Linear Algebra

held by Prof. Dr. Weitze-Schmithüsen

UNIVERSITÄT
DES
SAARLANDES

General Information

These lecture notes have been created by a student. Typesetting is no warrant for accuracy.

If you find any mistakes in the write-up, a hint via E-Mail would be greatly appreciated:

guenther@math.uni-sb.de

Contents

Part 1.

Linear Algebra I

Chapter I.

Basics

1. Prerequisites from Set Theory

For this lecture, we make due with naive set theory. A proper introduction to set theory, mathematical logic and the fundamental axioms requires a level of maturity that is not appropriate for this course.

1.1. Naive Set Theory

A set consists of objects. Those are called *elements*.

Example:
- The set of natural numbers $\mathbb{N}$ (excluding 0),
- the set of integers $\mathbb{Z}$,
- the set of rationals $\mathbb{Q}$,
- the set of real numbers $\mathbb{R}$,
- the set $M_1 = \{1, 2, 7, 11\}$,
- the set $M_2 = \{\text{Saarbrücken}, \text{Neunkirchen}, \text{Bexbach}, \text{Köln}\}$,
- the set $M_3 = \{\{1, 2\}, \{1, 7\}, \{1, 2, 7, 11\}\}$.

The notation '$x \in M$' means that x is an element of the set M. For typographical reasons, we will most often write 'x in M' instead of '$x \in M$'.

Two fundamental concepts for sets are the following:

- *Extensionality*: Two sets are identical if and only if they have the same elements. In other words: Let M_1 and M_2 be two sets. It holds $M_1 = M_2$ precisely if x belongs to M_1 if and only if x belongs to M_2.

- *Axiom Schema of Specification*: To any set M_1 and any assertion P on the elements of M_1 there is a set M_2 consisting of all the elements of M_1 for which P holds. That is

$$M_2 = \{x \in M_1 \mid P(x) \text{ holds}\}$$

exists and is indeed a set.

Example: Let $M_1 = \mathbb{R}$ and consider the assertion P that is true for a real number x, if it is positive. This defines the set

$$M_2 = \mathbb{R}_{>0} = \{x \in \mathbb{R} \mid P(x) \text{ holds}\} = \{x \in \mathbb{R} \mid x > 0\}.$$

1.2. Fundamentals of Propositional Logic

For this lecture, we understand a statement to be something that can either be true or false.[1]

To any statement A there is a negation $\neg A$ with the following property: If A is true, then $\neg A$ is false. If A is false, then $\neg A$ is true. From two statements A and B we can build new statements as follows:

(i) *The conjunction $A \wedge B$ ('A and B')*: 'A and B' is true if both A and B are true, and false otherwise. This is depicted in the following truth table:

$B \backslash A$	t	f
t	t	f
f	f	f

(ii) *The disjunction $A \vee B$ ('A or B')*: 'A or B' is false if A and B is false, and true otherwise, i.e.

$B \backslash A$	t	f
t	t	t
f	t	f

(iii) *The implication $A \Rightarrow B$ ('A implies B')*: 'A implies B' is false if A is true and B is false, and true otherwise. This yields the truth table

$B \backslash A$	t	f
t	t	t
f	t	f

[1] This is of course too simple. We circumvent giving a definition for what a statement should be in the first place, and the standard system of axioms is so complicated that there are statements for which it is undecidable if they are true or not.

The statement A is called *premiss* and B is called *conclusion* or *deduction*. In prose, 'A implies B' is often expressed as 'If A then B'.

(iv) *The equivalence $A \Leftrightarrow B$ ('A if and only if B')*: 'A if and only if B' is true if both A and B are true or false.

$B \backslash A$	t	f
t	t	f
f	f	t

Example: The statement 'If $2 = 5$, then 6 is uneven' is true.

1.3. Some Rules

For any statements A, B and C it holds:

(i) '$\neg(\neg A) \Leftrightarrow A$'.

(ii) '$A \wedge B \Leftrightarrow B \wedge A$' and '$A \vee B \Leftrightarrow B \vee A$'.

(iii) '$(A \wedge B) \wedge C \Leftrightarrow A \wedge (B \wedge C)$' and '$(A \vee B) \vee C \Leftrightarrow A \vee (B \vee C)$'.

(iv) '$(A \wedge B) \vee C \Leftrightarrow (A \vee C) \wedge (B \vee C)$' and '$(A \vee B) \wedge C \Leftrightarrow (A \wedge C) \vee (B \wedge C)$'.

(v) '$\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B$' and '$\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B$'.

The rules in (v) are known as de Morgan's laws. All of these can be checked by evaluating the respective truth tables using the rules from the previous subsection.

1.4. Statements on Sets using Quantifiers

Let M be a set and let $P(x)$ be a property depending on x. Using quantifiers new statements can be generated:

(i) *The universal quantifier $\forall$*: The statement '$\forall\, x \in M : P(x)$' is true, if for all elements x of M the property $P(x)$ holds.

(ii) *The existential quantifier $\exists$*: The statement '$\exists\, x \in M : P(x)$' is true, if there is (at least) an element of M such that $P(x)$ holds.

(iii) *The unique existential quantifier $\exists!$*: The statement '$\exists!\, x \in M : P(x)$' is true, if there is one and only one element x of M such that P applies to x.

1.5. Negation interchanges Quantifiers

The following equivalences hold:

$$\neg(\exists\, x \in M : P(x)) \Longleftrightarrow \forall\, x \in M : \neg P(x),$$
$$\neg(\forall\, x \in M : P(x)) \Longleftrightarrow \exists\, x \in M : \neg P(x).$$

2. Constructions in Set Theory

Definition I.2.1 (Subset): Let M_1 be a set and let M_2 be another set. If any x in M_2 also belongs to M_1, then M_2 is called a *subset of* M_1. In this case, we write $M_2 \subseteq M_1$.

The statement '$M_2 \subseteq M_1 \Leftrightarrow \forall\, x \in M_2 : x \in M_1$' expresses the previous definitions in terms of quantifiers.

Definition I.2.2 (Constructions in Set Theory): Let M_1 and M_2 be sets.

(i) The set $M_1 \cap M_2 = \{x \mid x \in M_1 \wedge x \in M_2\}$ is called *intersection of* M_1 *and* M_2.

(ii) The set $M_1 \cup M_2 = \{x \mid x \in M_1 \vee x \in M2\}$ is called *union of* M_1 *and* M_2.

(iii) The set $M_1 - M_2 = \{x \mid x \in M_1 \wedge x \notin M_2\}$ is called *difference of* M_1 *and* M_2. It also is denoted $M_1 \setminus M_2$.

(iv) The set $M_1 \times M_2 = \{(x, y) \mid x \in M_1, y \in M_2\}$ is called *cartesian product of* M_1 *and* M_2.

(v) For natural k, the set $M_1^k = \{(x_1, \ldots, x_k) \mid x_1 \in M_1 \wedge \cdots \wedge x_k \in M_1\}$ is called *k-fold cartesian power of* M_1 or *k-fold cartesian product of* M_1 *with itself.*

(vi) The set $\mathfrak{P}(M_1) = \{M \mid M \subseteq M_1\}$ of subsets of M_1 is called *power set of* M_1.

Example I.2.3: Consider the four sets $M_1 = \{1, 2\}$, $M_2 = \{1, 2, 3\}$, $M_3 = \varnothing$ and $M_4 = \{1, 7, a, b\}$. For these we may observe

(i) $M_3 \subseteq M_1 \subseteq M_2$.

(ii) $M_2 \cap M_4 = \{1\}$, $M_1 \cap M_2 = \{1, 2\}$, $M_2 \cap M_3 = \varnothing$.

(iii) $M_2 \cup M_4 = \{1, 2, 3, 7, a, b\}$, $M_1 \cup M_2 = \{1, 2\}$, $M_2 \cup M_3 = M_2$.

(iv) $M_1 \times M_4 = \{(1,1),(1,7),(1,a),(1,b),(2,1),(2,7),(2,a),(2,b)\}$.

(v) $M_2 - M_4 = \{2,3\}$.

(vi) $\mathfrak{P}(M_2) = \{\{1,2,3\},\{1,2\},\{1,3\},\{2,3\},\{1\},\{2\},\{3\},\varnothing\}$.

Notation I.2.4: Let M_1, M_2 and M_3 be sets. We define '$M_2 \supseteq M_1$' to mean the same as '$M_1 \subseteq M_2$', '$x \notin M$' to mean the same as '$\neg(x \in M)$' and '$M_1 \subsetneq M_2$' to mean '$M_1 \subseteq M_2 \wedge M_1 \neq M_2$'. If $M_1 \subsetneq M_2$, we call M_1 a *proper subset of* M_2. If M_1 is a subset of M_2, then the difference $M_2 - M_1$ is called *complement of* M_1 *in* M_2 and denoted $M_1^{\complement}$ or $C_{M_2}(M_1)$.

Remark I.2.5: Let M, M_1 and M_2 be sets. Then it holds:

(i) $M \subseteq M$.

(ii) It is $M_1 = M_2$ if and only if $M_1 \subseteq M_2$ and $M_2 \subseteq M_1$.

Proof: (i) We have to show that for any x in M it holds '$x \in M$'. This is clearly true.

(ii) This follows from extensionality. □

Proposition I.2.6: *Let M_1, M_2 and M_3 be sets. Then it holds:*

(i) $(M_1 \cup M_2) \cup M_3 = M_1 \cup (M_2 \cup M_3)$, $(M_1 \cap M_2) \cap M_3 = M_1 \cap (M_2 \cap M_3)$.

(ii) $M_1 \cup M_2 = M_2 \cup M_1$, $M_1 \cap M_2 = M_2 \cap M_1$.

(iii) $M_1 \cap (M_2 \cup M_3) = (M_1 \cap M_2) \cup (M_1 \cap M_3)$ *and* $M_1 \cup (M_2 \cap M_3) = (M_1 \cup M_3) \cap (M_1 \cup M_3)$.

Proof: Using extensionality, we exemplarily show the second assertion of (ii). Per definition we have the following equivalences:

$$\begin{aligned} x \in (M_1 \cap M_2) \cap M_3 &\iff x \in (M_1 \cap M_2) \wedge x \in M_3 \\ &\iff (x \in M_1 \wedge x \in M_2) \wedge x \in M_3 \\ &\iff x \in M_1 \wedge x \in M_2 \wedge x \in M_3 \qquad \text{(Subsection 1.3(iii))} \\ &\iff x \in M_1 \wedge (x \in M_2 \cap M_3) \\ &\iff x \in M_1 \cap (M_2 \cap M_3). \end{aligned}$$

The other assertions are shown in precisely the same manner. One says they are shown 'analogously'. □

Proposition I.2.7: *Let M be a set and let M_1, M_2 be subsets of M. Then it holds:*

(i) $M - (M - M_1) = C_M(C_M(M_1)) = (M_1^{\mathsf{c}})^{\mathsf{c}} = M_1$,

(ii) $M - M = \varnothing$,

(iii) $M - \varnothing = M$,

(iv) $(M_1 \cup M_2)^{\mathsf{c}} = M_1^{\mathsf{c}} \cap M_2^{\mathsf{c}}$,

(v) $(M_1 \cap M_2)^{\mathsf{c}} = M_1^{\mathsf{c}} \cup M_2^{\mathsf{c}}$.

Proof: Again, we don't show all assertions. Assetions (iv) and (v) are left as an exercise on one of the exercise sheets.

(i) The following equivalences hold:

$$\begin{aligned} x \in M - (M - M_1) &\iff x \in M \wedge x \notin M - M_1 \\ &\iff x \in M \wedge \neg(x \in M \wedge x \notin M_1) \\ &\iff x \in M \wedge (x \notin M \vee x \in M_1) \\ &\iff (x \in M \wedge x \notin M) \vee (x \in M \wedge x \in M_1) \\ &\iff x \in M \wedge x \in M_1 \\ &\iff x \in M_1. \end{aligned}$$

Here, we again used Subsection 1.3(iii) and in the last step we made use of the fact that M_1 is a subset of M. The above shows $M - (M - M_1) = M_1$.

(ii) Some x belongs to $M - M$ if and only if x belongs to M and does not belong to M. This is always wrong, hence $M - M = \varnothing$.

(iii) Using (ii) we may write $M - \varnothing = M - (M - M)$, and using (i) we can read off that $M - (M - M) = M = M - \varnothing$. □

Proposition I.2.8: *Let M_1, M_2 and M_3 be sets. Then it holds:*

(i) $M_1 \subseteq M_1 \cup M_2$,

(ii) $M_1 \cap M_2 \subseteq M_i$, $i = 1, 2$,

(iii) *If $M_1 \subseteq M_2$ and if $M_2 \subseteq M_3$, then $M_1 \subseteq M_3$,*

(iv) $M_1 \cup \varnothing = M_1$,

(v) $M_1 \cap \varnothing = \varnothing$,

(vi) $M_1 \subseteq M_2$ *if and only if* $M_1 \cap M_2 = M_1$,

(vii) $M_1 \subseteq M_2$ *if and only if* $M_1 \cup M_2 = M_2$.

These assertions should not pose a problem to you, if you try reasoning like in the previous examples. Some assertions will be exercises on an exercise sheet. It is a good idea to try convincing yourself that the remaining assertions are true, too.

3. Useful Proof Techniques

For this section, let A, B, M_1 and M_2 be sets.

3.1. Subset Verification

To show that A is a subset of B, we use the following technique: It holds $A \subseteq B$ if and only if each a in A belongs to B, too. For example, we might want to show $M_1 \cap M_2 \subseteq M_1$.

Let x be an element of $M_1 \cap M_2$. This is the case if and only if x belongs to M_1 and x belongs to M_2. In particular it holds '$x \in M_1$', i.e. $M_1 \cap M_2 \subseteq M_1$.

3.2. Equality of Sets

To show equality of sets A and B, we have to show that $A \subseteq B$ and $B \subseteq A$. For example, we might want to show for set M_1 and M_2 that, if M_1 is a subset of M_2, $M_1 \cap M_2 = M_1$.

We assume $M_1 \subseteq M_2$.

'$M_1 \cap M_2 \subseteq M_1$': We asserted in Proposition I.2.8(ii) that $M_1 \cap M_2 \subseteq M_1$.

'$M_1 \subseteq M_1 \cap M_2$': Let x be an element of M_1. Since M_1 is a subset of M_2, x then also belongs to M_2, i.e. it holds '$x \in M_1$' and '$x \in M_2$'. Thus x belongs to $M_1 \cap M_2$.

Since both inclusions hold, we have the equality $M_1 \cap M_2 = M_1$.

It is customary to abbreviate '$M_1 \cap M_2 \subseteq M_1$' in such a proof to '$\subseteq$'. Similarly for the other inclusion.

3.3. Equivalence of Assertions

Let A and B be assertions. Both assertions are equivalent, i.e. '$A \iff B$', if and only if '$A \implies B$' and '$B \implies A$' are true.

For example, we might want to show the equivalence of the statements '$M_1 \subseteq M_2$' and '$M_1 \cap M_2 = M_1$'. As for inclusions, one usually leaves out the respective statements and just notates the arrows.

'$\Longrightarrow$': If M_1 is a subset of M_2, then it follows $M_1 \cap M_2 = M_1$, as shown in the above subsection.

'$\Longleftarrow$': Suppose for the sets M_1 and M_2 that $M_1 \cap M_2 = M_1$ and let x be an element of M_1. By assumption x belongs to $M_1 \cap M_2$, i.e. x belongs to both M_1 and M_2. In particular, x is an element of M_2 and thus $M_1 \subseteq M_2$.

3.4. Proof by Contraposition

Let A and B be statements. The implication '$A \Longrightarrow B$' is true if and only if the implication '$\neg B \Longrightarrow \neg A$' is true. This can be shown using a truth table. Showing the latter implication rather than the direct implication is known as proof by contraposition. Examples of this proof technique will be presented in the exercise sessions.

3.5. Proof by Induction

Let $A(n)$ be a statement depending on a natural number n, let $S = \{n \in \mathbb{N} \mid A(n) \text{ is true}\}$ and let s_0 be an element of S. If we can show 'If n belongs to S, then so does $n+1$', then the statement $A(n)$ is true for all $n \geq s_0$. This principle is called 'proof by induction' and is rooted in the *Peano axioms for natural numbers*.

We might for example consider the following statement: For the natural number n, let $T(n) = 1 + 2 + \cdots + n$. Then $T(n) = \frac{1}{2}n(n+1)$. This formula for $T(n)$ is indeed true and can be shown via induction.

Firstly, $S = \{n \in \mathbb{N} \mid T(n) = \frac{1}{2}n(n+1)\}$. Since $T(1) = 1 = \frac{1}{2} \cdot 1 \cdot 2$, 1 belongs to S. If n is an element of S, i.e. if $T(n) = \frac{1}{2}n(n+1)$, then we may proceed as follows to show that $n+1$ also belongs to S:

$$\begin{aligned} T(n+1) &= 1 + 2 + \cdots + (n+1) \\ &= T(n) + (n+1) \\ &= \frac{n(n+1)}{2} + (n+1) \\ &= \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}. \end{aligned}$$

Hence the statement holds for any positive natural number n.

4. Maps

For this section let W, X, Y and Z be sets.

Notation I.4.1: If $M = \{a_1, \ldots, a_n\}$ is a finite set consisting of n elements, then n is called the *cardinality* or *size* of M and denoted $|M|$ or $\#M$.

Definition I.4.2: A *map* or *function* $f\colon X \to Y$ is an assignment that assigns to each x in X precisely one y in Y. In this situation we write

$$f\colon X \longrightarrow Y, \qquad x \longmapsto y = f(x).$$

The set X is called *domain of* f, and Y is called *range* or *image of* f. If no confusion is to be feared, the mention of f may be dropped. The set of all maps from X to Y is denoted

$$Y^X = \{f \mid f\colon X \to Y \text{ is a map}\}.$$

Remark I.4.3: (i) To be more precise, we can give the following set theoretic definition of a map: A map $f\colon X \to Y$ is a subset Γ_f of $X \times Y$ such that for all x in X there is one and only one y in Y with (x, y) in Γ_f. We write $x \mapsto y = f(x)$ if and only if (x, y) belongs to Γ_f. The set Γ_f is called (set theoretic) *graph of* f.

(ii) Two maps $f\colon X \to Y$ and $g\colon X \to Y$ are the same if and only if their graphs coincide, i.e. if for any x in X it holds $f(x) = g(x)$.

(iii) The set of maps $\varnothing^Y$ contains precisely one element f, whose graph Γ_f is the empty set. Here, $\varnothing$ is regarded as subset of $\varnothing \times Y = \varnothing$.

Example I.4.4: (i) For $X_1 = \{-1, 0, 1\}$ and $Y_1 = \{0, 1\}$, consider the two maps

$$f_1\colon X_1 \longrightarrow Y_1, \quad x \longmapsto x^3 - x, \qquad g_1\colon X_1 \longrightarrow Y_1, \quad x \longmapsto 0.$$

Then $f_1 = g_1$.

(ii) For $X_2 = \mathbb{R}$ and $Y_2 = \mathbb{R}_{\geq 0}$, the assignment $f_2\colon X_2 \to Y_2$, $x \mapsto x^2$ is an example for a map.

(iii) Let $X_3 = \mathbb{R}_{\geq 0}$ and $Y_3 = \mathbb{R}$. Then $f_3\colon X_3 \to Y_3$, $x \mapsto \sqrt{x}$ is a map.

(iv) Consider the two sets $X_4 = \{s \mid s \text{ is a student in this lecture hall}\}$ and $Y_4 = \{t \mid t \text{ is date of a day of the year}\}$. Then the assignment $f_4\colon X_4 \to Y_4$, $s \mapsto$ (Birthday of s) is a map between X_4 and Y_4.

Definition I.4.5 (Identity): Let M be a set. The map $\mathrm{id}_M\colon M \to M$, $x \mapsto x$ is called *identity of* M.

Definition I.4.6 (Preimage and Image): Let $f\colon X \to Y$ be a map. For a subset B of Y, $f^{-1}(B) = \{x \in X \mid f(x) \in B\}$ is called the *preimage of B under f*. For a subset A of X, $f(A) = \{f(x) \mid x \in A\}$ is called the *image of A under f*.

To be more precise, the map $f\colon X \to Y$ induces a map $f\colon \mathfrak{P}(X) \to \mathfrak{P}(Y)$, $A \mapsto f(A)$, which we—by abuse of notation—also denote by f. Same goes for $f^{-1}\colon \mathfrak{P}(Y) \to \mathfrak{P}(X)$. Note that firstly this is a safe praxis, because from context it is always clear which map is meant, and secondly that f^{-1} is not to be confused with an inverse to f. Taking preimages always makes sense, whereas inverses only exist in special circumstances which we will discuss later.

Example I.4.7: Going back to the maps from Example I.4.4, we have the following:

- $f_1^{-1}(\{0\}) = X_1$, $f_1^{-1}(\{1\}) = \varnothing$, $f_1^{-1}(Y_1) = X_1$, $f_1^{-1}(\varnothing) = \varnothing$ as well as $f_1(\{1\}) = f_1(\{0\}) = f_1(\{-1\}) = \{0\}$.
- $f_2^{-1}(\{y \in \mathbb{R} \mid y \geq 1\} = \{x \in \mathbb{R} \mid x \leq -1 \vee x \geq 1\}$.
- $f_3(\mathbb{R}_{\geq 0}) = \mathbb{R}_{\geq 0}$.

Definition I.4.8 (Composition and Restriction): Let $f\colon X \to Y$, $g\colon Y \to Z$ be two maps. The map $g \circ f\colon X \to Z$, $x \mapsto (g \circ f)(x) = g(f(x))$ is called *composition of f and g*.

Let $f\colon X \to Y$ be a map and let A be a subset of X. Then $f|_A\colon A \to Y$, $a \mapsto f(a)$ is called *restriction of f to A*.

Remark I.4.9 (Categorial Property): (i) Composition of maps is *associative*, i.e. for maps $f\colon W \to X$, $g\colon X \to Y$ and $h\colon Y \to Z$ it holds $h \circ (g \circ f) = (h \circ g) \circ f$. In the following, we thus write $h \circ g \circ f$ instead of either expression with parentheses.

(ii) The identity 'doesn't do anything' in compositions, that is for a map $f\colon X \to Y$ it holds $\mathrm{id}_Y \circ f = f = f \circ \mathrm{id}_X$.

Proof: To show equality of maps, one has to check that on any element of the domain both maps do the same. Using this strategy, we can verify both assertions.

(i) Let w be any element of W, the domain of $(h \circ g) \circ f$ and $h \circ (g \circ f)$. Then

$$\begin{aligned}(h \circ (g \circ f))(w) = h((g \circ f)(w)) &= h(g(f(w)) \\ &= (h \circ g)(f(w)) = ((h \circ g) \circ f)(w),\end{aligned}$$

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

(ii) For any x in X we have $(\mathrm{id}_Y \circ f)(x) = \mathrm{id}_Y(f(x)) = f(x)$, so $\mathrm{id}_Y \circ f = f$. The same argument shows $f = f \circ \mathrm{id}_X$. □

Example I.4.10: The composition of f_2 and f_3 from Example I.4.4 is given by $(f_3 \circ f_2)(x) = \sqrt{x^2} = |x|$.

Definition I.4.11 (Injective, Surjective, Bijective): Let $f\colon X \to Y$ be a map.

(i) If for any two x_1, x_2 in X with $f(x_1) = f(x_2)$ it holds $x_1 = x_2$, then f is called *injective*. This is the case if and only if for any y in Y, $\#f^{-1}(\{y\}) \leq 1$.

(ii) If for any y in Y there is some x in X with $f(x) = y$, then f is called *surjective*. This is the case if and only if for any y in Y, $\#f^{-1}(\{y\}) \geq 1$.

(iii) If for any y in Y there is one and only one x in X with $f(x) = y$, then f is called *bijective*. This is the case if and only if for any y in Y, $\#f^{-1}(\{y\}) = 1$.

A map is bijective if and only if it is both injective and surjective.

Definition I.4.12 (Inverse Map): Let $f\colon X \to Y$ and $g\colon Y \to X$ be maps. If for any x in X it holds $g(f(x)) = x$—i.e. if $g \circ f = \mathrm{id}_X$— and if for any y in Y it holds $f(g(y)) = y$—i.e. if $f \circ g = \mathrm{id}_Y$—then g is called *inverse map of* f. In this case, g is usually denoted f^{-1}.

Proposition I.4.13 (Properties of Bijective Maps): *Let $f\colon X \to Y$ be a map.*

(i) *There is an inverse map to f if and only if f is bijective.*

(ii) *If f has an inverse map, then it is uniquely determined by f and bijective itself. If this is the case, we denote it by f^{-1}.*

(iii) *If $f\colon X \to Y$ and $g\colon Y \to Z$ are bijective maps, then so is $g \circ f$. For its inverse it holds $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.*

Proof: (i) This assertion is an exercise on Sheet 1.

(ii) There are two parts to this assertion. We have to show bijectivity of the inverse, should it exist, and uniqueness of the inverse. First we takle uniqueness. Assume $g_1, g_2\colon Y \to X$ are both inverses to f, which means $g_i \circ f = \mathrm{id}_X$ and $f \circ g_i = \mathrm{id}_Y$. For any y in Y we then have

$$g_1(y) = (g_1 \circ \mathrm{id}_Y)(y) = (g_1 \circ (f \circ g_2))(y) = ((g_1 \circ f) \circ g_2) = (\mathrm{id}_X \circ g_2)(y) = g_2(y).$$

Here, we used a bunch of properties we showed before. We first used that the identity doesn't do anything (Remark I.4.9(ii)), then we used the definition of the inverse map, then associativity of composition (Remark I.4.9(i)), again the associativity of compositions and finally again the neutrality of the identity.

Now for bijectivity. Let g be the inverse map of f. This means it holds $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. As per definition, f is an inverse map to g, which makes g bijective by (i).

(iii) By assumption, f and g are bijective. Let f^{-1} and g^{-1} be the respective (unique) inverse maps which exist by (i). Then

$$(g \circ f) \circ (f^{-1} \circ g^{-1}) = g \circ f \circ f^{-1} \circ g^{-1} = g \circ \mathrm{id}_Y \circ g^{-1} = g \circ g^{-1} = \mathrm{id}_Z,$$

same for the other composition. This makes $f^{-1} \circ g^{-1}$ an inverse map to $g \circ f$, and by (ii) this inverse is unique, i.e. in fact $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. Assertion (i) now ensures bijectivity of $g \circ f$. □

Remark I.4.14: Let $f\colon X \to Y$ be a map and let b be an element of Y. If f is bijective, then f^{-1} has two meanings. There is the induced map $f^{-1}\colon \mathfrak{P}(Y) \to \mathfrak{P}(X)$ and there is the inverse map $f^{-1}\colon Y \to X$. For b as chosen, $f^{-1}(\{b\})$ means the set of preimages of b under f and $f^{-1}(b)$ is the unique element a of X such that $f(a) = b$. In this case, $f^{-1}(\{b\}) = \{f^{-1}(b)\}$, justifying our abuse of notation.

Example I.4.15 (Cartesian Products): Let k be a natural number. The k-fold cartesian power X^k of X with itself can be identified with $X^{\{1,\dots,k\}}$ by means of the following map:

$$\begin{aligned} F\colon X^k &\longrightarrow X^{\{1,\dots,k\}}, \\ (a_1, \dots, a_k) &\longmapsto (f\colon \{1, \dots, k\} \to X, \quad i \mapsto a_i) \end{aligned}$$

Because we can easily write down the inverse, namely the map G defined by $f \mapsto (f(1), \dots, f(k))$, we see that both maps are bijective.

The above example gives an insight into the origin of the notation '$X^{\{1,\dots,k\}}$' and it also justifies the following definition:

Definition I.4.16: We write X^0 for the one-point-set $X^{\varnothing}$ containing the empty map.

Definition I.4.17 (Permutations of a Set): The subset

$$\mathrm{Perm}(X) = \{f\colon X \to X \text{ bijective}\}$$

of X^X is called the *set of permutations of* X.

If $X = \{a_1, \ldots, n\}$ is a set with n elements, then $\mathrm{Perm}(X)$ consists of $n!$ maps. This can be shown using regular combinatorial arguments, e.g. using binomials and their properties.

5. Relations

For this section, let M be a set.

Definition I.5.1 (Relation): A subset R of $M \times M = M^2$ is called a *binary relation on M*. Instead of '$(x, y) \in R$' it is customary to write xRy or $x \sim_R y$.

Example I.5.2: (i) The relation $R_1 = \{(x, y) \in M \times M \mid x = y\}$ is called *equality relation on M*. This means xR_1y holds true if and only if $x = y$.

(ii) The following sets are relations on the set of real numbers $\mathbb{R}$:

$$R_2 = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \leq y\} \qquad R_3 = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x < y\}$$
$$R_4 = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \geq y\} \qquad R_5 = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x > y\}$$
$$R_6 = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \neq y\}$$

Definition I.5.3 (Properties of Relations): Let $R \subseteq M \times M$ be a relation.

(i) If for all x in M it holds xRx, then R is called *reflexive*.

(ii) If for any pair x, y in M with xRy it holds yRx, then R is called *symmetric*.

(iii) If for any pair x, y in M with xRy and yRx it holds $x = y$, then R is called *antisymmetric*.

(iv) If for any x, y, z in M with xRy and yRz it holds xRz, then R is called *transitive*.

Example I.5.4: For the relations in Example I.5.2 we have the following table:

	R_1 ('=')	R_2 ('$\leq$')	R_3 ('$<$')	R_6 ('$\neq$')
reflexive	✓	✓	–	–
symmetric	✓	–	–	✓
antisymmetric	✓	✓	✓	–
transitive	✓	✓	✓	–

Definition I.5.5 (Equivalence Relations and Orders): Let R be a binary relation on M. If R is reflexive, symmetric and transitive, then R is called an *equivalence relation on M*. If R is reflexive, antisymmetric and transitive, then R is called an *order on M*.

In literature—both german and english—relations that we call orders here are often referred to as *partial orders*. An order is then what we will call a total or linear order.

Example I.5.6: In Example I.5.2, R_1 is an equivalence relation and R_1, R_2 as well as R_4 are orders.

Example I.5.7 (Congruence): Let M be the set of integers $\mathbb{Z}$ and let n be a natural number. If for an integer a there is some integer k such that $a = kn$, we say n *divides* a. In this case, we write '$n \mid a$'. The relation

$$R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid n \text{ divides } a - b\}$$

is called *congruence modulo* n. Instead of 'aRb' one often writes '$a \equiv b \pmod{n}$' or '$a \equiv_n b$'. For example, $2 \equiv_5 12$, $2 \equiv_3 5$ or $2 \equiv_4 -2$.

Remark I.5.8: Congruence modulo n is an equivalence relation.

Proof: We have to show three properties: Reflexivity, symmetry and transitivity. The game is always the same: Write down assumptions, translate and look at what needs to be shown.

To show reflexivity, we have to show that each integer a is congruent to itself modulo n. This would mean that $a - a = 0$ is divisible by n. Clearly, for $k = 0$ we have $a - a = 0 = 0n$, so this is true.

For symmetry, we start with two integers a and b that are congruent modulo n. This means that $a - b$ is divisible by n, i.e. there is some integer k with $kn = a - b$. We now have to find an integer ℓ such that $\ell n = b - a = -(a - b)$. Hence $-k$ does the trick.

For transitivity, we again start with three integers a, b and c and assume that $a \equiv b \pmod{n}$ as well as $b \equiv c \pmod{n}$. We want to show that then, $a \equiv c \pmod{n}$. To get there, we break down our assumptions and where we want to go, then it will be clear how it is done.

By assumption we have integers k and ℓ such that $a - b = kn$ and $b - c = \ell n$. What we need is an integer m such that $a - c = mn$. Because of distributivity, we don't have to work hard for m: $a - c = (a - b) + (b - c) = kn + \ell n = (k + \ell)n$.□

Remark I.5.9: Let n be a natural number. Congruence modulo n gives a partition of $\mathbb{Z}$ into n sets $M_1 = \{a \in \mathbb{Z} \mid a \equiv_n 1\}$, $M_2 = \{a \in \mathbb{Z} \mid a \equiv_n 2\}$, $\ldots M_n = \{a \in \mathbb{Z} \mid a \equiv_n 0\}$. These sets are called *residue classes modulo* n.

