rank$(A)$ = c-rank$(A) = r$.*

Proof. Consider row-rank first. Let for $i = 1, \ldots, r$:

$$\mathbf{r}_i = (0, \ldots, 0, a_{ij_i}, \ldots, a_{in})$$

with $a_{ij_i} \neq 0$ the first non-zero component in row i, and such that $1 \leqslant j_1 < j_2 < \cdots < j_r \leqslant n$ (echelon form).

Clearly the $\mathbf{r}_i$ for $i = 1, \ldots, r$ are pairwise distinct. It suffices to show that they form a linearly independent set of r vectors.

Now if $\sum_i \lambda_i \mathbf{r}_i = \mathbf{0}$, then we find for $i = 1, \ldots, r$ in this order that $\lambda_i = 0$.

For $i = 1$: since $\mathbf{r}_1$ is the only vector among the $\mathbf{r}_i$ that has a non-zero entry in component j_1, namely $a_{1j_1} \neq 0$, we look at the j_1-component of the equation $\sum_i \lambda_i \mathbf{r}_i = \mathbf{0}$ and find that $\lambda_1 a_{1j_1} = 0$. This implies that $\lambda_1 = 0$.

Inductively, suppose that $\lambda_1 = \cdots = \lambda_{i-1} = 0$ is already established. Looking at the j_i-th component in the equation $\sum_i \lambda_i \mathbf{r}_i = \mathbf{0}$, we see that the only contribution is $\lambda_i a_{ij_i}$. And as $a_{ij_i} \neq 0$, $\lambda_i a_{ij_i} = 0$ implies that $\lambda_i = 0$ as well.

For column rank, we show that the column vectors $\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_r}$ are distinct and linearly independent. The argument is entirely similar to the above. □

This does not yet give us the desired equality between row-rank and column-rank for all matrices. We so far established this equality for matrices in echelon form only. The equality follows, however, if we can show that the transformations (T1-3) that transform any A into echelon form also preserve column rank.

Lemma 3.3.22 *Any application of the row transformations in the Gauß-Jordan procedure leaves the column-rank invariant. If A' is obtained from A by application of one of (T1), (T2), (T3), then* c-rank(A') = c-rank(A).

Proof. We find regular matrices $C \in \mathbb{F}^{(m,m)}$ such that $A' = CA$. Then the claim follows with Lemma 3.3.19 (a).

(T1): to exchange rows i and j use C with row vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{F}^m$ where

$$\mathbf{u}_k = \begin{cases} \mathbf{e}_k & \text{for } k \neq i, j \\ \mathbf{e}_i & \text{for } k = j \\ \mathbf{e}_j & \text{for } k = i \end{cases}$$

Check that this C operates as desired via $A \mapsto A' := CA$ and that C is regular.

(T2): to replace row $\mathbf{r}_i$ by $\lambda\mathbf{r}_i$, use C with row vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{F}^m$ where

$$\mathbf{u}_k = \begin{cases} \mathbf{e}_k & \text{for } k \neq i \\ \lambda\mathbf{e}_i & \text{for } k = i \end{cases}$$

Check that this C operates as desired via $A \mapsto A' := CA$ and that C is regular if $\lambda \neq 0$.

(T3): to replace row $\mathbf{r}_i$ by $\mathbf{r}_i + \lambda\mathbf{r}_j$ where $j \neq i$, use C with row vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{F}^m$ where

$$\mathbf{u}_k = \begin{cases} \mathbf{e}_k & \text{for } k \neq i \\ \mathbf{e}_i + \lambda\mathbf{e}_j & \text{for } k = i \end{cases}$$

Check that this C operates as desired via $A \mapsto A' := CA$ and that C is regular.

□

Corollary 3.3.23 *For any $A \in \mathbb{F}^{(m,n)}$:* $\text{c-rank}(A) = \text{r-rank}(A) = \text{rank}(\varphi_A)$.

Proof. We may apply a sequence of transformation steps (T1-3) to put A into echelon form. These steps do neither alter row-rank nor column-rank, by Lemmas 3.3.20 and 3.3.22. For the resulting echelon matrix we know that row-rank and column-rank are the same from Lemma 3.3.21. Agreement with the rank of φ_A is clear from Lemma 3.3.17 .

□

Exercise 3.3.10 Give a proof for Lemma 3.3.20 using suitable regular matrices $C \in \mathbb{F}^{(m,m)}$ and Lemma 3.3.19 (b).

Note that we can employ the Gauß-Jordan procedure to determine effectively

- the row-rank/column-rank of a matrix.
- the dimension of the span of a finite set of vectors in $\mathbb{F}^n$.
- a basis of the span of a finite set of vectors in $\mathbb{F}^n$.

For the second point just note that we may put the given vectors as the rows (or columns) of a matrix, and then determine the rank of it.

For the third point, one may use a matrix that contains the given vectors as rows, transform them according to Gauß-Jordan, and retain the non-zero rows as the desired basis (compare Lemmas 2.4.4 and 3.3.21).

The related analysis of systems of linear equations in terms of matrices and linear maps will be considered in the next chapter.

3.4 Aside: linear and affine transformations

We discuss *affine maps* [affine Abbildungen] in relation to linear maps, with a view to their geometrical significance. To exemplify the situation and to stay close to elementary geometrical intuition we explicitly consider the case of the real plane $\mathbb{R}^2$. In geometry one often uses the term *transformation* [Transformation] for bijective mappings.

The linear transformations (vector space automorphisms) of $\mathbb{R}^2$ – represented w.r.t. the standard basis $B = (\mathbf{e}_1, \mathbf{e}_2)$ say – give rise to the group $\mathrm{GL}_2(\mathbb{R})$ of regular 2×2 real matrices. Any such linear transformation fixes the null vector $\mathbf{0} = (0, 0)$ (the origin). Non-trivial translations

$$\begin{array}{rccc} \tau_{\mathbf{u}} \colon & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & \mathbf{v} & \longmapsto & \mathbf{v} + \mathbf{u} \end{array}$$

through some vector $\mathbf{u} \neq \mathbf{0}$ are not linear. Just as affine spaces are more general than vector spaces in that they do not have a fixed origin, affine maps are more general than linear maps in that they need not preserve this distinguished point. In particular, translations are perfectly admissible as affine transformations. Translations are gemometrically very natural and we recall how translations were built into the notion of affine spaces in section 2.3 (and how translations gave rise to the generalisation from linear to affine subspaces). In particular, translations *preserve lines* (one-dimensional affine subspaces) in the sense that the image of a line is again aline. The same is of course true of linear transformations.

Affine transformations combine linear maps with translations.

Definition 3.4.1 An affine transformation of $\mathbb{R}^2$ is a map $[\varphi, \mathbf{u}] = \tau_{\mathbf{u}} \circ \varphi$,

$$\begin{array}{rccl} [\varphi, \mathbf{u}]\colon & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & \mathbf{v} & \longmapsto & \varphi(\mathbf{v}) + \mathbf{u}, \end{array}$$

for any fixed linear transformation φ ($\mathbb{R}$-vector space automorphism of $\mathbb{R}^2$) and $\mathbf{u} \in \mathbb{R}^2$ (acting as a translation of $\mathbb{R}^2$).

