RAMP-UP FOR DATA SCIENCE – MATHEMATICAL FOUNDATIONS AND LINEAR ALGEBRA

TIMO DE WOLFF

Summer term 2024
— Version 1.0.0 —

Introduction

This is a script for the “Mathematical Foundations and Linear Algebra” part for the lecture

“*Ramp Up Course Mathematics*”,

taught at TU Braunschweig in Summer 2024. It is based on the books

- K.H. Rosen: “*Discrete Mathematics and its Applications*”, McGraw-Hill, 2012 (seventh edition),
- S.J. Leon “*Linear Algebra with Applications*”, Prentice Hall (eigth edition),
- G. Fischer “Lineare Algebra, Vieweg, 2003 (14th edition),
- S. Bosch: “*Algebra*”, Springer, 2004 (fifth edition),
- S. Lang: “*Algebra*”, Graduate Texts in Mathematics, Springer, 2004 (third edition).

This script is not officially published, but is only made available to the participants of the ramp-up course on data science. As such, it is only meant for personal use. The script is copyright protected; selling of copying of the content is prohibited.

If you recognize mistakes or parts that are unclear, please let me know in order to improve the lecture and the script.

1. Introduction to Algebra

Let us fix some general notation:

Definition 1.1.

- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ denotes the *natural*, *integral*, *rational*, *real* and *complex numbers*.
- $\mathbb{R}^n, \mathbb{C}^n, \ldots$ denotes the corresponding n-dimensional (real, complex) vector spaces.
- We use the short notation $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$ (analogously for $\mathbb{Z}, \mathbb{Q}, \ldots$).
- For all $n \in \mathbb{N}^*$ we use the notation $[n] := \{1, \ldots, n\}$ and $[n]_0 := \{0, \ldots, n\}$.
- Let M, N be sets. We denote
  - the *cardinality* of M as $\#M$.
  - the *complement* of M as M^c.
  - the *Cartesian Product* of M and N as $M \times N$.
- $\langle \cdot, \cdot \rangle$ denotes the *standard inner product* of vectors in their underlying vector space (assuming that it exists). Here, this will usually concern real vectors, i.e., $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \sum_{i=1}^n v_i w_i$.
- In this script, we follow the convention that vectors are printed boldly and have canonical length n. I.e., e.g., $\boldsymbol{x} = (x_1, \ldots, x_n)$.

⬡

Algebra is a fundamental and, at the same time, one of the oldest areas of mathematics. The word "algebra" is of Arab origin and refers to the calculation with equations (literally: "composing broken parts"). Classically, it considered indeed solving systems of equations like

$$\begin{aligned} x + y &= p \\ xy &= q \end{aligned}$$

for a given p, q. This is equivalent to solving the quadratic equation $x^2 + px + q$, which we all learned in high school (indeed, this was already investigated by Babylonians in 3000 BC and was mastered by Arab mathematicians from ca. 900 AD on). Throughout substantial parts of the Middle Ages decomposing algebraic equations into *radicals*, i.e., providing a symbolic solution for the zeros in dependence of the coefficients was an important question. In 1515, del Ferro found a solution for the cubic equation of the form $x^3 + ax = b$ with $a, b > 0$. However, he kept his solution as a secret; it was only published by Cardano in 1545. Today we know that it is impossible to decompose an algebraic equations of degree > 4 into radicals (the decomposition for degree 4 was solved by Cardano's student Ferrari in 1545). We do know, however, that every equation (every *univariate polynomial*) of degree k in one variable with complex coefficients has exactly k many solutions (counted with repetition). This theorem is called the *fundamental theorem of algebra*.

In modern times *abstract algebra* is not primarily concerned with solving abstract algebraic systems. Instead it analyzes a collection of abstract *algebraic structures*. These are given by sets with with operations (i.e., maps) combining (two) elements to a new one, which satisfy certain axioms. Prominent examples of such structures are *groups*, *rings*, *fields*, *vector spaces*, *modules* and *lattices*. Here we will in particularly talk about groups and rings with a short outlook towards fields.