Example I.5.10: Let n be the natural number 2. Then the sets M_1 and M_2 as declared in Remark I.5.9 are the sets of even respectively uneven integers.

Definition I.5.11: Let '$\sim$' be an equivalence relation on M and let x be an element of M. Then the subset $[x]_\sim = \{y \in M \mid x \sim y\}$ of M is called *equivalence class of x in M with respect to '$\sim$'.*

Example I.5.12: Let $M = \mathbb{Z}$, let n be a natural number, let '$\sim$' be '$\equiv_n$' and let $x = 1$. Then $[1]_\sim = \{y \in \mathbb{Z} \mid 1 \equiv_n y\} = \{1 + kn \mid k \in \mathbb{Z}\}$ is the equivalence class of x.

Theorem 1: *Let '$\sim$' be an equivalence relation on M. Then the following assertions hold:*

(i) *All equivalence classes are non-empty, i.e. for any x in M it holds $[x]_\sim \neq \varnothing$.*

(ii) *If x and y are in relation, then $[x]_\sim = [y]_\sim$.*

(iii) *The set M is the union of all equivalence classes, i.e. any element of M is contained in an equivalence class.*

(iv) *Any two distinct equivalence classes are disjoint, i.e. for any x and y in M it holds $[x]_\sim = [y]_\sim$ or $[x]_\sim \cap [y]_\sim = \varnothing$.*

Proof: (i) Let x be an element of M. Because '$\sim$' is reflexive, $x \sim x$. This means that x belongs to $[x]_\sim$ and thus $[x]_\sim \neq \varnothing$.

(ii) Let x and y be elements of M such that $x \sim y$. For any x' in $[x]_\sim$ it holds $x' \sim x$ by definition. Because '$\sim$' is transitive, $x' \sim x$ and $x \sim y$ imply that $x' \sim y$, which means that x' belongs to $[y]_\sim$. This shows $[x]_\sim \subseteq [y]_\sim$. By changing names, $[y]_\sim \subseteq [x]_\sim$ follows and thus $[x]_\sim = [y]_\sim$.

(iii) Because the equivalence classes are subset of M we get for free that their union is contained in M. The other inclusion follows from our arguments in (i).

(iv) Let x and y be elements of M such that $[x]_\sim \cap [y]_\sim \neq \varnothing$. Because the intersection is non-empty, we find some z in $[x]_\sim \cap [y]_\sim$. By (ii) we obtain $[x]_\sim = [z]_\sim$ and $[z]_\sim = [y]_\sim$ so that $[x]_\sim = [z]_\sim$. □

Definition I.5.13 (Quotient by Equivalence Relation): Let '$\sim$' be an equivalence relation on M. Then $M_\sim = M/\sim = \{[x]_\sim \mid x \in M\}$ is called *set of equivalence classes in M with respect to '$\sim$'* or *M modulo $\sim$*. The map $\pi\colon M \to M_\sim$, $x \mapsto [x]_\sim$ is called *canonical projection.*

Definition I.5.14 (General Intersections and Unions): Let I be a set and for any element i of I let M_i be a set. Then we define

$$\bigcap_{i\in I} M_i = \{x \mid \text{For all } i \text{ in } I : x \in M_i\}, \quad \bigcup_{i\in I} M_i = \{x \mid \text{There is } i \text{ in } I : x \in M_i\}.$$

Note that in literature, I in the above context is sometimes called *index set*. This does not specify the set any further, it is just meant to convey that the elements of I are to be understood as 'indices'. Our definition works for anything from finite sets to uncountable unordered sets.

Definition I.5.15 (Partition): Let P be a subset of $\mathfrak{P}(M)$. If the empty set is not contained in P, if $\bigcup_{A\in P} A = M$ and if any two distinct A and B in P are disjoint, P is called a *partition*.

Corollary I.5.16 (of Theorem 1): *Let M be a set and let '$\sim$' be an equivalence relation on M. Then the quotient $M/\sim$ is a partition.*

On an exercise sheet you will show that in fact equivalence relations and partitions are 'the same', that is using a partition, one can declare an equivalence relation.

Heuristically, this is what you should picture when thinking of equivalence relations. They are meant to partition a set into sets of elements that share a certain property. Congruence modulo n partitions the set of integers $\mathbb{Z}$ into sets of integers that leave the same remainder when dividing with remainder by n.

6. Addendum and Outlook

For a formal introduction to set theory, see for example the books ...

Set theory is founded on axioms, i.e. rules that are meant to hold for sets are defined. Different axiomatic systems are in use, mostly depending on the 'respective understanding of infinity'; we use ZFC in this lecture. Among the ZFC axioms are the *axiom of extensionality*, the *axiom schema of specification*, the *empty set axiom*, which ensures the existence of the empty set, and the *axiom of choice*. Surprisingly, the axiom of choice can't be deduced from the other axioms—otherwise it needn't be an axiom—and it is intriguing to think through which statements do need the axiom of choice.

Chapter II.

Vector Spaces and Systems of Linear Equations

For typographical reasons, in these notes we use the notation

$$(x_1, \ldots, x_n)^t = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

for column vectors, mostly in running text. This notation makes sense from a methodical point of view, too, as we will learn later. In the following, we write $\mathbb{R}^n = \{(x_1, \ldots, x_n)^t \mid x_1, \ldots, x_n \in \mathbb{R}\}$.

1. Motivation

Example II.1.1: Consider the following system of linear equations:

$$\begin{aligned} 2x_1 + 6x_2 + 4x_3 &= 8, \\ x_2 - 2x_3 &= 6, \\ x_1 + 4x_2 &= 10. \end{aligned} \tag{II.1}$$

We are interested in the solution set

$$\mathbb{S} = \{x = (x_1, x_2, x_3) \in \mathbb{R}^3 \mid x \text{ satisfies } Eq.\ (\text{II.1})\}.$$

Important questions, among others, are: 'Does Eq. (II.1) have a solution?', 'If Eq. (II.1) has a solution, how many are there? 1, 2, 3? Infinitely many?', 'If there are infinitely many solutions, how can they be stated? What structure does $\mathbb{S}$ have? How "big" is $\mathbb{S}$?', 'Is there a general solution procedure for systems of linear equations to determine $\mathbb{S}$?'

To all those questions, we will have given a satisfactory answer by the end of this chapter.

Idea II.1.2: (i) *Matrix notation*: The system of linear equations Eq. (II.1) is determined by the data

$$A = \begin{pmatrix} 2 & 6 & 4 \\ 0 & 1 & -2 \\ 1 & 4 & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} 8 \\ 6 \\ 10 \end{pmatrix},$$

so that Eq. (II.1) may be expressed as augmented matrix $(A|b)$.

(ii) *Reduction to homogeneous system of linear equations*: If $x' = (x'_1, x'_2, x'_3)^t$ is a solution for Eq. (II.1), then, for any other solution $x = (x_1, x_2, x_3)^t$ of Eq. (II.1) it holds

$$2(x'_1-x_1)+6(x'_2-x_2)+4(x'_3-x_3) = 2x'_1+6x'_2+4x'_3-(2x_1+6x_2+4x_3) = 8-8 = 0,$$

analogously for the other 2 equations in Eq. (II.1). This means that the vector $y = x' - x = (x'_1 - x_1, x'_2 - x_2, x'_3 - x_3)^t$ is a solution for

$$\begin{aligned} 2x_1 + 6x_2 + 4x_3 &= 0, \\ x_2 - 2x_3 &= 0, \\ x_1 + 4x_2 &= 0. \end{aligned} \tag{II.2}$$

Thus, for any solution x of Eq. (II.1), it holds $x = x' + y$, where y is a solution of Eq. (II.2). We call Eq. (II.2) the *homogeneous system of linear equations to Eq.* (II.1). This means there is merit in studying solution sets of homogeneous systems of linear equations.

(iii) *Structure of the solution set of a homogeneous system of linear equations*: Let $x = (x_1, x_2, x_3)^t$ and $y = (y_1, y_2, y_3)^t$ be solutions of Eq. (II.2), and let λ be a real number. For their sum $v = (v_1, v_2, v_3)^t = (x_1 + y_1, x_2 + y_2, x_3 + y_3)^t$ and for the multiple $w = (w_1, w_2, w_3)^t = (\lambda x_1, \lambda x_2, \lambda x_3)^t$ it then holds

$$\begin{aligned} 2v_1 + 6v_2 + 4v_3 &= 2(x_1 + y_1) + 6(x_2 + y_2) + 4(x_3 + y_3) \\ &= 2x_1 + 6x_2 + 4x_3 + 2y_1 + 6y_2 + 4y_3 = 0 + 0 = 0 \end{aligned}$$

and

$$2w_1 + 6w_2 + 4w_3 = 2(\lambda x_1) + 6(\lambda x_2) + 4(\lambda x_3) = \lambda(2x_1 + 6x_2 + 4x_3) = \lambda 0 = 0;$$

same for the other two equations from Eq. (II.2). Hence v and w are also solutions to Eq. (II.2). We just showed that for the solution set $\mathbb{S}_h$ of Eq. (II.2) the following is true: If x and y belong to $\mathbb{S}_h$ and if λ is a real number, then $x + y$ and λx lie in $\mathbb{S}_h$, too.

2. Vector Spaces

In this section, we want to introduce the concept of an $\mathbb{R}$-vector space, generalising $\mathbb{R}^n$ to something more abstract. The structure of $\mathbb{R}^n$ is determined by *vector addition*, *scalar multiplication* and the rules that hold for both the laws of composition individually as well as their interplay.

Reminder II.2.1: Let $x = (x_1, \ldots, x_n)^t$, $y = (y_1, \ldots, y_n)^t$ be elements of $\mathbb{R}^n$ and let λ be a real number. Then

$$x + y = (x_1 + y_1, \ldots, x_n + y_n)^t, \qquad \lambda x = (\lambda x_1, \ldots, \lambda x_n)^t.$$

For any elements x, y and z of $\mathbb{R}^n$ it holds

(i) $(x + y) + z = x + (y + z)$.

(ii) $x + y = y + x$,

Law (i) is called *associative property of addition*, law (ii) is called *commutative property of addition*.

Definition II.2.2 (Real Vector Space): Let V be a set, let $\mathbf{0} = \mathbf{0}_V$ be an element of V, and suppose there are two maps

$$\begin{aligned} +\colon V \times V &\longrightarrow V, \qquad (v, w) \longmapsto +(v, w) = v + w, \\ \cdot\colon \mathbb{R} \times V &\longrightarrow V, \qquad (\lambda, w) \longmapsto \cdot(\lambda, w) = \lambda \cdot w. \end{aligned}$$

If for '$+$' it holds

(A1) For any x, y, z in V, $(x + y) + z = x + (y + z)$.

(A2) For any x, y in V, $x + y = y + x$.

(A3) For any x in V, $x + \mathbf{0} = x = \mathbf{0} + x$, i.e. $\mathbf{0}$ is a *zero vector*.

(A4) For any x in V there is one and only one y in V such that $x+y = \mathbf{0} = y+x$. This unique element y is called *additive inverse of* x and is notated $-x$.

and if for '$\cdot$' it holds

(M1) For any x in V, $1 \cdot x = x$.

(M2) For any real numbers λ_1, λ_2 and x in V, $(\lambda_1 + \lambda_2)x = \lambda_1 x + \lambda_2 x$.

(M3) For any real number λ and any x, y in V, $\lambda(x + y) = \lambda x + \lambda y$.

(M4) For any real numbers λ_1, λ_2 and x in V, $\lambda_1(\lambda_2 \cdot x) = (\lambda_1 \lambda_2) \cdot x$.

then the 4-tuple $(V, \mathbf{0}_V, +, \cdot)$ is called a *real vector space*. The maps '$+$' and '$\cdot$' are called *laws of composition*. The statements (A1)–(A4), (S1)–(S4) are called *vector space axioms*, (A1) is called *associative property of addition*, (A2) is called *commutative property of addition*, (M1) is called *unital property*, and (M2)–(M4) are called *distributive properties of scalar multiplication*.

Most often, the zero vector and the laws of composition do not get specified, because they are clear from context. For example, on $\mathbb{R}^n$ one almost always considers the laws of composition mentioned in Reminder II.2.1, and belonging to those is the zero vector $\mathbf{0}_{\mathbb{R}^n} = (0, \ldots, 0)^t$. If context prevents ambiguity, one imprecisely calls just V a real vector space.

Example II.2.3: (i) The field of real numbers $\mathbb{R}$ with '$+$' and '$\cdot$' as usual and $\mathbf{0}_{\mathbb{R}} = 0$ is an $\mathbb{R}$-vector space.

(ii) The n-fold cartesian product $\mathbb{R}^n$ together with the the laws of composition from Reminder II.2.1 and the zero vector $\mathbf{0} = (0, \ldots, 0)^t$ is an $\mathbb{R}$-vector space.

(iii) Let M be a set. Then $V = \mathbb{R}^M$ together with the laws of composition

$$\begin{aligned} +\colon \mathbb{R}^M \times \mathbb{R}^M &\longrightarrow \mathbb{R}^M, \qquad (f_1, f_2) \longmapsto \Big(g\colon M \to \mathbb{R}, \quad m \mapsto f_1(m) + f_2(m)\Big), \\ \cdot\colon \mathbb{R} \times \mathbb{R}^M &\longrightarrow \mathbb{R}^M, \qquad (\lambda, f) \longmapsto \Big(h\colon M \to \mathbb{R}, \quad m \mapsto \lambda f(m)\Big) \end{aligned}$$

turns into a real vector space. They are often called *pointwise operations*. Constructions like this are always possible, when the range carries 'enough structure'. In this case, images of elements of M are real numbers and those can be added as well as multiplied. The corresponding zero vector is $\mathbf{0}_{\mathbb{R}^M}\colon M \to \mathbb{R}$, $m \mapsto 0$.

If $M = \{1, \ldots, n\}$, then $\mathbb{R}^M$ can be identified with $\mathbb{R}^n$, see Example I.4.15.

(iv) Similarly, for any set M and real vector space W, W^M can be made into an $\mathbb{R}$-vector space like above.

We want to show that in (iv) we indeed constructed a new real vector space. For the axioms (A1)-(A4), we essentially have to argue why the properties of '$+$' on W transfer. For example, commutativity of addition of functions is shown as follows: Let $f, g\colon M \to W$ be functions. We want to show that the newly defined functions '$f + g$' and '$g + f$' agree, i.e. we have to show that any element m of M is mapped to the same element of W under $f + g$ and $g + f$. Because addition of W has the commutative property, for any m in M we find

$$(f + g)(m) = f(m) + g(m) = g(m) + f(m) = (g + f)(m).$$

This means that addition of functions as defined has the commutative property. All the other properties are shown similarly.

The hard part about this proof will probably be convincing oneself that it shows what needs to be shown. It does.

Proposition II.2.4: *Let $(V, +, \cdot, \mathbf{0}_V)$ be a real vector space. Then it holds:*

(i) *For any v, w in V there is one and only one x in V, such that $v + x = w$. We define $w - v = x$. In particular, $\mathbf{0}_V$ is the only element of V that satisfies (A3).*

(ii) *For any v, w in V, $w - v = w + (-v)$.*

(iii) *For any real number λ and any v in V, $\lambda\mathbf{0}_V = \mathbf{0}_V = 0v$. Furthermore, $-\mathbf{0}_V = \mathbf{0}_V$ and for any v in V, $\mathbf{0} - v = -v$.*

(iv) *For any real number λ and any v in V, $\lambda(-v) = (-\lambda)v$. In particular, $-v = (-1)v$.*

(v) *For any real number λ and any v, w in V, $\lambda(v - w) = \lambda v - \lambda w$.*

(vi) *For any real numbers λ_1, λ_2 and any v in V, $(\lambda_1 - \lambda_2)v = \lambda_1 v - \lambda_2 v$.*

The proof of these assertions is left as an exercise to the reader.

In the following, let $(V, +, \cdot, \mathbf{0}_V)$ always be a real vector space.

Definition II.2.5 (Linear Subspace): Let U be a subset of V. If $\mathbf{0}_V$ belongs to U, if for any x, y in U also $x + y$ lies in U and if for any real number λ and any x in U, λx belongs to U, then U is called a *vector subspace* or *linear subspace*.

That for x, y in U also $x + y$ belongs to U is often referred to as U being closed under addition. Same goes for scalar multiplication.

Proposition II.2.6 (Linear Subspaces are Vector Spaces): *Let U be a vector subspace of $(V, +, \cdot, \mathbf{0}_V)$. Then U together with $+|_{U\times U}$, $\cdot|_{\mathbb{R}\times U}$ and $\mathbf{0}_U = \mathbf{0}_V$ is a real vector space.*

Proof: We have to show two things: The restrictions of the laws of composition need to map to U, and the vector space axioms have to hold for U.

That the laws of composition restrict to proper maps $+|_{U\times U} \colon U \times U \to U$ and $\cdot|_{\mathbb{R}\times U} \colon \mathbb{R} \times U \to U$ are precisely the requirements for U to be a vector subspace.

The axioms (A1), (A2), (A3) and (S1)–(S4) hold automatically, because they hold in the bigger space V. It remains to show that additive inverses of elements of U are in fact found in U.

Let u be an element of U. Because u then in particular is an element of V, there is a unique additive inverse $-u$ such that $u + (-u) = \mathbf{0}_V = (-u) + u$. By Proposition II.2.4(iv) we know that indeed $-u = (-1)u$ and because U is closed under scalar multiplication, $-u$ does belong to U. □

Proposition II.2.7 (Intersections of Vector Subspaces are Vector Subspaces): *Let I be a non-empty set and for each i in I, let U_i be a vector subspace of V. Then $U = \bigcap_{i \in I} U_i$ is a vector subspace of V, too.*

Proof: The proof consists of unravelling the definition of intersections over index sets and linear subspaces. As a reminder, u belongs to U if and only if u belongs to every U_i.

First, we need to show that $\mathbf{0}_V$ belongs to U. Because every U_i is a linear subspace of V by assumption, $\mathbf{0}_V$ lies in every U_i and hence in U.

Let now u and w be elements of U and let λ be a real number. We need to show closedness of U under addition and scalar multiplication. By definition of U, u and w lie in every U_i and because every U_i is closed under addition and scalar multiplication, $u + w$ and λu belong to every U_i. Again, by definition of the intersection, $u + w$ and λu lie in U. □

Proposition II.2.8 (Sums of Vector Spaces): *Let U_1 and U_2 be linear subspaces of V. Then, the subset $U_1 + U_2 = \{x + y \mid x \in U_1, y \in U_2\}$ of V is a vector subspace.*

Proof: As to the zero vector: Because U_1 and U_2 are linear subspaces of V, $\mathbf{0}_V$ lies in U_1 and U_2. Due to $\mathbf{0}_V = \mathbf{0}_V + \mathbf{0}_V$, the zero vector belongs to $U_1 + U_2$.

As to closedness under addition: Let u and w be two elements of $U_1 + U_2$. Per Definition of $U_1 + U_2$ there are elements x_1, x_2 of U_1 and y_1, y_2 of U_2, such that $u = x_1 + y_1$ and $w = x_2 + y_2$. Now

$$u + w = (x_1 + y_1) + (x_2 + y_2) = (x_1 + x_2) + (y_1 + y_2)$$

and because U_1, U_2 are linear subspaces of V, we have $x_1 + x_2$ belonging to U_1 as well as $y_1 + y_2$ belonging to U_2. Hence $u + w$ is of the required form and lies in $U_1 + U_2$. Closedness under scalar multiplication is shown completely analogously. □

Definition II.2.9 (Sums of Linear Subspaces): Let n be a natural number and let $U_1, \ldots, U_n$ be linear subspaces of V. Then

$$U_1 + \cdots + U_n = \{x_1 + \cdots + x_n \mid x_1 \in U_1, \ldots, x_n \in U_n\}$$

is called *sum of* $U_1, \ldots, U_n$.

From Proposition II.2.8 it follows that $U_1 + \cdots + U_n$ is again a linear subspace of V.

3. Matrices

Let a and b be integers such that $a \leq b$ and let f be a function defined on the integers mapping to some set 'that has an addition, a zero, multiplication and a one'. Then we write

$$\sum_{i=a}^{b} f(i) = f(a) + f(a+1) + \cdots + f(b), \quad \prod_{i=a}^{b} f(i) = f(a) \cdot f(a+1) \cdots f(b)$$

If $a > b$, we define $\sum_{i=a}^{b} f(i)$ to have the value zero and $\prod_{i=a}^{b} f(i)$ to have the value one.

For this section, let p, q, m and n be natural numbers.

Example: The following 'things' are matrices:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \mathbb{R}^{2\times 3}, \quad \begin{pmatrix} 0 & -3.75 \\ 1.01 & 2.79 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \quad \begin{pmatrix} 1 & 2 \\ 7 & 42 \\ \pi & 0 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

Definition II.3.1 (Matrix): A map $A\colon \{1, \ldots, p\} \times \{1, \ldots, q\} \to \mathbb{R}$ is called a *real $p \times q$ matrix*. The number p is called *number of rows of A* and q is called the *number of columns of A*. As with sequences, A is identified with the family of its images, and one writes $a_{i,j}$ for $A((i,j))$. Suggestively, the matrix A is written as

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,q} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,q} \\ \vdots & \vdots & \cdots & \vdots \\ a_{p,1} & a_{p,2} & \cdots & a_{p,q} \end{pmatrix}.$$

If $p = q$, A is said to be *quadratic*.

The set $\mathbb{R}^{p\times q} = \{A \mid A \text{ is real } p \times q \text{ matrix}\}$ is called *set of real $p \times q$ matrices*. Alternative notations are $\mathrm{Mat}(p \times q, \mathbb{R})$ or $\mathrm{Mat}_{\mathbb{R}}(p, q)$. We write $\mathbb{R}^p = \mathbb{R}^{p\times 1}$.

Remark II.3.2: In Definition II.3.1, $p = 0$ or $q = 0$ could be allowed. We will for the most part stick to natural numbers.

Example II.3.3: (i) The matrix A in $\mathbb{R}^{p\times q}$ with entries $a_{i,j} = A((i,j)) = 0$ for $1 \leq i \leq p$, $1 \leq j \leq q$ is called *zero matrix* and is notated $\mathbf{0}_{p\times q}$ or, briefly, $\mathbf{0}$.

(ii) Let A be the square $p \times p$ matrix with entries

$$a_{i,j} = A((i,j)) = \delta_{i,j} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{else.} \end{cases}$$

The symbol $\delta_{i,j}$ as defined above is called *Kronecker delta*. The matrix A is called *unit matrix* or *identity matrix* and is notated I_p, E_p or $\mathbb{1}_p$.

Definition II.3.4 (Operations for Matrices): Let A, B be real $p \times q$ matrices and let λ be a real number.

(i) The matrix $C = A + B$ with entries $C((i,j)) = A((i,j)) + B((i,j))$ is called *sum of A and B*.

(ii) The matrix $D = \lambda A$ with entries $D((i,j)) = \lambda A((i,j))$ is called *λ-fold multiple of A*.

In the following, to improve readability, we write $A(i,j)$ instead of $A((i,j))$.

Remark II.3.5: The laws of composition defined in Definition II.3.4 coincide with those from Example II.2.3(iii) and together with the corresponding zero matrix, they turn $\mathbb{R}^{p\times q}$ into a real vector space.

Definition II.3.6 (Matrix Multiplication): Let A be a real $p \times q$ matrix and let B be a real $q \times r$ matrix. Then, the real $p \times r$ matrix C with entries

$$C(i,k) = \sum_{j=1}^{q} A(i,j)B(j,k)$$

is called *matrix product of A and B*. We write $C = A \cdot B$.

Example II.3.7:

$$\begin{pmatrix} 1 & 2 & 3 \\ -3 & -2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1-2+6 & 0+0+3 \\ -3+2-2 & 0+0-1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ -3 & -1 \end{pmatrix}$$

Definition II.3.8 (Transpose): Let A be a real $n \times m$ matrix. The real $m \times n$ matrix with entries $B(i,j) = A(j,i)$, $1 \le i \le n$, $1 \le j \le m$, is called the *transpose of A*. This matrix is often written A^t or $A^\top$. Occasionally, one also finds ${}^tA$.

Example II.3.9:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \qquad A^t = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Proposition II.3.10: *We have the following rules for the laws of composition for matrices, where numbers of rows and colums are such that the rules makes sense:*

(i) $A(BC) = (AB)C$.

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

(iii) $(A + B)C = AC + BC$.

(iv) *For any real number* r, $A(rB) = (rA)B = r(AB)$.

(v) $(A + B)^t = A^t + B^t$, $(AB)^t = B^t A^t$ *and* $(A^t)^t = A$.

(vi) *For a suitably sized unit matrix* I, $IA = A$, $BI = B$.

Rule (i) shows that matrix multiplication is associative, the rules (ii) and (iii) are distributive properties of matrix multiplication over matrix addition. Note that we need two of those, because multiplication of matrices is not commutative. Also take note that transposing a product reverses the order of factors.

Example II.3.11 (Matrix Multiplication not commutative): Clearly, we don't have a chance to evaluate two products AB and BA, if A and B are not square. Here, commutativity can't even be hoped for. But even if A and B are square matrices the products most often don't coincide:

$$\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 0 & -1 \end{pmatrix} \neq \begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}.$$

Proof (of Proposition II.3.10): As an example, we show the first property. The other properties are shown similarly. Let thus A in $\mathbb{R}^{p\times q}$, B in $\mathbb{R}^{q\times m}$ and C in $\mathbb{R}^{m\times n}$. Then $(AB)C$ and $A(BC)$ are real $p \times n$ matrices.

We defined matrices as maps, hence for equality we have to check that the maps $A(BC)$ and $(AB)C$ are the same, i.e. we have to check for any (a, b) in $\{1, \dots, p\} \times \{1, \dots, n\}$ that $(A(BC))(a, b) = ((AB)C)(a, b)$, which is to say that the matrices have to agree entry-wise. For such a tuple (a, b), we find

$$\begin{aligned}
(A(BC))(a,b) &= \sum_{x=1}^{q} A(a,x)(BC)(x,b) \\
&= \sum_{x=1}^{q} A(a,x)\Big(\sum_{y=1}^{m} B(x,y)C(y,b)\Big) \\
&= \sum_{x=1}^{q}\sum_{y=1}^{m} A(a,x)B(x,y)C(y,b) \\
&= \sum_{y=1}^{m}\sum_{x=1}^{q} \Big(A(a,x)B(x,y)\Big)C(y,b) \\
&= \sum_{y=1}^{m}(AB)(a,y)C(y,b) = ((AB)C)(a,b). \qquad \square
\end{aligned}$$

4. Invertible Matrices

For this section, let p, q, n, m be natural numbers.

Definition II.4.1 (Invertible Matrices): Let A be a real $n \times n$ matrix. If there is a real $n \times n$ matrix B such that $AB = I_n = BA$, then A is called *invertible* or *regular*. We denote the set of invertible real $n \times n$ matrices by

$$\mathrm{GL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is invertible}\}.$$

The 'GL' abbreviates 'general linear group'.

Definition II.4.2 (Inverse Matrix): Let A be a regular real $n \times n$ matrix. Then there is one and only one real $n \times n$ matrix B such that $AB = I_n = BA$. This matrix is called *inverse matrix of* A or *inverse of* A and denoted A^{-1}.

For uniqueness of the inverse, assume B and B' are real $n \times n$ matrices such that $AB = BA = I_n = AB' = B'A$. Then $B = BI_n = BAB' = I_nB' = B'$.

Example II.4.3 (Special Invertible Matrices): Let $\alpha, \beta, a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real numbers.

(i) Let $A = \left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right)$. We can check that $B = \left(\begin{smallmatrix} 1 & -\alpha \\ 0 & 1 \end{smallmatrix}\right)$ is the inverse of A:

$$AB = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}.$$

Thus A in invertible with $A^{-1} = B$. Multiplying a matrix with A from the left has the following effect:

$$\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a_1 + \alpha b_1 & a_2 + \alpha b_2 & \cdots & a_n + \alpha b_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix}.$$

This means: Multiplying a matrix by A from the left adds the α-fold multiple of the second row to the first row. Multiplying a matrix by A from the right adds the α-fold multiple of the first column to the second column:

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix} \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_1 & \alpha a_1 + b_1 \\ a_2 & \alpha a_2 + b_2 \\ \vdots & \vdots \\ a_n & \alpha a_n + b_n \end{pmatrix}.$$

(ii) For $V = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ we have $VV = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, i.e. V is invertible and self-inverse. Such matrices are called involutions. Multiplying a matrix by V from the left interchanges rows,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & \cdots & b_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix},$$

and multiplying by V from the right exchanges columns,

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b_1 & a_1 \\ b_2 & a_2 \\ \vdots & \vdots \\ b_n & a_n \end{pmatrix}.$$

(iii) Let α and β both be non-zero. Then $D = \left(\begin{smallmatrix} \alpha & 0 \\ 0 & \beta \end{smallmatrix}\right)$ is invertible with inverse $E = \left(\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & \beta^{-1} \end{smallmatrix}\right)$. Multiplying another matrix by D from the left yields

$$\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} \alpha a_1 & \alpha a_2 & \cdots & \alpha a_n \\ \beta b_1 & \beta b_2 & \cdots & \beta b_n \end{pmatrix},$$

i.e. the rows are scaled by α respectively β, and multiplying by D from the right scales columns:

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} = \begin{pmatrix} \alpha a_1 & \beta b_1 \\ \alpha a_2 & \beta b_2 \\ \vdots & \vdots \\ \alpha a_n & \beta b_n \end{pmatrix}.$$

Definition II.4.4 (Matrix Unit): Let $1 \leq i \leq m$ and $1 \leq j \leq n$ be natural numbers. Then, by

$$E_{i,j} \colon \{1, \ldots, m\} \times \{1, \ldots, n\} \longrightarrow \mathbb{R}, \qquad (k, \ell) \longmapsto \begin{cases} 1, & \text{if } k = i \text{ and } \ell = j, \\ 0, & \text{else} \end{cases}$$

we denote the matrix with the only non-zero entry 1 at position (i, j). The matrices $E_{i,j}$ are called *matrix units*.

Note that matrix units are not units in the sense of ring theory, they are not invertible.

Example II.4.5 (Some Matrix Units): The matrices

$$E_{1,2} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 4},$$

$$E_{3,1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 2}, \qquad E_{1,2} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

are matrix units.

Remark II.4.6 (Matrices as Sums of Matrix Units): Let $A = (a_{i,j})$ be a real $m \times n$ matrix. Then A may be written as

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j} E_{i,j}.$$

Proposition II.4.7 (Rules for Matrix Units): *Let $E_{i,j}$ be matrix unit in $\mathbb{R}^{q\times m}$.*

(i) *If $E_{k,\ell}$ is another matrix unit, then*

$$E_{i,j}E_{k,\ell} = \delta_{k,j}E_{i,\ell} = \begin{cases} E_{i,\ell}, & \text{if } k = j, \\ \mathbf{0}_{p\times q}, & \text{else.} \end{cases}$$

(ii) *For a matrix M in $\mathbb{R}^{m\times n}$ it holds*

$$E_{i,j}M = \sum_{b=1}^{n} M(j,b)E_{i,b} \in \mathbb{R}^{q\times n}.$$

This means: The matrix $E_{i,j}M$ is the matrix whose i-th row is the j-th row of M and all other entries are zero.

(iii) *For a matrix M in $\mathbb{R}^{p\times q}$ it holds*

$$ME_{i,j} = \sum_{a=1}^{p} M(a,i)E_{a,j} \in \mathbb{R}^{p\times m}.$$

This means: The matrix $ME_{i,j}$ is the matrix whose j-th column is the i-th column of M and all other entries are zero.

Proof: (i) Let A be the product $E_{i,j}E_{k,\ell}$, which lives in $\mathbb{R}^{q\times n}$. For the (a,b)-th entry of A we find

$$A(a,b) = \sum_{x=1}^{m} E_{i,j}(a,x)E_{k,\ell}(x,b).$$

This entry is zero, unless $i = a$, $j = x = k$ and $\ell = b$. In that case, the above entry is one. Thus $A(a,b) = 1$ if and only if $(a,b) = (i,\ell)$ and $k = j$, otherwise it is zero.