Observation 3.4.2 *The affine transformations of $\mathbb{R}^2$ form a group with composition and $\mathrm{id}_{\mathbb{R}^2} (= [\mathrm{id}_{\mathbb{R}^2}, \mathbf{0}])$ as its neutral element.*

Exercise 3.4.1 For any two affine transformations, represent their composition $[\varphi_2, \mathbf{u}_2] \circ [\varphi_1, \mathbf{u}_1]$ in the form $[\varphi, \mathbf{u}]$ for suitable φ and $\mathbf{u}$ that are to be determined in terms of $\varphi_i, \mathbf{u}_i$. With the help of this, verify the group axioms.

Definition 3.4.3 Three points of $\mathbb{R}^2$ are *collinear* [kollinear] iff they are contained in a common line.

Exercise 3.4.2 Show that three distinct points $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ are collinear iff $\{\mathbf{v}_1 - \mathbf{v}_0, \mathbf{v}_2 - \mathbf{v}_0\}$ is linearly dependent. [Hint: compare section 2.5.3].

Exercise 3.4.3 Let $[\varphi, \mathbf{u}]$ be an affine transformation of $\mathbb{R}^2$. Show that

(i) the image set of a line in $\mathbb{R}^2$ under $[\varphi, \mathbf{u}]$ is again a line.

(ii) any three non-collinear points are mapped by $[\varphi, \mathbf{u}]$ into three non-collinear points.

(iii) $[\varphi, \mathbf{u}]$ is linear iff $[\varphi, \mathbf{u}](\mathbf{0}) = \mathbf{0}$ iff $\mathbf{u} = 0$.

Lemma 3.4.4 *An affine transformation of $\mathbb{R}^2$ is uniquely determined by its image on any three non-collinear points in $\mathbb{R}^2$. For any non-collinear $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2$ and non-collinear $\mathbf{v}'_0, \mathbf{v}'_1, \mathbf{v}'_2$ there is precisely one affine transformation $[\varphi, \mathbf{u}]$ such that $[\varphi, \mathbf{u}](\mathbf{v}_i) = \mathbf{v}'_i$ for $i = 0, 1, 2$.*

Proof. Put $\mathbf{a} := \mathbf{v}_1 - \mathbf{v}_0$, $\mathbf{b} := \mathbf{v}_2 - \mathbf{v}_0$ and $\mathbf{a}' := \mathbf{v}'_1 - \mathbf{v}'_0$, $\mathbf{b}' := \mathbf{v}'_2 - \mathbf{v}'_0$. By assumption, $(\mathbf{a}, \mathbf{b})$ forms a basis of $\mathbb{R}^2$, and so does $(\mathbf{a}', \mathbf{b}')$.

Existence: let $\varphi \in \mathrm{Aut}(\mathbb{R}^2)$ be the unique linear map that maps $\mathbf{a}$ to $\mathbf{a}'$ and $\mathbf{b}$ to $\mathbf{b}'$. Let $\mathbf{u} := \mathbf{v}'_0 - \varphi(\mathbf{v}_0)$. Then $[\varphi, \mathbf{u}]$ is as desired (check this!).

Uniqueness: let $[\varphi_1, \mathbf{u}_1]$ and $[\varphi_2, \mathbf{u}_2]$ be as required. We need to show that $[\varphi_1, \mathbf{u}_1] = [\varphi_2, \mathbf{u}_2]$. Consider $[\varphi, \mathbf{u}] := [\varphi_2, \mathbf{u}_2]^{-1} \circ [\varphi_1, \mathbf{u}_1]$, which is again an affine transformation. It suffices to show that $[\varphi, \mathbf{u}] = \mathrm{id}_{\mathbb{R}^2}$. Note that $[\varphi, \mathbf{u}](\mathbf{v}_i) = \mathbf{v}_i$ for $i = 0, 1, 2$. This implies that $\varphi(\mathbf{a}) = \mathbf{a}$ and $\varphi(\mathbf{b}) = \mathbf{b}$, and as $(\mathbf{a}, \mathbf{b})$ is a basis, $\varphi = \mathrm{id}_{\mathbb{R}^2}$. $[\varphi, \mathbf{u}](\mathbf{v}_0) = \mathbf{v}_0 + \mathbf{u} = \mathbf{v}_0$ then implies that $\mathbf{u} = \mathbf{0}$, and hence $[\varphi, \mathbf{u}] = [\mathrm{id}_{\mathbb{R}^2}, \mathbf{0}] = \mathrm{id}_{\mathbb{R}^2}$.

□

We state the following without proof.

Proposition 3.4.5 *The affine transformations of $\mathbb{R}^2$ are precisely the line-preserving transformations: any bijective function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ with the property that the image set of any line in $\mathbb{R}^2$ under f is again a line, is an affine transformation.*

Chapter 4

Matrix Arithmetic

4.1 Determinants

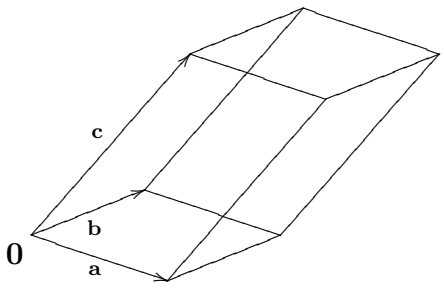
Consider a square matrix $A \in \mathbb{F}^{(n,n)}$.

We are interested in a simple test for regularity of A. Ideally we would like to have an easy to compute function of the individual entries a_{ij} in A whose value directly tells us whether A is regular or not.

Equivalently, thinking of A as a tuple of n column vectors $\mathbf{a}_j \in \mathbb{F}^n$ for $j = 1, \ldots, n$ [or as a tuple of n row vectors $\mathbf{r}_i$ for $i = 1, \ldots, n$], we want a simple computational test for whether the $\mathbf{a}_j$ [or the $\mathbf{r}_i$] are mutually distinct and linearly independent, i.e., whether they span all of $\mathbb{F}^n$.

Geometrically in $\mathbb{R}^n$ the question whether n vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{F}^n$ span $\mathbb{R}^n$ is the same as the question whether the volume of the parallelepiped [Parallelepiped, Spat] $\mathrm{P}(\mathbf{a}_1, \ldots, \mathbf{a}_n) := \left\{ \sum_{i=1}^n \lambda_i \mathbf{a}_i \colon 0 \leqslant \lambda_i \leqslant 1 \text{ for } i = 1, \ldots, n \right\} \subseteq \mathbb{R}^n$ is different from 0. [1]

$P(\mathbf{a}, \mathbf{b}, \mathbf{c}) \subseteq \mathbb{R}^3$



[1] Any figure contained inside a hyper-plane in $\mathbb{R}^n$ (i.e., an object of lower dimension than n) has zero volume in $\mathbb{R}^n$.