On the one hand, this sounds (and is!) very abstract. On the other hand, these structures are the foundation of a collection of very concrete applications. Some examples:

(1) The necessity to solve polynomial systems of equations (in several variables) in an effective way is also in today's applications from robotics to statistics an omnipresent problem. Algebra, more specifically its subfield *algebraic geometry*, is the basis for methods which are developed further algorithmically in the context of *computational algebraic geometry* to tackle such types of problems. For further details visit our lecture "Computational Algebraic Geometry".

(2) *Polynomial optimization* concerns a very general but complicated class of optimization problems with applications in for example chemical reaction networks, theoretical computer science, and portfolio optimization. The abstract foundation for solution strategies for these problems are located in *real algebraic geometry*. For further details visit our lecture "Nonnegativity and Polynomial Optimization".

(3) *Coding theory* concerns the effective and robust transmitting and storing of data, where e.g. the channel that is used might be noisy, or the data might have to be compressed. The related area of *cryptography* concerns the questions, how messages can be exchanged such that a third party cannot learn its content even it he or she intercepts the message. Both coding theory and cryptography are build on algebraic methods in a fundamental way using groups, fields, rings, and lattices.

(4) A very classical application of algebra is geometry such as e.g. the question whether it is possible to construct certain objects with straightedge and compass. Gauß proved for example that it is possible to construct a regular polygon with 17 vertices, which had been an open problem for a long time.

(5) Finally, a crucial tool for analyzing manifolds is an investigation of their *homology* and *homotopy groups*. This has very concrete applications for example in tomography or material science. During the last years, this topic gained a lot of momentum in data science under the label *topological data analysis*. Homology and homotopy are areas in algebraic topology and thus part of algebra.

All of you have studied linear algebra as part of your previous studies, which is also an important part of the more general area of algebra: vector spaces are an algebraic structures and correspond to analyzing linear systems of equations of the form:

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

An option of a generalization would be to delete the linearity constraint. While this leads to important problems already mentioned in the context of computational algebraic geometry, this approach would still be too specialized for the context of modern algebra. Instead, our goal is to introduce abstract general structures – groups, rings, fields, etc. – which can be applied to many different instances in the mathematical universe.

Understanding the behavior of these very general instances thus gives us insight about the behavior of many specialized instances of interests in mathematics and their applications.

2. Revision of Mathematical Foundations

In this section we recall a collection of mathematical foundations.

2.1. **Logic.** We revise of some logical foundations. We start with elementary logic in the form of *propositional logic* (or *zero-th order logic*), which only consists of *elementary propositions* and *connectives*, which are used to compose more complex propositions.

Definition 2.1.

(1) A *(logical elementary) proposition* is a declarative (i.e., fact describing) sentence, which is true or false, but not both.

(2) The two values *true* and *false* that a proposition can take are called *truth values* (or *Boolean values*). We use the short notation "**T**" and "**F**".

⬡

This definition makes use the two axioms, which we highlight here for clarity:

Axiom 2.2.

(1) Every proposition is true or false.

(2) No proposition is true and false simultaneously.

We recall the connectives of propositional logic.

Definition 2.3. Let A and B logical propositions. Our zero-th order logic has the following *connectives* and rules for their applications:

(1) The *negation* of A, short: $\neg A$. The resulting statement is: "*Not A*", i.e.: "*It is not the case that A is true.*"

(2) The *conjunction* of A and B, short: $A \wedge B$. An application of the conjunction yields the statement: "*A and B.*", i.e., the proposition is true if both A and B are true.

(3) The *disjunction* of A and B, short: $A \vee B$. An application of the disjunction yields the statement: "*A or B.*" (non-exclusively), i.e., a proposition which is true if A or B or both A and B are true.

(4) The *implication* from A to B, short: $A \rightarrow B$. An application of the implication yields the proposition: "*If A (is the case), then B (is also true).*" (and if A is not true, then no constraints are given regarding the truth value of B).". I.e., the statement is true unless A is true, but B is false.