(ii) Let M be a real $m \times n$ matrix. Then

$$E_{i,j}M = E_{i,j}\Big(\sum_{a=1}^{m}\sum_{b=1}^{n} M(a,b)E_{a,b}\Big) = \sum_{a=1}^{m}\sum_{b=1}^{n} M(a,b)E_{i,j}E_{a,b},$$

where we used Proposition II.3.10(ii) for the last equality. Because we have $E_{i,j}E_{a,b} = 0$ if and only if $j \neq a$, we obtain the claimed equality.

(iii) Absolutely analogous to (ii). □

Definition II.4.8 (Elementary Matrices): Let $a, \alpha_1, \ldots, \alpha_n$ be real numbers and let $1 \leq i, j \leq n$ be natural numbers with $i \neq j$. Then we define the following three types of real $n \times n$ matrices:

(i) *Addition matrices* $A_{i,j}^{\alpha} = I_n + \alpha E_{i,j}$. All entries on the diagonal of A are 1, the entry at position (i,j) is α, all other entries are zero.

(ii) *Permutation matrices* $V_{i,j} = I_n - E_{i,i} - E_{j,j} + E_{i,j} + E_{j,i}$. The matrix $V_{i,j}$ results from the unit matrix by setting to zero the diagonal entries at positions (i,i) and (j,j) and setting to one the values at positions (i,j) and (j,i).

(iii) *Diagonal matrices* $\mathrm{diag}(\alpha_1, \ldots, \alpha_n) = \sum_{i=1}^{n} \alpha_i E_{i,i}$. The entries on the diagonal are $\alpha_1, \ldots, \alpha_n$, all other entries are zero.

Proposition II.4.9 (Regularity of Elementary Matrices): *The matrices from Definition II.4.8 are invertible, i.e. lie in* $\mathrm{GL}_n(\mathbb{R})$.

Proof: It is easy to guess the respective inverses and to verify that they indeed are inverses. For example, for the addition matrix $A_{i,j}^{\alpha}$, we find

$$A_{i,j}^{\alpha}A_{i,j}^{-\alpha} = (I_n + \alpha E_{i,j})(I_n - \alpha E_{i,j}) = I_n - \alpha E_{i,j} + \alpha E_{i,j} - \alpha^2 E_{i,j}E_{j,i}.$$

By assumption we have $i \neq j$, such that Proposition II.4.7(i) shows that $E_{i,j}E_{j,i} = \mathbf{0}$. This means that $A_{i,j}^{\alpha}A_{i,j}^{-\alpha} = I_n$ and thus $A_{i,j}^{-\alpha}$ is the inverse of $A_{i,j}^{\alpha}$.

For the permutation matrix $V_{i,j}$, one can show $V_{j,i}$ to be the inverse, and for the diagonal matrix $\mathrm{diag}(a_1, \ldots, a_n)$, provided $a_1, \ldots, a_n$ are all different from zero, the inverse is $\mathrm{diag}(a_1^{-1}, \ldots, a_n^{-1})$. □

Proposition II.4.10 (Elementary Matrices and Row Transformations): *The matrices from Definition II.4.8 have, via multiplication from the left, the following effect on an $n \times m$ matrix M:*

(i) *$A_{i,j}^{\alpha} M$ emerges from M via addition of the α-fold of the j-th row of M to the i-th row of M.*

(ii) *$V_{i,j} M$ emerges from M via exchange of i-th and j-th column.*

(iii) *$\mathrm{diag}(a_1, \ldots, a_n) M$ emerges from M via scaling of the k-th row by a_k, $1 \leq k \leq n$.*

Proof: These assertions are easily shown using the rules from Proposition II.4.7.

(i) Writing out the product, we find

$$\begin{aligned} A_{i,j}^{\alpha} M &= (I_n + \alpha E_{i,j}) M \\ &= M + \alpha E_{i,j} M = M + \alpha \Big(\sum_{b=1}^{m} M(j,b) E_{i,b} \Big) = M + \sum_{b=1}^{m} \alpha M(j,b) E_{i,b} \end{aligned}$$

as claimed.

(ii) Similarly,

$$\begin{aligned} V_{i,j} M &= (I_n - E_{i,i} - E_{j,j} + E_{i,j} + E_{j,i}) M \\ &= M - \Big(\sum_{b=1}^{n} M(i,b) E_{i,b} \Big) - \Big(\sum_{b=1}^{n} M(j,b) E_{j,b} \Big) + \Big(\sum_{b=1}^{n} M(j,b) E_{i,b} \Big) \\ &\quad + \Big(\sum_{b=1}^{n} M(i,b) E_{i,b} \Big) \\ &= M - \Big(\sum_{b=1}^{n} \big(M(j,b) - M(i,b) \big) E_{i,b} \Big) + \Big(\sum_{b=1}^{n} \big(M(i,b) - M(j,b) \big) E_{j,b} \Big), \end{aligned}$$

so in the i-th row we have the entries $M(j,b)$ and in the j-th row we have the entries $M(i,b)$.

(iii) Using $E_{i,i} E_{a,b} = \mathbf{0}$ whenever $i \neq a$, we find

$$\mathrm{diag}(a_1, \ldots, a_n) M = \Big(\sum_{i=1}^{n} a_i E_{i,i} \Big) \Big(\sum_{a=1}^{n} \sum_{b=1}^{m} M(a,b) E_{a,b} \Big) = \sum_{i=1}^{n} \sum_{b=1}^{m} \alpha_i M(i,b) E_{i,b},$$

which we wanted to show. □

Proposition II.4.11 (Elementary Matrices and Column Transformations): *The matrices from Definition II.4.8 have, via multiplication from the right, the following effect on an $m \times n$ matrix M:*

(i) *$MA^{\alpha}_{i,j}$ emerges from M via addition of the α-fold of the i-th column to the j-th column of M.*

(ii) *$MV_{i,j}$ emerges from M via exchange of i-th and j-th column.*

(iii) *$M\,\mathrm{diag}(a_1, \ldots, a_n)$ emerges from M by scaling of the k-th column by a_k, $1 \leq k \leq n$.*

Proof: Using $(MA)^t = A^t M^t$ allows us to retreat to the proof of Proposition II.4.10. By way of example, we show (i); the other arguments are analogous. We have $(MA^{\alpha}_{i,j})^t = (A^{\alpha}_{i,j})^t M^t = A^{\alpha}_{j,i} M^t$ and $A^{\alpha}_{j,i} M^t$ emerged from M^t through addition of the α-fold of the i-th row to the j-th row. Hence $MA^{\alpha}_{i,j}$ emerges from M through addition of the α-fold of the i-th column to the j-th column, as claimed. □

Example II.4.12 (Inverses of Regular 2×2 Matrices): Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be a real matrix. Then A is invertible if and only if $ad - bc \neq 0$, and in this case

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Lemma II.4.13 (Multiplication of Block Matrices): *Let M_1 and M_2 be real $p \times q$ respectively $q \times m$ matrices of the shape*

$$M_1 = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad M_2 = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

with blocks A in $\mathbb{R}^{p_1 \times q_1}$, B in $\mathbb{R}^{p_1 \times q_2}$, C in $\mathbb{R}^{p_2 \times q_1}$, D in $\mathbb{R}^{p_2 \times q_2}$, E in $\mathbb{R}^{q_1 \times m_1}$, F in $\mathbb{R}^{q_1 \times m_2}$, G in $\mathbb{R}^{q_2 \times m_1}$ and H in $\mathbb{R}^{q_2 \times m_2}$ and natural numbers $p_1, p_2, q_1, q_2, m_1, m_2$ such that $p_1 + p_2 = p$, $q_1 + q_2 = q$ and $m_1 + m_2 = m$. Then

$$M_1 M_2 = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix}.$$

This assertion follows from the definition of matrix multiplication. If in doubt, try checking it yourself.

Proposition II.4.14: *Suppose that $p < n$ and let A be the real $n \times n$ matrix*

$$A = \begin{pmatrix} I_p & B \\ \mathbf{0} & D \end{pmatrix}$$

with blocks I_p in $\mathbb{R}^{p\times p}$, $\mathbf{0}$ in $\mathbb{R}^{(m-p)\times p}$, B in $\mathbb{R}^{p\times(n-p)}$ and D in $\mathbb{R}^{(n-p)\times(n-p)}$. The matrix A is invertible if and only if the block D is invertible, and in this case,

$$A^{-1} = \begin{pmatrix} I_p & -BD^{-1} \\ \mathbf{0} & D^{-1} \end{pmatrix}.$$

Proof: '$\Longleftarrow$': This is a mere calculation using Lemma II.4.13.

'$\Longrightarrow$': Suppose A were invertible, i.e. there were a matrix real $n \times n$ matrix M such that $AM = MA = I_n$. This matrix M could be written as block matrix $M = \left(\begin{smallmatrix} E & F \\ G & H \end{smallmatrix}\right)$ with blocks E in $\mathbb{R}^{p\times p}$, F in $\mathbb{R}^{p\times(n-p)}$, G in $\mathbb{R}^{(n-p)\times p}$ and H in $\mathbb{R}^{(n-p)\times(n-p)}$. From Lemma II.4.13 we read off that

$$I_n = MA = \begin{pmatrix} E & EB + FD \\ G & GB + HD \end{pmatrix},$$

such that $E = I_p$, $G = \mathbf{0}$, $I_{n-p} = GB + HD = HD$ needed to hold. Similarly,

$$I_n = AM = \begin{pmatrix} I_p & F + BH \\ \mathbf{0} & DH \end{pmatrix}$$

enforced $DH = I_{n-p}$, i.e. D were invertible with inverse H. □

Proposition II.4.15 (Inverse of Product): *Let A and B be real invertible $n \times n$ matrices. Then AB has the inverse $(AB)^{-1} = B^{-1}A^{-1}$; in particular AB is also invertible.*

Proof: We just need to check. $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$ and $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I$. □

5. Systems of Linear Equations

For this section, let n and m be natural numbers.

Definition II.5.1: Let $a_{i,j}$, $1 \le i \le n$ and $1 \le j \le m$, be real numbers. A set of equations of the form

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,m}x_m &= b_1, \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,m}x_m &= b_2, \\ \vdots \qquad\quad \vdots \qquad\quad \cdots \qquad\quad \vdots \qquad\quad &\vdots \\ a_{n,1}x_1 + a_{1,2}x_2 + \cdots + a_{n,m}x_m &= b_n \end{aligned}$$

is called a *real system of linear equations in n equations and m variables*. Instead of 'variables', 'unknowns' is also used occasionally. The numbers $a_{i,j}$ are called *coefficients* and the x_i are called *variables* respectively *unknowns*. Schematically, the system of linear equations is written as

$$\left(\begin{array}{ccc|c} a_{1,1} & \cdots & a_{1,m} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & \cdots & a_{n,m} & b_n \end{array}\right).$$

The matrix $A = (a_{i,j})$ is called *coefficient matrix* and $(A|b)$ is called *augmented matrix* to the system of linear equations. The set

$$\mathbb{S} = \mathbb{S}(A, b) = \{x = (x_1, \ldots, x_m)^t \in \mathbb{R}^m \mid x_1, \ldots, x_m \text{ satisfy the equations}\}$$

is called *solution set* and $\mathbb{S}^h = \mathbb{S}(A, \mathbf{0})$ is called *solution set to the homogeneous system of linear equations*. For brevity, we will call a system of linear equations a *linear system* and a homogeneous system of linear equations a *homogeneous linear system*.

In the following, let A be a real $n \times m$ matrix and let b be a vector from $\mathbb{R}^n$. We consider the system of linear equations with schematic description $(A|b)$.

Remark II.5.2: It holds $Ax = b$ if and only if x belongs to $\mathbb{S}(A|b)$.

Proposition II.5.3 (Structure of the Solution Set):

(i) *The solution set* $\mathbb{S}^h = \mathbb{S}(A, \mathbf{0})$ *is a linear subspace of* $\mathbb{R}^n$.

(ii) *If* $x^{(s)}$ *is in* $\mathbb{S}(A, b)$, *then* $\mathbb{S}(A, b) = x^{(s)} + \mathbb{S}^{(h)} = \{x^{(s)} + v \mid v \in \mathbb{S}^{(h)}\}$.

An element of $\mathbb{S}(A, b)$ is sometimes called a 'special solution' to indicate that a solution for the system of linear equations has been chosen.

Proof: (i) Let x and y be solutions of the homogeneous linear system, i.e. $Ax = \mathbf{0}$ and $Ay = \mathbf{0}$. Then we have $A(x + y) = Ax + Ay = \mathbf{0} + \mathbf{0} = \mathbf{0}$ and thus $x + y$ lies in $\mathbb{S}^{(h)}$.

If λ is a real number and x is a solution of the homogeneous linear system, then $A(\lambda x) = \lambda(Ax) = \lambda\mathbf{0} = \mathbf{0}$, which shows that λx lies in $\mathbb{S}^{(h)}$.

(ii) '$\subseteq$': For x in $\mathbb{S}(A, b)$ it holds $A(x - x^{(s)}) = Ax - Ax^{(s)} = b - b = \mathbf{0}$, which means that $v = x - x^{(s)}$ is a solution of the homogeneous linear system. Furthermore, $x = x^{(s)} + v$ is of the claimed form.

'$\supseteq$': Let v from $\mathbb{S}^{(h)}$ be given. Then $x = x^{(s)} + v$ belongs to $\mathbb{S}(A, b)$, because $Ax = A(x^{(s)} + v) = Ax^{(s)} + Av = b + \mathbf{0} = b$. □

Proposition II.5.4 (Solution Strategy for Linear Systems): *Let C be a real regular $n \times n$ matrix. A vector x in $\mathbb{R}^n$ satisfies $Ax = b$ if and only if $CAx = Cb$.*

Proof: As for '$\Longleftarrow$': If $CAx = Cb$, then $Ax = C^{-1}CAx = C^{-1}Cb = b$. '$\Longrightarrow$' is clear. □

Corollary II.5.5 (Elementary Row Transformations): *If in Proposition II.5.4 one chooses for C one of the matrices from Definition II.4.8, that is*

(i) $C = A^{\alpha}_{i,j}$, *α some non-zero real number,*

(ii) $C = V_{i,j}$, *or*

(iii) $C = \mathrm{diag}(a_1, \ldots, a_n)$ *for non-zero real numbers $a_1, \ldots, a_n$,*

then one obtains the elementary row transformations

(i) *addition of the α-fold of equation j to equation i,*

(ii) *exchange of equations i and j,*

(iii) *scaling of equation i by a_i for $1 \leq i \leq n$.*

The elementary row transformations do not change the solution set of a linear system.

Definition II.5.6 (Standard-Basis Vectors): Let i be a natural number with $1 \leq i \leq n$. Identifying $\mathbb{R}^n \cong \mathbb{R}^{n\times 1}$, we call the matrix $e_i \colon \{1, \ldots, n\} \times \{1\} \to \mathbb{R}$ with $e_i(j, 1) = \delta_{i,j}$ the *i-th standard-basis vector*. Under the identification, e_i corresponds to $(\delta_{i,j})^t_{1\leq j\leq n}$.

Note that, much like with matrix units, each vector $x = (x_1, \ldots, x_n)^t$ in $\mathbb{R}^n$ can be expressed as a sum of standard-basis vectors, namely $x = \sum_{i=1}^n x_i e_i$.

If $A = (a_{i,j})$ is a real $m \times n$-matrix, then $Ae_j = \sum_{k=1}^n a_{k,j} e_k$ is the j-th column of A. (This is a direct computation using the definition of standard-basis vectors.)

Definition II.5.7 (Row Echelon Form and Rank): Let T be a real $n \times m$ matrix. If there are a non-negative integer r and natural numbers $s_1, \ldots, s_r$ with $1 \leq s_1 < \cdots < s_r \leq m$ such that

(i) For $1 \leq i \leq r$, $t_{i,s_i} = 0$, for $k \neq i$, $t_{k,s_i} = 0$ and for $k < s_i$, $t_{i,k} = 0$;

(ii) For $i \geq r + 1$ and $1 \leq j \leq m$, $t_{i,j} = 0$;

then T is in *row echelon form*. The number r is called *rank of* T, and the indices $s_1, \ldots, s_r$ are called *column indices*.

A matrix T is in row echelon form, if the s_i-th column of T is i-th standard-basis vector e_i of $\mathbb{R}^n$, if left of the respective entry 1 at position (i, j) there are no non-zero entries in row i, and if all rows from the $r + 1^{\text{st}}$ onwards are rows of zeros.

Note that in literature there are whole zoo of different definitions of row echelon forms. In the end, it is not decisive for the theory, which definition is chosen. There are good reasons for most conventions. It is just important to stick to one.

In the following, let $T = (t_{i,j})$ be a real $n \times m$ matrix in row echelon form of rank r and let $s_1, \ldots, s_r$ be the column indices of T.

Lemma II.5.8 (Special Solution for Row Echelon Form): *Let b be an element of $\mathbb{R}^n$. The linear system with schematic description $(T|b)$ has a solution if and only if $b_{r+1} = \cdots = b_n = 0$. In this case, the vector $x^{(s)} = \sum_{i=1}^{r} b_i e_{s_i}$ from $\mathbb{R}^n$ is a solution.*

Proof: '$\Longrightarrow$': An equation of the form $0 = b_i$ for $b_i \neq 0$ has no solution.

'$\Longleftarrow$': Using $b_{r+1} = \cdots = b_n = 0$ and thus $b_i e_i = 0$ for $i > r$, we find

$$Tx^{(s)} = T\Big(\sum_{i=1}^{r} b_i e_{s_i}\Big) = \sum_{i=1}^{r} b_i T e_{s_i} = \sum_{i=1}^{r} b_i e_i = \sum_{i=1}^{n} b_i e_i = b$$

as claimed. □

Example II.5.9 (A Linear System in Row Echelon Form): We consider the linear system

$$\begin{aligned} x_2 + 0x_3 + \alpha x_4 + 0x_5 + \beta x_6 &= b_1 \\ x_3 + \gamma x_4 + 0x_5 + \delta x_6 &= b_2 \\ x_5 + \varepsilon x_6 &= b_3 \\ 0 &= b_4 \end{aligned}$$

This linear system has rank 3 and column indices $s_1 = 2$, $s_2 = 3$ and $s_3 = 5$. By Lemma II.5.8 this linear system has a solution if and only if $b_4 = 0$, in which case

$$x^{(s)} = \sum_{i=1}^{r} b_i e_{s_i} = b_1 e_2 + b_2 e_3 + b_3 e_5 = (0, b_1, b_2, 0, b_3)^t$$

is one such solution. Note that by naming conventions, the corresponding augmented matrix has a leading zero column.

Example II.5.10 (Homogeneous Solutions): For Example II.5.9, we obtain the following solutions to the corresponding homogeneous linear system:

$$\begin{aligned} x_2 &= -\alpha x_4 - \beta x_6 \\ x_3 &= -\gamma x_4 - \delta x_6 \\ x_5 &= \qquad - \varepsilon x_6 \end{aligned}$$

'Reasonable' values for x_1, x_4 and x_6 yield the following special solutions to the homogeneous linear system:

(i) Choosing $x_1 = 1$, $x_4 = 0$, $x_6 = 0$, we get $x_2 = x_3 = x_5 = 0$ and $F^{(1)} = e_1$ is a solution of $(T|\mathbf{0})$.

(ii) Choosing $x_1 = 0$, $x_4 = 1$, $x_6 = 0$, the equations give $x_2 = -\alpha$, $x_3 = -\gamma$ and $x_5 = 0$. The vector $F^{(4)} = e_4 - \alpha e_2 - \gamma e_3 + 0e_5$ is a solution of $(T|\mathbf{0})$.

(iii) Choosing $x_1 = 0$, $x_4 = 0$ and $x_6 = 1$, $x_2 = -\beta$, $x_3 = -\delta$ and $x_5 = -\varepsilon$. Then $F^{(6)} = e_6 - \beta e_2 - \delta e_3 - \varepsilon e_5$ solves the system $(T|\mathbf{0})$.

Lemma II.5.11 (Solution Set to Homogeneous Linear System): *Let J be the set $\{1, \ldots, m\} - \{s_1, \ldots, s_r\}$.*

(i) *For $j \in J$, $F^{(j)} = e_j - \sum_{i=1}^{r} t_{i,j} e_{s_i}$ is a solution to the homogeneous linear system $Tx = \mathbf{0}$. The vectors $F^{(j)}$ are called* fundamental solutions.

(ii) *For the homogeneous solution set $\mathbb{S}^{(h)} = \mathbb{S}(T, \mathbf{0})$, it holds $\mathbb{S}^{(h)} = \{\sum_{j \in J} \lambda_j F^{(j)} \mid \lambda_j \in \mathbb{R}\}$. Moreover, for any v in $\mathbb{S}^{(h)}$, the representation $v = \sum_{j \in J} \lambda_j F^{(j)}$ is unique.*

Proof: (i) Because of $t_{i,j} = 0$ for $i \geq r + 1$, we have

$$TF^{(j)} = Te_j - \sum_{i=1}^{r} t_{i,j} Te_{s_i} = Te_j - \sum_{i=1}^{r} t_{i,j} e_i = Te_j - \sum_{i=1}^{n} t_{i,j} e_i = \mathbf{0}.$$

(ii) The inclusion '$\supseteq$' follows from the combination of (i) and the fact that $\mathbb{S}^{(h)}$ is a linear subspace of $\mathbb{R}^m$.

As for '$\subseteq$': Let $x = \sum_{i=1}^{m} x_i e_i$ be a solution to the homogeneous system and put $v = x - \sum_{j \in J} x_j F^{(j)}$. For this v we then also have $Tv = \mathbf{0}$, i.e. v belongs to $\mathbb{S}^{(h)}$. If we can now show that $v = 0$, then we have shown that each x can be uniquely written in the claimed form.

We already know that $v_j = 0$ for each j in J and we want to use that $Tv = \mathbf{0}$. The i-th entry of Tv is $\sum_{k=1}^{m} t_{i,k} v_k$. For $1 \leq i \leq r$, it holds $t_{i,s_i} = 1$, and it holds $t_{i,k} = 0$ if $k \notin J$ and $k \neq s_i$, thus $0 = \sum_{k=1}^{m} t_{i,k} v_k = v_{s_i}$ for each $1 \leq i \leq r$. This means that v has to be the zero vector and we are done. $\square$

Theorem 2: *Let T be a real $n \times m$ matrix in row echelon form of rank r with column indices $s_1, \ldots, s_r$ and let b be a vector from $\mathbb{R}^n$. Writing $x^{(s)} = \sum_{i=1}^{r} b_i e_{s_i}$, $J = \{1, \ldots, m\} - \{s_1, \ldots, s_r\}$ and $F^{(j)} = e_j - \sum_{i=1}^{r} t_{i,j} e_{s_i}$ as before, for the solution set $\mathbb{S} = \mathbb{S}(T, b)$ it holds*

$$\mathbb{S} = x^{(s)} + \Big\{ \sum_{j \in J} \lambda_j F^{(j)} : \lambda_j \in \mathbb{R} \Big\}.$$

Proof: Follows from Proposition II.5.3, Lemma II.5.8 and Lemma II.5.11. □

Now that we have conquered the solution theory for matrices in row echelon form, we will try to develop a general solution theory by reducing back to row echelon form.

Lemma II.5.12 (Gauss Algorithm): *Let A be a real $n \times m$ matrix. Then, there is a real regular $n \times n$ matrix C such that CA is in row echelon form.*

Proof: We show this assertion via induction on the number n of rows of A.

As base case, we consider $n = 1$, i.e. A consists of one row only. If A contains only zeros, we are done, since the zero matrix is in row echelon form and $C = (1)$ does the trick.

If not every entry of A is zero, we put $s_1 = \min\{j \in \{1, \ldots, m\} \mid a_{i,j} \neq 0\}$ (i.e. we take as column index the column with the first non-zero entry), choose C to be the matrix (a_{1,s_1}^{-1}) and then CA is in row echelon form.

For the induction step, we suppose the assertion holds for the natural number $n - 1$. We want to show that it is true for n, too. If A is the zero matrix, we are done. If A is not the zero matrix, we again pick out the first non-zero column from the left, i.e. we put

$$s_1 = \min\{j \in \{1, \ldots, m\} \mid \text{There is } i \text{ such that } a_{i,j} \neq 0\}.$$

Furthermore, we put $i_0 = \min\{i \in \{1, \ldots, n\} \mid a_{i,s_i} \neq 0\}$, that is, we find the row index of the first non-zero entry in the ${s_1}^{\text{th}}$ column of A. This means A is of the form

$$A = \begin{pmatrix} 0 & \cdots & 0 & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & 0 & \vdots & & \vdots \\ \vdots & & \vdots & a_{i_0,s_1} & \vdots & & \vdots \\ \vdots & & \vdots & * & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & * & * & \cdots & * \end{pmatrix}.$$

Thus, $A_1 = V_{1,i_0}(\prod_{i\neq i_0} A_{i,i_0}(-a_{i,s_1})\operatorname{diag}(1,\dots,1,a_{i_0,s_1}^{-1},1,\dots,1)A$ is of the form

$$A_1 = \left(\begin{array}{cccc|c} 0 & \cdots & 0 & 1 & z \\ \hline 0 & \cdots & 0 & 0 & \\ \vdots & & \vdots & \vdots & \tilde{A} \\ 0 & \cdots & 0 & 0 & \end{array}\right).$$

By induction hypothesis there is a real regular $(n-1)\times(n-1)$ matrix $\tilde{C}$ such that $\tilde{C}\tilde{A}$ is in row echelon form $\tilde{T}$ of rank $\tilde{r}$ and with column indices $\tilde{s}_1,\dots,\tilde{s}_{\tilde{r}}$. By Proposition II.4.14, the real $n\times n$ matrix $\left(\begin{smallmatrix}1 & 0\\ 0 & \tilde{C}\end{smallmatrix}\right)$ is regular and by Lemma II.4.13,

$$\begin{pmatrix}1 & 0\\ 0 & \tilde{C}\end{pmatrix} A_1 = \left(\begin{array}{cccc|c} 0 & \cdots & 0 & 1 & z \\ \hline 0 & \cdots & 0 & 0 & \\ \vdots & & \vdots & \vdots & \tilde{T} \\ 0 & \cdots & 0 & 0 & \end{array}\right) = A_2.$$

It remains to arrange for $z_{\tilde{s}_j}$ to be zero, when $1 \leq i \leq \tilde{r}$. The matrix $T = \prod_{i=2}^{\tilde{r}+1} A_{1,i}(-z_{\tilde{s}_{i-1}+s_1})A_2$ now has row echelon form and the matrix

$$C = \prod_{i=2}^{\tilde{r}+1} A_{1,i}(-z_{\tilde{s}_{i-1}+s_1}) \begin{pmatrix}1 & 0\\ 0 & \tilde{C}\end{pmatrix} V_{1,i_0} \prod_{i\neq i_0} A_{i,i_0}(-a_{i,s_1})\operatorname{diag}(1,\dots,1,a_{i_0,s_1}^{-1},1,\dots,1)$$

does the trick. □

Remark II.5.13: Lemma II.5.12 states an algorithm for bringing a linear system into row echelon form.

Example II.5.14: In the following, we determine the solution set of a linear

system according to our algorithm:

$$\left(\begin{array}{ccccc|c} 0&0&0&0&0&0\\ 0&2&4&2&0&6\\ 0&3&6&-3&-6&-3\\ 0&0&0&1&1&2 \end{array}\right) \xrightarrow{\frac{1}{2}\cdot \mathrm{II}} \left(\begin{array}{ccccc|c} 0&0&0&0&0&0\\ 0&1&2&1&0&3\\ 0&3&6&-3&-6&-3\\ 0&0&0&1&1&2 \end{array}\right)$$

$$\xrightarrow{\mathrm{I}\leftrightarrow\mathrm{II}} \left(\begin{array}{ccccc|c} 0&1&2&1&0&3\\ 0&0&0&0&0&0\\ 0&3&6&-3&-6&-3\\ 0&0&0&1&1&2 \end{array}\right) \xrightarrow{\mathrm{III}-3\cdot\mathrm{I}} \left(\begin{array}{ccccc|c} 0&1&2&1&0&3\\ 0&0&0&0&0&0\\ 0&0&0&-6&-6&-12\\ 0&0&0&1&1&2 \end{array}\right)$$

$$\xrightarrow{-\frac{1}{6}\cdot\mathrm{III}} \left(\begin{array}{ccccc|c} 0&1&2&1&0&3\\ 0&0&0&0&0&0\\ 0&0&0&1&1&2\\ 0&0&0&1&1&2 \end{array}\right) \xrightarrow{\mathrm{II}\leftrightarrow\mathrm{III}} \left(\begin{array}{ccccc|c} 0&1&2&1&0&3\\ 0&0&0&1&1&2\\ 0&0&0&0&0&0\\ 0&0&0&1&1&2 \end{array}\right)$$

$$\xrightarrow{\mathrm{IV}-\mathrm{II}} \left(\begin{array}{ccccc|c} 0&1&2&1&0&3\\ 0&0&0&1&1&2\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0 \end{array}\right) \xrightarrow{\mathrm{I}-\mathrm{II}} \left(\begin{array}{ccccc|c} 0&1&2&0&-1&1\\ 0&0&0&1&1&2\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0 \end{array}\right)$$

By the assertions before, the solution set is

$$\mathbb{L} = \begin{pmatrix} 0\\1\\0\\2\\0 \end{pmatrix} + \left\{ \lambda_1 \begin{pmatrix} 1\\0\\0\\0\\0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0\\-2\\1\\0\\0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0\\1\\0\\-1\\1 \end{pmatrix} : \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} \right\}.$$

Lemma II.5.15 (Multiplication with Regular Matrices): *Let A be a real regular $n \times n$ matrix and let v be any element of $\mathbb{R}^n$. If v is non-zero, then so is Av.*

Proof: If $Av = \mathbf{0}$, then $v = A^{-1}Av = A^{-1}\mathbf{0} = \mathbf{0}$. □

Lemma II.5.16 (Uniqueness of Row Echelon Form): *Let T and T' be two real $n \times m$ matrices in row echelon form. If there is some real regular $n \times n$ matrix D with $T' = DT$, then $T' = T$.*

Proof: We show the assertion using induction.

For the base case, let $n = 1$. Thus T and T' consist of one row each, i.e. belong to $\mathbb{R}^{1\times m}$, and there is some non-zero real number d such that $T' = dT$. In particular, T has no non-zero entries if and only if T' doesn't have any

non-zero entries. Otherwise, the first non-zero entry of T and T' has to be in the same position. Because T and T' are in row echelon form, this entry is 1 both times. Hence $d = 1$ is enforced and $T = T'$.

Suppose the assertion holds for the natural number n. Suppose further that r is the rank of T, that $s_1, \ldots, s_r$ are the column indices of T, that r' is the rank of T' and that $s'_1, \ldots, s'_r$ are the column indices of S'. The matrices T and T' thus are of the form

$$T = \begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ \vdots & & \vdots & 0 & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & * & \cdots & * \end{pmatrix}, \qquad T' = \begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ \vdots & & \vdots & 0 & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & * & \cdots & * \end{pmatrix}$$

In the following, we denote the i^{th} column of a matrix A by $A^{(i)}$, that is we write $T = (t^{(1)}, \ldots, t^{(m)})$ and $T' = (t'^{(1)}, \ldots, t'^{(m)})$. Then it holds:

(i) $t^{(j)} = \mathbf{0}$ for $1 \leq j \leq s_1 - 1$ and $t'^{(j)} = \mathbf{0}$ for $1 \leq j \leq s'_1 - 1$,

(ii) $t^{(s_k)} = e_k$ for $1 \leq k \leq r$ and $t'^{(s'_k)} = e_k$ for $1 \leq k \leq r'$,

(iii) ${t^{(j)}}_i = T(i,j) = 0$ for $i > r$ and ${t'^{(j)}}_i = T'(i,j) = 0$ for $i > r'$,

(iv) $t'^{(j)} = Dt^{(j)}$ for $1 \leq j \leq m$.