The value of the determinant, $\det(A) \in \mathbb{F}$, also written $|A|$, of a matrix A will address all these issues:

$$\det(A) \neq 0 \quad \text{iff} \quad A \text{ regular},$$

equivalently:

$$\begin{array}{lll} \det(A) = 0 & \text{iff} & \dim\big(\text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)\big) < n \\ & \text{iff} & \dim\big(\text{span}(\mathbf{r}_1, \ldots, \mathbf{r}_n)\big) < n \\ & \text{iff} & \text{rank}(A) < n. \end{array}$$

In the case of $\mathbb{R}^n$, $\det(A)$ will moreover be the *oriented volume* of the parallelepiped given by the column vectors $\mathbf{a}_j$ [or the row vectors $\mathbf{r}_i$].

An oriented volume is a real number whose absolute value gives the (non-negative) volume in the usual sense, and whose sign accounts for the *orientation* relative to the standard basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. In $\mathbb{R}^3$ for instance, the parallelepiped $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is the standard unit cube, and has (oriented) volume 1. Its mirror images are attributed oriented volume -1. This correspondings to the geometric observation that a reflection in a plane in 3-dimensional space inverts the orientation. For instance, the following are given oriented volume -1: the reflection of $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ w.r.t. the 2, 3-plane, $P(-\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$; or w.r.t. the plane spanned by $\mathbf{e}_1 + \mathbf{e}_2$ and $\mathbf{e}_3$, $P(\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3)$.

4.1.1 Determinants as multi-linear forms

To connect the postulates that will characterise the determinant with the geometric intuition of an oriented volume in $\mathbb{R}^n$, we collect some basic properties of the oriented volume function

$$\begin{array}{rcl} \mu\colon (\mathbb{R}^n)^n & \longrightarrow & \mathbb{R} \\ (\mathbf{a}_1, \ldots, \mathbf{a}_n) & \longmapsto & \mu(\mathbf{a}_1, \ldots, \mathbf{a}_n) := \text{ the oriented volume of } \mathrm{P}(\mathbf{a}_1, \ldots, \mathbf{a}_n). \end{array}$$

Definition 4.1.1 Consider a function $f\colon (\mathbb{F}^n)^n \to \mathbb{F}$.

(i) f is a *multi-linear form* [Multilinearform] iff for $i = 1, \ldots, n$ and any fixed $(n-1)$-tuple of arguments $\mathbf{a}_j \in \mathbb{F}^n$ for $j \neq i$, the map

$$\begin{array}{rcl} \mathbb{F}^n & \longrightarrow & \mathbb{F} \\ \mathbf{a} & \longmapsto & f(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n) \end{array}$$

is linear.

(ii) f is called *antisymmetric* [antisymmetrisch] iff, for any $1 \leqslant i < j \leqslant n$ and any $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{F}^n$:

$$\mathbf{a}_i = \mathbf{a}_j \quad \Rightarrow \quad \mu(\mathbf{a}_1, \ldots, \mathbf{a}_n) = 0.$$

(iii) f is called *alternating* [alternierend] iff, for any $1 \leqslant i < j \leqslant n$ and any $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{F}^n$:

$$\begin{aligned}\mu(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_i, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_{j-1}, \mathbf{a}_j, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n) \;=\\ -\mu(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_j, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_{j-1}, \mathbf{a}_i, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n).\end{aligned}$$

Observation 4.1.2 *Let $f\colon (\mathbb{F}^n)^n \to \mathbb{F}$ be multi-linear. If $1 \neq -1$ in $\mathbb{F}$, then f is antisymmetric iff it is alternating.* [2]

Proof. Consider some fixed $1 \leqslant i < j \leqslant n$ and fixed arguments $\mathbf{a}_k$ for $k \neq i, j$. Put

$$\begin{array}{rcl} g\colon \mathbb{F}^n \times \mathbb{F}^n & \longrightarrow & \mathbb{F} \\ (\mathbf{a}, \mathbf{b}) & \longmapsto & f(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n). \end{array}$$

For the claim of the observation we want to show that then g is antisymmetric ($g(\mathbf{a}, \mathbf{a}) = 0$ for all $\mathbf{a}$) iff g is alternating ($g(\mathbf{a}, \mathbf{b}) = -g(\mathbf{b}, \mathbf{a})$ for all $\mathbf{a}, \mathbf{b}$).

If g is alternating, then $g(\mathbf{a}, \mathbf{a}) = -g(\mathbf{a}, \mathbf{a})$ implies that $g(\mathbf{a}, \mathbf{a}) = 0$.

If g is antisymmetric, then

$$\begin{array}{rcl} 0 \;=\; g(\mathbf{a}+\mathbf{b}, \mathbf{a}+\mathbf{b}) & = & g(\mathbf{a}, \mathbf{a}) + g(\mathbf{a}, \mathbf{b}) + g(\mathbf{b}, \mathbf{a}) + g(\mathbf{b}, \mathbf{b}) \\ & = & g(\mathbf{a}, \mathbf{b}) + g(\mathbf{b}, \mathbf{a}) \end{array}$$

implies that $g(\mathbf{a}, \mathbf{b}) = -g(\mathbf{b}, \mathbf{a})$.

□

Observation 4.1.3 *Let $f\colon (\mathbb{F}^n)^n \to \mathbb{F}$ be multi-linear and antisymmetric. Then $f(\mathbf{a}_1, \ldots, \mathbf{a}_n) = 0$ for any linearly dependent tuple of vectors $\mathbf{a}_i$.*

Proof. Let for instance $\mathbf{a}_1 = \sum_{i=2}^n \lambda_i \mathbf{a}_i$. Then

$$f(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \sum_{i=2}^{n} \lambda_i f(\mathbf{a}_i, \mathbf{a}_2, \ldots, \mathbf{a}_n).$$

But $f(\mathbf{a}_i, \mathbf{a}_2, \ldots, \mathbf{a}_n) = 0$ for $i = 2, \ldots, n$ by antisymmetry.

□

[2] If $1 = -1$ in $\mathbb{F}$ (fields of characteristic 2, e.g., $\mathbb{F}_2$), then still antisymmetry implies alternation for any multi-linear function; but alternation rather becomes a symmetry criterion in this case!

Definition 4.1.4 Let $n \geqslant 1$. A *determinant function* on $\mathbb{F}^n$ is a function $\det\colon (\mathbb{F}^n)^n \to \mathbb{F}$ which

(D1) is multi-linear,

(D2) is antisymmetric,

(D3) has value 1 on the standard labelled basis of $\mathbb{F}^n$: $\det(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$.

One also may think of the determinant as a function $\det\colon \mathbb{F}^{(n,n)} \to \mathbb{F}$, if $(\mathbb{F}^n)^n$ is identified with $\mathbb{F}^{(n,n)}$ via the column vectors that make up a matrix. In this context one often prefers the notation $|A|$ for $\det(A)$.

We shall show next that for any $n \geqslant 1$ there is a unique function det on $(\mathbb{F}^n)^n$ satisfying (D1), (D2) and (D3).

4.1.2 Permutations and alternating functions

Recall from Definition 1.2.2 the group S_n of permutations of the n-element set $\{1, \ldots, n\}$, i.e., the group of bijections $\sigma\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ with composition.

Any alternating function is sensitive to swaps of any two arguments: any such swap changes the sign. We want to analyse the behaviour of an n-ary alternating function f under arbitrary permutations of the arguments. For an arbitrary $\sigma \in S_n$, i.e., a bijection $\sigma\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$, what is

$$f(\mathbf{a}_{\sigma(1)}, \mathbf{a}_{\sigma(2)}, \ldots, \mathbf{a}_{\sigma(n)})$$

in terms of $f(\mathbf{a}_1, \ldots, \mathbf{a}_n)$?