(5) The *equivalence*) of A and B, short: $A \leftrightarrow B$. It yields the statement: "A if and only if B.", i.e., a proposition which is true if both A and B are true or both A and B are false.

(6) The *exclusive disjunction* of A and B, short: $A \nleftrightarrow B$. It yields the proposition: "Either A or B.", i.e., a proposition which is true if and only if A is true and B is false or vice versa.

A logical proposition is called *compound*) if it contains connectives, and *elementary / atomic proposition* otherwise. ⬡

We summarize the truth values of propositions given by the application of connectives in the following truth tabular:

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \rightarrow B$	$A \leftrightarrow B$	$A \nleftrightarrow B$
T	**T**	**F**	**T**	**T**	**T**	**T**	**F**
T	**F**	**F**	**F**	**T**	**F**	**F**	**T**
F	**T**	**T**	**F**	**T**	**T**	**F**	**T**
F	**F**	**T**	**F**	**F**	**T**	**T**	**F**

It turns out that in propositional logic is a language which is not rich enough in expressivity for most applications. We give an example.

Example 2.4. Consider the propositions: $A :=$ "*Every natural number can be decomposed into prime factors.*" and $B :=$ "*There exist cities with animals in their coat of arms.*". In zero-th order logic both propositions are elementary (and in this case also both true). However, the statements are complex with respect to their content: A considers the set of all natural numbers, and B considers the set of all cities. In particular, A should (in our usual logical understanding of language and the world) imply the statement $C :=$ "*The number 33 is decomposable into prime factors.*". Similarly, the proposition $D :=$ "*The city of Braunschweig has a lion in its coat of arms.*" should jointly with the proposition $E :=$ "*All lions are animals.*" imply the proposition B. Propositional logic is, however, not rich enough as a language to express this. ⬡

We would thus like to extend our language in two ways:

(1) We need variables, which can be instantiated as objects.
(2) We wish to be able to control statements of the form "*It holds for all that...*" and. "*There exist...*" in an adequate way.

Definition 2.5. As stated at the beginning of the chapter, we call the union of all objects, which we talk about (here: in first order logic) our *universe* (or *domain*), and we denote it as $\mathbb{U}$. The alphabet of first order logic consists of:

(1) All connectives, which we already learned about in zero-th order logic.
(2) *Constants* $a, b, c, a_1, \dots, a_\ell$: every constant represents a particular object in the universe $\mathbb{U}$ (distinct constants may refer to the same object).
(3) *(Free) variables* $x, y, z, x_1, \dots, x_k$: They represent an "empty spot", which can refer to an arbitrary object in the universe. This reference has, however, **not** been instantiated yet.
(4) *Propositional functions*: $P(x), Q(x, y), R(x_1, \dots, x_n)$. A (propositional) function consists of one or more variables and a *predicate* P (which we can interpret as a property). The propositional function – predicate joint with variable(s) – form an *expression* – as statement, which depends on one or several objects of the universe (indicated by variables in the statements). Propositional functions, which depend on free variables, do **not** have a truth value, unless all free variables are replaced by constants or bound by quantifiers.

(5) The *universal quantifier* $\forall$ and the *existential quantifier* $\exists$.
(6) Auxiliary / technical symbols: $($, $)$, $=$, $\in$, $\models$.

⬡

2.2. **Sets.**

2.2.1. *Elementary Definitions.*

Definition 2.6. A *set* is a (non-ordered) collection of objects (in our underlying universe), which we denote as *elements* of the set. We write short: $a \in S$ (say: "a is element of S") to express that a is in the set S. Analogously, we write $a \notin S$ if a is no element of S, i.e., if $\neg(a \in S)$.

If a set S consists of finitely many elements $a_1, \ldots, a_n$, we write

$$S \;=\; \{a_1, \ldots, a_n\}.$$

Another short notation (particularly for large sets) is

$$S = \{\text{'superior set'} \;:\; \text{'condition'}\}.$$

⬡

An example for such a set is e.g.,

$$\{x \in \mathbb{N} \;:\; x \text{ is even } \wedge x \geq 150\},$$

which denotes all natural numbers, which are larger or equal than 150. Note that sets also can contain other sets as elements.