In particular, for $1 \leq j \leq s_1 - 1$ it holds $t'^{(j)} = Dt^{(j)} = D\mathbf{0} = \mathbf{0}$, and by Lemma II.5.15, $t'^{(s_1)} = Dt^{(s_1)} = De_1 \neq \mathbf{0}$, that is $s'_1 = s_1$ and so $e_1 = t'^{(s_1)} = Dt^{(s_1)} = De_1$. This means that D is of the form

$$D = \left(\begin{array}{c|c} 1 & z \\ \hline 0 & \\ \vdots & \hat{D} \\ 0 & \end{array}\right)$$

for some real $(n-1) \times (n-1)$ matrix $\hat{D}$ and some z in $\mathbb{R}^{1\times(n-1)}$. Because D is invertible, $\hat{D}$ is invertible by Proposition II.4.14. Furthermore, we may write T and T' as

$$T = \begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ \vdots & & \vdots & 0 & & & \\ \vdots & & \vdots & \vdots & & \hat{T} & \\ 0 & \cdots & 0 & 0 & & & \end{pmatrix}, \qquad T' = \begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ \vdots & & \vdots & 0 & & & \\ \vdots & & \vdots & \vdots & & \hat{L} & \\ 0 & \cdots & 0 & 0 & & & \end{pmatrix}.$$

Due to Lemma II.4.13, $\hat{T}' = \hat{D}\hat{T}$, and per induction hypothesis, $\hat{T} = \hat{T}'$. Thus the assertion follows. □

Theorem 3 (on the Gauss Normal Form): *For any real $n \times m$ matrix A there is one and only one real $n \times m$ matrix T in row echelon form, such that there is a real regular $n \times n$ matrix C with $T = CA$.*

Note that the row echelon form T of A is unique, but the regular matrix C is absolutely not!

Proof: The existence of such a matrix T follows from Lemma II.5.12.

As for uniqueness: Suppose T_1 and T_2 are real $n \times m$ matrices in row echelon form and suppose that C_1, C_2 are real regular $n \times n$ matrices, such that $T_1 = C_1A$ as well as $T_2 = C_2A$. Then $A = C^{-1}T_1$ and $T_2 = C_2C_1^{-1}T_1$. Because C_2C^{-1} is regular by Proposition II.4.15, Lemma II.5.16 yields that indeed $T_1 = T_2$. □

The matrix T from Theorem 3 is also called *Gauss normal form of A*.

Definition II.5.17 (Rank of Matrix): Let A be a real $n \times m$ matrix and let T be the corresponding Gauss normal form of A. The rank r of T is called *rank of A* and denoted $\operatorname{rank}(A)$.

Remark II.5.18: If A is a real $n \times m$ matrix and if D is a real regular $n \times n$ matrix, then the Gauss normal forms of A and DA are equal. In particular, $\operatorname{rank}(A) = \operatorname{rank}(DA)$.

Let C be a regular real $n \times n$ matrix such that $CA = T$ is in row echelon form. Then $CD^{-1}(DA) = T$, that is A and DA have the same row echelon form.

Conclusion II.5.19: Given a linear system $Ax = b$ consisting of a real $n \times m$ matrix A and a vector b from $\mathbb{R}^n$, our strategy is the following to determine the solution set:

(i) Determine some real regular $n \times n$ matrix C such that $CA = T$ is in row echelon form as described in Lemma II.5.12.

(ii) Determine the solution set $\mathbb{S}(Tx, Cb)$ using Theorem 2, which by Proposition II.5.4 then also is the solution set of $Ax = b$.

Example II.5.20: We consider the following linear system in row echelon form:

$$(T|b) = \left(\begin{array}{cccccc|c} 0 & 1 & 0 & \alpha & 0 & \beta & b_1 \\ 0 & 0 & 1 & \gamma & 0 & \delta & b_2 \\ 0 & 0 & 0 & 0 & 1 & \varepsilon & b_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_5 \end{array}\right).$$

We observe the following:

(i) $\operatorname{rank}(T|b) = 3$ if and only if $b_4 = b_5 = 0$; otherwise $\operatorname{rank}(T|b) = 4 > 3 = \operatorname{rank}(T)$. We have a solution if and only if $\operatorname{rank}(T|b) = 3$, i.e. if and only if $b_4 = b_5 = 0$.

(ii) If $b_4 = b_5 = 0$, we have a unique solution if and only if there are no fundamental solutions, that is if $m - r = 0$, where $m = 6$ is the number of columns of T.

Corollary II.5.21 (of Theorem 3): *Let A be a real $n \times m$ matrix, let b be a vector from $\mathbb{R}^n$ and let $Ax = b$ be the corresponding linear system.*

(i) *There is a solution to $Ax = b$ if and only if* $\operatorname{rank}(A) = \operatorname{rank}(A|b)$.

(ii) *If the linear system has a solution, there is a unique solution if and only if* $\operatorname{rank}(A) = m$.

(iii) *The linear system $Ax = c$ is solvable for any c in $\mathbb{R}^n$ if and only if* $\operatorname{rank}(A) = n$.

Proof: Choose a real regular $n \times n$ matrix C such that $CA = T$ is in row echelon form. In particular, $\operatorname{rank}(A) = \operatorname{rank}(T) = r$.

(i) The linear system $(A|b)$ is solvable if and only if $(CA|Cb)$ is solvable, i.e. if $(T|Cb)$ is solvable. By Theorem 2, $(T|Cb)$ is solvable if and only if $\operatorname{rank}(T|Cb) = \operatorname{rank}(T)$. Because of $\operatorname{rank}(A) = \operatorname{rank}(T)$ we have $\operatorname{rank}(T|Cb) = \operatorname{rank}(T)$ if and only if $\operatorname{rank}(A) = \operatorname{rank}(A|b)$, and we are done.

(ii) Suppose $(A|b)$ has a solution. Then so does $(T|Cb)$. There is a unique solution to $(A|b)$ if and only if the same goes for $(T|Cb)$, which is the case if and only if $m = r$.

(iii) The linear system $(T|c)$ is solvable for any c in $\mathbb{R}^n$ if and only if T doesn't have rows without non-zero entries, hence if $n = r$. □

Corollary II.5.22: *Let A be a real $n \times n$ matrix. Then the following are equivalent:*

(i) *A is regular,*

(ii) $\operatorname{rank}(A) = n$,

(iii) *There is a real regular $n \times n$ matrix S with $AS = I_n$.*

Proof: '(i) $\Longrightarrow$ (ii)': If A is regular, then there is some real regular $n \times n$ matrix B with $BA = I_n$, which is to say that I_n is the Gauss normal form of A. Thus $\operatorname{rank}(A) = \operatorname{rank}(I_n) = n$.

'(ii) $\Longrightarrow$ (i)': We want to use that I_n is the only row echelon form in $\mathbb{R}^{n\times n}$ of rank n. We thus have: If $\operatorname{rank}(A) = n$, then there is a real regular $n \times n$ matrix C such that $CA = I_n$, that is $A = C^{-1}CA = C^{-1}$. In particular, A is regular.

'(i) $\Longrightarrow$ (iii)' is clear. As to '(iii) $\Longrightarrow$ (i)': If $AS = I_n$, then, for any c in $\mathbb{R}^n$, $ASc = c$. This means that the linear system $Ax = c$ has a solution for any c in $\mathbb{R}^n$, and thus $\operatorname{rank}(A) = n$ follows from Corollary II.5.21. □

Remark II.5.23: The Gauss algorithm provides a method to determine invertibility of a real $n \times n$ matrix A.

Example II.5.24: Suppose we are given the real 3×3 matrix

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

We know that A is regular if and only if A has the Gauss normal form I_3. Taking a look at the proof of Corollary II.5.22 '(ii) $\Longrightarrow$ (i)' again, we see that determining the row echelon form of A and applying the respective elementary row transformations to I_3 yields the inverse, if there is one. We thus may as well try:

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -1 & 1 \\ 0 & 1 & 0 & 2 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right).$$

Definition II.5.25: Let A be a real $n \times m$ matrix. We denote the corresponding map $\mathbb{R}^m \to \mathbb{R}^n$, $x \mapsto Ax$ by ϕ_A.

The associated map ϕ_A has the following properties: For any x and y from $\mathbb{R}^n$ it holds $\phi_A(x + y) = A(x + y) = Ax + Ay = \phi_A(x) + \phi_A(y)$; for any real number λ and any x in $\mathbb{R}^n$ we have $\phi_A(\lambda x) = A(\lambda x) = \lambda(Ax) = \lambda\phi_A(x)$. A map between vector spaces with these properties is called a *linear map* or *homomorphism of vector spaces*. More on this later.

Remark II.5.26: From Corollary II.5.22 we read off: The map ϕ_A is surjective if and only if $\operatorname{rank}(A) = n$ and ϕ_A is injective if and only if $\operatorname{rank}(A) = m$.

Chapter III.

Mathematical Structures - Groups, Rings, Fields

1. Groups

Definition III.1.1 (Law of Composition): Let M be a non-empty set. A map $\star\colon M \times M \to M$ is called a *law of composition on* M. Usually, we write $m_1 \star m_2$ instead of $\star(m_1, m_2)$.

If for all a, b and c from M it holds $(a \star b) \star c = a \star (b \star c)$, then '$\star$' is called *associative*.

If for all a and b from M it holds $a \star b = b \star a$, then '$\star$' is called *commutative*.

Definition III.1.2 (Group): Let G be a non-empty set together with a law of composition $\star$. If '$\star$' is associative, and if there is an element e of G such that for any g in G it holds $e \star g = g = g \star e$, and if for any g in G there is some h in G with $g \star h = e = h \star g$, then the tuple $(G, \star)$ is called a *group*. If no confusion of the law of composition in question is to be feared, we imprecisely call G a group.

The element e is then called a *neutral element*, and an element h such that $g \star h = e = h \star g$ is called *inverse of* g and is usually denoted g^{-1}.

Note that neutral elements are unique and that inverses are unique. To show uniqueness of neutral elements, suppose e' were another neutral element in G. Then we had $e' = e \star e' = e$.

As to uniqueness of inverses, suppose h_1 and h_2 were inverses of g in G, i.e. suppose h_1 and h_2 were such that $h_1 \star g = e = g \star h_1$ and $h_2 \star g = e = g \star h_2$. We then had $h_1 = h_1 \star e = h_1 \star (g \star h_2) = (h_1 \star g) \star h_2 = e \star h_2 = h_2$. This justifies our notation 'g^{-1}' for *the* inverse of g.

Definition III.1.3 (Are you Abel?): Let G be a group with law of composition $\star$. If '$\star$' is commutative, then G is called a *commutative group* or *abelian group*.

Example III.1.4 (Some Groups): The following are groups:

(i) $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$,

(ii) $\mathbb{R}$-vector spaces with their addition,

(iii) $(\mathbb{Q} - \{0\}, \cdot)$ and $(\mathbb{R} - \{0\}, \cdot)$,

(iv) $\mathrm{GL}_n(\mathbb{R})$ with multiplication of matrices,

(v) $\mathrm{Perm}(M)$ for a non-empty set M, together with composition of maps as law of composition (see Definition I.4.17). For the special set $M = \{1, \ldots, n\}$, one typically denotes $\mathrm{Perm}(M)$ by S_n and calls it *symmetric group on n letters*.

The groups from (i)–(iii) are all abelian, while those from (iv) and (v) in general are not.

Example III.1.5 (Group of Congruence Classes): Let n be a natural number. Then, as shown in Remark I.5.9, '$\equiv_n$' declares an equivalence relation on the set of integers $\mathbb{Z}$. We write $\mathbb{Z}/n\mathbb{Z}$ for the set of equivalence classes with respect to '$\equiv_n$', i.e. $\mathbb{Z}/n\mathbb{Z} = \{[0], [1], \ldots, [n-1]\}$. On this set of equivalence classes, we can declare a new law of composition as follows:

$$+ \colon \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z}, \qquad [a] + [b] = [a + b].$$

This indeed does define a law of composition and turns $\mathbb{Z}/n\mathbb{Z}$ into an abelian group.

Proof: To show '+' yields a law of composition amounts to checking that no property of a map is violated. That is, the image is contained in $\mathbb{Z}/n\mathbb{Z}$, every tuple in $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ is assigned something, and that the definition does not depend on the chosen representative of one of the equivalence classes. In this case, the above assignment would assign more than one element to one element of the domain, which is illegal.

The first two points are of no concern to us. Adding two representatives and taking the respective equivalence class gives an element of $\mathbb{Z}/n\mathbb{Z}$, and we can do this for any equivalence class, since there are no empty equivalence classes.

As to the independence of the chosen representatives, let a' and b' be integers such that $a' \equiv_n a$ and $b' \equiv_n b$. We need to show that $a + b \equiv_n a' + b'$.

For two integers z and z', being equivalent with respect to '$\equiv_n$' means that $z' - z$ is divided by n, or in other words that there is some integer k such that $z' - z = kn$. Our different representatives a' and b' may thus be written as $a + kn$ respectively $b + \ell n$, for suitable integers k and ℓ. Now

$$\begin{aligned}(a' + b') - (a + b) &= a' + b' - a - b \\ &= (a' - a) + (b' - b) \\ &= \Big((a + kn) - a\Big) + \Big((b + \ell n) - b\Big) = (k + \ell)n,\end{aligned}$$

hence $a' + b' \equiv_n a + b$.

It remains to show that our new law of composition '$+$' is associative and commutative, and that there are a neutral element as well as inverses.

Because we calculate with representatives which are integers, and because addition of integers is both associative and commutative, there is nothing to be shown here.

With a keen eye, we spot $[0]$ as the neutral element of $\mathbb{Z}/n\mathbb{Z}$, since for any equivalence class $[a]$ we have $[a] + [0] = [a + 0] = [a]$. Equally clear is that the inverse of $[a]$ from $\mathbb{Z}/n\mathbb{Z}$ is $[-a]$. □

In the following, we will, depending on context, use $[a]$ or $\overline{a}$ to denote the equivalence class of the integer a with respect to '$\equiv_n$'.

Definition III.1.6 (Subgroup): Let $(G, \star)$ be a group and let H be a non-empty subset of G. If H contains the neutral element of G, if H is closed under $\star$, i.e. if for any h_1 and h_2 in H also $h_1 \star h_2$ belongs to H, and if H is closed under taking inverses, then H is called a *subgroup of* G.

Remark III.1.7: In the situation of Definition III.1.6, $\star|_{H \times H} \colon H \times H \to H$ is a law of composition, turning H into a group in its own right.

Proposition III.1.8 (Subgroup Criterion): *Let $(G, \star)$ be a group and let H be a subset of G. Then H is a subgroup of G if and only if H is non-empty, and if for any h_1, h_2 in H, also $h_1 \star h_2^{-1}$ belongs to H.*

Proof: As usual, let e denote the neutral element of G.

'$\Longrightarrow$': Suppose that H is a subgroup of G. Then e belongs to H, so H is in particular non-empty. Given any elements h_1 and h_2, as per definition of a subgroup, h_2^{-1} is an element of H, and $h_1 \star (h_2^{-1})$ belongs to H.

'$\Longleftarrow$': Suppose that H is a non-empty subset of G such that for any h_1, h_2 in H, $h_1 \star h_2^{-1}$ lies in H. Because H is non-empty, there is some h in H. By assumption, $h \star h^{-1} = e$ lies in H. In particular, for any h in H we thus have that $e \star h^{-1} = h^{-1}$ belongs to H. If we now have h_1 and h_2 in H, then h_2^{-1} lies in H and $h_1 \star (h_2^{-1})^{-1} = h_1 \star h_2$ is an element of H. □

Remark III.1.9: Let n be a natural number. Then $n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$ is a subgroup of $(\mathbb{Z}, +)$, which essentially follows from the distributive properties of multiplication of integers.

In the following, let $(G, \star)$ be a group with neutral element e.

Remark III.1.10 (Intersection of Subgroups): Let I be a non-empty index set and let $(G_i)_{i \in I}$ be a family of subgroups of $(G, \star)$. Then also $\bigcap_{i \in I} G_i$ is a subgroup of $(G, \star)$.

Proof: The intersection is non-empty, because every subgroup G_i contains the neutral element e of G.

Let now g_1 and g_2 be elements of $\bigcap_{i \in I} G_i$. Then, for any $i \in I$, g_1 and g_2 belong to G_i, and since G_i is a subgroup, so does $g_1 \star g_2^{-1}$. Per definition of the intersection, $g_1 \star g_2^{-1}$ thus lies in it. □

Definition III.1.11 (Subroups Generated by Subsets): Let M be a subset of G and let I be the set $\{H \subseteq G \mid H \text{ is subgroup of } (G, \star) \text{ and } M \subseteq H\}$. Then

$$\langle M \rangle = \bigcap_{H \in I} H$$

is called the *subgroup generated by* M. It is the smallest subgroup of G with respect to inclusion that contains M. If M consists of one element only, say $M = \{m\}$, then $\langle M \rangle = \langle \{m\} \rangle$ is called *cyclic*. Somewhat imprecise, we most often write $\langle m \rangle$ instead of $\langle \{m\} \rangle$.

Note that cyclic groups are always abelian.

Example III.1.12: Let $(G, \star)$ be the group $(\mathbb{Z}/10\mathbb{Z}, +)$. Then, for example, $\langle [1] \rangle = \mathbb{Z}/10\mathbb{Z}$, $\langle [2] \rangle = \{[2], [4], [6], [8], [10]\}$ and $\langle [3] \rangle = \mathbb{Z}/10\mathbb{Z}$.

Proposition III.1.13 (Characterisation of Cyclic Groups): *Let g be an element of G. Then $\langle g \rangle = \{g^k \mid k \in \mathbb{Z}\}$, where*

$$g^k = \begin{cases} g \star g \star \cdots \star g, & \text{if } k > 0, \\ e, & \text{if } k = 0, \\ (g^{-k})^{-1}, & \text{if } k < 0. \end{cases}$$

Definition III.1.14 (Order of Group): The number of elements of G is called *order of* G and is denoted $|G|$, $\mathrm{ord}(G)$ or $\#G$.

If g is an element of G, then the order of $\langle g \rangle$ is called the *order of* g.

Example III.1.15: Let again $(G, \star) = (\mathbb{Z}/10\mathbb{Z}, +)$. Then G is cyclic, since $\langle [1] \rangle = G$. Furthermore, $\mathrm{ord}(G) = 10$, $\mathrm{ord}([1]) = 10$, $\mathrm{ord}([2]) = 5$ and $\mathrm{ord}([5]) = 2$.

Lemma III.1.16: *Let H be a subgroup of $(G, \star)$ and consider the relation on G defined by '$g_1 \sim g_2$ if $g_1 \star g_2^{-1}$ lies in H'. Then it holds:*

(i) *'$\sim$' is an equivalence relation on G, and the equivalence class of $[g]$ is $H \star g = \{h \star g \mid h \in H\}$.*

(ii) *For g in G, the map $F_g \colon [e_G] = H \to [g] = H \star g$, $h \mapsto h \star g$ is a bijection.*

Proof: (i) This is a routine calculation, see e.g. Remark I.5.9.

(ii) Given an element $h \star g$ in $H \star g$, $F_g(h) = h \star g$. Thus F_g is surjective. If h_1 and h_2 are elements of H with $F_g(h_1) = F_g(h_2)$, then $h_1 \star g = h_2 \star g$, and multiplication with g^{-1} from the right yields $h_1 = h_2$, so F_g is injective, too. □

Definition III.1.17 (Right Cosets): Let G be a group, let H be a subgroup of G and let '$\sim$' be the equivalence relation on G declared via '$g_1 \sim g_2$ if $g_1 \star g_2^{-1}$ in H'. Then $H \star g = [g]$ is called the *right coset of g in G*.

Theorem 4 (Lagrange): *Let $(G, \star)$ be a finite group and let H be any subgroup of G. Then* $\mathrm{ord}(G)$ *is divided by* $\mathrm{ord}(H)$.

Proof: The group H declares an equivalence relation on G, as discussed in Lemma III.1.16. By assertion (ii) of this lemma, each right coset has the same number of elements, i.e. for any g in G, $\#(H \star g) = \#(H)$. Theorem 1 ensures that G is the union of the equivalence classes, that is $G = \bigcup_{g \in G} H \star g$, and either $H \star g_1 = H \star g_2$, or $H \star g_1 \cap H \star g_2 = \varnothing$. Since G is finite, there are only finitely many distinct equivalence classes, say k many. If we choose elements $g_1, \ldots, g_k$, one from each right coset, we obtain a disjoint union $G = \bigcup_{i=1}^{k} H \star g_i$ and

$$\mathrm{ord}(G) = \sum_{i=1}^{k} \#(H \star g_i) = \sum_{i=1}^{k} \#(H) = k\#(H),$$

which we wanted to show. □

Corollary III.1.18: *Let $(G, \star)$ be a finite group such that* $\mathrm{ord}(G)$ *is a prime number p. Then G is cyclic.*

Proof: Let g be an element of $G - \{e\}$ and let $H = \langle g \rangle$. Then $\mathrm{ord}(H) \geq 2$, and by Theorem 4, $\mathrm{ord}(G) = p$ is divided by $\mathrm{ord}(H)$. Because p is prime, $\mathrm{ord}(G) = \mathrm{ord}(H)$ follows and thus $G = \langle g \rangle$ is cyclic. □

Remark III.1.19: Let $(G, \star)$ be a group and let g_1, g_2 be elements of G. Then $(g_1 \star g_2)^{-1} = g_2^{-1} \star g_1^{-1}$. Because inverses are unique, it is sufficient to check that the asserted inverse is indeed an inverse.

2. Group Homomorphisms

Definition III.2.1: Let $(G, \star)$ and $(H, \bullet)$ be groups and let $\varphi\colon G \to H$ be a map. If for all g_1, g_2 in G it holds $\varphi(g_1 \star g_2) = \varphi(g_1) \bullet \varphi(g_2)$, then φ is called a *homomorphism of groups* or *group homomorphism*. By

$$\mathrm{Hom}(G, H) = \{\varphi\colon G \to H \text{ homomorphism of groups}\}$$

we denote the set of group homomorphisms from G to H.

For clarity, we may write $\varphi\colon (G, \star) \to (H, \bullet)$ to indicate that φ is a group homomorphism between G and H when equipped with the respective laws of composition.

Example III.2.2 (First Group Homomorphisms): (i) Given a matrix A in $\mathbb{R}^{n\times m}$, the associated map $\varphi_A\colon (\mathbb{R}^m, +) \to (\mathbb{R}^n, +)$ defined by $x \mapsto Ax$ is a homomorphism of groups.

(ii) Due to the functional identity of the exponential function, $\exp\colon (\mathbb{R}, +) \to (\mathbb{R} - \{0\}, \cdot)$, $x \mapsto \exp(x)$ is a homomorphism of groups.

(iii) For any group $(G, \star)$ and any element g of G, $\varphi_g\colon (\mathbb{Z}, +) \to G$, $k \mapsto g^k$ is a homomorphism of groups.

(iv) For the group $(S_n, \circ)$, $(S_n, \circ) \to (\mathrm{GL}_n(\mathbb{R}), \cdot)$, $\sigma \mapsto A_{\sigma^{-1}}$ with

$$A_\sigma(i, j) = \begin{cases} 1, & \text{if } j = \sigma(i), \\ 0, & \text{else,} \end{cases}$$

declares a homomorphism of groups. A group homomorphism into the general linear group of some vector space (in this case $\mathbb{R}^n$) is called a *representation*. This particular representation is also known as *fundamental representation of the symmetric group* S_n.

(v) For any groups $(G, \star)$ and $(H, \bullet)$, $\varphi\colon (G, \star) \to (H, \bullet)$, $g \mapsto e_H$ is a homomorphism of groups, also called *trivial homomorphism*.

The only non-obvious example is the fundamental representation of S_n. So let's go about verifying that it is indeed a homomorphism. Let permutations σ_1 and σ_2 be given, and let i, j be elements of $\{1, \ldots, n\}$. Then

$$(A_{\sigma_1} A_{\sigma_2})(i,j) = \sum_{k=1}^{n} A_{\sigma_1}(i,k) A_{\sigma_2}(k,j) = \left.\begin{cases} 1, & \text{if } j = \sigma_2(\sigma_1(i)), \\ 0, & \text{else} \end{cases}\right\} = A_{\sigma_2 \circ \sigma_1}(i,j),$$

since $A_{\sigma_1}(i,k) A_{\sigma_2}(k,j) = 1$ if and only if $k = \sigma_1(i)$ and $j = \sigma_2(k)$; zero otherwise. Hence $A_{\sigma_1} A_{\sigma_1^{-1}} = A_{\mathrm{id}} = I_n$, such that A_{σ_1} is indeed regular for any permutation σ_1. Furthermore, for any σ_1, σ_2 in S_n,

$$\varphi(\sigma_1 \circ \sigma_2) = A_{(\sigma_1 \circ \sigma_2)^{-1}} = A_{\sigma_2^{-1} \circ \sigma_1^{-1}} = A_{\sigma_2^{-1}} A_{\sigma_1^{-1}} = \varphi(\sigma_1)\varphi(\sigma_2),$$

making φ a homomorphism of groups.

In the following, we forgo explicit mentioning of the laws of composition in question.

Proposition III.2.3 (Group Homomorphisms and Inverses): *Let G and H be groups and let $\varphi\colon G \to H$ be a group homomorphism. Then it holds:*

(i) *For the neutral element e_G of G, $\varphi(e_G) = e_H$.*

(ii) *For any g in G, $\varphi(g^{-1}) = \varphi(g)^{-1}$.*

Proof: (i) Because e_G is a neutral element, it holds $e_G = e_G \star e_G$. Thus $\varphi(e_G) = \varphi(e_G \star e_G) = \varphi(e_G) \star \varphi(e_G)$. This makes $\varphi(e_G)$ a neutral element of H, and neutral elements are unique.

(ii) Lets check that $\varphi(g^{-1})$ is an inverse to $\varphi(g)$. We have

$$\varphi(g) \bullet \varphi(g^{-1}) = \varphi(g \star g^{-1}) = \varphi(e_G) = e_H$$

and similarly we convince ourselves of $\varphi(g^{-1}) \bullet \varphi(g) = e_H$. □

A decisive characteristic datum of a homomorphism is its kernel.

Definition III.2.4 (Kernel of Homomorphism): Let G and H be groups and let $\varphi\colon G \to H$ be a homomorphism. Then $\ker \varphi = \{g \in G \mid \varphi(g) = e_H\}$ is called the *kernel of* φ.

The kernel decides if the homomorphism is injective or not. Furthermore, an equivalence relation can be declared using the kernel, partitioning the elements of G according to their image under φ.

Proposition III.2.5 (Kernel vs. Injectivity): *Let G and H be groups and let $\varphi\colon G \to H$ be a group homomorphism . It holds $\ker\varphi = \{e_G\}$ if and only if φ is injective.*

Proof: '$\Longleftarrow$': We know that $\varphi(e_G) = e_H$. Because φ is injective by assumption, $\ker\varphi = \{e_G\}$.

'$\Longrightarrow$': Assume $\ker\varphi = \{e_G\}$ and let g_1, g_2 be elements of G such that $\varphi(g_1) = \varphi(g_2)$. Then $e_H = \varphi(g_2) \bullet \varphi(g_1)^{-1} = \varphi(g_2) \bullet \varphi(g_1^{-1}) = \varphi(g_2 \star g_1^{-1})$, i.e. $g_2 \star g_1^{-1}$ belongs to $\ker\varphi$. This means $g_2 \star g_1^{-1} = e_G$, and thus $g_1 = g_2$, showing the injectivity of φ. □

Proposition III.2.6 (Image and Preimage of Subgroups): *Let G and H be groups, and let $\varphi\colon G \to H$ be a homomorphism of groups. If G_1 is a subgroup of G, then $\varphi(G_1)$ is a subgroup of H. If H_1 is a subgroup of H, then $\varphi^{-1}(H)$ is a subgroup in G. In particular, $\operatorname{im}\varphi = \varphi(G)$ and $\ker\varphi = \varphi^{-1}(\{e_H\})$ are subgroups of the respective groups.*

Proof: Because G_1 is a subgroup, e_G lies in G_1. By Proposition III.2.3 e_G is mapped to e_H by φ, meaning that e_H belongs to $\varphi(G_1)$. Remains to show that $\varphi(G_1)$ is closed with respect to '$\bullet$'. Let x_1, x_2 be elements of $\varphi(G_1)$. Per definition, there are g_1, g_2 in G_1 such that $\varphi(g_i) = x_i$. As G_1 is a group, $g_1 \star g_2^{-1}$ belongs to G_1, and $\varphi(g_1 \star g_2^{-1}) = \varphi(g_1) \bullet \varphi(g_2)^{-1} = x_1 \bullet x_2^{-1}$. We have thus written $x_1 \bullet x_2^{-1}$ as the image of something in G_1.

Since H_1 is a subgroup of H, e_H belongs to H_1 and again by Proposition III.2.3, $e_G \subseteq \varphi^{-1}(\{e_H\})$, i.e. $\varphi^{-1}(H_1)$ is non-empty. Now for the closedness with respect to $\star$: Let x_1, x_2 be elements of $\varphi^{-1}(H_1)$. We have to show that $x_1 \star x_2^{-1}$ belongs to $\varphi^{-1}(H_1)$ as well, that is we have to find something in H_1 such that $x_1 \star x_2^{-1}$ is mapped to it. By definition, $\varphi(x_1)$ and $\varphi(x_2)$ certainly lie in H_1. H_1 is a subgroup of H, thus $\varphi(x_1) \bullet \varphi(x_2)^{-1} = \varphi(x_1 \star x_2^{-1})$ belongs to it and we are done. □

Example III.2.7 (Continuation of Example III.2.2): In the order of appearance, we have the following images and kernels:

(i) $\ker\varphi = \mathbb{S}(A, \mathbf{0})$, $\operatorname{im}\varphi = \{b \in \mathbb{R}^n \mid \mathbb{S}(A, b) \neq \varnothing\}$.

(ii) $\ker\varphi = \{0\}$, $\operatorname{im}\varphi = \mathbb{R}_{>0}$.

(iii) $\ker\varphi_g = \{0\}$ respectively $\operatorname{ord}(g)\mathbb{Z}$, if $\operatorname{ord}(g)$ is finite, $\operatorname{im}\varphi_g = \langle g \rangle$.

(iv) $\ker\varphi = \{\mathrm{id}\}$, $\operatorname{im}\varphi$ is called *group of permutation matrices.*

(v) $\ker\varphi = G_1$, $\operatorname{im}\varphi = \{e_{G_2}\}$.

Definition III.2.8 (Types of Homomorphisms): Let G and H be groups and let $\varphi\colon G \to H$ be a homomorphism of groups. If φ is injective, φ is called a *monomorphism*. If φ is surjective, φ is called *monomorphism*. If there is a group homomorphism $\psi\colon H \to G$ such that $\psi \circ \varphi = \mathrm{id}_G$, $\varphi \circ \psi = \mathrm{id}_H$, then φ is called an *isomorphism*. If $G = H$, then φ is called an *endomorphism*. If φ is both an endomorphism and an isomorphism, then φ is called an *automorphism*.

Proposition III.2.9 (Characterisation of Isomorphisms): *Let G and H be groups and let $\varphi\colon G \to H$ be a homomorphism of groups. Then φ is an isomorphism if and only if φ is bijective.*

Proof: '$\Longrightarrow$': By definition, ψ is in particular an inverse map to φ, making φ bijective.

'$\Longleftarrow$': We have to show that the inverse map φ^{-1} is a homomorphism of groups. Let therefore h_1 and h_2 be elements of H. Then

$$\psi(h_1 \bullet h_2) = \psi(\varphi(\psi(h_1)) \bullet \varphi(\psi(h_2))) = \psi(\varphi(\psi(h_1) \star \psi(h_2))) = \psi(h_1) \star \psi(h_2)$$

because $\psi \circ \varphi = \mathrm{id}_G$ and $\varphi \circ \psi = \mathrm{id}_H$. □

Proposition III.2.10 (Kernel as Normal Subgroup): *Let G and H be groups and let $\varphi\colon G \to H$ be a group homomorphism. For any g in G and any h in* $\ker\varphi$, *ghg^{-1} lies in* $\ker\varphi$.

Proof: We have to check that ghg^{-1} is mapped to e_H under φ. It holds $\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g)^{-1} = \varphi(g)\varphi(g)^{-1} = e_H$ because h is mapped to e_H by assumption. □

Remark III.2.11 (Composition of Homomorphisms): Let G_1, G_2 and G_3 be groups and let $\varphi\colon G_1 \to G_2$, $\psi\colon G_2 \to G_3$ be homomorphisms. Then so is $\psi \circ \varphi\colon G_1 \to G_3$.