Definition 4.1.5 (i) A permutation in $\tau \in S_n$ is called a *transposition* [Transposition] if it swaps two distinct elements and leaves the rest alone: there are $1 \leqslant i < j \leqslant n$ such that $\tau(i) = j$, $\tau(j) = i$, and $\tau(k) = k$ for all $k \neq i, j$. The transposition that swaps i and j is denoted $\tau = (i, j)$. Note that $\tau^2 = \tau \circ \tau = \mathrm{id}$ for any transposition τ.

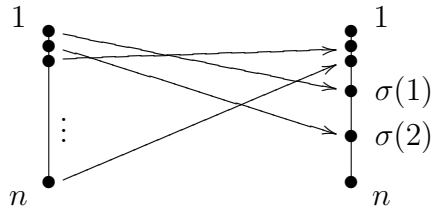
(ii) A transposition of the form $(i, i+1)$ for $1 \leqslant i < n$ is called a *next neighbour transposition*, nnt for short.

(iii) A permutation $\sigma \in S_n$ is called *even* [gerade] iff it can be written as a composition of an even number of nnts; it is *odd* [ungerade] iff it can be written as a composition of an odd number of nnts.

With any $\sigma \in S_n$, let us associate the number of pairs that are put in reverse order by σ:

$$\nu(\sigma) := |\{(i,j)\colon 1 \leqslant i < j \leqslant n \text{ and } \sigma(j) < \sigma(i)\}|.$$

This number is pictorially represented by the number of cross-overs between arrows in the diagram



Note that $0 \leqslant \nu(\sigma) \leqslant \frac{n(n-1)}{2}$ and that $\nu(\sigma) = 0$ only if $\sigma = \mathrm{id}$.

Proposition 4.1.6 *Let $n \geqslant 2$. Any permutation $\sigma \in S_n$ can be represented as a composition of nnts. The parity of the number of nnts in any such representation is fully determined by σ and has value $\nu(\sigma) \bmod 2$. In particular, any $\sigma \in S_n$ is either even or odd, and not both.*

Proof. By induction on $\nu(\sigma)$, we firstly show that any $\sigma \in S_n$ can be written as a composition

$$\sigma = \mathrm{id} \circ \tau_1 \circ \cdots \circ \tau_{\nu(\sigma)}$$

with exactly $\nu(\sigma)$ many nnts τ_i.

Base case, $\nu(\sigma) = 0$: $\sigma = \mathrm{id}$ (0 nnts).

Induction step, suppose the claim is true for all σ with $\nu(\sigma) < k$; we need to show the claim for σ with $\nu(\sigma) = k$. Let $\nu(\sigma) = k > 0$. Let i be maximal with the property that $\sigma(i) > \sigma(i+1)$ (there are such i as $\sigma \neq \mathrm{id}$). Let $\sigma' := \sigma \circ (i, i+1)$. Note that $\nu(\sigma') = \nu(\sigma) - 1$. By the induction hypothesis, $\sigma' = \tau_1 \circ \cdots \circ \tau_{\nu(\sigma')}$ for suitable nnt τ_j. But then $\sigma = \tau_1 \circ \cdots \circ \tau_{\nu(\sigma')} \circ (i, i+1)$ as desired.

It remains to show that any σ is exclusively either a composition of an even or of an odd number of nnts. For this, we observe that composition with a single nnt always changes ν by $+1$ or -1. In fact, for any nnt $\tau = (i, i+1)$:

$$\nu(\sigma \circ \tau) = \begin{cases} \nu(\sigma) - 1 & \text{if } \sigma(i) > \sigma(i+1) \\ \nu(\sigma) + 1 & \text{if } \sigma(i) < \sigma(i+1) \end{cases}$$

It follows that $\nu(\sigma)$ is even for an even σ, and odd for an odd σ.

□

Definition 4.1.7 The *sign* [Signatur] of $\sigma \in S_n$ is defined to be $\mathrm{sign}(\sigma) := (-1)^{\nu(\sigma)} \in \{-1, +1\}$. Equivalently,

$$\mathrm{sign}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}$$

Lemma 4.1.8 *For any alternating function $f\colon (\mathbb{F}^n)^n \to \mathbb{F}$, $\sigma \in S_n$ and any $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{F}^n$:*

$$f(\mathbf{a}_{\sigma(1)}, \ldots, \mathbf{a}_{\sigma(n)}) = \mathrm{sign}(\sigma) f(\mathbf{a}_1, \ldots, \mathbf{a}_n).$$

Proof. By induction on the number of transpositions required to represent σ.

□

Exercise 4.1.1 Let $n \geqslant 1$. Show that the set of even permutations,

$$A_n = \{\sigma \in S_n \colon \sigma \text{ even } \} \subseteq S_n$$

also forms a group with composition. It is called the *alternating group* [alternierende Gruppe], a subgroup of the symmetric group S_n. Determine the size of A_n in terms of n.

4.1.3 Existence and uniqueness of the determinant

The last lemma and antisymmetry of any determinant function imply that for arbitrary $i_1, \ldots, i_n \in \{1, \ldots, n\}$:

$$\det(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) = \begin{cases} \mathrm{sign}(\sigma) & \text{if the map } j \mapsto i_j \text{ is a} \\ & \text{permutation } \sigma \in S_n \\ 0 & \text{else.} \end{cases} \qquad (*)$$

Together with multi-linearity, the stipulation of values on all tuples of standard basis vectors according to $(*)$ fully determines the function det.

Let $\mathbf{a}_j = (a_{1j}, \ldots, a_{nj})$ for $j = 1, \ldots, n$. Then

$$\begin{aligned}
&\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \\
&\quad = \det(\textstyle\sum_{i_1=1}^{n} a_{i_1 1}\, \mathbf{e}_{i_1}, \mathbf{a}_2, \ldots, \mathbf{a}_n) \\
&\quad = \sum_{i_1=1}^{n} a_{i_1 1} \det(\mathbf{e}_{i_1}, \mathbf{a}_2, \ldots, \mathbf{a}_n) \\
&\quad = \sum_{i_1=1}^{n} a_{i_1 1} \det(\mathbf{e}_{i_1}, \sum_{i_2=1}^{n} a_{i_2 2}\, \mathbf{e}_{i_2}, \mathbf{a}_3, \ldots, \mathbf{a}_n) && \text{[linearity in } \mathbf{a}_1\text{]} \\
&\quad = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} a_{i_1 1} a_{i_2 2} \det(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \mathbf{a}_3, \ldots, \mathbf{a}_n) && \text{[linearity in } \mathbf{a}_2\text{]} \\
&\quad = \cdots \\
&\quad = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \ldots \sum_{i_n=1}^{n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \det(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_n}) \\
&\quad = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}. && \text{[by } (*)\text{!]}
\end{aligned}$$

Therefore, if there is a determinant function, it is uniquely determined.