We use sets for some short notation: for all $n \in \mathbb{N}^*$ we denote $[n] := \{1, \ldots, n\}$.

Definition 2.7.

(1) Let S, T be sets. Then S and T are *equal*, short: $S = T$, if and only if the contain the same elements. Formally:
$$S = T \;\Leftrightarrow\; \forall x : x \in S \leftrightarrow x \in T.$$
(2) A set, which does not contain an element, is called the *empty set*; we denote it as $\emptyset$.

⬡

Example 2.8. We have:

$$\{1, 3, 5\} \;=\; \{3, 5, 1\} \;=\; \{1, 1, 3, 5, 5, 3, 3, 5\}.$$

⬡

The definition of equality of sets has a fundamental meaning: Sets are **solely** characterized by their elements. This fact is called *principle of extensionality*, and it is one of the axioms of the *Zermelo-Fraenkel-set theory*, which modern mathematics and logic is based on. Now, we can also give a more precise interpretation of predicates, which we encountered in first order logic: A predicate is nothing else than a set, which collects objects (or tuples of objects, which we discuss in what follows). Hence, following the definition of equality of sets, two predicates are equal (in the sense of first order logic), if the statements given by them hold for exactly the same objects of the universe.

Definition 2.9. A set P is called *subset* of a set Q, short: $P \subseteq Q$, if and only if every element of P is also an element of Q. Formally:

$$P \subseteq Q \;:=\; \forall x : x \in P \to x \in Q.$$

⬡

It follows from the definition above that the empty set is a subset of every other set. We point out that a naive (non-axiomatic) set theory leads to paradoxes like the (→ *Russel-Antinomy*), which must be avoided (and is avoided via the axioms of Zermelo-Fraenkel).

Definition 2.10. If a set S contains exactly $n \in \mathbb{N}$ elements, we call S *finite* and say, *S has cardinality n*, and we write: $\#S = n$ (or $|S| = n$). A set, which is not finite, is called *infinite*. ⬡

Definition 2.11. An *(ordered) n-tuple* $(a_1, \dots, a_n)$ is an ordered collection of n elements, such that a_1 is the first, a_2 is the second element, etc.. A 2-tuple is called a *pair*. Formally, we can define 2-tuples via sets as follows:

$$(a, b) \;:=\; \{\{a\}, \{a, b\}\};$$

analogously for n-tuples. ⬡

Remark 2.12. *Note:*

1. *We have in general $(a, b) \neq (b, a)$ (in contrast to $\{a, b\} = \{b, a\}$), and $(a, a) \neq (a)$.*
2. *In the context of programming languages (or general data structures)an n-tuple is the equivalent of a list or a (on dimensional) array respectively.*

We introduce Cartesian products.

Definition 2.13. Let $A_1, \dots, A_n$ be sets. We define the *Cartesian product* as:

$$\bigtimes_{i=1}^{n} A_i \;:=\; A_1 \times \cdots \times A_n \;:=\; \{(a_1, \dots, a_n) \;:\; \forall i \in [n] : a_i \in A_i\}.$$

⬡

The Cartesian product of the sets $A_1, \dots, A_n$ thus is a set itself, namely the set of all n-tuples, for which the i-th element belongs to the set A_i.

2.3. **Set operations.** Sets can be combined in various ways. For the rest of this section let A, B, C be arbitrary sets.

Definition 2.14.

1. The *union* of the sets A, B, short: $A \cup B$, is the set of all elements, which are contained in A or in B or both. Formally:

$$A \cup B \;:=\; \{x \;:\; x \in A \vee x \in B\}.$$

2. The *intersection* of the sets A, B, short: $A \cap B$, is the set of all elements, which are contained in A and in B. Formally:

$$A \cap B \;:=\; \{x \;:\; x \in A \wedge x \in B\}.$$

(3) The *difference* of A and B, short: $A \setminus B$ is the set of all elements, which are contained in A, but not in B. Formally:

$$A \setminus B \;:=\; \{x \;:\; x \in A \wedge x \notin B\}.$$

⬡

Note: $A \setminus B \neq B \setminus A$ in general.