The set $\mathrm{Aut}(G)$ of automorphisms of a group G is a subgroup of $\mathrm{Perm}(G)$ with respect to composition of maps.

Remark III.2.12 (Characterisation of Cyclic Subgroup): Let G be a group and let g be an element of G. We have introduced $\langle g\rangle$ as the smallest subgroup of G with respect to inclusion that contains g. In Example III.2.7(iii) we noted that $\langle g\rangle$ is the image of $\varphi_g\colon \mathbb{Z} \to G$, $k \mapsto g^k$. If the order k_0 of g is finite, then we have the following statements on $\langle g\rangle$ and k_0:

(i) The subgroup generated by g is the set $\{e_G, g, g^2, \ldots, g^{k_0-1}\}$.

(ii) The order k_0 is characterised by $k_0 = \min\{k \in \mathbb{N} \mid g^k = e_G\}$.

(iii) It holds $g^n = e_G$ if and only if n is divided by k_0. To see this, one could for example consider $n \pmod{k_0}$.

3. The Symmetric Group

Let M in the following always be a set. We will study the group of permutations of M in this section. Many results here are needed for Section 3 in Chapter V, but fit the current schedule because we are currently getting comfortable with groups.

Reminder III.3.1: We introduced $\mathrm{Perm}(M) = \{f\colon M \to M \text{ bijective}\}$ in Definition I.4.17. We had also introduced the special permutation group $S_n = \mathrm{Perm}(\{1, \dots, n\})$ in Example III.1.4. Both are groups when equipped with composition of maps as law of composition.

Example III.3.2 (The Symmetric Group S_3): As stated in Example III.1.4, S_3 has $3! = 3 \cdot 2 = 6$ elements, which we listed in Table III.1. In particular, having listed 6 maps, the list is exhaustive. We chose τ_i for the respective

	1	2	3
id	1	2	3
τ_1	1	3	2
τ_2	2	1	3
τ_3	3	2	1
ζ_1	2	3	1
ζ_2	3	1	2

Table III.1.: The elements of the symmetric group S_3.

elements, because they are what we will call *transpositions* later on, that is they exchange only two elements. For this section, we will fix the names given here for the elements of S_3.

Remark III.3.3: As already remarked in Definition I.4.17, $\mathrm{ord}(S_n) = n!$. This can be shown using standard combinatorial arguments.

Notation III.3.4 (Cauchy's Two-Line Notation): Let σ be an element of S_n. Then we write σ as

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}.$$

Example III.3.5: The permutations id, ζ_1 and τ_2 look like this in Cauchy's two-line notation for permutations:

$$\mathrm{id} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \zeta_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

Definition III.3.6 (Support of a Permutation): Let σ be a permutation of M. The set $\operatorname{supp}(\sigma) = \{x \in M \mid \sigma(x) \neq x\}$ is called *support of* σ. Two permutations σ_1 and σ_2 with $\operatorname{supp}(\sigma_1) \cap \operatorname{supp}(\sigma_2) = \varnothing$ are called *disjoint*.

Example III.3.7: For the permutations from Example III.3.2 we have, for example, $\operatorname{supp}(\xi_1) = \{1, 2, 3\}$, $\operatorname{supp}(\tau_2) = \{1, 2\}$ and $\operatorname{supp}(\mathrm{id}) = \varnothing$.

Remark III.3.8 (on Commutation of Disjoint Permutations): If σ_1 and σ_2 are disjoint permutations of a set M, then $\sigma_1 \circ \sigma_2 = \sigma_2 \circ \sigma_1$. Being disjoint is mandatory for this assertion. For example, for the permutations ξ_1 and τ_2 from Example III.3.2 it holds

$$\xi_1 \circ \tau_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \tau_2 \circ \xi_1.$$

Definition III.3.9 (Cycles and Transpositions): Let $x_1, \ldots, x_k$ be elements of the set M. We denote the permutation

$$\xi \colon M \longrightarrow M, \qquad x \longmapsto \begin{cases} x_{i+1}, & \text{if } x = x_i \text{ and } i \in \{1, \ldots, k-1\}, \\ x_1, & \text{if } x = x_k, \\ x, & \text{else} \end{cases}$$

by $(x_1 \ldots x_k)$ and call such a permutation a *k-cycle*. In this case, k is called the *length of* ξ.

If σ is a permutation of M and if there is some k such that σ is a k-cycle, we call σ a *cycle*. Cycles of length 2 are called *transpositions*.

Example III.3.10: Looking back at Example III.3.2, τ_1, τ_2 and τ_3 are transpositions, and ξ_1, ξ_2 are 3-cycles.

The permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix}$$

of $\{1, \ldots, 5\}$ can be written as composition of cycles, namely $\sigma = (1\,2) \circ (3\,4\,5)$.

The 4-cycle $\eta = (2\,5\,7\,3)$ in S_7 can be written as composition of transpositions, e.g. $\sigma = (2\,5) \circ (5\,7) \circ (3\,7)$.

Remark III.3.11 (Natural Decomposition into Transpositions): Suppose we are given a k-cycle $(x_1 \ldots x_k)$. Then there is a natural decomposition of this cycle into a product of transpositions:

$$(x_1 \ldots x_k) = (x_1\, x_2) \circ (x_2\, x_3) \circ \cdots \circ (x_{k-1}\, x_k).$$

This is seen by plugging in the elements of $\operatorname{supp}(\xi) = \{x_1, \ldots, x_k\}$ into both permutations, remembering that compositions are evaluated from right to left.

Example III.3.12: Consider the permutation σ from S_{12} with Cauchy two-line notation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 1 & 12 & 8 & 2 & 7 & 4 & 11 & 10 & 9 & 3 & 5 & 6 \end{pmatrix}.$$

Even though $\sigma = (1) \circ (2\,12\,6\,4) \circ (3\,8\,10) \circ (5\,7\,11) \circ (9)$ is not a cycle itself, σ is a composition of disjoint cycles.

Theorem 5 (on Decomposition of Permutations into Cycles): *Let M be a finite set. For any permutation σ in $\mathrm{Perm}(M)$, there are disjoint cycles $\xi_1, \ldots, \xi_N$, such that $\sigma = \xi_1 \circ \cdots \circ \xi_N$. It holds $\mathrm{supp}(\xi_i) \subseteq \mathrm{supp}(\sigma)$. The empty product, i.e. $N = 0$, is admitted, hence $\sigma = \mathrm{id}$ is covered.*

Proof: We show this assertion via induction on the cardinality of $\mathrm{supp}(\sigma)$. If $\#\,\mathrm{supp}(\sigma) = 0$, then $\sigma = \mathrm{id}$ and the assertion holds, so we have the base case covered.

For the induction step, we suppose the assertion holds true for any permutation σ' in $\mathrm{Perm}(M)$ with $\#\,\mathrm{supp}(\sigma') < \#\,\mathrm{supp}(\sigma)$. In order to make use of this assumption, we have to split off a cycle from σ. For this, we choose an element of $\mathrm{supp}(\sigma)$, i.e. an element of M that is moved somewhere by σ, and keep track of what σ does to it. Let's write this down formally.

Let x_0 be an element of $\mathrm{supp}(\sigma)$ and let $k_0 = \min\{k \in \mathbb{N} \mid \sigma^k(x_0) = x_0\}$. Clearly, $k \leq \mathrm{ord}(\sigma)$ (since $\sigma^{\mathrm{ord}(\sigma)} = \mathrm{id}$ and thus $\sigma^{\mathrm{ord}(\sigma)}(x_0) = x_0$). We obtain the set $Z_1 = \{x_0, \sigma(x_0), \ldots, \sigma^{k_0-1}(x_0)\}$ with cardinality k_0, and define the k_0-cycle $\xi_1 = (x_0\, \sigma(x_0) \ldots \sigma^{k_0-1}(x_0))$. For any x in Z_1 it then holds $\xi_1(x) = \sigma(x)$ by construction, and $\mathrm{supp}(\xi_1) = Z_1$, i.e. $\xi_1|_{M-Z_1} = \mathrm{id}\,|_{M-Z_1}$.

Now we split off ζ_1 by defining $\sigma_1 = \zeta_1^{-1} \circ \sigma$:

$$\sigma_1(x) = \begin{cases} \zeta_1^{-1}(\sigma(x)) = x, & \text{if } x \in Z_1, \\ \zeta_1^{-1}(\sigma(x)) = \sigma(x), & \text{if } x \in M - Z_1. \end{cases}$$

In particular, $\mathrm{supp}(\sigma_1) \subset \mathrm{supp}(\sigma) - Z_1$ and $\#\,\mathrm{supp}(\sigma_1) < \#\,\mathrm{supp}(\sigma)$, because Z_1 contains at least two elements by choice of x_0.

By induction hypothesis, σ_1 has a decomposition into disjoint cycles $\xi_2, \ldots, \xi_N$ with $\mathrm{supp}(\xi_i) \subseteq \mathrm{supp}(\sigma_1)$, such that $\mathrm{supp}(\xi_1) \cap \mathrm{supp}(\xi_i) = \varnothing$ for $2 \leq i \leq N$. In total this yields the decomposition $\sigma = \xi_1 \circ \sigma_1 = \xi_1 \circ (\xi_2 \circ \cdots \circ \xi_N)$, and the cycles have the asserted traits. □

In the following, we sometimes write permutations as products of disjoint cycles and thus obtain a second way of notating permutations.

Corollary III.3.13 (of Theorem 5): *Let σ be an element of S_n. Then there are a non-negative integer m and transpositions $\tau_1, \ldots, \tau_m$ such that $\sigma = \tau_1 \circ \cdots \circ \tau_m$.*

We can decompose σ into a product of disjoint cycles, and we have already seen a possibility to decompose a cycle into a product of transpositions.

Be careful: We neither claim m nor $\tau_1, \ldots, \tau_m$ to be uniquely determined; this would also be wrong. Because transpositions are self-inverse, a given decomposition could be bloated by pre- or postcomposing that product with the square of a transposition. We will show however that the parity of m is uniquely determined, i.e. if m is even or odd.

Definition III.3.14 (Signum): Let σ be an element of S_n. Then

$$\operatorname{sgn}(\sigma) = \prod_{1 \le i < j \le n} \frac{\sigma(j) - \sigma(i)}{j - i}$$

is called *signum* or *sign of* σ. Here, $\prod_{1 \le i < j \le n}(\cdot)$ means $\prod_{i=1}^{n} \prod_{j=i+1}^{n}(\cdot)$.

Example III.3.15: (i) Let's use the definition to compute the signum of a permutation (and see that this is no fun in general): For $\sigma = (1\,2\,3\,4)$ in S_4 we obtain

$$\begin{aligned}\operatorname{sgn}(\sigma) &= \frac{\sigma(2)-\sigma(1)}{2-1} \cdot \frac{\sigma(3)-\sigma(1)}{3-1} \cdot \frac{\sigma(4)-\sigma(1)}{4-1} \cdot \frac{\sigma(3)-\sigma(2)}{3-2} \\ &\quad \cdot \frac{\sigma(4)-\sigma(2)}{4-2} \cdot \frac{\sigma(4)-\sigma(3)}{4-3} \\ &= \frac{3-2}{2-1} \cdot \frac{4-2}{3-1} \cdot \frac{4-3}{3-2} \cdot \frac{1-3}{4-2} \cdot \frac{1-4}{4-3} = (-1) \cdot (-1) \cdot (-1) = -1.\end{aligned}$$

Looking at this example, it become plausible that $\operatorname{sgn}(\sigma) = -1^k$ for the non-negative integer $k = \#\{(i,j) \mid i < j \wedge \sigma(i) > \sigma(j)\}$.

(ii) Consider the transposition $\tau = (1\,2)$ in S_n. Then all factors in the definition of $\operatorname{sgn}(\tau)$ are 1, except those where i is 1 or 2. Hence

$$\operatorname{sgn}(\tau) = \prod_{j=2}^{n} \frac{\sigma(j)-2}{j-1} \prod_{j=3}^{n} \frac{\sigma(j)-1}{j-2} = \frac{\sigma(2)-2}{2-1} \prod_{j=3}^{n} \frac{j-2}{j-1} \prod_{j=3}^{n} \frac{j-1}{j-2} = -1$$

where we have used that $\sigma(2) = 1$ and that $\sigma(j) = j$ for $j \ge 3$.

Remark III.3.16 (Number of Inversions and Effect of Permutations): As in Example III.3.15 one can show for any permutation σ in S_n that the number k as defined in (i) is such that $\operatorname{sgn}(\sigma) = (-1)^k$. The tuples that are counted

by k are called *inversions of* σ and thus k is the number of inversions of σ. In particular, the signum of and permutation is either 1 or -1, which at first sight is not at all obvious.

If π is another permutation of S_n, then

$$\operatorname{sgn}(\sigma) = \prod_{1 \le i < j \le n} \frac{\sigma(j) - \sigma(i)}{j - i} = \prod_{1 \le i < j \le n} \frac{\sigma(\pi(j)) - \sigma(\pi(i))}{\pi(j) - \pi(i)}$$

which is due to the fact that π just permutes factors in this product. More precisely, the factor to the tuple (i, j) on the left-hand side corresponds to the factor to the tuple (i', j') on the right-hand side, where

$$(i', j') = \begin{cases} (\pi(i), \pi(j)), & \text{if } \pi(i) < \pi(j), \\ (\pi(j), \pi(i)), & \text{if } \pi(i) > \pi(j). \end{cases}$$

Proposition III.3.17: *Suppose σ_1 and σ_2 are permutations in S_n. Then it holds* $\operatorname{sgn}(\sigma_1 \circ \sigma_2) = \operatorname{sgn}(\sigma_1) \operatorname{sgn}(\sigma_2)$.

Proof: Using the preceding remark, we get

$$\begin{aligned} \operatorname{sgn}(\sigma_1 \circ \sigma_2) &= \prod_{1 \le i < j \le n} \frac{\sigma_1(\sigma_2(j)) - \sigma_1(\sigma_2(i))}{j - i} \\ &= \prod_{1 \le i < j \le n} \frac{\sigma_1(\sigma_2(j)) - \sigma_1(\sigma_2(i))}{\sigma_2(j) - \sigma_1(j)} \prod_{1 \le i < j \le n} \frac{\sigma_2(j) - \sigma_2(i)}{j - i} \\ &= \operatorname{sgn}(\sigma_1) \operatorname{sgn}(\sigma_2). \end{aligned}$$

□

Proposition III.3.18 (Conjugation Trick): *Let M be a finite set with at least two elements , and let a, b, a' and b' be elements of M such that $a \neq a'$ as well as $b \neq b'$. For a permutation π with $\pi(a') = a$ and $\pi(b') = b$ it holds*

$$(a'\, b') = \pi^{-1} \circ (a\, b) \circ \pi.$$

Proof: This is shown by evaluation. For some x in M we get

$$(\pi^{-1} \circ (a\, b) \circ \pi)(x) = \begin{cases} b', & \text{if } x = a', \\ a', & \text{if } x = b', \\ x, & \text{if } x \notin \{a', b'\}. \end{cases}$$

□

Theorem 6 (on the Sign of Permutations): *Let n be a natural number.*

(i) *The signum function* $\operatorname{sgn}\colon S_n \to \{\pm 1\}$ *is a homomorphism of groups.*

(ii) *If τ is a transposition, then* $\operatorname{sgn}(\tau) = -1$.

(iii) *If ξ is an ℓ-cycle, then*

$$\operatorname{sgn}(\xi) = (-1)^{\ell-1} = \begin{cases} -1, & \text{if } \ell \text{ is even,} \\ 1, & \text{if } \ell \text{ is odd.} \end{cases}$$

Proof: (i) This is a paraphrasing of Proposition III.3.17.

(ii) By the Conjugation Trick Proposition III.3.18 there is a transposition π such that $\tau = \pi \circ (1\,2) \circ \pi^{-1}$. In Example III.3.15 we have computed that $\operatorname{sgn}(1\,2) = -1$ and due to (i) this means $\operatorname{sgn}(\tau) = \operatorname{sgn}(\pi)(-1)\operatorname{sgn}(\pi^{-1}) = -1$.

(iii) Let $\xi = (x_1 \ldots x_\ell)$ be an ℓ-cycle with $x_1, \ldots, x_\ell$ in $\{1, \ldots, n\}$. As we noted in Remark III.3.11, $\xi = (x_1\,x_2) \circ (x_2\,x_3) \circ \cdots \circ (x_{\ell-1}\,x_\ell)$ and, again by (i), $\operatorname{sgn}(\xi) = (-1)^{\ell-1}$. □

4. Rings

Groups in some sense generalise the pair of real numbers together with addition that is well-known and familiar from school. They are an abstract concept of what is necessary to make reasonable computations in a set.

In this section we want to generalise something that is also known from school: Real numbers can be added and multiplied. In a way that both laws of composition 'fit' respectively 'play nicely', meaning the distributive property of multiplication over addition. The corresponding algebraic structure is called a *ring*.

Definition III.4.1: Let R be a non-empty set together with two laws of composition '$+$' and '$\cdot$'. If $(R, +)$ is an abelian group, if '$\cdot$' is associative, if there is a neutral element 1_R with respect to '$\cdot$' which is different from 0_R, and if for all x, y and z in R it holds

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z), \qquad (y + z) \cdot x = (y \cdot x) + (z \cdot x)$$

then $(R, +, \cdot)$ is called a *unital ring*. Is '$\cdot$' commutative in addition, then R is called a *commutative unital ring*.

The law of composition '$+$' is called *addition*, '$\cdot$' is called *multiplication*. Most often, we will not mention the laws of composition and just call R a unital

ring. In the following, we will shorten 'unital ring' to just 'ring' respectively 'commutative unital ring' to 'commutative ring'.

The neutral elements with respect to both laws of composition are denoted 0_R and 1_R, or just 0 and 1.

We agree on the notation $x - y = x + (-y)$, and we write $x \cdot y + z$ if we mean $(x \cdot y) + z$. Often, we leave out multiplication signs, i.e. we might write xy if we mean $x \cdot y$.

In different branches of mathematics there are different views towards a sensible notion of 'ring'. Correspondingly, there are many different definitions of 'ring', which makes it imperative to stick with the convention set by your context, e.g. this lecture.

For context: In commutative algebra and algebraic geometry, a ring usually is what we defined to be a commutative unital ring. A ring without a unit (i.e. a neutral element for multiplication) is often called a 'rng' to indicate that the ring is missing something.

In functional analysis and more specifically the theory of operator algebras, it is very natural to work with non-commutative rings (think of $\mathbb{R}^{n\times n}$). Also rings without a unit are commonplace, like special spaces of functions.

Example III.4.2: (i) $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$ and $(\mathbb{R}, +, \cdot)$ are rings.

(ii) $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ is a ring. Here, multiplication is declared by $[a] \cdot [b] = [a \cdot b]$. That this is in fact a well-defined law of composition is shown just like for addition.

(iii) Let R be any ring and let M be a non-empty set. Then R^M, together with the following laws of composition, turns into a ring:

$$f{+}g\colon M \longrightarrow R, \quad m \longmapsto f(m){+}g(m), \qquad f{\cdot}g\colon M \longrightarrow R, \quad m \longmapsto f(m){\cdot}g(m),$$

where f and g are maps from M to R. The constant zero map $\mathbf{0}\colon M \to R$ is the neutral element with respect to addition, and the constant one map $\mathbf{1}\colon M \to R$ is the neutral element with respect to multiplication.

Proposition III.4.3: *Let $(R, +, \cdot)$ be a ring. Then it holds:*

(i) *For any x in R, $0_R \cdot x = 0_R = x \cdot 0_R$.*

(ii) *For any x, y in R, $(-x) \cdot y = -(x \cdot y) = x \cdot (-y)$.*

Proof: (i) It holds $0_R \cdot x = (0_R + 0_R) \cdot x = 0_R \cdot x + 0_R \cdot x$, i.e. $0_R \cdot x$ is a neutral element for addition. Because neutral elements are unique, $0_R \cdot x = 0_R$. The same goes for $x \cdot 0_R$.

(ii) We have to show that $(-x) \cdot y$ respectively $x \cdot (-y)$ is additively inverse to $x \cdot y$. By distributivity we find $x \cdot y + (-x) \cdot y = (x + -x) \cdot y = 0_R \cdot y = 0_R$, similarly for the other case. □

Definition III.4.4 (Ring Homomorphism): Let $(R, +, \cdot)$ and $(S, +, \cdot)$ be rings and let $\phi\colon R \to S$ be a map. If for any x, y in R it holds $\phi(x +_R y) = \phi(x) +_S \phi(y)$ as well as $\phi(x \cdot_R y) = \phi(x) \cdot_S \phi(y)$ and if additionally, $\phi(1_R) = 1_S$, then ϕ is called a *homomorphism of unital rings*.

In the following, we shorten *homomorphism of unital rings* to *homomorphism of rings* respectively *ring homomorphism* in keeping with our convention of calling unital rings just rings.

By

$$\mathrm{Hom}_{\mathrm{Ring}}(R, S) = \{\phi\colon R \to S \text{ ring homomorphism}\}$$

we denote the set of ring homomorphisms from R to S.

If $\phi\colon R \to S$ is a ring homomorphism, it is in particular a homomorphism of the groups $(R, +)$ and $(S, +)$. In particular,

$$\ker \phi = \{r \in R \mid \phi(r) = 0_S\}, \qquad \mathrm{im}\, \phi = \{\phi(r) \mid r \in R\}$$

make sense again. However, $\ker \phi$ is no ring in the sense of our definition, since 1_R doesn't belong to it. Just like one did for groups one shows that ϕ is injective if and only if $\ker \phi = \{0_R\}$.

Remark III.4.5 (Canonical Homomorphism): Let R be a ring. Then there is a canonical homomorphism $\chi_R\colon \mathbb{Z} \to R$ such that $\chi_R(0) = 0_R$, $\chi_R(n+1) = \chi_R(n) + 1$ for non-negative integers n and $\chi_R(-n) = -\chi_R(n)$ for natural numbers n. That is, it holds

$$\phi(z) = \begin{cases} \sum_{i=1}^{z} 1_R, & \text{if } z > 0, \\ 0_R, & \text{if } z = 0; \\ \sum_{i=1}^{-z} -1_R, & \text{if } z < 0. \end{cases}$$

This homomorphism we call *canonical homomorphism for R*. Its image is a ring in R that is at the same time a cyclic group with respect to '+'.

Definition III.4.6 (Characteristic of a Ring): Let R be a ring. Then

$$\mathrm{char}(R) = \begin{cases} 0, & \text{if } \sum_{i=1}^{k} 1_R \neq 0_R \text{ for all } k, \\ \min\{k \in \mathbb{N} \mid \sum_{i=1}^{k} 1_R = 0_R\}, & \text{else ,} \end{cases}$$

is called the *characteristic of R*.

The canonical homomorphism to R, χ_R, is injective if and only if $\operatorname{char}(R) = 0$. In this case, we may identify $\chi_R(\mathbb{Z}) = \operatorname{im} \chi_R$ with the ring of integers $\mathbb{Z}$ and regard $\mathbb{Z}$ as subring of R.

If $\operatorname{char}(R)$ is a natural number n, then—in the same spirit—we may regard $\mathbb{Z}/n\mathbb{Z}$ as a subring of R.

Definition III.4.7 (Group of Units): Let R be a ring and let x be an element of R. If there is another element y of R such that $xy = 1_R = yx$, then x is called *invertible*. If x is invertible, the element y is uniquely determined and often denoted x^{-1}.

The set $R^\times = \{x \in R \text{ invertible}\}$ is called *group of units* or *unit group of* R and indeed is a group with respect to multiplication.

The uniqueness argument was already covered at the beginning on the section on groups.

That $R^\times$ is a group is true, as 1_R belongs to it, and because the product of two invertible elements x_1, x_2 is again invertible with inverse $((x_1x_2)^{-1} = x_2^{-1}x_1^{-1}$. This was shown in Proposition II.4.15 for matrices.

A polynomial $a_0 + a_1X + \cdots + a_NX^N$ with coefficients in a ring R is a family $(a_n)_{0\le n\le N}$ respectively a sequence $(a_n)_{n\in\mathbb{N}}$ of elements of R, where $a_n = 0_R$ for $n > N$.

We recall that a map $a\colon I \to R$ from an index set I to R is called a *family in* R, and that usually one denotes $a_n = a(n)$, identifying the map a with the tuple of images $(a(n))_{n\in I}$. For the special index set $I = \mathbb{N}_0$ respectively $I = \mathbb{N}$ one speaks of sequences.

If $(a_n)_{n\in\mathbb{N}}$ is a sequence in R and if there are an element r in R and a natural number N such that $a_n = r$ for any $n \ge N$, then $(a_n)_{n\in\mathbb{N}_0}$ is called *eventually constant*.

Definition III.4.8 (Polynomial): Let R be a ring and let $(a_n)_{n\in\mathbb{N}}$ be a sequence. If $(a_n)_{n\in\mathbb{N}}$ is eventually constant with $a_n = 0$ for all n larger than an index N, then $(a_n)_{n\in\mathbb{N}}$ is called a *polynomial* p *over* R.

If the sequence is constant with value zero, we call it the *zero polynomial* and denote it by $\mathbf{0}$.

The number

$$\deg(p) = \begin{cases} \max\{i \in \mathbb{N}_0 \mid a_i \neq 0\}, & \text{if } p \neq \mathbf{0}, \\ -\infty, & \text{if } p = \mathbf{0} \end{cases}$$

is called *degree of the polynomial* p. Polynomials of degree 0 are called *constant polynomials*.

Definition III.4.9 (Polynomial Ring): Let R be a ring. We denote

$$R[X] = R^{(\mathbb{N}_0)} = \{(a_n)_{n\in\mathbb{N}_0} \text{ is polynomial over } R\}$$

and call this the *polynomial ring over* R. We write X^m for the sequence $(\delta_{m,n})_{n\in\mathbb{N}_0}$, where $\delta_{m,n}$ denotes the Kronecker delta, and identify the polynomial $p = (a_n)_{n\in\mathbb{N}_0}$ with the sum $p = \sum_{i=0}^{\deg(p)} a_i X^i$. Then $R[X]$ turns into a ring with the laws of composition

$$p + q = \sum_{i=0}^{\max\{\deg(p),\deg(q)\}} (a_i + b_i)X^i, \qquad pq = \sum_{i=0}^{\deg(p)+\deg(q)} c_i,$$

where $p = \sum_{i=0}^{\deg(p)} a_i X^i$ and $q = \sum_{i=0}^{\deg(q)} b_i X^i$ are polynomials over R, and $c_i = \sum_{k=0}^{i} a_k b_{i-k}$ for $0 \le i \le \deg(p) + \deg(q)$. The neutral element with respect to addition is the zero polynomial, and the neutral element with respect to multiplication is the polynomial $p = (a_n)_{n\in\mathbb{N}_0}$ with $a_0 = 1_R$, $a_n = 0_R$ for $n \ge 1$.

Note that via multiplication as defined above, $X = X^1$, $X^i X^j = X^{i+j}$ and $X^0 = 1_R$. The polynomial X is often called *indeterminate*.

Definition III.4.10 (Field): Let R be a ring. If $R^\times = R - \{0\}$, i.e. if any non-zero element of R is invertible, then R is called a *field*.

Example III.4.11: (i) The sets of rational respectively real numbers are fields with the usual laws of composition.

(ii) Neither the ring of integers $\mathbb{Z}$ nor $\mathbb{R}^{n\times n}$ together with addition and multiplication of matrices are fields.

(iii) For a natural number n, the ring $\mathbb{Z}/n\mathbb{Z}$ as defined before is not a field in general. For example, if $n = 6$, then $\overline{2}\cdot\overline{3} = \overline{0}$. If there were s in $\mathbb{Z}/6\mathbb{Z}$ such that $s\cdot\overline{2} = \overline{1}$, then we had $\overline{3} = \overline{1}\cdot\overline{3} = s\cdot\overline{2}\cdot\overline{3} = s\cdot\overline{0} = \overline{0}$, which is absurd.

Proposition III.4.12: *Let p be a prime number. Then $\mathbb{Z}/p\mathbb{Z}$ is a field, most often denoted $\mathbb{F}_p$.*

Proof: Because $\mathbb{Z}/p\mathbb{Z} = \{\overline{0}, \overline{1}, \ldots, \overline{p-1}\}$ is a commutative ring, which we showed in Example III.4.2, it remains to show that any non-zero element is invertible. Let thus a be a non-zero element of $\mathbb{Z}/p\mathbb{Z}$. The subgroup of $(\mathbb{Z}/p\mathbb{Z}, +)$ that is generated by a contains $\overline{0}$ and a, i.e. is the whole of $\mathbb{Z}/p\mathbb{Z}$ by Theorem 4. Hence, also $\overline{1}$ lies in $\langle a\rangle$, which means that there is some natural number k with $\overline{1} = \sum_{i=1}^{k} a = \overline{k}a$. □

Example III.4.13 (Field of Complex Numbers): Let $\mathbb{C}$ denote the cartesian product $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. For elements $c_1 = (u_1, v_1)$ and $c_2 = (u_2, v_2)$, we put

$$c_1 + c_2 = (u_1 + u_2, v_1 + v_2), \qquad c_1 c_2 = (u_1 u_2 - v_1 v_2, u_1 v_2 - u_2 v_1),$$

thereby declaring an addition and a multiplication on $\mathbb{C}$. These laws of composition make $\mathbb{C}$ into a field. We write $i = (0, 1)$, and we identify a real number r with the tuple $(r, 0)$. Hence, c_1 may be written as $u_1 + \mathrm{i}v_1$. With this notation, $i^2 = (0, 1)(0, 1) = (-1, 0) = -1$.

Chapter IV.

Vector Spaces and Dimension Theory

In this chapter, we want to generalise the concept of an $\mathbb{R}$-vector space, introduced in Chapter II, to the notion of a vector space over an arbitrary field. We will then examine the properties of such vector spaces.

1. Vector Spaces

Let K denote an arbitrary field throughout this section.

Definition IV.1.1 (K-Vector Space): Let $(V, +)$ be an abelian group. If there is a law of composition $\cdot\colon K \times V \to V$ such that

(i) For any v in V, $1_K v = v$,

(ii) For any λ_1, λ_2 in K and any v in V, $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$,

(iii) For any λ in K and v_1, v_2 in V, $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$,

(iv) For any λ_1, λ_2 in K and v in V, $\lambda_1(\lambda_2 v) = (\lambda_1 \lambda_2)v$,

then $(V, +, \cdot)$ is called a *K-vector space* or vector space over K.

As usual—if the laws of composition are clear from context—we often imprecisely call V a K-vector space.

Example IV.1.2: The set $K^n = \{(x_1, \ldots, x_n)^t \mid x_1, \ldots, x_n \in K\}$ turns into an abelian group with pointwise addition. Together with component-wise scalar multiplication, it is a vector space over K.

Remark IV.1.3: (i) For $K = \mathbb{R}$, Definition IV.1.1 agrees with Definition II.2.2.

(ii) We can formulate Definition II.3.1 with entries in K, defining $(p \times q)$-matrices with entries in K as maps $\{1, \ldots, p\} \times \{1, \ldots, q\} \to K$. In doing so, we obtain the K-vector space $K^{p \times q}$.

(iii) All assertions on real vector spaces, matrices and systems of linear equations from sections 2 through 5 from Chapter II hold just the same over any field K.

In the following, by '$+$' we will always denote the addition of a vector space, and by '$\cdot$' we denote the scalar multiplication of a vector space. For an element v of a K-vector space V and an element λ of K, we usually write λv instead of $\lambda \cdot v$.

Definition IV.1.4 (Vector Space Homomorphism): Let V and W be vector spaces over a field K and let $\phi\colon V \to W$ be a map. If for any u and v in V it holds $\phi(u+v) = \phi(u)+\phi(v)$ and if for any λ in K, v in V it holds $\phi(\lambda v) = \lambda\phi(v)$, then ϕ is called a *K-vector space homomorphism*, *homomorphism of K-vector spaces* or a *K-linear map*.

If ϕ is K-linear and if there is a K-linear map $\psi\colon W \to V$ such that $\psi \circ \phi = \mathrm{id}_V$ and $\phi \circ \psi = \mathrm{id}_W$, then ϕ is called an *isomorphism*. In this case, the vector spaces V and W are called *isomorphic*. The existence of an isomorphism from V to W is often expressed via '$V \cong W$'.