Moreover, taking the last line as a definition of det, we actually obtain a determinant function. Recall that we view a determinant function as a function on $n \times n$ matrices or on n-tuples of (the corresponding column) vectors in $\mathbb{F}^n$.

Proposition 4.1.9 *Let*

$$\begin{array}{rcl} \det\colon \mathbb{F}^{(n,n)} & \longrightarrow & \mathbb{F} \\ A = (a_{ij}) & \longmapsto & \det(A) = |A| := \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}. \end{array}$$

This function satisfies the postulates (D1-3) of an n-dimensional determinant function; hence it is the unique n-dimensional determinant function over $\mathbb{F}$.

Proof. Uniqueness was addressed above. We need to verify that det as defined satisfies (D1), (D2) and (D3). For (D3) see Example 4.1.10 below.

We check (D1) by showing linearity of $(\mathbf{a}_1, \ldots, \mathbf{a}_n) \mapsto \det(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ in its first argument $\mathbf{a}_1$. Linearity in the other arguments can either be shown in the same way, or be inferred via (D2).

Let $\mathbf{a}_1 = (a_{11}, \ldots, a_{n1})$, $\mathbf{a}'_1 = (a'_{11}, \ldots, a'_{n1})$, and $\lambda \in \mathbb{F}$. We want to show that

$$\begin{aligned} \det(\lambda\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) &= \lambda\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \\ \det(\mathbf{a}_1 + \mathbf{a}'_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) &= \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) + \det(\mathbf{a}'_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) \end{aligned}$$

These follow straight from the definition:

$$\begin{aligned} \det(\lambda\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)(\lambda a_{\sigma(1)1})a_{\sigma(2)2}\cdots a_{\sigma(n)n} \\ &= \lambda\det(\mathbf{a}_1, \ldots, \mathbf{a}_n); \\ \det(\mathbf{a}_1 + \mathbf{a}'_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)(a_{\sigma(1)1} + a'_{\sigma(1)1})a_{\sigma(2)2}\cdots a_{\sigma(n)n} \\ &= \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) + \det(\mathbf{a}'_1, \mathbf{a}_2, \ldots, \mathbf{a}_n). \end{aligned}$$

It remains to show (D2): det is antisymmetric. Let $1 \leqslant i < j \leqslant n$, $\mathbf{a}_i = \mathbf{a}_j$. Let $\tau = (i, j)$ the transposition that exchanges i and j. Then

$$\begin{aligned} &\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \\ &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\, a_{\sigma(1)1}\cdots a_{\sigma(n)n} \\ &= \sum_{\sigma:\, \sigma(i)<\sigma(j)} \operatorname{sign}(\sigma)\, a_{\sigma(1)1}\cdots a_{\sigma(n)n} \;+ \sum_{\sigma:\, \sigma(i)>\sigma(j)} \operatorname{sign}(\sigma)\, a_{\sigma(1)1}\cdots a_{\sigma(n)n} \\ &= \sum_{\sigma:\, \sigma(i)<\sigma(j)} \operatorname{sign}(\sigma)\, a_{\sigma(1)1}\cdots a_{\sigma(n)n} \;+ \sum_{\sigma:\, \sigma(i)<\sigma(j)} \operatorname{sign}(\sigma\circ\tau)\, a_{\sigma(1)1}\cdots a_{\sigma(n)n} \\ &= \sum_{\sigma:\, \sigma(i)<\sigma(j)} \operatorname{sign}(\sigma)\, a_{\sigma(1)1}\cdots a_{\sigma(n)n} \;+ \sum_{\sigma:\, \sigma(i)<\sigma(j)} (-1)\operatorname{sign}(\sigma)\, a_{\sigma(1)1}\cdots a_{\sigma(n)n} \\ &= 0, \end{aligned}$$

where we split up S_n according to whether a permutation reverses the order of i and j or not. Note that $\mathbf{a}_i = \mathbf{a}_j$ implies that $a_{\sigma\circ\tau(i)i} = a_{\sigma(j)i} = a_{\sigma(j)j}$ and $a_{\sigma\circ\tau(j)j} = a_{\sigma(i)j} = a_{\sigma(i)i}$, whence $a_{\sigma\circ\tau(i)i}a_{\sigma\circ\tau(j)j} = a_{\sigma(i)i}a_{\sigma(j)j}$.

□

Example 4.1.10 Here are some cases of particular matrices whose determinant is easy to compute from scratch.

(i) The determinant of the unit matrix E_n has value 1:

$$|E_n| = \det(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1.$$

This follows from the observation that the only permutation $\sigma \in S_n$ for which all positions $(\sigma(i), i)$ are on the diagonal is $\sigma = \mathrm{id}$.

(ii) Similarly, if $A \in \mathbb{F}^{(n,n)}$ is a *diagonal matrix* [Diagonalmatrix], i.e., with off-diagonal entries $a_{ij} = 0$ for all $i \neq j$, then its determinant is just the product of the diagonal entries a_{ii}:

$$\begin{vmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & a_{nn} \end{vmatrix} = a_{11} \cdots a_{nn}.$$

(iii) More generally, also for an echelon (upper triangle) matrix $A \in \mathbb{F}^{(n,n)}$, the determinant is just the product of its diagonal entries. Note that for an echelon matrix, $a_{ij} = 0$ for $n \geqslant i > j \geqslant 1$. Therefore, any $\sigma \in S_n$ which has any values $\sigma(i) > i$ will not contribute, and only $\sigma = \mathrm{id}$ remains:

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & a_{nn} \end{vmatrix} = a_{11} \cdots a_{nn}.$$

Example 4.1.11 Consider $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{F}^{(2,2)}$. The above definition reduces to the familiar:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \sum_{\sigma = \mathrm{id}, (12)} \mathrm{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} = a_{11}a_{22} - a_{12}a_{21}.$$

Exercise 4.1.2 Similarly considering all six permutations in S_3, show that for $A \in \mathbb{F}^{(3,3)}$:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{cases} \quad a_{11}\big(a_{22}a_{33} - a_{23}a_{32}\big) \\ -\, a_{21}\big(a_{12}a_{33} - a_{13}a_{32}\big) \\ +\, a_{31}\big(a_{12}a_{23} - a_{13}a_{22}\big). \end{cases}$$

The pattern according to which a 3×3 determinant reduces to a sum of 3 determinants of the order 2×2, is generalised in section 4.1.5 below with the idea of expanding a determinant.

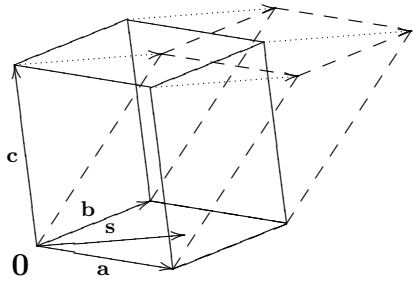
4.1.4 Further properties of the determinant

The following is known as the *shear invariance* [Scherinvarianz] of the determinant. This is another property that is obvious for a parallelepiped volume in $\mathbb{R}^n$. Its proof, as a simple consequence of multi-linearity and antisymmetry, is left as an exercise.