Example 2.15. Let $A := \{1, 3, 4\}$ and $B := \{2, 3, 5\}$. Then we have:

$$\begin{array}{rclrcl} A \cap B &=& \{3\} & A \cup B &=& \{1, 2, 3, 4, 5\} \\ A \setminus B &=& \{1, 4\} & B \setminus A &=& \{2, 5\}. \end{array}$$

⬡

Definition 2.16. Let $\mathbb{U}$ be our universe. We we call $A^c := \mathbb{U} \setminus A$ the *complement* of A. ⬡

We give a series of identities for sets, which will partially be tackled in the homework.

Identities	$A \cap \mathbb{U} = A, \qquad A \cup \emptyset = A$
Dominance	$A \cup \mathbb{U} = \mathbb{U}, \qquad A \cap \emptyset = \emptyset$
Further identities	$A \cup A = A, \qquad A \cap A = A$
Complements	$(A^c)^c = A, \qquad A \cup A^c = \mathbb{U}, \qquad A \cap A^c = \emptyset$
Commutativity	$A \cup B = B \cup A, \qquad A \cap B = B \cap A$
Associativity	$A \cup (B \cup C) = (A \cup B) \cup C,$
	$A \cap (B \cap C) = (A \cap B) \cap C$
Distributivity	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$
	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
de Morgan's laws	$(A \cap B)^c = A^c \cup B^c, \qquad (A \cup B)^c = A^c \cap B^c$
Absorbation	$A \cup (A \cap B) = A. \qquad A \cap (A \cup B) = A$

Table 1. Identities for sets.

For intersections and unions of several sets $A_1, \ldots, A_n, \ldots$ we use the following simplified notation:

$$\bigcup_{i=1}^{n} A_i \;:=\; A_1 \cup \cdots \cup A_n \qquad \bigcap_{i=1}^{n} A_i \;:=\; A_1 \cap \cdots \cap A_n.$$

In many applications it is useful to consider infinite unions and intersections. Let I be an arbitrary index set. We denote:

$$\bigcup_{i=1}^{\infty} A_i \qquad \text{and} \qquad \bigcup_{i \in I} A_i \text{ respectively.}$$

2.4. **Functions.** The concept of a *function* is, next to sets, likely the most fundamental one in mathematics (and in computer science). A function is a map of elements from one set to the elements of another set, such that a couple of conditions are satisfied.

Definition 2.17. Let A, B be two sets. A *function* or *map* from A to B is an assignment of elements in A to the elements in B such that **every** element in A is assigned to **exactly one** element in B. We denote functions usually with lowercase letters. If f is a function, which maps from A to B, we write $f : A \to B$.

For a function $f : A \to B$ we call the set A the *domain* and B *co-domain* of f. If $x \in A$, we call $y := f(x) \in B$ the *image* of x and we call x the *preimage* of y. We denote the set of all images of elements in A as the *image* or *range* of f; short: $\mathrm{im}(f)$ or $f(A)$ (note that $\mathrm{im}(f) \subseteq B$). We say f *maps A to B*. If $C \subseteq A$ is a subset of A, we define

$$f(C) \;:=\; \{y \in B \;:\; \exists x \in C : f(x) = y\}.$$

For a given function $f : A \to B$ we define the *graph* as

$$\mathrm{graph}(f) \;:=\; \{(a, b) \;:\; a \in A, b = f(a) \in \mathrm{im}(f) \subseteq B, \}$$

⬡

Definition 2.18. Let $f : A \to B$ be a function. Then we call f:

(1) *injective* or *one-to-one*, if for all $x_1, x_2 \in A$ we have: $f(x_1) = f(x_2)$ implies $x_1 = x_2$.
(2) *surjective* or *onto*, if for all $y \in B$ there exists an $x \in A$, such that $f(x) = y$. In other words: if $\mathrm{im}(f) = B$.
(3) *bijective*, if f is both injective and surjective.