If ϕ is K-linear and $V = W$, then ϕ is called an *endomorphism*. An endomorphism that is at the same time an isomorphism is called an *automorphism*.

In particular, if ϕ is a homomorphism of K-vector spaces, then ϕ is a homomorphism of the abelian groups V and W.

Example IV.1.5: (i) A matrix A in $K^{p \times q}$ defines the linear map $\varphi_A\colon K^q \to K^p$, $v \mapsto Av$.

(ii) The transposition map ${}^t\colon K^{p \times q} \to K^{q \times p}$, $A \mapsto A^t$ is a linear map. This was argued in Proposition II.3.10.

(iii) The zero map $V \to W$, $v \mapsto 0_W$ is a linear map.

Remark IV.1.6: Let V and W be K-vector spaces and let $\phi\colon V \to W$ be a K-linear map. The kernel $\ker\phi = \{v \in V \mid \phi(v) = \mathbf{0}_W\} = \phi^{-1}(\mathbf{0}_W)$ is a linear subspace of V. The homomorphism ϕ is injective if and only if $\ker\phi = \{\mathbf{0}_V\}$. It is an isomorphism if and only if it is a bijective homomorphism. Compositions on linear maps are linear maps.

From Proposition III.2.6 we already know that $\ker\phi$ is an abelian subgroup of V. If now λ is an element of K and v belongs to $\ker\phi$, then $\phi(\lambda v) = \lambda\phi(v) = \lambda\mathbf{0}_W = \mathbf{0}_W$, hence $\ker\phi$ is a linear subspace of V.

The characterisation of injectivity for homomorphisms was done in Proposition III.2.5 for group homomorphisms and carries over immediately. Similarly, the proof of Proposition III.2.3 generalises. The last assertion is directly verified with a straight forward calculation.

We want to recall that $W^V = \{\phi\colon V \to W \text{ is a map}\}$ is a K-vector space.

Proposition IV.1.7: *The subset*

$$\mathrm{Hom}_K(V,W) = \{\phi\colon V \to W \textit{ is a linear map}\} \subseteq W^V$$

is a linear subspace. We briefly write $\mathrm{Hom}(V,W) = \mathrm{Hom}_K(V,W)$.

2. Bases and Linear Independence

Let K be a field. For any element v of K^n there are uniquely determined $\lambda_1,\ldots,\lambda_n$ from K such that $v = \sum_{i=1}^n \lambda_i e_i$. Here, e_i denotes the i-th standard basis vector $e_i = (\delta_{i,j})_{1\le j\le n}$.

If $F^{(0)},\ldots,F^{(r)}$ denote the fundamental solutions of the system of linear equations $Ax = \mathbf{0}$ and if v is an element if $\mathbb{S}(A|\mathbf{0})$, then v may be uniquely expressed as a sum $v = \sum_{i=1}^r \alpha_i F^{(i)}$, i.e. the coefficients $\alpha_1,\ldots,\alpha_r$ are uniquely determined like above.

In this section we want to introduce and study the notions of 'basis' and 'dimension'. We will then show that any K-vector space (with a certain caveat) has a basis, and that all bases have the same cardinality. The dimension of a vector space is thus defined as the common cardinality of its bases.

Definition IV.2.1 (Linear Combination): Let V be a K-vector space, let M be a subset of V and let v be an element of V. If there are a natural number n, elements $v_1,\ldots,v_n$ of M and elements $\lambda_1,\ldots,\lambda_n$ of K such that $v = \sum_{i=1}^n \lambda_i v_i$, then v is called a *linear combination of* M.

If in the above definition $n = 0$, then the sum is empty and the empty sum by definition has the value $\mathbf{0}_V$.

The zero vector $\mathbf{0}_V$ is linear combination of any subset of a K-vector space V. Furthermore, it is the only linear combination of the empty set.

The definition of linear combination just given is easy to grasp but a bit cumbersome in practice. Linear combinations of different subsets of V are still elements of V and may as such be added. This makes them linear combinations

of the union of the two involved sets. Expressing this in notation is annoying because one has to 'fill up' sums by zeroes. Note that also, the involved sets were not required to be finite. To address both issues, we give a more elegant definition of the concept.

Definition IV.2.2 (Linear Combination, revised): Let M be a set and let $f\colon M \to K$ be a map. Then $\operatorname{supp}(f) = \{m \in M \mid f(m) \neq 0\}$ is called *support of* f. By $K^{(M)}$ we denote the set of maps from M to K whose support is a finite set.

For a subset M of a K-vector space V and a map $\lambda\colon M \to K$ with finite support, we call $\sum_{m\in M} \lambda(m)m = \sum_{m\in\operatorname{supp}(\lambda)} \lambda(m)m$ a *linear combination of* M.

Because the support of λ was required to be finite, the above sum is in fact a finite sum. This is important, because we otherwise needed to give some sense to an infinite sum.

Definition IV.2.3 (Basis): Let V be a K-vector space and let B be a subset of V. If any element v of V can uniquely be represented as linear combination of B, i.e. if for any v in V there is a uniquely determined λ in $K^{(B)}$ such that $v = \sum_{b\in\operatorname{supp}(\lambda)} \lambda(b)b$, then B is called a *basis of* V.

As a first milestone, we will show that any 'finitely generated' vector space has a basis.

Definition IV.2.4 (Linear Independence): Let V be a K-vector space and let M be a subset of V. If the only map λ in $K^{(M)}$ that combines to the zero vector, i.e. $\sum_{m\in\operatorname{supp}(\lambda)} \lambda(m)m = \mathbf{0}_V$, is the zero map, then M is called *linearly independent*. Any linear combination that evaluates to the zero vector is called a *linear relation* among the involved vectors.

Linear independence demands that the zero vector be a unique linear combination of M. In fact, this enforces any linear combination of M to be unique.

Example IV.2.5: Let K be the field of real numbers $\mathbb{R}$.

(i) Consider the subset $M_1 = \{v_1 = (1,2,3)^t, v_2 = (1,-1,1)^t\}$ of K^3 and the vector $v = (1,5,5)^t$. Then v is a linear combination of M_1, as $v = 2v_1 + 3v_2$.

(ii) The subset $M_2 = \{(1,0)^t, (1,1)^t\}$ is a linearly independent subset of K^2. For if $\lambda_1(1,0)^t + \lambda_2(1,1)^t = (0,0)^t$, then both $\lambda_1 + \lambda_2 = 0$ and $\lambda_2 = 0$ must hold. Hence, also $\lambda_1 = 0$. The corresponding map $\lambda\colon M_2 \to K$ is given by $(1,0)^t \mapsto 0$ and $(1,1)^t \mapsto 0$, that is λ is the zero map.

(iii) If we add the vector v to the set M_1 from (i), we end up with a linearly dependent set. Clearly. As we convinced ourselves that $v = 2v_1 + 3v_2$, or in different words, $\mathbf{0}_V = v - 2v_1 - 3v_2$ and here, not all coefficients are zero.

Remark IV.2.6 (Linear Independence for Finite Sets): Let V be a K-vector space and let $M = \{v_1, \ldots, v_n\}$ be a finite subset of V. Then M is linearly independent if and only if for any linear relation $\sum_{i=1}^{n} \lambda_i v_i = \mathbf{0}_V$ it follows $\lambda_1 = \cdots = \lambda_n = 0$.

Definition IV.2.7 (Linear Span): Let V be a K-vector space and let M be a subspace of V. Then $\mathrm{Lin}(M) = \{v \text{ is linear combination of } M\}$ is called the *linear hull*, *linear span* or briefly *span of* M. We may also write $\langle M \rangle = \mathrm{Lin}(M)$ or $\langle v_1, \ldots, v_n \rangle = \mathrm{Lin}(\{v_1, \ldots, v_n\})$.

Example IV.2.8 (First Spans): (i) Consider the subset $M = \{e_1, e_2\}$ of K^3. Then

$$\begin{aligned}\mathrm{Lin}(M) &= \left\langle \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix} \right\rangle \\ &= \left\{ \lambda_1 \begin{pmatrix}1\\0\\0\end{pmatrix} + \lambda_2 \begin{pmatrix}0\\1\\0\end{pmatrix} : \lambda_1, \lambda_2 \in K \right\} = \left\{ \begin{pmatrix}\lambda_1\\\lambda_2\\0\end{pmatrix} : \lambda_1, \lambda_2 \in K \right\}\end{aligned}$$

(ii) For any K-vector space V, $M = \varnothing$ is subset and $\mathrm{Lin}(M) = \{\mathbf{0}_V\}$.

Proposition IV.2.9 (Properties of Spans): *Let V be a K-vector space and let M be a subset of V. Then it holds:*

(i) *The set M is contained in its span* $\mathrm{Lin}(M)$.

(ii) *The span* $\mathrm{Lin}(M)$ *is a linear subspace of V.*

(iii) *The span* $\mathrm{Lin}(M)$ *is characterised by*

$$\mathrm{Lin}(M) = \bigcap (U \subseteq V \text{ linear subspace with } M \subseteq U).$$

(iv) *If M' is a subset of M, then* $\mathrm{Lin}(M')$ *is a subset of* $\mathrm{Lin}(M)$.

(v) *The subset M is a linear subspace of V if and only if $M = \mathrm{Lin}(M)$. In particular,* $\mathrm{Lin}(\mathrm{Lin}(M)) = \mathrm{Lin}(M)$.

(vi) *For two linear subspace U_1, U_2 of V it holds* $\mathrm{Lin}(U_1 \cup U_2) = U_1 + U_2$. *Here, $U_1 + U_2 = \{v + w \mid v \in U_1, w \in U_2\}$.*

Proof: (i) If x is an element of M, then $x = 1x$ expresses x as a linear combination of M.

(ii) As per definition, $\mathbf{0}_V$ belongs to $\mathrm{Lin}(M)$. If v and w are elements of $\mathrm{Lin}(M)$, then there are functions λ and μ in $K^{(M)}$ such that $v = \sum_{x\in\mathrm{supp}(\lambda)} \lambda(x)x$ and $w = \sum_{y\in\mathrm{supp}(\mu)} \mu(y)y$. Because $K^{(M)}$ is a K-vector space, $\lambda + \mu$ makes sense as a function, and because the laws of composition is point-wise, the sum $v + w$ corresponds to the linear combination associated to the function $\lambda + \mu$. Similarly, one show closedness under scalar multiplication.

(iii) '$\supseteq$': Because $\mathrm{Lin}(M)$ is a linear subspace of V that contains M, $\mathrm{Lin}(M)$ appears in the intersection on the right. Hence, the intersection is contained in $\mathrm{Lin}(M)$.
'$\subseteq$': Let v be an element of $\mathrm{Lin}(M)$. Then there is a function λ in $K^{(M)}$ such that $v = \sum_{x\in\mathrm{supp}(\lambda)} \lambda(x)x$. By definition, $\mathrm{supp}(\lambda)$ is a finite subset of M, hence all elements of $\mathrm{supp}(\lambda)$ are contained in every vector space involved in the intersection. Because vector spaces are closed under linear combination, the sum $v = \sum_{x\in\mathrm{supp}(\lambda)} \lambda(x)x$ is contained in every subspace containing M and thus lies in their intersection.

(iv) This is clear from the definitions.

(v) '$\Longleftarrow$' follows from (ii) and '$\Longrightarrow$' follows from (iii).

(vi) '$\subseteq$': As $U_1 + U_2$ is a linear subspace of V containing U_1 and U_2, i.e. $U_1 \cup U_2$, $\mathrm{Lin}(U_1 \cup U_2)$ is contained in $U_1 + U_2$.
'$\supseteq$': Let w be an element of $U_1 + U_2$. Then w is of the form $w = u_1 + u_2$ for some u_i in U_i. In particular, the u_i belong to $U_1 \cup U_2$, hence w is an element of $\mathrm{Lin}(U_1 \cup U_2)$. □

Theorem 7 (Criterion I for Bases): *Let V be a K-vector space. A subset B of V is a basis if and only if B is linearly independent and* $\mathrm{Lin}(B) = V$.

Proof: The implication '$\Longrightarrow$' is a direct consequence of Definition IV.2.3.
For '$\Longleftarrow$' let v be an element of V. Since $V = \mathrm{Lin}(B)$ by assumption, v is a linear combination of B. What remains to show is uniqueness. Suppose λ, μ are maps in $V^{(B)}$ such that $v = \sum_{w\in B} \lambda(w)w = \sum_{w\in B} \mu(w)w$. Then

$$\mathbf{0}_V = v - v = \sum_{w\in B} (\lambda(w) - \mu(w))w.$$

Because B was assumed to be linearly independent, this implies $\lambda(w) - \mu(w) = 0$ for any w in B, i.e. the maps λ and μ agree, which shows uniqueness. □

Example IV.2.10: Let V be the $\mathbb{R}$-vector space $\mathbb{R}^2$ and consider the subset $B = \{b_1 = (1,1)^t, b_2 = (1,-1)^t\}$ of V. Is this set B a basis of V? Theorem 7 states that we have to show two things: Firstly that B is linearly independent and secondly that B spans V.

As for linear independence: Suppose λ_1 and λ_2 are real numbers such that $\lambda_1 b_1 + \lambda_2 b_2 = \mathbf{0}$. This yields the equations $\lambda_1 + \lambda_2 = 0$ and $\lambda_1 - \lambda_2 = 0$ for the first respectively the second coordinate of this linear combination, which is equivalent to the linear system of equations

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

If we denote the above matrix by A, we know that the set B is linearly independent if and only if the homogeneous system $Ax = \mathbf{0}$ only has the trivial solution $x = \mathbf{0}$. By Corollary II.5.21 this is equivalent with the rank of A being equal to the number of rows of A, i.e. with $\operatorname{rank}(A) = 2$. The rank of A can be determined by direct calculation:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \xrightarrow{\text{II}-\text{I}} \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \xrightarrow{\frac{1}{2}\text{II}} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \xrightarrow{\text{I}-\text{II}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Hence B is indeed a basis of $\mathbb{R}^2$.

Proposition IV.2.11 (Criterion for Bases in K^n): *Let n and m be natural numbers, let K be a field, let $v_1, \ldots, v_m$ be elements of K^n and let $A = (v_1|\ldots|v_m)$.*

(i) *The set $\{v_1, \ldots, v_m\}$ is linearly independent if and only if* $\operatorname{rank}(A)$ *equals the number m of columns of A.*

(ii) *The set $\{v_1, \ldots, v_m\}$ is a generating set for K^n if and only if* $\operatorname{rank}(A)$ *equals the number n of rows of A.*

This follows from our theory for systems of linear equations discussed in Chapter II.

Corollary IV.2.12 (Dimension of K^n): *Let n be a natural number and let K be a field. Then any basis of K^n has precisely n elements.*

Proof: Let B be a basis of K^n. By Proposition IV.2.11(i), the cardinality of B is at most n. In particular, B is finite. Hence there are a natural number m and elements $v_1, \ldots, v_m$ such that $B = \{v_1, \ldots, v_m\}$. If we write $A = (v_1|\ldots|v_m)$, then Proposition IV.2.11(ii) tells us that $m = \operatorname{rank}(A) = n$, and we are done.□

Theorem 8 (Coordinate Map): *Let V be a K-vector space.*

(i) *If B is a basis of V, then $V \cong K^{(B)}$.*

(ii) *If B is a finite set, say $B = \{v_1, \ldots, v_n\}$, then $V \cong K^n$.*

Proof: (i) The map $\Gamma\colon K^{(B)} \to V$, $\lambda \mapsto \sum_{b\in B} \lambda(b)b$ is linear since the laws of composition on $K^{(B)}$ are point-wise and because of the calculation rules for finite sums. By assumption, $\mathrm{Lin}(B) = V$, which makes Λ surjective. The linear independence of B means that $\ker(\Lambda) = \{\mathbf{0}\}$ and so Λ is injective, i.e. an isomorphism.

(ii) By (i) we have the isomorphies $V \cong K^{(B)} \cong K^{(\{e_1,\ldots,e_n\})} \cong K^n$, where $\{e_1, \ldots, e_n\}$ denotes the standard basis of K^n. □

Definition IV.2.13 (Dimension): Let n be a natural number, let V be a K-vector space of dimension n and let $B = \{v_1, \ldots, v_n\}$ be a basis of V. Then any basis of V has n elements and we call $\dim V = n$ the *dimension of V*. Vector spaces whose dimension is a natural number are called *finite-dimensional*.

Theorem 9 (Criterion II for Bases): *Let V be a K-vector space and let B be a subset of V. Then the following are equivalent:*

(i) *The set B is a basis of V.*

(ii) *The set B is a linearly independent subset of V that is maximal with respect to inclusion, i.e. any subset M of V that contains B as a proper subset is not linearly independent.*

(iii) *The set B is a generating set of V which is minimal with respect to inclusion, i.e. there is no proper subset N of B with $\mathrm{Lin}(N) = V$.*

A set as described in (ii) is called a maximal linearly independent subset of V, and one as described in (iii) is called minimal generating set of V.

Proof: '(i) $\Longrightarrow$ (ii)': As B is assumed to be a basis of V, B is in particular linearly independent. If M is a subset of V that contains B as a proper subset, then there is some element v in $M - B$. Because B is a basis, there is a map λ in $K^{(B)}$ with $v = \sum_{b\in B} \lambda(b)b$. The map $\lambda'\colon M \to K$ declared by

$$\lambda'(b) = \begin{cases} \lambda(b), & \text{if } b \in B, \\ -1, & \text{if } b = v, \\ 0, & \text{otherwise} \end{cases}$$

has finite support, that is corresponds to a linear combination of M. More precisely, $\sum_{m\in M}\lambda'(m)m = \sum_{b\in B}\lambda(b)b - 1v = \mathbf{0}$, so λ' corresponds to a non-trivial linear relation among the elements of M, making M linearly dependent.

'(ii) $\Longrightarrow$ (iii)': Let B be a maximal linearly independent subset of V. We want to show that then, it already holds $\mathrm{Lin}(B) = V$ and that B is a minimal generating set for V.

To show that $\mathrm{Lin}(B) = V$ let v in V be given. If v belongs to B, then it already belongs to $\mathrm{Lin}(B)$. Otherwise, let $M = B \cup \{v\}$. Because B is maximal linearly independent and as B is properly contained in M, M must be linearly dependent. Thus there is some non-zero map λ in $K^{(N)}$ such that $\sum_{m\in M}\lambda(m)m = \mathbf{0}_V$. If $\lambda(v) \neq 0$, then B were linearly dependent. Hence, we could divide $\lambda(v)$ to obtain $v = -\sum_{m\in B}\lambda(v)^{-1}\lambda(m)m$, so v belongs to $\mathrm{Lin}(B)$.

Now for the minimality of B as a generating set: If there were a proper subset N of B with $\mathrm{Lin}(N) = \mathrm{Lin}(B)$, then B were linearly dependent contrary to our assumption.

'(iii) $\Longrightarrow$ (i)': Let B be a minimal generating set for V. We want to show that B then has to be linearly independent. Assume, B were linearly dependent. Then there were a non-zero map λ in $K^{(B)}$ with $\sum_{w\in B}\lambda(w)w = \mathbf{0}_V$. As λ is non-zero, there is some element v_0 of B such that $\lambda(v_0) \neq 0$. Like in the proof of '(ii) $\Longrightarrow$ (iii)', it then held $v_0 = -\sum_{w\in B-\{v_0\}}\lambda(v_0)^{-1}\lambda(w)w$, i.e. we had $\mathrm{Lin}(B - \{v_0\}) = \mathrm{Lin}(B) = V$. □

Corollary IV.2.14: *Let V be a K-vector space and let M be a non-empty subset of V.*

(i) *If M is linearly dependent, then there is a superfluous v in M, i.e. a v such that v lies in $\mathrm{Lin}(M - \{v\})$.*

(ii) *If M is linearly independent and if v belongs to $V - \mathrm{Lin}(M)$, then $M \cup \{v\}$ is linearly independent, too.*

Proof: Assertion (i) is a consequence of '(iii) $\Longrightarrow$ (i)' in the proof of Theorem 9, and assertion (ii) follows from '(ii) $\Longrightarrow$ (iii)' in the proof of Theorem 9. □

Theorem 10: *Let V be a K-vector space and suppose V has a finite generating set. Then it holds:*

(i) *The vector space V has a basis.*

(ii) *Any generating set of V contains a basis of V.*

(iii) *Any linearly independent subset of V can through union with a finite subset of V be completed to a basis of V.*

Proof: (i) Proving (ii) entails this assertion.

(ii) By Corollary IV.2.14(i) and the finiteness of the generating set it is possible to obtain a minimal generating set of V in finitely many steps. By Theorem 9 this then is a basis of V.

(iii) By (ii), V has a finite basis. According to Proposition IV.2.11 and Theorem 9, any linearly independent subset of V is finite. Given one such linearly independent subset, through finitely many applications of Corollary IV.2.14(ii) we end up with a basis of V. □

Definition IV.2.15 (Finite-dimensional Vector Space): Let V be a K-vector space. If V has a finite generating set, then V is called *finite-dimensional*. Otherwise, V is called *infinite-dimensional*. If V is finite-dimensional and B is any basis of V, then we call $\dim_K(V) = \#(B)$ the *dimension of V over K*.

3. Linear Extension and Transformation Matrix

Due to the calculation rules for finite sums, linear maps are uniquely determined by their images on basis vectors of the domain. We will see that this allows us to describe the behaviour of linear maps between finite-dimensional vector spaces through matrices.

Theorem 11 (Linear Extension Theorem): *Let V and W be K-vector spaces and let B be a basis of V. Then it holds:*

(i) *Any linear map $\Phi\colon V \to W$ is uniquely determined by its restriction $\phi = \Phi|_B\colon B \to W$.*

(ii) *Any map $\phi\colon B \to W$ can be uniquely extended to a linear map $\Phi\colon V \to W$, i.e. there is one and just one linear map $\Phi\colon V \to W$ such that $\Phi|_B = \phi$.*

Proof: (i) Let v be an element of V. Since B is a basis of V, there is some map λ_v in $K^{(B)}$ with $v = \sum_{b\in B} \lambda_v(b)b$ and this allows us to write

$$\Phi(v) = \Phi\Big(\sum_{b\in B} \lambda_v(b)b\Big) = \sum_{b\in B} \lambda_v(b)\Phi(b) = \sum_{b\in B} \lambda_v(b)\phi(b).$$

(ii) Any element v of V is a unique linear combination of B, that is there is a unique map λ_v in $K^{(B)}$ such that $v = \sum_{b\in B} \lambda_v(b)b$. Thus we may define $\Phi\colon V \to W$ by $\Phi(v) = \sum_{b\in B} \lambda_v(b)\phi(b)$. Due to the calculation rules for finite sums, this defines a linear map and the uniqueness of this extension follows from (i). □

Corollary IV.3.1: *Let V be a K-vector space and let B be a basis. Then the map*

$$H\colon \operatorname{Hom}_K(V,W) \longrightarrow B^{(W)}, \qquad \Phi \longmapsto \Phi|_B$$

is an isomorphism of K-vector spaces.

As we have seen in Theorem 8, for a K-vector space V with basis $B = \{b_1, \ldots, b_n\}$, there is a an isomorphism

$$\Lambda_B\colon K^n \longrightarrow V, \qquad (x_1, \ldots, x_n)^t \longmapsto \sum_{i=1}^{n} x_i b_i.$$

We denote the corresponding inverse map from V to K^n by D_B. The inverse map is determined by $D_B(b_i) = e_i$.

In what follows, an order for the elements of B will be necessary. We will thus consider *ordered bases*, which are tuples in V^n. If n and m are natural numbers and A is a matrix in $K^{n\times m}$, then we denote $L_A\colon K^m \to K^n$, $x \mapsto Ax$.

Theorem 12 (Transformation Matrix of a Linear Map): *Let V and W be K-vector spaces with ordered bases $B = (b_1, \ldots, b_n)$ respectively $C = (c_1, \ldots, c_m)$ and let $\phi\colon V \to W$ be a linear map. Then there is one and only one matrix A in $K^{n\times m}$ such that $D_C \circ \phi = L_A \circ D_B$, or put differently: There is one and only one matrix A that renders commutative the diagram*

$$\begin{array}{ccc} V & \xrightarrow{\ \phi\ } & W \\ {\scriptstyle D_B}\downarrow & & \downarrow{\scriptstyle D_C} \\ K^m & \xrightarrow[L_A]{} & K^n \end{array}$$

The entries of the matrix $A = (a_{i,j})$ are determined by $\phi(b_j) = \sum_{i=1}^{n} \lambda_{i,j} c_i$. We write $D_{C,B}(\phi) = A$ and call this the transformation matrix of ϕ with respect to the bases B and C.

Proof: First, we sketch the situation for $V = K^m$ and $W = K^n$, both equipped with the standard bases. Here, the transformation matrix of a linear map $\psi\colon K^m \to K^n$ is given by $A = (\psi(e_1)|\ldots|\psi(e_n))$, since Ae_i yields the i-th column of A and for any $v = (v_1, \ldots, v_m)^t$ in K^m we have

$$\psi(v) = \psi\Big(\sum_{i=1}^{m} v_i e_i\Big) = \sum_{i=1}^{m} v_i \psi(e_i) = \sum_{i=1}^{m} v_i A e_i = A\Big(\sum_{i=1}^{m} v_i e_i\Big) = Av,$$

which just means $L_A = \psi$.

On the level of abstract K-vector spaces, we do not have standard bases at our disposal. Thus, we need to translate to the known situation using our coordinate maps. Looking at the diagram and remembering Theorem 11, the maps we aim to compare are uniquely determined by their images on the b_i.

Because $D_B(b_i)$ yields the i-th standard basis vector e_i and as $L_A e_i$ returns the i-th column of A, we know that the i-th column of A needs to contain the coordinates of $\phi(b_i)$ with respect to C. But this corresponds precisely to the equations we gave in the statement of the theorem. □

Remark IV.3.2 (Determining Equation for the Transformation Matrix): In the situation of Theorem 12 for any v in V it holds $D_C(\phi(v)) = D_{C,B}(\phi) D_B(v)$.

Example IV.3.3: Let $V = W = \mathbb{R}^2$ and let $\phi\colon V \to V$ be the reflection at the diagonal $y = x$. Consider the ordered basis $B = (b_1 = (1,1)^t, b_2 = (1,-1)^t)$. If we express v in $\mathbb{R}^2$ as $v = \lambda_1 b_1 + \lambda_2 b_2$, then $\phi(v) = \lambda_1 b_1 - \lambda_2 b_2$. The linearity of ϕ is obvious. Furthermore,

$$D_B(\phi(v)) = \begin{pmatrix} \lambda_1 \\ -\lambda_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} D_B(v)$$

which determines the transformation matrix of ϕ with respect to the basis B.

Proposition IV.3.4 (Transformation Matrices of Compositions): *Let V_1, V_2 and V_3 be K-vector spaces of dimensions n_1, n_2 and n_3, respectively, let $\phi_1\colon V_1 \to V_2$ and $\phi_2\colon V_2 \to V_3$ be linear maps and let B_i be an ordered basis of V_i. For the corresponding transformation matrices it then holds*

$$D_{B_3,B_1}(\phi_2 \circ \phi_1) = D_{B_3,B_2}(\phi_2) D_{B_2,B_1}(\phi_1).$$

This means: Multiplication of matrices corresponds to composition of linear maps and vice versa.

Proof: We are in the situation

$$\begin{array}{ccccc} V_1 & \xrightarrow{\phi_1} & V_2 & \xrightarrow{\phi_2} & V_3 \\ {\scriptstyle D_{B_1}}\downarrow & & {\scriptstyle D_{B_2}}\downarrow & & \downarrow{\scriptstyle D_{B_3}} \\ K^{n_1} & \xrightarrow[L_1]{} & K^{n_2} & \xrightarrow[L_2]{} & K^{n_3} \end{array}$$

where L_1, L_2 are the corresponding linear maps

$$\begin{aligned} L_1\colon K^{n_1} &\longrightarrow K^{n_2}, \qquad x \longmapsto D_{B_2,B_1}(\phi_1)x, \\ L_2\colon K^{n_2} &\longrightarrow K^{n_3}, \qquad x \longmapsto D_{B_3,B_2}(\phi_2)x. \end{aligned}$$

Both squares of the above diagram are commutative by Theorem 12. The entire diagram is commutative, as for any v_1 in V_1 it holds

$$(D_{B_3} \circ \phi_2 \circ \phi_1)(v_1) = (L_2 \circ D_{B_2} \circ \phi_1)(v_1) = (L_2 \circ L_1 \circ D_{B_1})(v_1).$$

Furthermore, by definition of the maps L_1 and L_2, for any x in K^{n_1} it holds $L_2 \circ L_1(v_1) = D_{B_3,B_2}(\phi_2) D_{B_2,B_1}(\phi_1)(x)$. By the uniqueness of the transformation matrix it now follows $D_{B_3,B_1}(\phi_2 \circ \phi_1) = D_{B_3,B_2}(\phi_2) D_{B_2,B_1}(\phi_1)$. □

Definition IV.3.5 (Change-of-Basis Matrix): Let n be a natural number and let V be an n-dimensional K-vector space. If B and B' are ordered bases of V, then $D_{B',B} = D_{B',B}(\mathrm{id})$ is called *change-of-basis matrix* or *transition matrix from B to B'*.

Note that there are different customs as to the direction of this transition. By definition, for any vector v in V the transition matrix from above is characterised by the following equation:

$$D_{B'}(v) = D_{B',B}(\mathrm{id}) D_B(v) = D_{B',B} D_B(v).$$

This means, $D_{B',B}$ converts coordinates with respect to B into coordinates with respect to B'. It is therefore also called *change-of-coordinates matrix* from B to B'. On the other hand, the entries of $D_{B',B} = (\lambda_{i,j})$ are determined by the equations $b_j = \sum_{i=1}^{n} \lambda_{i,j} b'_i$, i.e. the entries of $D_{B',B}$ express the elements of B' in terms of linear combinations of the elements of B. Therefore, *change-of-basis matrix from B' to B* also is a justifiable name. The custom just explained here is especially wide-spread among physicists and differential geometricians.

Keep this in mind when reading other literature on this subject. Check the conventions used in the respective work.

Proposition IV.3.6 (Transformation Matrices under Changes of Bases): *Let V and W be finite-dimensional K-vector spaces with ordered bases $B = (b_1, \ldots, b_n)$ respectively $C = (c_1, \ldots, c_m)$ and let $\phi\colon V \to W$ be a linear map. Let furthermore B' and C' be additional ordered bases. Then it holds*

$$D_{C',B'}(\phi) = D_{C',C} D_{C,B}(\phi) D_{B,B'}.$$

Proof: The above situation is captured in the diagram

$$\begin{array}{ccccccc}
V & \xrightarrow{\mathrm{id}} & V & \xrightarrow{\phi} & W & \xrightarrow{\mathrm{id}} & W' \\
{\scriptstyle D_{B'}}\downarrow & & {\scriptstyle D_B}\downarrow & & \downarrow{\scriptstyle D_C} & & \downarrow{\scriptstyle D_{C'}} \\
K^m & \xrightarrow[L_{D_{B,B'}}]{} & K^m & \xrightarrow[L_{D_{C,B}(\phi)}]{} & K^n & \xrightarrow[L_{D_{C',C}}]{} & K^n
\end{array}$$

Since $\mathrm{id} \circ \phi \circ \mathrm{id} = \phi$, the outer arrows of this diagram represent the transformation matrix of ϕ with respect to B' and C'. The composition of the arrows on the bottom row yield the map $x \mapsto D_{C',C} D_{C,B}(\phi) D_{B,B'} x$ as claimed. □

Theorem 13 (Inverses of Change-of-Basis Matrices): *Let n be a natural number and let V be an n-dimensional K-vector space.*

(i) *Let B and B' be ordered bases of V. Then, the change-of-basis matrix $D_{B,B'}$ is regular and it holds $D_{B',B} = D_{B,B'}^{-1}$.*

(ii) *If $V = K^n$ and $E = (e_1, \ldots, e_n)$ denotes the ordered standard basis, then $D_{E,B} = (b_1 | \ldots | b_n)$.*

(iii) *For ordered bases B and B' of K^n it holds*

$$D_{B',B} = D_{B',E} D_{E,B} = D_{E,B'}^{-1} D_{E,B}.$$

Proof: (i) This follows from Proposition IV.3.6 for $\phi_1 = \phi_2 = \mathrm{id}$ and $B_1 = B$, $B_2 = B'$, $B_3 = B$.