Lemma 4.1.12 *For any* $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{F}^n$, $1 \leqslant i \leqslant n$ *and any linear combination* $\mathbf{s} = \sum_{j \neq i} \lambda_j \mathbf{a}_j$ *of the arguments* $\mathbf{a}_j$, $j \neq i$: *:*

$$\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \det(\mathbf{a}_1, \ldots, \mathbf{a}_i + \mathbf{s}, \ldots, \mathbf{a}_n) = \det(\mathbf{a}_1, \ldots, \mathbf{a}_n).$$

$P(\mathbf{a}, \mathbf{b}, \mathbf{c})$ and $P(\mathbf{a}, \mathbf{b}, \mathbf{c} + \mathbf{s})$
$\mathbf{s} \in \operatorname{span}(\mathbf{a}, \mathbf{b})$



Definition 4.1.13 For $A \in \mathbb{F}^{(n,n)}$ let its *transpose* [Transponierte] A^T be obtained by exchanging entries across the $(i = j)$-diagonal:

$$\text{for } A = (a_{ij}): \qquad A^T = (a^T_{ij}) \text{ where } a^T_{ij} := a_{ji}.$$

Note that transposition swaps the roles between row vectors $\mathbf{r}_i$ and column vectors $\mathbf{a}_i$ in A. As another consequence of the explicit definition of the determinant we find the following property.

Lemma 4.1.14 $|A| = |A^T|$, *i.e.,*

$$\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \det(\mathbf{r}_1, \ldots, \mathbf{r}_n),$$

if the $\mathbf{a}_j$ *are the column vectors and the* $\mathbf{r}_i$ *the row vectors of* $A \in \mathbb{F}^{(n,n)}$.

Proof.

$$\begin{aligned}
&\det(\mathbf{r}_1, \ldots, \mathbf{r}_n) \\
&\quad = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \\
&\quad = \sum_{\sigma' \in S_n} \operatorname{sign}((\sigma')^{-1}) a_{\sigma'(1)1} \cdots a_{\sigma'(n)n} \qquad [\text{as } \sigma' := \sigma^{-1} \colon \sigma(i) \mapsto i] \\
&\quad = \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \qquad [\text{as } \operatorname{sign}(\sigma^{-1}) = \operatorname{sign}(\sigma)].
\end{aligned}$$

□

Multi-linearity directly shows that the column transformation of replacing $\mathbf{a}_i$ by $\mathbf{a}_i + \lambda \mathbf{a}_j$ for some $i \neq j$ leaves $\det(A)$ invariant. For instance,

$$\begin{aligned}
\det(\mathbf{a}_1 + \lambda \mathbf{a}_2, \mathbf{a}_2, \ldots, \mathbf{a}_n) &= \det(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) + \lambda \det(\mathbf{a}_2, \mathbf{a}_2, \mathbf{a}_2, \ldots, \mathbf{a}_n) \\
&= \det(\mathbf{a}_1, \ldots, \mathbf{a}_n).
\end{aligned}$$

With the last lemma, the same applies to the corresponding row transformations.

Corollary 4.1.15 *The value of* $\det(A)$ *is invariant under the following row and column transformations:*

(R) replacing a row $\mathbf{r}_j$ *by* $\mathbf{r}_j + \lambda \mathbf{r}_i$ *for some* $\lambda \in \mathbb{F}$ *and* $i \neq j$.

(C) replacing a column $\mathbf{a}_j$ *by* $\mathbf{a}_j + \lambda \mathbf{a}_i$ *for some* $\lambda \in \mathbb{F}$ *and* $i \neq j$.

Exercise 4.1.3 (i) Detail the effect of the other row transformations in Gauß-Jordan (applied to a square matrix $A \in \mathbb{F}^{(n,n)}$) on the value of the determinant.

(ii) Use Example 4.1.10, part (iii), for the value of the determinant of a square echelon matrix. Combine these insights to obtain a method for the computation of $|A|$ for arbitrary $A \in \mathbb{F}^{(n,n)}$.

The determinant is compatible with matrix multiplication in the following very nice way.

Proposition 4.1.16 *For* $A, B \in \mathbb{F}^{(n,n)}$*:* $|AB| = |A||B|$.

Proof. Let $C = AB$. Let the row and column vectors of A be $\mathbf{r}_i$ and $\mathbf{a}_j$ as usual. Let similarly $\mathbf{b}_j$ and $\mathbf{c}_j$ be the column vectors of B and C, respectively.

From the rules of matrix multiplication we find that the column vectors $\mathbf{c}_j$ of C are the matrix products $A\mathbf{b}_j$ between A and the column vectors $\mathbf{b}_j$ of B:

$$\begin{aligned} \mathbf{c}_j = C\mathbf{e}_j = \textstyle\sum_i c_{ij}\mathbf{e}_i = AB\mathbf{e}_j = A\mathbf{b}_j \\ = \quad A(\textstyle\sum_i b_{ij}\mathbf{e}_i) = \sum_i b_{ij}A\mathbf{e}_i = \sum_i b_{ij}\mathbf{a}_i. \end{aligned}$$

Therefore

$$\begin{aligned} |C| &= \det(\mathbf{c}_1, \ldots, \mathbf{c}_n) = \det\big(A\mathbf{b}_1, \ldots A\mathbf{b}_n\big) \\ &= \det_n\big(\textstyle\sum_k b_{k1}\mathbf{a}_k, \sum_k b_{k2}\mathbf{a}_k, \ldots, \sum_k b_{kn}\mathbf{a}_k\big) \\ &= \sum_{k_1=1}^{n}\sum_{k_1=2}^{n}\cdots\sum_{k_n=1}^{n}\det(b_{k_1 1}\mathbf{a}_{k_1}, b_{k_2 2}\mathbf{a}_{k_2}, \ldots, b_{k_n n}\mathbf{a}_{k_n}). \end{aligned}$$

In the last nested sum only those contributions can be non-zero for which $i \mapsto k_i$ is a permutation; in all other cases the determinant has two linearly dependent arguments. Therefore

$$\begin{aligned} |C| &= \textstyle\sum_{\sigma\in S_n} b_{\sigma(1)1}\cdots b_{\sigma(n)n}\det(\mathbf{a}_{\sigma(1)}, \ldots, \mathbf{a}_{\sigma(n)}) \\ &= \textstyle\sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\; b_{\sigma(1)1}\cdots b_{\sigma(n)n}\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \\ &= |B||A|. \end{aligned}$$

□

Exercise 4.1.4 Give an alternative proof of the assertion of the proposition, in the case where $|A| \neq 0$. Show that for any fixed A, the function

$$\begin{aligned} f\colon \mathbb{F}^{(n,n)} &\longrightarrow \mathbb{F} \\ B &\longmapsto \tfrac{|AB|}{|A|} \end{aligned}$$

satisfies (D1-3). Hence it must be *the* determinant function with value $|B|$ on B.

The behaviour of determinants on matrix products implies that for a regular matrix A with inverse A^{-1}, $1 = |E_n| = |AA^{-1}| = |A| \cdot |A^{-1}|$. We obtain two corollaries.

Corollary 4.1.17 *If A is regular then $|A| \neq 0$.*

Together with Observation 4.1.3 above, we have thus established non-vanishing determinant as a criterion for regularity.