⬡

Finally we introduce inverse functions and compositions of functions.

Definition 2.19. Let $f : A \to B$ be a bijection. Then we define the *inverse function* as

$$f^{-1} : B \to A, \text{ such that } \forall x \in A, y \in B : f^{-1}(y) = x \Leftrightarrow f(x) = y.$$

⬡

Definition 2.20. Let $f : A \to B$ and $g : B \to C$ be functions. Then we define the *composition* $g \circ f : A \to C$ (say: “g after f) by defining for all $x \in A$:

$$(g \circ f)(x) \;:=\; g(f(x)).$$

⬡

3. Relations

3.1. **Introduction of Relations and Equivalence Relations.** We introduce *relations*, which can be considered a generalization of functions:

Definition 3.1. Let $A, B, A_1, \ldots, A_n$ be sets. Then

(1) a *binary relation* from A to B is a subset of $A \times B$,
(2) an *n-ary relation* of $A_1, \ldots, A_n$ is a subset of $A_1 \times A_2 \times \cdots \times A_n$,
(3) a *relation on A* is a subsets of $A \times A$.

If $R \subseteq A \times B$ is a relation, one also write *aRb* for $(a, b) \in R$. We say *a is related to b (by R)*. ○

Every function $f : A \to B$ is a relation from A to B, where the relation is given by $\text{graph}(f) \subseteq A \times B$. The inverse is not the case, since a relation $R \subseteq A \times B$ allows that for $a \in A$ and $b_1, b_2 \in B$ we have $(a, b_1), (a, b_2) \in R$. Moreover, it is not requested that for all $a \in A$ there exists a $b \in B$ such that $(a, b) \in R$. That means, we drop exactly those restrictions for relations, which we required for functions.

We discuss some further properties of relations:

Definition 3.2. Let R be a relation on A. Then R is called

1. *reflexive*, if for all $a \in A$ it holds that: $(a, a) \in R$.
2. *symmetric*, if for all $a, b \in A$ it holds that : $(a, b) \in R \to (b, a) \in R$.
3. *anti-symmetric*, if for all $a, b \in A$ it holds that: $(a, b) \in R \wedge (b, a) \in R \to a = b$.
4. *asymmetric*, if for all $a, b \in A$ it holds that: $(a, b) \in R \to (b, a) \notin R$.
5. *transitive*, if for all $a, b, c \in A$ it holds that: $(a, b) \in R \wedge (b, c) \in R \to (a, c) \in R$.

○

We introduce some important instances of relations:

Definition 3.3. A relation R on a set A is called and *equivalence relation*, if it is reflexive, symmetric, and transitive. In this case we say that for $(a, b) \in R \subseteq A \times A$ that *a is equivalent to b (w.r.t. R)*). It is common to denote an equivalence relation as $\sim$ instead of R. We denote briefly *$a \sim b$* instead of $(a, b) \in\sim$. ○

We already know a couple of structures, which yield equivalence relations. Let us consider a few examples:

Example 3.4.

1. Every set A has two trivial equivalence relations:
   (a) $\sim_1$: $\forall a, b \in A : a \sim_1 b :\Leftrightarrow a = b$,
   (b) $\sim_2$: $\forall a, b \in A : a \sim_2 b$.
2. On the set $\mathbb{R}$ of the real numbers the floor function $\sim_3$ induces an equivalence relation via
$$\forall a, b \in R : a \sim_3 b :\Leftrightarrow \lfloor a \rfloor = \lfloor b \rfloor.$$
3. The absolute value function $|\cdot| : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is an equivalence relation $\sim_4$ on $\mathbb{R}$ given by
$$\forall a, b \in R : a \sim_4 b :\Leftrightarrow |a| = |b|.$$
4. The rational numbers are given by an equivalence relation, which we discuss in detail in Definition 3.12 and Lemma 3.13.