(ii) With $B' = E = (e_1, \ldots, e_n)$ we obtain the determining equations $b_j = \sum_{i=1}^n \lambda_{i,j} e_i$ for the coefficients of the change-of-basis matrix, i.e. the $\lambda_{i,j}$ are precisely the coordinates of the b_j with respect to the standard basis.

(iii) Follows from Proposition IV.3.6 and (i). □

4. Sums of Subspaces and Quotient Spaces

If V is a K-vector space and $U_1, \ldots, U_n$ are linear subspaces of V, then we have already shown that

$$\sum_{i=1}^n U_i = U_1 + \cdots + U_n = \{x_1 \ + \cdots + x_n \mid x_1 \in U_1, \ldots, x_n \in U_n\}$$

is another linear subspace of V. We call $U_1 + \cdots + U_n$ the *sum of* $U_1, \ldots, U_n$.

Definition IV.4.1 (Direct Sum): Let V be a K-vector space and let $U_1, \ldots, U_n$ be linear subspaces of V. If for any linear relation $u_1 + \cdots + u_n = \mathbf{0}$ with u_i in U_i it follows that $u_1 = \cdots = u_n = \mathbf{0}$, then the sum $U_1 + \cdots + U_n$ is called *direct*. In this case we write $\bigoplus_{i=1}^n U_i = \sum_{i=1}^n U_i$.

Remark IV.4.2: Let V be a K-vector space and let $U_1, \dots, U_n$ be linear subspaces of V. If the sum $\sum_{i=1}^n U_i$ is direct, then for any indices $i \neq j$ it holds $U_i \cap U_j = \{\mathbf{0}\}$. For if v belonged to $U_i \cap U_j$, then $v + -v + \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}$, which enforces $v = \mathbf{0}$ by definition of the direct sum.

The converse of the above statement is false. Let for example $V = \mathbb{R}^2$, $U_1 = \langle (1,0)^t \rangle$, $U_2 = \langle (0,1)^t \rangle$ and $U_3 = \langle (1,1)^t \rangle$. Then $U_i \cap U_j = \{\mathbf{0}\}$ for $i \neq j$, however

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

such that $\sum_{i=1}^3 U_i$ is not a direct sum.

Theorem 14 (on Direct Sums): *Le V be a K-vector space and let $U_1, \dots, U_n$ be linear subspaces of V.*

(i) *Let $B_1, \dots, B_n$ be bases of the $U_1, \dots, U_n$. If the sum $\sum_{i=1}^n U_i$ is direct, then $B = \bigcup_{i=1}^n B_i$ is a basis of $\bigoplus_{i=1}^n U_i$.*

(ii) *Suppose $U_1, \dots, U_n$ are finite-dimensional. Then their sum is direct if and only if $\dim(\sum_{i=1}^n U_i) = \sum_{i=1}^n \dim U_i$.*

Proof: (i) By definition of $\bigoplus_{i=1}^n U_i$ the set B is a generating set. Remains to show that B is linearly independent. To this end, let λ be a map in $K^{(B)}$ such that $\mathbf{0} = \sum_{b \in B} \lambda(b) b$. For $1 \leq i \leq n$ let $u_i = \sum_{b \in B_i} \lambda(b) b$. Then $\sum_{i=1}^n u_i = \mathbf{0}$. But directness of the sum enforces $u_1 = \cdots = u_n = \mathbf{0}$. Because the B_i are bases, for $1 \leq i \leq n$ we have $\lambda|_{B_i} = \mathbf{0}$. Altogether, λ is the zero map, and thus B is linearly independent.

(ii) '$\Longrightarrow$' follows from (i). As to '$\Longleftarrow$': Because the U_i are finite-dimensional, each of these vector spaces has a basis, say B_i is a basis of U_i. Let $B = \bigcup_{i=1}^n B_i$. Then we have

$$\#B \leq \sum_{i=1}^n \#B_i = \sum_{i=1}^n \dim U_i = \dim \Big(\sum_{i=1}^n U_i \Big).$$

Since B is a generating set for $\sum_{i=1}^n U_i$, we must have $\dim \sum_{i=1}^n U_i \leq \#B$. By Theorem 9, B must be a basis for $\sum_{i=1}^n U_i$ and the linear independence of B yields the directness of the sum. □

Definition IV.4.3 (Equivalence Modulo a Linear Subspace): Let V be a vector space over K and let U be a linear subspace of V. Via '$v_1 \sim v_2$ if $v_1 - v_2$ lies in U', an equivalence relation is declared on V. The equivalence relation is called *equivalence modulo U*. For v in V, $[v] = \{w \in V \mid v \sim w\}$ denotes the equivalence class of v and $V/U = V/\sim = \{[v] \mid v \in V\}$ denotes the set of equivalence classes modulo U.

In the above situation, for v in V, one usually denotes $v+U = \{v+u \mid u \in U\}$. It can be shown (and you should in fact do so) that $v + U = [v]$.

Proposition IV.4.4 (Quotient by Linear Subspace): *Let V be a K-vector space, let U be a linear subspace of V and let V/U by the set of equivalence classes modulo U. Then,*

$$[v] + [w] = [v + w], \qquad \lambda[v] = [\lambda v]$$

declares laws of composition on V/U that make V/U into a K-vector space itself. The resulting K-vector space is called quotient of V by U *or just a* quotient space. *The map $\pi\colon V \to V/U$, $v \mapsto [v]$ is called* canonical projection. *This map is surjective, linear and satisfies* $\ker \pi = U$.

Proof: Since computations with equivalence classes using the laws of composition defined above reduce to computations with representatives in a K-vector space, we know immediately that they make V/U into a K-vector space as soon as we have shown their well-definedness. To this end we have to show that the results do not depend on chosen representatives. For demonstration of the arguments, we show well-definedness for addition. Let v, v' and w, w' be elements of V such that $v \sim v'$ and $w \sim w'$. Then $v' - v$ and $w' - w$ belong to U, which means there are elements u_1 and u_2 of U such that $v' - v = u_1$ and $w' - w = u_2$. Now,

$$v' + w' = (v + u_1) + (w + u_2) = (v + w) + (u_1 + u_2)$$

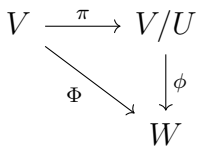
which means that $v + w$ and $v' + w'$ belong to the same equivalence class.

We turn to the canonical projection. The surjectivity of π follows from the fact that V decomposes into the disjoint union of the equivalence classes. To show linearity, let v and w be elements of V and let α be an element of K. By the definition of the K-vector space structure on V/U we find

$$\pi(v + \alpha w) = [v + \alpha w] = [v] + [\alpha w] = [v] + \alpha[w] = \pi(v) + \alpha\pi(w).$$

Finally, the characterisation $[0] = 0 + U = U$ shows that $\ker \pi$ is indeed U. □

Theorem 15 (Fundamental Theorem on Homomorphisms): *Let V and W be vector spaces over K, let $\Phi\colon V \to W$ be linear and let U be a linear subspace of V such that $U \subseteq \ker \Phi$. Then there is one and only one linear map $\phi\colon V/U \to W$ rendering commutative the diagram*



i.e. $\Phi = \phi \circ \pi$. *One says the map* Φ factorises through V/U. *If even* $U = \ker \Phi$, *then* ϕ *is injective and in particular,* $V/\ker \Phi \cong \operatorname{im} \Phi$ *via* ϕ.

Proof: We need to have for v in V that $\Phi(v) = \phi(\pi(v)) = \phi([v])$, i.e. we have no choice but to define $\phi([v]) = \Phi(v)$. It remains to show that this is well-defined. Let thus v and v' be representatives of $[v]$ in V/U. Then there is u in U such that $v = v'+u$ and as U is contained in $\ker \Phi$, $\Phi(v') = \Phi(v+u) = \Phi(v)+\varphi(u) = \Phi(v)$.

We have already seen that U is the neutral element of V/U. If $U = \ker \Phi$, then ϕ is injective per construction. Trivially, $\operatorname{im} \Phi = \operatorname{im} \phi$—again because V is the disjoint union of the equivalence classes—and thus ϕ is surjective if considered as map $\phi\colon V/\ker \Phi \to \operatorname{im} \Phi$, making ϕ an isomorphism. □

Proposition IV.4.5 (Basis of Quotient Space): *Let* V *be a* K*-vector space and let* U *be a linear subspace. If* B' *is a basis of* U *and* B *is a basis of* V *containing* B', *then* $C = \{[b] \mid b \in B - B'\}$ *is a basis of* V/U.

Proof: First we show that $\operatorname{Lin}(C) = V/U$. To this end, let v in V be given. Because B is a basis of V there is a map λ in $K^{(B)}$ such that $v = \sum_{b\in B} \lambda(b)b$. Since π is linear,

$$[v] = \Big[\sum_{b\in B} \lambda(b)b\Big] = \sum_{b\in B-B'} \lambda(b)[b] + \sum_{b\in B'} \lambda(b)b = \sum_{b\in B-B'} \lambda(b)[b].$$

Now we show the linear independence of C. Let thus a map λ in $K^{(C)}$ with $\sum_{c\in C} \lambda([c])[c] = [0]$ be given. Writing $u = \sum_{b\in B-B'} \lambda(b)b$, it holds $[u] = [0]$, i.e. u belongs to U. We know that B' is a basis for U', hence there is a map λ_u in $K^{(B')}$ with $u = \sum_{b\in B'} \lambda_u(b)b$. Thus

$$\sum_{b\in B-B'} \lambda(b)b - \sum_{b\in B'} \lambda_u(b)b = \mathbf{0}.$$

Because B is a basis of V, λ must be the zero map. □

Theorem 16 (Rank-Nullity Theorem): *Let* V *be a finite-dimensional* K*-vector space with* $\dim V = n$.

(i) *If* U *is a linear subspace of* V, *then* $\dim V/U = \dim V - \dim U$.

(ii) *If* U_1 *and* U_2 *are linear subspaces of* V, *then*

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

(iii) *If* W *is another* K*-vector space and* $\phi\colon V \to W$ *is linear, then*

$$\dim V = \dim \ker \phi + \dim \operatorname{im} \phi.$$

Proof: (i) In Proposition IV.4.5 we showed how one can obtain a basis for V/U. In particular, we determined the dimension of V/U.

(ii) We want to use the Fundamental Theorem of Homomorphisms to show this assertion. To this end we pick a suitable surjective map with the correct kernel, namely

$$\alpha\colon U_1 \times U_2 \longrightarrow U_1 + U_2, \qquad (u_1, u_2) \longmapsto u_1 - u_2.$$

For a given $u_1 + u_2$ in $U_1 + U_2$, $(u_1, -u_2)$ is a pre-image under α. Hence α is surjective. The kernel of α is $\ker\alpha = \{(u_1, u_2) \mid u_1 = u_2\}$. Thus the map $U_1 \cap U_2 \to \ker\alpha$ defined by $u \mapsto (u, u)$ is an isomorphism of vector spaces.

Finally, we can relate the dimensions of $U_1 \times U_2$ and $U_1 + U_2$. If B_1 is a basis of U_1 and if B_2 is a basis of U_2, then $B = \{(b, 0) \mid b \in B_1\} \cup \{(0, b) \mid b \in B_2\}$ yields a basis of $U_1 \times U_2$ of cardinality $\dim U_1 + \dim U_2$. Together with assertion (iii) we deduce

$$\begin{aligned}\dim U_1 + \dim U_2 &= \dim(U_1 \times U_2)\\ &= \dim\ker\alpha + \dim\operatorname{im}\alpha = \dim(U_1 \cap U_2) + \dim(U_1 + U_2).\end{aligned}$$

(iii) It was asserted in Theorem 15 that $\operatorname{im}\phi \cong V/\ker\phi$. In the exercises you will show that this entails that both vector spaces need to have the same dimension, hence $\dim\operatorname{im}\phi = \dim(V/\ker\phi) = \dim V - \dim\ker\phi$. □

Definition IV.4.6 (Rank and Kernel): Let V be a finite-dimensional K-vector space, let W be another K-vector space and let $\phi\colon V \to W$ be linear. Then $\operatorname{rank}\phi = \dim\operatorname{im}\phi$ is called the *rank of* ϕ.

Let n and m be natural numbers and let A in $K^{n\times m}$. Then $\ker A = \{x \in K^m \mid Ax = \mathbf{0}\}$ is called *kernel of the matrix* A.

Theorem 17 (on Ranks): *Let V and W be finite-dimensional K-vector spaces and let $\phi\colon V \to W$ be linear. Then it holds* $\dim V = \dim\ker\phi + \operatorname{rank}\phi$. *If ordered bases $B = (b_1, \ldots, b_m)$ of V and $C = (c_1, \ldots, c_n)$ of W are chosen, and if A denotes the transformation matrix $D_{C,B}(\phi)$, then it holds:*

(i) *The kernel of ϕ is isomorphic to the kernel of A, i.e. both definitions of kernel fit together.*

(ii) *If $\{e_1, \ldots, e_m\}$ denotes the standard basis of K^m, then the image of ϕ is isomorphic to* $\operatorname{Lin}(Ae_1, \ldots, Ae_m)$, *which is the linear span of the columns of A.*

(iii) *The ranks of A and ϕ agree, that is* $\operatorname{rank} A = \operatorname{rank}\phi$.

(iv) *If $s_1, \dots, s_m$ denote the columns of A and $z_1, \dots, z_n$ denote the rows of A, then*

$$\operatorname{rank} A = \dim \operatorname{Lin}(s_1, \dots, s_m) = \dim \operatorname{Lin}(z_1, \dots, z_n).$$

Proof: The first assertion is nothing else but Theorem 16(iii) formulated using the definition of rank. Now for the numbered assertions.

(i) By definition of the transformation matrix, the diagram

$$\begin{array}{ccc} V & \xrightarrow{\phi} & W \\ {\scriptstyle D_B}\downarrow & & \downarrow{\scriptstyle D_C} \\ K^m & \xrightarrow[\varphi_A]{} & K^n \end{array}$$

commutes. Via the isomorphism D_B, the linear subspace $\ker \phi$ of V is mapped to a subspace U of K^n. By commutativity of the diagram, for an element v of $\ker \phi$ it holds $\mathbf{0}_{K^n} = D_C(\phi(v)) = \varphi_A(D_B(v)) = AD_B(v)$, hence $U \subseteq \ker \varphi_A$. Since the coordinate homomorphisms are isomorphisms, we also have $\varphi_A = D_C \circ \phi \circ D_B^{-1}$. For w in $\ker \varphi_A$ it thus holds $\mathbf{0}_{K^n} = \varphi_A(w) = D_C(\phi(D_B^{-1}(w)))$. Injectivity of the coordinate homomorphisms yields that $D_B^{-1}(v)$ belongs to $\ker \phi$. It follows $D_B^{-1}(\ker \varphi_A) \subseteq \ker \phi$, so altogether $D_B(\ker \phi) = \ker \varphi_A$. In particular, the two linear subspaces are isomorphic to each other.

(ii) This is shown in just the same way as the previous assertion.

(iii) We know that $\operatorname{rank} A = m - \dim \ker A$. By (ii), $\dim \ker A = \dim \ker \phi$, such that $\operatorname{rank} A = m - \dim \ker \phi$. Theorem 16(iii) shows that this is precisely $\operatorname{rank} \phi$.

(iv) The image of the linear map $\varphi_A \colon K^m \to K^n$, $x \mapsto Ax$ is the span of the columns of A. Thus (iii) gives $\operatorname{rank} A = \operatorname{rank} \phi = \dim\langle s_1, \dots, s_m\rangle$.

Let now T be the row echelon form of A. The row echelon form of A results from A through elementary row transformations. More precisely, there are row transformations $Z_1, \dots, Z_N$ (i.e. $Z_k = A_{i,j}^{\alpha}$, $Z_k = V_{i,j}$ or $Z_k = \operatorname{diag}(\alpha_1, \dots, \alpha_n)$ for suitable parameters) such that $T = Z_1 \cdots Z_N A$.

Let $A_k = Z_{k+1} \cdots Z_N A$. For $A_{k+1} = Z_k A_k$ the span of the row vectors of A_k is the span of the row vectors of A_{k+1}, i.e. the span of the rows of A equals the span of the rows of T. But then also $\dim \operatorname{Lin}(z_1, \dots, z_n) = \operatorname{rank} T = \operatorname{rank} A$. □

Chapter V.

Endomorphisms of Vector Spaces

1. Endomorphisms and Chances of Bases

Remark V.1.1 (Change of Bases): Let V be a finite-dimensional vector space over K with ordered basis $B = (b_1, \ldots, b_n)$, let $\phi\colon V \to V$ be linear and let $B' = (b'_1, \ldots, b'_n)$ be another ordered basis of V. If we denote $A = D_{B,B}(\phi)$, $A' = D_{B',B'}(\phi)$ and $S = D_{B',B}$, then Proposition IV.3.6 gives

$$A' = D_{B',B'}(\phi) = D_{B',B} D_{B,B}(\phi) D_{B,B'} = SAS^{-1}.$$

Definition V.1.2 (Similarity): Let n be a natural number and let A_1, A_2 in $K^{n\times n}$. If there is an S in $\mathrm{GL}_n(K)$ such that $A_2 = SAS^{-1}$, then A_1 and A_2 are called *similar*.

In other words: Two square matrices A_1 and A_2 are similar if and only if they are transformation matrices of the same endomorphism, possibly with respect to different bases.

Proposition V.1.3 (Rank as Similarity Invariant): *Let n be a natural number and let A in $K^{n\times n}$. If B in $K^{n\times n}$ is another matrix that is similar to A, then* $\operatorname{rank} A = \operatorname{rank} B$.

2. Eigenvalues and Eigenvectors

Definition V.2.1 (Eigenvectors, Eigenvalues): Let V be a finite-dimensional K-vector space, let $\phi\colon V \to V$ be linear and let A in $K^{n\times n}$. Let λ be an element of K.

(i) If there is a v in $V - \{\mathbf{0}\}$ such that $\phi(v) = \lambda v$, then v is called an *eigenvector of* ϕ and λ is called the corresponding *eigenvalue of* ϕ. The

set $\mathrm{Eig}(\phi, \lambda) = \{v \in V \mid \phi(v) = \lambda v\}$ is called the corresponding *eigenspace of* ϕ.

(ii) If there is x in $K^n - \{\mathbf{0}\}$ such that $Ax = \lambda x$, then x is called an *eigenvector of* A and λ is called the corresponding *eigenvalue of* A. The set $\mathrm{Eig}(A, \lambda) = \{x \in K^n \mid Ax = \lambda x\}$ is called the corresponding *eigenspace of* A.

The set of eigenvalues $\mathrm{Spec}\,\phi = \{\lambda \in K \text{ eigenwert of } \phi\}$ is called *spectrum of* ϕ and $\mathrm{Spec}\,A = \{\lambda \in K \text{ eigenvalue of } A\}$ is called the *spectrum of* A.

Remark V.2.2: Let V be an n-dimensional vector space over K with ordered basis $B = (b_1, \ldots, b_n)$ and let $\phi\colon V \to V$ be linear.

(i) If $A = D_{B,B}(\phi)$, then $\mathrm{Spec}\,\phi = \mathrm{Spec}\,A$. Furthermore, v in V is an eigenvector of ϕ for the eigenvalue λ if and only if $x = D_B(v)$ is an eigenvector of A for the eigenvalue λ.

(ii) A λ in K belongs to $\mathrm{Spec}\,\phi$ respectively $\mathrm{Spec}\,A$ if and only if the corresponding eigenspace is non-trivial.

(iii) We have the equivalencies

$$Av = \lambda v \iff (A - \lambda I_n)v = \mathbf{0} \iff v \in \ker(A - \lambda I_n),$$

i.e. $\mathrm{Eig}(A, \lambda) = \ker(A - \lambda I_n)$. Analogously, $\mathrm{Eig}(\phi, \lambda) = \ker(\phi - \lambda\,\mathrm{id})$. In particular, the eigenspaces of A respectively the eigenspaces of ϕ are linear subspaces.

Example V.2.3: (i) Let $\lambda_1, \ldots, \lambda_n$ be elements of K, let $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Then $\mathrm{Spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$, furthermore the eigenspaces can easily be determined. Namely $\mathrm{Eig}(A, \lambda_i) = \mathrm{Lin}(e_i)$.

(ii) Let $\phi\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection at the diagonal (refer to Example IV.3.3). With respect to the basis $B = \{(1,1)^t, (1,-1)^t\}$ of $\mathbb{R}^2$, the transformation matrix of ϕ is

$$D_{B,B}(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

such that $\mathrm{Spec}(\phi) = \{\pm 1\}$.

Definition V.2.4 (Diagonalisability): Let n be a natural number, let V be a K-vector space of dimension n, let $\phi\colon V \to V$ be an endomorphism of V and let A be a square $n \times n$-matrix with entries in K.

(i) If there are an ordered basis B' of V and elements $\lambda_1, \dots, \lambda_n$ of K such that $D_{B',B'}(\phi) = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, then ϕ is called *diagonalisable*.

(ii) If there are a matrix S in $\mathrm{GL}_n(K)$ and elements $\lambda_1, \dots, \lambda_n$ of K such that $SAS^{-1} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, then A is called *diagonalisable*.

An endomorphism is diagonalisable if and only if its transformation matrix with respect to any basis is diagonalisable, see Remark V.1.1.

Proposition V.2.5 (Sum of Eigenspaces is Direct): *Let V be a finite-dimensional K-vector space, let ϕ be an endomorphism of V and let $\lambda_1, \dots, \lambda_k$ be pairwise distinct eigenvalues of ϕ with corresponding eigenspaces* $\operatorname{Eig}(\phi, \lambda_i)$. *Then the sum of these eigenspaces is direct, i.e.* $\sum_{i=1}^{k} \operatorname{Eig}(\phi, \lambda_i) = \bigoplus_{i=1}^{k} \operatorname{Eig}(\phi, \lambda_i)$.

Proof: We show the assertion via induction on the number of eigenvalues k. For $k = 0, 1$ the assertion trivially holds true.

Assume now the assertion holds for $k-1$ and let $\mathbf{0} = u_1 + \dots + u_k$ for u_i in $\operatorname{Eig}(\phi, \lambda_i)$. Application of ϕ to this linear relation yields $\mathbf{0} = \phi(\mathbf{0}) = \sum_{i=1}^{k} \lambda_i u_i$. Furthermore, since $\sum_{i=1}^{k} u_i = \mathbf{0}$, we also have $\sum_{i=1}^{k} \lambda_k u_i = \mathbf{0}$. The difference of both linear relations reads

$$\mathbf{0} = (\lambda_1 - \lambda_k)u_1 + \dots + (\lambda_{k-1} - \lambda_k)u_{k-1} + (\lambda_k - \lambda_k)u_k.$$

By assumption, the sum $\sum_{i=1}^{k-1} \operatorname{Eig}(\phi, \lambda_i)$ is direct. Hence each summand is zero, and because we also assumed the eigenvalues to be pairwise distinct, no difference is zero, forcing $u_1 = \dots = u_{k-1} = \mathbf{0}$. Finally, the first $k-1$ eigenvectors being zero reduces the original linear relation to $u_k = \mathbf{0}$ and the induction step is done. □

Let λ be an eigenvalue of ϕ. Then, for v in $\operatorname{Eig}(\phi, \lambda)$ it holds $\phi(v) = \lambda v$, or in different words $\phi(\operatorname{Lin}(v)) \subseteq \operatorname{Lin}(v)$. By linearity of ϕ it thus follows that $\phi|_{\operatorname{Eig}(\phi,\lambda)}$ is an endomorphism of $\operatorname{Eig}(\phi, \lambda)$.

If ϕ is a diagonalisable endomorphism, for any eigenvalue λ we have the decomposition $V = \operatorname{Eig}(\phi, \lambda) \oplus \bigoplus_{\lambda \neq \mu \in \operatorname{Spec}(\phi)} \operatorname{Eig}(\phi, \mu) = U \oplus W$ and $\phi(U) \subseteq U$, $\phi(W) \subseteq W$.

Corollary V.2.6: *Let V be a finite-dimensional K-vector space and let ϕ be an endomorphism of V with spectrum* $\operatorname{Spec}(\phi) = \{\lambda_1, \dots, \lambda_k\}$.

(i) *It holds* $\#\operatorname{Spec}(\phi) \leq \dim V$.

(ii) *The endomorphism ϕ is diagonalisable if and only if* $V = \bigoplus_{i=1}^{k} \operatorname{Eig}(\phi, \lambda_i)$ *and it holds* $V = \bigoplus_{i=1}^{k} \operatorname{Eig}(\phi, \lambda_i)$ *if and only if* $\dim V = \sum_{i=1}^{k} \dim \operatorname{Eig}(\phi, \lambda_i)$.

Proof: (i) An element λ of K belongs to the spectrum of ϕ if and only if $\dim \operatorname{Eig}(\phi, \lambda) \geq 1$. See Remark V.2.2. As the sum of eigenspaces to pairwise distinct eigenvalues is direct and because dimension is monotonous, we can't have $k > \dim V$.

(ii) '$\Longleftarrow$' is clear.

'$\Longrightarrow$': If ϕ is diagonalisable, then there is an ordered basis $B = (b_1, \ldots, b_n)$ of V such that the transformation matrix $A = D_{B,B}(\phi)$ of ϕ with respect to B is a diagonal matrix. Per construction of the transformation matrix and since the transformation matrix is diagonal by assumption,

$$\phi(b_i) = D_B^{-1} \circ \varphi_A \circ D_B(b_i) = D_B^{-1}\varphi_A(e_i) = D_B^{-1}(A_{i,i}e_i) = A_{i,i}b_i$$

which means that the elements of B are eigenvectors of ϕ. Those elements of B that correspond to the same diagonal entries of A form linearly independent subsets in their respective eigenspaces. Because of the remark preceding the corollary, they also generate the eigenspaces.

The second equivalence is immediately seen to be a formulation of the first in terms of dimensions. □

3. The Determinant

In preparation of this section, we briefly recall properties of the homomorphism of groups $\operatorname{sgn}\colon S_n \to \{\pm 1\}$. For a k-cycle σ it holds $\operatorname{sgn}(\sigma) = (-1)^k$. In particular, transpositions have sign -1. Any permutation can be decomposed into k-cycles, and any k-cycle may be decomposed into transpositions. Consequently, any permutation can be expressed as a product of transpositions.

Definition V.3.1 (Determinant): Let n be a natural number and let A be a square $n \times n$-matrix with entries in K. Then

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \tag{V.1}$$

is called the *determinant of A*.

Example V.3.2: Let $n = 2$ and let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. The group S_2 has two elements, namely id and (12), and so Eq. (V.1) evaluates to

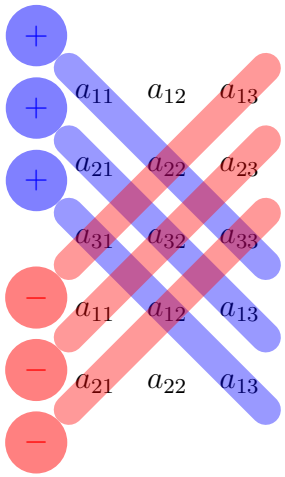
$$\det A = a_{1,1}a_{2,2} - a_{1,2}a_{2,1} = ad - bc.$$

In the exercises it was shown that $\det A$ in this case is a crucial feature: It determines if A is invertible ($\det A \neq 0$) or not ($\det A = 0$).

Example V.3.3: Let $n = 3$ and $A = (a_{i,j})_{1\leq i,j\leq 3}$ in $K^{3\times 3}$. The symmetric group on three letters is $\{\mathrm{id}, (123), (132), (12), (13), (23)\}$, the permutations having sign -1 and the identity as well as the 3-cycles having sign 1. Accordingly, Eq. (V.1) in this case reads

$$\begin{aligned}\det A = {} & a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} \\ & -a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,2}a_{3,1} - a_{1,1}a_{2,3}a_{3,2}.\end{aligned}$$

The above formula is also known as 'Rule of Sarrus'. The is a useful mnemonic illustration for it:



For $n \geq 4$, an evaluation of Eq. (V.1) is impractical. For these cases, other techniques are employed. However, Eq. (V.1) still bears theoretical value for arbitrary n.

Proposition V.3.4 (Properties of the Determinant): *Let n be a natural number and let $D\colon \prod_{i=1}^{n} K^n \to K$, $(x_1, \ldots, x_n) \mapsto \det(x_1|\ldots|x_n)$.*

(i) *If $v_1, \ldots, v_n$ and v_i' are elements of K^n, then*

$$\begin{aligned}&D(v_1, \ldots, v_{i-1}, v_i + v_i', v_{i+1}, \ldots, v_n) \\ &\quad = D(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n) + D(v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_n).\end{aligned}$$

(ii) *If $v_1, \ldots, v_n$ are elements of K^n and if λ belongs to K, then*

$$D(v_1, \ldots, v_{i-1}, \lambda v_i, v_{i+1}, \ldots, v_n) = \lambda D(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n).$$

(iii) *If $v_1, \ldots, v_n$ belong to K^n and if there are indices $i \neq j$ such that $v_i = v_j$, then $D(v_1, \ldots, v_n) = 0$.*

(iv) *If $e_1, \dots, e_n$ denotes the standard basis of K^n, then $D(e_1, \dots, e_n) = 1$.*

Proof: Let $v_1, \dots, v_n$ be elements of K^n and let $A = (v_1 | \dots | v_n)$.

(i) Let i be an index in $\{1, \dots, n\}$, let $v_i' = (t_1, \dots, t_n)$ be a vector in K^n, let $A' = (v_1 | \dots | v_{i-1} | v_i + v_i' | v_{i+1} | \dots | v_n)$ and $A'' = (v_1 | \dots | v_{i-1} | v_i' | v_{i+1} | \dots | v_n)$. Then

$$\begin{aligned}\det A' &= \sum_{\sigma \in S_n} \prod_{k=1}^{n} a'_{k,\sigma(k)} \\ &= \sum_{\sigma \in S_n} \prod_{\substack{k=1 \\ \sigma(k) \neq i}}^{n} a_{k,\sigma(k)} (a_{\sigma^{-1}(i),i} + t_{\sigma^{-1}(i),i)}) = \det A + \det A''.\end{aligned}$$

(ii) Let λ be an element of K and let $A' = (v_1 | \dots | v_{i-1} | \lambda v_i | v_{i+1} | \dots | v_n)$. The factor λ appears once in every factor of $\det A'$, thus $\det A' = \lambda \det A$.

(iii) Let k and ℓ be distinct indices such that $v_k = v_\ell$ and let σ_0 be the transposition $(k\ell)$ in S_n. Since transpositions have negative sign, for any permutation σ in S_n it holds $\operatorname{sgn}(\sigma \circ \sigma_0) = \operatorname{sgn}(\sigma_0 \circ \sigma) = -\operatorname{sgn}(\sigma)$. Write $A_n = \{\sigma \in S_n \mid \operatorname{sgn}(\sigma) = 1\}$, we may express composition with σ_0 as the a bijection $A_n \to S_n - A_n$. Finally, for a permutation σ in S_n it holds

$$\sigma \circ \sigma_0(i) = \begin{cases} \sigma(i), & \text{if } i \neq \{k, \ell\}, \\ \sigma(\ell), & \text{if } i = k, \\ \sigma(k), & \text{if } i = \ell. \end{cases}$$

Because $v_k = v_\ell$, we obtain $\prod_{i=1}^n a_{i,\sigma\circ\sigma_0^{-1}(i)} = \prod_{i=1}^n a_{i,\sigma^{-1}(i)}$ such that

$$\begin{aligned}\det A &= \sum_{\sigma \in A_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} + \sum_{\sigma \in S_n - A_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \\ &= \sum_{\sigma \in A_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} + \sum_{\sigma \in A_n} \operatorname{sgn}(\sigma \circ \sigma_0) \prod_{i=1}^{n} a_{i,\sigma'(i)} \\ &= \sum_{\sigma \in A_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} - \sum_{\sigma \in A_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} = 0.\end{aligned}$$

(iv) Let $A = (e_1 | \dots | e_n)$. Since $a_{i,j} = \delta_{i,j}$, any permutation that is different from the identity results in a summand that doesn't contribute to the sum. Hence $\det A = \prod_{i=1}^n a_{i,i} = 1$. □

Example V.3.5: Let $A = (v_1|v_2|v_3)$ in $K^{3\times 3}$ be given.