Corollary 4.1.18 *Similar matrices have the same determinant: if C is regular with inverse C^{-1}, then*

$$|CAC^{-1}| = |A|.$$

Recall that a similarity class of matrices in $\mathbb{F}^{(n,n)}$ precisely corresponds to all the representations $[\![\varphi]\!]_B^B$ of some endomorphism $\varphi \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^n)$, for arbitrary choices of bases B for $\mathbb{F}^n$. Invariance of the determinant across entire similarity classes means that $|A|$ is fully determined by the underlying endomorphism φ – any representational matrix $[\![\varphi]\!]_B^B$ for φ will produce the same value for its determinant.

Observation 4.1.19 *The map that associates with any endomorphism $\varphi \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^n)$ the value of the determinant $|A_\varphi^{BB}|$ for any labelled basis B of $\mathbb{F}^n$, is well defined* (determinant of an endomorphism).

4.1.5 Computing the determinant

The following technique is known as *expanding a determinant* [Entwicklung]. Let $A = (a_{ij}) \in \mathbb{F}^{(n,n)}$ with entries a_{ij} forming the column vectors $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{F}^n$.

We give a typical example of this technique at first. "Expanding $\det(A)$ w.r.t. the first column $\mathbf{a}_1$" means that we use linearity of $|A| = \det(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ in its first argument $\mathbf{a}_1$ to reduce the problem of calculating $|A|$ to the calculation of several determinants for smaller sub-matrices of A. From linearity in the first argument:

$$\begin{aligned} |A| &= \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \\ &= \sum_{i=1} a_{i1} \det(\mathbf{e}_i, \mathbf{a}_2, \ldots, \mathbf{a}_n) \\ &= \sum_{i=1} (-1)^{i+1} a_{i1} \, |A_{[i1]}|, \end{aligned}$$

where $A_{[i1]}$ is the $(n-1)\times(n-1)$ matrix obtained from A be deleting the first column and the i-th row:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{(i-1)1} & a_{(i-1)2} & \cdots & a_{(i-1)n} \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \qquad A_{[i1]} = \left(\begin{array}{c|ccc} a_{12} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{(i-1)2} & \cdots & a_{(i-1)n} \\ \hline a_{(i+1)2} & \cdots & a_{(i+1)n} \\ \vdots & & \vdots \\ a_{n2} & \cdots & a_{nn} \end{array}\right)$$

The justification of the formula is based on the observation that $\det(\mathbf{e}_i, \mathbf{a}_2, \ldots, \mathbf{a}_n) = (-1)^{i+1}\det(A_{[i1]})$. For this observe that, by repeated application of Corollary 4.1.15 and alternation:

$$\det(\mathbf{e}_i, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{(i-1)2} & \cdots & a_{(i-1)n} \\ 1 & a_{i2} & \cdots & a_{in} \\ 0 & a_{(i+1)2} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

$$= \begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{(i-1)2} & \cdots & a_{(i-1)n} \\ 1 & 0 & \cdots & 0 \\ 0 & a_{(i+1)2} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = (-1)^{i+1} \begin{vmatrix} 1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & A_{[i1]} & \\ 0 & & & \end{vmatrix}$$

Note that in the last transformation we perform $(i-1)$ swaps of (adjacent) row vectors to bring the i-th row to the top. This introduces a factor of $(-1)^{i-1} = (-1)^{i+1}$. For the evaluation of that determinant, finally, note that the permutations $\sigma \in S_n$ for which we get non-zero contributions are precisely those with $\sigma(1) = 1$. For these we get exactly the summand corresponding to $\det(A_{[i1]})$.

We sum up these observations, which extend to other columns and (via passage to the transpose) to rows.

Proposition 4.1.20 *Let $A \in \mathbb{F}^{(n,n)}$ have entries a_{ij} in the i-th row and j-th column. For $1 \leqslant i, j \leqslant n$ let $A_{[ij]}$ stand for the matrix in $\mathbb{F}^{(n-1,n-1)}$ obtained from A by deletion of the i-th row and j-th column.*

For any choice of a row or column index $1 \leqslant k \leqslant n$, $\det(A)$ can be expanded w.r.t. the k-th column according to

$$|A| = \sum_i (-1)^{i+k} a_{ik} |A_{[ik]}|,$$

or w.r.t. the k-th row according to

$$|A| = \sum_j (-1)^{k+j} a_{kj} |A_{[kj]}|.$$

Exercise 4.1.5 Check the above rules explicitly for $n = 3$ and compare with the previous Exercise 4.1.2.

4.2 Inversion of matrices

The determinants of the reduced matrices $A_{[ij]}$ also play an important role in the explicit computation of the inverse of a regular matrix $A \in \mathbb{F}^{(n,n)}$ (inverse w.r.t. matrix multiplication, i.e., in the group $\mathrm{GL}_n(\mathbb{F})$).

We know from the last section that A is regular iff $\det(A) \neq 0$.

Consider now the matrix A' with entries

$$a'_{ij} := (-1)^{i+j} |A_{[ji]}|.$$

Note that the passage from A to A' involves the reduced matrix in *transposed positions*, and an application of a chequer board *sign pattern* of $+/-$ weights (here for $n = 5$):

$$\begin{pmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{pmatrix}$$

We calculate the matrix product $C := A'A$, whose entries we denote c_{ij}:

$$c_{ij} = \sum_k a'_{ik} a_{kj} = \sum_k (-1)^{k+i} |A_{[ki]}|\, a_{kj} = \sum_k a_{kj} (-1)^{k+i} |A_{[ki]}|.$$

For $i = j$ this is the expansion of $\det(A)$ w.r.t. the i-th column, and we find that

$$c_{ii} = |A|.$$

All other (non-diagonal) entries c_{ij} for $i \neq j$ are in fact 0, as they correspond to the determinant of a matrix $A_{i/j}$ in which the i-th column $\mathbf{a}_i$ of A has been replaced by a copy of the j-th column $\mathbf{a}_j$. But any such determinant is 0 by antisymmetry! Therefore, we have found that

$$A'A = |A|\, E_n,$$

whence the inverse of $A \in \mathrm{GL}_n(\mathbb{F})$ is the matrix $|A|^{-1}A'$, with entries

$$\begin{pmatrix} & {\scriptstyle (j)} & \\ & \vdots & \\ {\scriptstyle (i)} \quad \cdots & (-1)^{i+j}\frac{|A_{[ji]}|}{|A|} & \end{pmatrix}$$

Exercise 4.2.1 Explicitly compute the entries in the inverse matrix of the general regular 2×2 matrix, and similarly for regular 3×3 matrices.

4.3 Systems of linear equations revisited

We are now equipped for a comprehensive treatment of systems of linear equations. The theory we have developed so far allows us

- to connect systems of linear equations with linear maps,
- to solve them using matrix arithmetic, and
- to determine solvability and the dimension of the solution space in terms of ranks and determinants of matrices.

Recall that a system of m linear equations over $\mathbb{F}^n$ has the form

$$E = E[A, \mathbf{b}]\colon \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n = b_3 \\ \qquad\qquad\qquad\qquad\qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \end{cases}$$

where we regard this system as generated from the matrix $A = (a_{ij})$ of coefficients on the left-hand side and the column vector $\mathbf{b}$ of coefficients on the right-hand side.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad \text{with column vectors } \mathbf{a}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Note that the entire system E can thus be written as

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \qquad \text{or just: } A\mathbf{x} = \mathbf{b}.$$

The associated *homogeneous system* E^* is obtained by replacing $\mathbf{b}$ by $\mathbf{0}$, i.e., $E^* = E[A, \mathbf{0}]$.