(i) For some λ in K and $A' = (v_1 + \lambda v_2|v_2|v_3)$ it holds $\det A = \det A'$, since

$$\det A' = \det(v_1|v_2|v_3) + \lambda \det(v_2|v_2|v_3) = \det A.$$

Here, we used properties (i) and (ii) of Proposition V.3.4 for the first equality sign, and property (iv) for the second equality sign.

(ii) For $A' = (v_3|v_2|v_1)$ we have $\det A = \det A'$, because

$$\begin{aligned} 0 &= \det(v_1 + v_3|v_2|v_1 + v_3) \\ &= \det(v_1|v_2|v_1 + v_3) + \det(v_3|v_2|v_1 + v_3) \\ &= \det(v_1|v_2|v_1) + \det(v_1|v_2|v_3) + \det(v_3|v_2|v_1) + \det(v_3|v_2|v_3) \\ &= \det(v_1|v_2|v_3) + \det(v_3|v_2|v_1). \end{aligned}$$

Hence $\det A = \det(v_1|v_2|v_3) = -\det(v_3|v_2|v_1) = -\det A'$.

(iii) For some λ in K and $A' = (v_1|\lambda v_2|v_3)$, property (ii) of Proposition V.3.4 shows $\det A' = \lambda \det A$.

Elementary row and column transformations are realised by addition matrices $A^\alpha_{k,\ell} = I_n + \alpha E_{k,\ell}$, permutation matrices $V_{k,\ell} = I_n - E_{k,k} - E_{\ell,\ell} + E_{k,\ell} + E_{\ell,k}$ and diagonal matrices $\operatorname{diag}(\alpha_1, \ldots, \alpha_n)$ with non-zero diagonal entries. Those matrices will be called *special matrices* in what follows.

Proposition V.3.6 (Determinant and Special Matrices): *Let n be a natural number and let A be a square $n \times n$-matrix with entries in K. Then it holds:*

(i) *For $A' = AA^\alpha_{k,\ell}$, $\det A' = \det A$.*

(ii) *For $A' = AV_{k,\ell}$, $\det A' = -\det A$.*

(iii) *For $A' = A\operatorname{diag}(\alpha_1, \ldots, \alpha_n)$, $\det A' = \alpha_1 \cdots \alpha_n \det A$.*

The same calculations as in Example V.3.5 show the assertions.

Remark V.3.7: Let n be a natural number. The $n \times n$-special matrices $A^\alpha_{k,\ell}$, $V_{k,\ell}$ and $\operatorname{diag}(\alpha_1, \ldots, \alpha_n)$ have the determinants $\det A^\alpha_{k,\ell} = 1$, $\det V_{k,\ell} = -1$ and $\det \operatorname{diag}(\alpha_1, \ldots, \alpha_n) = \alpha_1 \cdots \alpha_n$.

That $\det V_{k,\ell} = -1$ can be seen directly from the properties of the map D in Proposition V.3.4, then $\det A^\alpha_{k,\ell} = 1$ follows from that and the last determinant can be evaluated with Leibniz' formula.

In consequence, Proposition V.3.6 shows for any square $n \times n$-matrix A with entries in K and any $n \times n$-special matrix X that $\det(AX) = \det A \det X$.

Remark V.3.8: Let n be natural number and let A in $K^{n\times n}$ be given. We have the following rules for elementary column operations and their effect on the determinant of A:

(i) If A' emerges from A via addition of the λ-fold of the k-th column to the ℓ-th column, where k and ℓ are distinct, then $\det A' = \det A$.

(ii) If A' emerges from A via permutation of the k-th and ℓ-th column, again for k and ℓ distinct, then $\det A' = -\det A$.

(iii) If A' emerges from A via multiplication of the k-th column by λ, then $\det A' = \lambda \det A$.

(iv) If a column of A is the zero vector, then $\det A = 0$.

Theorem 18 (Properties of the Determinant): *Let n be a natural number and let A, A_1 and A_2 be square $n \times n$-matrices with entries in K.*

(i) *It holds* $\det A \neq 0$ *if and only if A belongs to* $\mathrm{GL}_n(K)$.

(ii) *It holds* $\det A = \det A^t$.

(iii) *It holds* $\det(A_1 A_2) = \det A_1 \det A_2$.

(iv) *If A belongs to* $\mathrm{GL}_n(K)$, *then* $\det A^{-1} = (\det A)^{-1}$.

(v) *The determinant is invariant under similarity, i.e. for any S in* $\mathrm{GL}_n(K)$, $det(SAS^{-1}) = \det A$.

(vi) *Some λ in K is eigenvalue of A if and only if* $\det(A - \lambda I_n) = 0$.

Proof: (i) Suppose A were not invertible. Then there were a row echelon form T' with $\det T' = 0$ and special matrices $X_1, \ldots, X_k$ such that $A^t = X_1 \ldots X_k T'$. We then had $A = T'^t X_k^t \cdots X_1^t$ and furthermore, $X_1^t, \ldots, X_k^t$ were special matrices, too. From Proposition V.3.6 it followed $\det A = 0$.

Suppose A were invertible. Then, I_n were the row echelon form of A and the same arguments as in the other case showed that $\det A \neq 0$.

(ii) Suppose A were not invertible. Then A^t also weren't invertible, and by (i) we had $\det A = 0 = \det A^t$.

Suppose A were invertible. Then, as I_n were the row echelon form of A, there were special matrices $X_1, \ldots, X_\ell$ such that $A = X_1 \cdots X_\ell$. By Proposition II.3.10, $A^t = X_\ell^t \cdots X_1^t$. Again, from Proposition V.3.4 it followed that $\det A = \prod_{i=1}^{\ell} \det X_i = \prod_{i=1}^{\ell} \det X_i^t = \det A^t$.

(iii) If A_1 and A_2 are invertible, then both are products of special matrices and the assertion follows from Proposition V.3.6.

If A_1 is invertible and A_2 is not, then $\operatorname{rank} A_2) \leq n-1$ and $\dim\ker A_2 \geq 1$ by Theorem 16. Hence there is v in $K^n - \{\mathbf{0}\}$ such that $A_2 v = \mathbf{0}$. Then even more so, $A_1 A_2 v = \mathbf{0}$ and thus $\dim\ker(A_1 A_2) \geq 1$, too. This means that $A_1 A_2$ is not invertible, and by (i) we get $0 = \det(A_1 A_2) = \det A_1 \cdot 0 = \det A_1 \det A_2$.

If A_2 is invertible but A_1 is not, then we can reduce to the previous case by transposing the product:

$$\det(A_1 A_2) = \det(A_2^t A_1^t) = \det A_2^t \det A_1^t = \det A_2 \det A_1 = \det A_1 \ \det A_2.$$

(iv) If A is invertible, then $\det A^{-1} \det A = \det(A^{-1}A) = \det I_n = 1$ which shows the claim.

(v) By (iv), $\det(SAS^{-1}) = (\det S) \det A (\det S)^{-1} = \det A$.

(vi) '$\Longrightarrow$': Suppose λ is an eigenvalue of A. Then there is a vector v in $K^n - \{\mathbf{0}\}$ such that $Av = \lambda v$. This means $\mathbf{0} = Av - \lambda v = (A - \lambda I_n)v$. Hence, $\ker(A - \lambda I_n)$ is a proper subspace of K^n and $A - \lambda I_n$ is not injective, hence not invertible.

'$\Longleftarrow$': If $\det(A - \lambda I_n) = 0$, then $A - \lambda I_n$ is not invertible. By Theorem 16, $A - \lambda I_n$ is surjective if it is injective and vice versa. Thus, $A - \lambda I_n$ is not injective and $\ker(A - \lambda I_n)$ is a proper subspace of K^n. □

Because $\det A = \det A^t$, Remark V.3.8 has an analogue for elementary row transformations. The replacements in the formulation are obvious.

Definition V.3.9 (Determinant of an Endomorphism): Let V be a finite-dimensional vector space over K with ordered basis $B = (b_1, \ldots, b_n)$ and let $\phi\colon V \to V$ be linear. Then $\det\phi = \det D_{B,B}(\phi)$ is called *determinant of* ϕ.

This notion of determinant, even though defined through some transformation matrix, is well-defined because the determinant is invariant under similarity and any two transformation matrices of ϕ are similar to each other.

Definition V.3.10 (Characteristic Polynomial): Let n be a natural number, let A be a square $n \times n$-matrix with entries in K, let V be a finite-dimensional K-vector space with ordered basis $B = (b_1, \ldots, b_n)$ and let $\phi\colon V \to V$ be linear.

(i) $\chi_A = \det(A - \lambda I_n)$ is called *characteristic polynomial of* A.

(ii) $\chi_\phi = \det(\phi - \lambda\,\mathrm{id})$ is called *characteristic polynomial of* ϕ.

Example V.3.11: For $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ it is

$$\chi_A = \det \begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2.$$

Because zeroes of the characteristic polynomial are precisely the eigenvalues of A, $\operatorname{Spec} A = \{1\}$.

4. Laplace Expansion

Definition V.4.1: Let n be a natural number and let A be a square $n \times n$-matrix with entries in K. For indices $1 \leq i, j \leq n$, $M_{i,j}$ denotes the square $(n-1) \times (n-1)$-matrix obtained from A by removing the i-th row and j-th column.

Example V.4.2: Let $n = 3$ and consider

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

For $(i, j) = (1, 2)$, $(i, j) = (2, 2)$ und $(i, j) = (3, 2)$ we obtain the 2×2-matrices

$$M_{1,2} = \begin{pmatrix} 3 & 1 \\ 2 & 3 \end{pmatrix}, \qquad M_{2,2} = \begin{pmatrix} 1 & 3 \\ 2 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}.$$

Theorem 19 (Laplace Expansion Theorem): *Let n be a natural number and let A be an $n \times n$-matrix with entries in K. Furthermore, let k be an element of $\{1, \ldots, n\}$.*

(i) *The Laplace expansion along the k-th row of A is*

$$\det A = \sum_{j=1}^{n} (-1)^{j+k} a_{k,j} \det(M_{k,j}).$$

(ii) *The Laplace expansion along the k-th column of A is*

$$\det A = \sum_{j=1}^{n} (-1)^{i+k} a_{i,k} \det(M_{i,k}).$$

Example V.4.3: For the matrix from Example V.4.2 and $k = 2$, we obtain the following expansion along the second column:

$$\det A = -2 \det \begin{pmatrix} 3 & 1 \\ 2 & 3 \end{pmatrix} + 2 \det \begin{pmatrix} 1 & 3 \\ 2 & 3 \end{pmatrix} - \det \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = -12.$$

To prove Theorem 19, which we want to show via induction on the size of the matrix, we go through some preparatory statements on block matrices.

Proposition V.4.4 (Determinants for Block Matrices): *Let n be a natural number, let k be an index in $\{1,\dots,n\}$, let X in $K^{n\times n}$, Y in $K^{(n-k)\times k}$ and Z in $K^{(n-k)\times(n-k)}$ be given. Then*

$$\det\begin{pmatrix} X & \mathbf{0} \\ Y & Z \end{pmatrix} = \det X \det Y = \begin{pmatrix} X & Y \\ \mathbf{0} & Z \end{pmatrix}.$$

Proof: First, consider the special case where $X = I_k$ and $Y = \mathbf{0}$. Then, $\det(\begin{smallmatrix} I_k & \mathbf{0} \\ \mathbf{0} & Z \end{smallmatrix}) = \det Z$. This can be seen as follows: First, we use column transformations to transform Z into the transpose of a row echelon form, then, these column transformations make $(\begin{smallmatrix} I_k & 0 \\ 0 & Z \end{smallmatrix})$ into the transpose of a row echelon form, and then the assertion follows.

Now we consider the general case, where X is any $k \times k$-matrix. Using the product expansion

$$\begin{pmatrix} X & \mathbf{0} \\ \mathbf{0} & Y \end{pmatrix} = \begin{pmatrix} I_k & \mathbf{0} \\ \mathbf{0} & Z \end{pmatrix} \begin{pmatrix} X & \mathbf{0} \\ \mathbf{0} & I_{n-k} \end{pmatrix}$$

we can reduce this case to the first case we considered, and our assertion holds true here as well.

Finally, we come to the assertion from our proposition. If $\det X = 0$, we can, without changing the determinant, achieve using special matrices that the block in which X sits contains a zero column. By design, the entire matrix then has a zero column and its determinant is zero. In this case, our assertion holds.

If $\det X \neq 0$, then X is regular. Since

$$H = \begin{pmatrix} I_k & \mathbf{0} \\ -YX^{-1} & I_{n-k} \end{pmatrix}$$

is a triangular matrix, we can read off that $\det H = 1$. If we multiply A by H from the left, we obtain the block matrix $HA = (\begin{smallmatrix} X & \mathbf{0} \\ \mathbf{0} & Z \end{smallmatrix})$. By multiplicativity of the determinant this yields

$$\det\begin{pmatrix} X & \mathbf{0} \\ Y & Z \end{pmatrix} = \det(HA) = \det H \det A = \det X \det Z$$

where we used the cases from before, which establishes the claim. □

Remark V.4.5 (Laplace Expansion along the First Row): Let A in $K^{n\times n}$ be given and write A as

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ s_1 & s_2 & \cdots & s_n \end{pmatrix}$$

where $s_1, \ldots, s_n$ denote the columns of A without their respective first entries. Then

$$\begin{aligned}\det A &= \det \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 \\ s_1 & s_2 & \cdots & s_n \end{pmatrix} + \cdots + \det \begin{pmatrix} 0 & \cdots & 0 & a_{1,n} \\ s_1 & \cdots & s_{n-1} & s_n \end{pmatrix} \\ &= \det \begin{pmatrix} a_{1,1} & \mathbf{0} \\ s_1 & M_{1,1} \end{pmatrix} - \det \begin{pmatrix} a_{1,2} & \mathbf{0} \\ s_2 & M_{1,2} \end{pmatrix} \\ &\quad + \det \begin{pmatrix} a_{1,3} & \mathbf{0} \\ s_3 & A_{1,3} \end{pmatrix} + \cdots + (-1)^{n+1} \begin{pmatrix} a_{1,n} & \mathbf{0} \\ s_n & M_{1,n} \end{pmatrix} \\ &= \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} \det M_{1,j}.\end{aligned}$$

Proof (of Theorem 19): We only show (i), (ii) then follows from the behaviour of determinants under transposition.

We write A as vector of its column vectors, i.e. $A = (z_1 | \ldots | z_n)^t$. Via $(i-1)$ transpositions of rows we can achieve that the i-th row of A moves to the first position, i.e.

$$\det A = (-1)^{i-1} \det(z_i | z_1 | \ldots | z_{i-1} | z_1 | z_{i+1} | \ldots | z_n)^t.$$

The preceding remark sketches how Proposition V.4.4 is used to derive the formula for Laplace expansion from this equation. □

Chapter VI.

Inner Products and Spectral Theorems—An Outlook

In this chapter, we want to study 'vector spaces with geometry'. More precisely, we want to examine vector spaces that provide a concept of distance, length and potentially angles.

1. Euclidean and Unitary Vector Spaces

In school, you probably got acquainted with $\mathbb{R}^3$. On this special vector space, $\langle x, y\rangle = x_1y_1 + x_2y_2 + x_3y_3$ declares an 'inner product', $\|x\| = \langle x, x\rangle^{1/2}$ defines a 'norm', and $d(x, y) = \|x - y\|$ gives a concept of distance. Maybe, via the cosine law the relationship $\langle x, y\rangle = \|x\|\|y\| \cos\alpha$ was established.

In this section we want to abstract from this inner product on $\mathbb{R}^3$ to inner products on $\mathbb{R}$-vector spaces and even $\mathbb{C}$-vector spaces and study, what this entails.

Definition VI.1.1: Let V be a real vector space and let $\beta\colon V \times V \to \mathbb{R}$ be a map. If for any v_1, v_2, w from V and real numbers λ it holds

$$\beta(v_1 + \lambda v_2, w) = \beta(v_1, w) + \lambda\beta(v_2, w), \quad \beta(w, v_1 + \lambda v_2) = \beta(w, v_1) + \lambda\beta(w, v_2),$$

if for any v, w in V it holds $\beta(v, w) = \beta(w, v)$ and if for any v in V it holds $\beta(v, v) \geq 0$ with equality if and only if $v = 0$, then β is called an *inner product*. The tuple (V, β) is then called an *euclidean vector space*.

In order of appearance, the properties of β are called *bilinearity*, *symmetry* and *positive definiteness*.

As always, we will imprecisely call V a euclidean vector space if no ambiguity with regard to the inner product is to be feared.

Example VI.1.2: Let $V = \mathbb{R}^n$. If for v in V, we denote the components with respect to the standard basis by v_i, that is $v = \sum_{i=1}^n v_i e_i$, then $\langle v, w\rangle = \sum_{i=1}^n v_i w_i$ declares an inner product on V. This inner product is called *standard inner product*.

Definition VI.1.3 (Associated Norm and Metric): Let V be a euclidean vector space with inner product $\langle\cdot,\cdot\rangle$. Then $\|v\| = \langle v, v\rangle^{1/2}$ is called *norm of* v, defining a function $\|\cdot\|\colon V \to \mathbb{R}$ called *norm*.

For v and w in V, $d(v, w) = \|v - w\|$ is called *distance between* v *and* w. This defines a function $d\colon V \times V \to \mathbb{R}$, called a *metric on* V.

Proposition VI.1.4 (Properties of Norm and Metric): *Let V be a euclidean vector space with inner product $\langle\cdot,\cdot\rangle$.*

The norm $\|\cdot\|$ satisfies the following properties: For any v in V, $\|v\| \geq 0$ and equality holds if and only if $v = 0$, for any v in V and any λ in $\mathbb{R}$, $\|\lambda v\| = |\lambda|\|v\|$ and for any two v and w in V, $\|v + w\| \leq \|v\| + \|w\|$.

The metric d that is induced by $\|\cdot\|$ satisfies the following properties: For any v and w in V, $d(v, w) \geq 0$ with equality if and only if $v = w$, for any u, v and w in V it holds $d(v, w) = d(w, v)$, and for any v and w in V, $d(u, w) \leq d(u, v) + d(v, w)$.

In order of appearance, the properties of $\|\cdot\|$ are called *positive definiteness*, *homogenity* and *triangular inequality*. Similarly, the properties of d are called *positive definiteness*, *symmetry* and *triangular inequality*.

Any function $q\colon V \to \mathbb{R}_{\geq 0}$ that is positive definite, homogeneous and satisfies the triangular inequality is called a *norm*. Similarly, any function $d'\colon V \times V \to \mathbb{R}_{\geq 0}$ that is positive definite, symmetric and satisfies the triangular inequality is called a *metric*.

While a norm needs to be defined on a vector space, a metric can more generally be defined on any set—be aware that the convention that a set together with a metric is called a metric space implies no further structure.

Theorem 20 (Cauchy Schwarz Inequality): *Let V be a euclidean vector space with inner product $\langle\cdot,\cdot\rangle$. For any v and w in V it holds*

$$|\langle v, w\rangle| \leq \langle v, v\rangle\langle w, w\rangle.$$

We have equality in the above inequality if and only if v and w are linearly dependent.

Definition VI.1.5 (Angles): Let V be a euclidean vector space with inner product $\langle \cdot, \cdot \rangle$. For v and w from V it holds by Theorem 20 that

$$-1 \leq \frac{\langle v, w \rangle}{\|v\| \|w\|} \leq 1,$$

that is there is a uniquely determined α in $[0, \pi)$ such that $\cos(\alpha) = \langle v, w \rangle (\|v\| \|w\|)^{-1}$. This α is called the *angle enclosed by v and w* and gets denoted $\measuredangle(v, w)$ occasionally. If $\alpha = \pi/2$, i.e. if $\langle v, w \rangle = 0$, we call v and w *perpendicular* or *orthogonal*.

Definition VI.1.6: Let V be a euclidean vector space with inner product $\langle \cdot, \cdot \rangle$ and basis $B = \{b_1, \ldots, b_n\}$. If for any distinct indices i, j it holds $\langle b_i, b_j \rangle = 0$, then B is called an *orthogonal basis*. If furthermore $\langle b_i, b_i \rangle = 1$ for $1 \leq i \leq n$, then B is called an *orthonormal basis*.

Proposition VI.1.7 (Properties of Orthonormal Bases): *Let V be a euclidean vector space with inner product $\langle \cdot, \cdot \rangle$.*

(i) *If $B = \{b_1, \ldots, b_n\}$ is an orthonormal basis of V, then for any v in V it holds $v = \sum_{i=1}^{n} \langle v, b_i \rangle b_i$. This is called* Fouriers formula.

(ii) *If $B = \{v_1, \ldots, v_n\}$ is any basis of V, then the recursively defined vectors*

$$b_1 = v_1, \ldots, b_\ell = v_\ell - \sum_{i=1}^{\ell-1} \frac{\langle w_i, v_\ell \rangle}{\langle v_i, v_i \rangle} v_i$$

form an orthogonal basis of V. This procedure is called Gram-Schmidt process *or* Gram-Schmidt orthogonalisation.

In particular, any euclidean vector space has an orthogonal basis, and via normalisation even an orthonormal basis.

Definition VI.1.8 (Unitary Vector Space): Let V be a complex vector space and let $\beta \colon V \times V \to \mathbb{C}$ be a map. If for any v_1, v_2, w in V and complex numbers λ it holds

$$\beta(v_1 + \lambda v_2, w) = \beta(v_1, w) + \lambda \beta(v_2, w), \quad \beta(w, v_1 + \lambda v_2) = \beta(w, v_1) + \overline{\lambda} \beta(w, v_2),$$

if for any v and w in V it holds $\beta(v, w) = \overline{\beta(w, v)}$ and if for any v in V it holds $\beta(v, v) \geq 0$ with equality if and only if $v = 0$, then β is called an *inner product*. The tuple (V, β) is called a *unitary vector space*.

In order of appearance, the properties of β are called *sesquilinearity*, *Hermitianity* and *positive definiteness*.

If β is Hermitian, then the values $\beta(v,v)$ are all real numbers, as $\beta(v,v) = \overline{\beta(v,v)}$. This enables the concept of positive definiteness to make sense for Hermitian forms. The sesquilinearity is necessary to be able to induce a norm with the inner product:

$$\langle \lambda v, \lambda v\rangle = (\lambda\overline{\lambda}\langle v,v\rangle)^{1/2} = (|\lambda|^2\langle v,v\rangle)^{1/2} = |\lambda|\langle v,v\rangle^{1/2}.$$

Example VI.1.9 (Standard Inner Product): Let $V = \mathbb{C}^n$. If the coordinates of v in V with respect to the standard basis are denoted $v_1,\dots,v_n$, then $\langle v,w\rangle = \sum_{i=1}^n v_i\overline{w}_i$ declares an inner product on V. This inner product is called *standard inner product*.

2. The Gram Matrix

In the following, $\mathbb{K}$ denotes the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$. In those instances where substantial differences between the real and complex concept exist, we will specify.

Example VI.2.1 (The Universal Example): Let A be a real $n\times n$-matrix. Then $\beta(v,w) = v^tAw$ declares a bilinear form on $\mathbb{R}^n$. Similarly, if A is a complex $n\times n$-matrix, then $\beta(v,w) = v^tA\overline{w}$ declares a sesquilinear form on $\mathbb{C}^n$. Here, $\overline{w}$ means the vector whose components with respect to the standard basis are the complex conjugates to those of w.

Definition VI.2.2 (Gram Matrix): Let V be a $\mathbb{K}$-vector space, let $\beta\colon V\times V\to\mathbb{K}$ be bilinear respectively sesquilinear, and let $B = (b_1,\dots,b_n)$ be a basis of V. The matrix $G_B(\beta) = G = (g_{i,j})_{1\le i,j\le n}$ with entries $g_{i,j} = \beta(b_i,b_j)$ in $\mathbb{K}$ is called *Gram matrix* or *fundamental matrix of β with respect to B*.

Proposition VI.2.3: *Let V be a $\mathbb{K}$-vector space, let $\beta\colon V\times V\to\mathbb{K}$ be bilinear respectively sesquilinear and let $B = (b_1,\dots,b_n)$ be an ordered basis of V. Then, β is completely determined by $G_B(\beta)$, i.e. for any v and w in V it holds*

$$\beta(v,w) = \begin{cases} D_B(v)^tG_B(\beta)D_B(w), & \text{if } \mathbb{K}=\mathbb{R},\\ D_B(v)^tG_B(\beta)\overline{D_B(w)}, & \text{if } \mathbb{K}=\mathbb{C}.\end{cases}$$

If B' is another ordered basis of V, then the Gram matrix of β changes as follows under the changes of basis from B to B':

$$G_{B'}(\beta) = \begin{cases} D_{B,B'}^tG_B(\beta)D_{B,B'}, & \text{if } \mathbb{K}=\mathbb{R},\\ D_{B,B'}^tG_B(\beta)\overline{D_{B,B'}}, & \text{if } \mathbb{K}=\mathbb{C}.\end{cases}$$

The basis B is orthonormal with respect to β, i.e. $\beta(b_i, b_j) = \delta_{i,j}$, if and only if $G_B(\beta) = I_n$.

Remark VI.2.4 (Form vs. Gram Matrix): Let V be a $\mathbb{K}$-vector space with bilinear form respectively sesquilinear form $\beta\colon V \times V \to \mathbb{K}$ and let B be an ordered basis of V.

(i) If $\mathbb{K} = \mathbb{R}$, the form β is symmetric if and only if $G_B(\beta)$ is symmetric. If $\mathbb{K} = \mathbb{C}$, the form β is Hermitian if and only if $G_B(\beta) = \overline{G_B(\beta)}^t = G^*$. In this case, the matrix $G_B(\beta)$ is called Hermitian also.

(ii) The form β is positive definite if and only if for any v in $\mathbb{K}^n - \{\mathbf{0}\}$ it holds $x^t G_B(\beta)\overline{x} > 0$. In this case, the matrix is also called positive definite.

Completely analogously to 'positive definite' one defines the notions 'negative definite', 'positive semidefinite', 'negative semidefinite' and 'indefinite'. The 'semidefinite' properties replace strict inequalities with inequalities, for an indefinite form, $\beta(v, v)$ is non-negative for some v and negative for others. A form β is negative definite if and only if $-\beta$ is positive definite. Same goes for its Gram matrix.

Proposition VI.2.5 (Criteria for Positive Definiteness): *Let G in $\mathbb{K}^{n\times n}$ be a symmetric respectively Hermitian matrix.*

(i) *If G is positive definite, then* $\det G > 0$.

(ii) *The matrix G is positive definite if and only if* $\operatorname{Spec} G \subseteq \mathbb{R}_{>0}$.

(iii) *The matrix G is positive definite if and only if the leading principal minors of G are positive.*

The third criterion is known as 'Hurwitz criterion' and is included for completeness sake. The leading principal minors are the determinants of those square sub matrices that emerge from the matrix by deleting the last k rows and columns.

3. Orthogonal and Unitary Endomorphisms

Definition VI.3.1: Let V be a $\mathbb{K}$-vector space with inner product $\langle\cdot,\cdot\rangle$ and let $\phi\colon V \to V$ be linear. If for any v and w in V it holds $\langle\phi(v), \phi(w)\rangle = \langle v, w\rangle$, then ϕ is called *orthogonal* respectively *unitary* for the real respectively complex case.

Remark VI.3.2 (Geometric Properties): If ϕ is an orthogonal endomorphism, then by definition, ϕ preserves the inner product and consequently angles. Additionally, for $v = w$, the definition yields $\|\phi(v)\| = \|v\|$ for the induced norm $\|\cdot\|$, i.e. ϕ perserves lengths.

For unitary vector spaces, there is no concept of measured angle between any two vectors. There only is the concept of orthogonality. Here, ϕ just preserves inner products and consequently lengths.

If ϕ is orthogonal respectively unitary, then ϕ is injective because only the zero vector has length zero. But because of Theorem 16, this means any orthogonal respectively unitary map is invertible.

Compositions of orthogonal respectively unitary maps are orthogonal respectively unitary.

Definition VI.3.3 (Orthogonal and Unitary Matrices): A real $n \times n$-matrix A with $AA^t = I_n$ is called *orthogonal*. A complex $n \times n$-matrix with $AA^* = I_n$ is called *unitary*.

Remark VI.3.4: Let V be a $\mathbb{K}$-vector space with inner product $\langle \cdot, \cdot \rangle$ and let $B = (b_1, \ldots, b_n)$ be an ordered basis of V. An endomorphism ϕ is orthogonal respectively unitary if and only if $D_{B,B}(\phi)$ is. For this to hold true, it is necessary that B is an orthonormal basis.

Proposition VI.3.5 (Characterisation of Unitary Matrices): *Let A be a square $n \times n$-matrix with entries in $\mathbb{K}$. The following are equivalent:*

(i) *A is orthogonal respectively unitary.*

(ii) *The columns of A form and orthonormal basis of $\mathbb{K}^n$.*

(iii) *The rows of A form an orthonormal basis of $\mathbb{K}^n$.*

Definition VI.3.6: Let n be a natural number. The sets

$$O(n) = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid AA^t = I_n\}, \qquad U(n) = \{A \in \mathrm{GL}_n(\mathbb{C}) \mid AA^* = I_n\}$$

are called *orthogonal group* respectively *unitary group*.

Both of these sets really are subgroups of $\mathrm{GL}_n(\mathbb{K})$. The identity matrix lies in both of them, and multiplication of matrices corresponds to composition of endomorphisms. We also showed that orthogonal respectively unitary endomorphisms are invertible. Remains to show that their inverses are orthogonal respectively unitary. Let thus ϕ be unitary and let ψ be the corresponding inverse. Then, for any v and w in V, we have vectors v' and w' in V such that $v' = \phi(v)$, $w' = \phi(w)$ and

$$\langle \psi(v'), \psi(w') \rangle = \langle \psi(\phi(v)), \psi(\phi(w)) \rangle = \langle v, w \rangle = \langle \phi(v), \phi(w) \rangle = \langle v', w' \rangle.$$

4. The Spectral Theorem

Definition VI.4.1: Let A be a complex $n \times n$-matrix. If $AA^* = A^*A$, then A is called *normal*.

Theorem 21 (Spectral Theorem): *Let A be a complex $n \times n$-matrix. If A is normal, and if $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of A, then there is a unitary change-of-basis matrix S such that $A = S\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)S^{-1}$.*

Normal matrices are not only diagonalisable, but there even is an orthonormal basis of eigenvectors with respect to which the transformation matrix of A is diagonal.

Remark VI.4.2 (Examples for Normal Matrices): If A is orthogonal or unitary, then $AA^* = I_n = A^*A$. In particular, A is normal. If A is symmetric or Hermitian, then it even holds $A = A^*$. Such matrices are hence also normal.

Remark VI.4.3: Let A be a Hermitian matrix. Then, all eigenvalues of A are real . If namely λ is an eigenvalue of A with eigenvector v, then

$$\langle\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Av, v\rangle = \langle v, A^*v\rangle = \langle v, Av\rangle = \overline{\lambda}\langle v, v\rangle.$$

The equality $\langle Av, w\rangle = \langle v, A^*w\rangle$ can be checked explicitly in coordinates. There even is an intrinsic reason for this to be true, but we can't give this reason here.

Remark VI.4.4 (Special Cases of the Spectral Theorem): (i) If A is an Hermitian matrix, then all eigenvalues are real and there is a unitary matrix S such that $A = S\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)S^{-1}$.

(ii) If A is symmetric, then there is an orthogonal matrix S such that $A = S\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)S^{-1}$.

(iii) If A is unitary, then all eigenvalues of A have complex absolute value 1 and there is a unitary matrix S such that $A = S\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)S^{-1}$.

(iv) If A is orthogonal, then there is an orthogonal matrix S such that

$$SAS^{-1} = \begin{pmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & -1 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & -1 & & & & \\ & & & & & & \cos\theta_1 & -\sin\theta_1 & & \\ & & & & & & \sin\theta_1 & \cos\theta_1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & \cos\theta_d & -\sin\theta_d \\ & & & & & & & & \sin\theta_d & \cos\theta_d \end{pmatrix}$$