4.3.1 Using linear maps and matrices

Regarding the system $E = E[A, \mathbf{b}]$ as a vector equation $A\mathbf{x} = \mathbf{b}$ in $\mathbb{F}^m$, we may consider the left-hand side of that equation as describing the image of a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$ under the linear map

$$\begin{aligned} \varphi = \varphi_A \colon \mathbb{F}^n &\longrightarrow \mathbb{F}^m \\ (x_1, \ldots, x_n) &\longmapsto A\mathbf{x} = \textstyle\sum_j x_j \mathbf{a}_j, \end{aligned}$$

where the $\mathbf{a}_j$ are the column vectors of the matrix A. In this sense, $E = E[A, \mathbf{b}]$ becomes the equation $\varphi_A(\mathbf{x}) = \mathbf{b}$ while the homogeneous system E^* is equivalent to $\varphi_A(\mathbf{x}) = \mathbf{0}$.

For the homogeneous system, the solution space therefore is

$$S(E^*) = \big\{\mathbf{x} \in \mathbb{F}^n : \varphi_A(\mathbf{x}) = \mathbf{0}\big\} = \ker(\varphi_A).$$

We thus find, with Theorem 3.1.12, that the dimension of the solution space of E^* is

$$\dim(S(E^*)) = n - \dim(\text{image}(\varphi_A)) = n - \text{rank}(A).$$

This also confirms again that there must be non-trivial solutions in case $m < n$, simply because $\text{rank}(A) \leqslant \min(m, n)$.

For the corresponding solution sets of the original inhomogeneous system E we know that it is either empty, or an affine subspace generated by any fixed solution to E and the subspace $S(E^*)$. If non-empty, the dimension of this affine subspace is the same as the dimension of $S(E^*)$.

In terms of the map φ_A, moreover, we can say that E has a solution iff $\mathbf{b} \in \text{image}(\varphi_A)$ iff

$$\mathbf{b} \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n).$$

Lemma 4.3.1 *Let $E = E[A, \mathbf{b}]$ and let $[A, \mathbf{b}]$ be the matrix with column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{b}$, $[A, \mathbf{b}] \in \mathbb{F}^{(m,n+1)}$. Then $S(E) \neq \emptyset$ iff $\mathbf{b} \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ iff*

$$\text{rank}([A, \mathbf{b}]) = \text{rank}(A).$$

Proof. Recall that the rank of a matrix is the dimension of the span of its column vectors. Therefore $\text{rank}([A, \mathbf{b}]) = \text{rank}(A)$ iff

$$\dim(\text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{b})) = \dim(\text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n))$$

iff these two spans are equal, i.e., iff $\mathbf{b} \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \text{image}(\varphi_A)$. □

Note that $\mathbf{b} \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \text{image}(\varphi)$ will be the case for any $\mathbf{b}$ if φ_A is surjective. But φ_A is surjective (an epimorphism) iff $\text{rank}(\varphi_A) = \text{rank}(A) = m$.

On the other hand, we know that any solution (if it exists) is unique iff φ_A is injective (a monomorphism) iff $\ker(\varphi_A) = \{\mathbf{0}\}$ iff $\dim(\text{image}(\varphi_A)) = n$ iff $\text{rank}(\varphi_A) = \text{rank}(A) = n$.

Observation 4.3.2 *For $E = E[A, \mathbf{b}]$ where $A \in \mathbb{F}^{(m,n)}$ and $\mathbf{b} \in \mathbb{F}^m$:*

(i) *$E[A, \mathbf{b}]$ is solvable for every right-hand side $\mathbf{b} \in \mathbb{F}^m$ iff $\text{rank}(A) = m$. (NB: this can only happen if $m \leqslant n$.)*

(ii) *$E[A, \mathbf{b}]$ has at most one solution for every choice of a right-hand side $\mathbf{b} \in \mathbb{F}^m$ iff $\text{rank}(A) = n$. (NB: this can only happen if $n \leqslant m$.)*

(iii) *$E[A, \mathbf{b}]$ has precisely one solution for every choice of a right-hand side $\mathbf{b} \in \mathbb{F}^m$ iff $\text{rank}(A) = n = m$.*

Note that if $\mathrm{rank}(A) < m$ then row transformations according to Gauß-Jordan will allow us to transform $E = E[A, \mathbf{b}]$ into an equivalent system $\hat{E} = E[\hat{A}, \hat{\mathbf{b}}]$ such that

(i) either $\hat{E}$ obviously has no solutions, because it contains equations of the form $0 = \hat{b}_i$ for $\hat{b}_i \neq 0$.

(ii) or $\mathrm{rank}(\hat{A}) = \mathrm{rank}(A) = \hat{m}$, the new number of rows; this variant is achieved by dropping the trivial rows $0 = 0$ in the outcome of Gauß-Jordan.

4.3.2 Solving regular systems

We call a system of linear equantions regular if its left-hand side matrix of coefficients A is a regular square matrix.

In this case we know that also every inhomogeneous system $E[A, \mathbf{b}]$ has a unique solution. In fact there is a simple formula for obtaining this solution involving the determinant.

First observe that the inverse matrix A^{-1} of A (A is regular) holds the key to the solution of the inhomogeneous system $E[A, \mathbf{b}]$ as

$$S(E[A, \mathbf{b}]) = \big\{\mathbf{x} \in \mathbb{F}^n \colon \varphi_E(\mathbf{x}) = \mathbf{b}\big\} = \varphi_E^{-1}(\mathbf{b}) = \big\{A^{-1}\mathbf{b}\big\},$$

since the inverse matrix represents the inverse map.

However, the solution $A^{-1}\mathbf{b}$ is even more directly available in a manner known as *Cramer's rule* [Cramersche Regel].

Lemma 4.3.3 *Let $A \in \mathbb{F}^{(n,n)}$ be regular, $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{F}^n$. Then the unique solution to the inhomogeneous system of linear equations $E[A, \mathbf{b}]$ is obtained as $\mathbf{x} = (x_1, \ldots, x_n)$ where*

$$x_i = |A|^{-1} \det\big(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n\big).$$

Proof. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the solution. That means that

$$\mathbf{b} = \sum_j x_j \mathbf{a}_j.$$

Feeding this into $\det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n)$ we get

$$\begin{aligned} & \det\big(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \textstyle\sum_j x_j \mathbf{a}_j, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n\big) \\ = \; & \textstyle\sum_j x_j \det\big(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_j, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n\big). \end{aligned}$$

The only non-zero contribution in the sum, if any, occurs for $j = i$ (anti-symmetry!), which yields

$$\begin{aligned} & \det\big(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n\big) \\ = \;& x_i \det\big(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_i, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n\big) \;=\; x_i\,|A|, \end{aligned}$$

i.e., $x_i = |A|^{-1} \det\big(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n\big)$ as claimed.

□

Index