○

Definition 3.5. Let $\sim$ be an arbitrary equivalence relation on A and let $a \in A$ be arbitrary. The we call the $[a]_\sim$ of all $b \in A$, which are equivalent to a (w.r.t. $\sim$) the *equivalence class* *(of a)*. I.e.,

$$[a]_\sim \ := \ \{b \in A \ : \ a \sim b\}.$$

The element a is called the *representative* of the equivalence class $[a]_\sim$ (where every element $b \in [a]_\sim$ can chosen as representative of $[a]_\sim$). ⬡

Proposition 3.6. *Let $\sim$ be an equivalence relation on a set A. The for all $a, b \in A$ the following statements are equivalent:*

(1) $a \sim b$,
(2) $[a]_\sim = [b]_\sim$,
(3) $[a]_\sim \cap [b]_\sim \neq \emptyset$.

We leave the proof for the exercises.

Definition 3.7. Let A be an arbitrary set and let $A_1, \ldots, A_r \subseteq A$ be subsets of A with $A_i \neq \emptyset$ for all $i \in [r]$. The we call $A_1, \ldots, A_r$ a *partition* of A if:

(1) $\bigcup_{j=1}^r A_j = A$ and
(2) $A_i \cap A_j = \emptyset$ for all $i, j \in [r]$ with $i \neq j$.

⬡

Remark 3.8. *The definition of a partition can be extended to an infinite set of (non-empty) subsets $A_i \subseteq A$, $i \in I$ for an index set I. The previous conditions hold analogously.*

Now, we relate equivalence classes and partitions.

Corollary 3.9. *Every equivalence relation $\sim$ on a set A induces a partition on A via the equivalence classes of $\sim$.*

Proof. Follows immediately from Proposition 3.6. □

Example 3.10. Let $A := [8] \cup \{0\}$. Then $\{0, 3, 6\}, \{1, 4, 7\}, \{2, 5, 8\}$ is a partition of A. It is given by calculation modulo 3 on A. That means, elements on A belong to the same equivalence class if and only if division modulo 3 yields the same rest. ⬡

It is also the case that every partition on A induces an equivalence relation on A.

Proposition 3.11. *Let A be a set and let $(A_i)_{i \in I}$ be a partition on A. The $(A_i)_{i \in I}$ is an equivalence relation $\sim$ on A given via*

$$\forall a, b \in A : a \sim b \ :\Leftrightarrow \ \exists i \in I : a \in A_i \wedge b \in A_i.$$

We leave the proof as exercise.

As a last step in this subsection, we give a rigorous definition of the rational numbers $\mathbb{Q}$. A naive definition of the rational numbers as set of all fractions is not accurate, since it omits that we want to consider two rational numbers as equal, if they satisfy

$$\frac{a}{b} \ = \ \frac{k \cdot a}{k \cdot b},$$

for $a, k \in \mathbb{Z}$ and $b \in \mathbb{Z}^*$. So far, we had no possibility, to define a set satisfying this property, but now we can solve this issue using equivalence relations.

Definition 3.12. As a first step, we define

$$M \ := \ \{(a,b) \ : \ a \in \mathbb{Z}, b \in \mathbb{Z}^*\},$$

and we define an equivalence relation $\sim$ on M via

$$(a,b) \sim (a',b') \ :\Leftrightarrow \ ab' \ = \ a'b.$$

We define the *rational numbers* as

$$\mathbb{Q} \ := \ M/\sim,$$

i.e. as the set of the equivalence classes on M w.r.t. $\sim$. We denote for every $(a,b) \in M$ the corresponding equivalence class in $\mathbb{Q}$ as $\frac{a}{b}$ or a/b. The corresponding addition and multiplication on $\mathbb{Q}$ is given by:

$$\frac{a}{b} \,\widetilde{+}\, \frac{a'}{b'} \ = \ \frac{ab'+a'b}{bb'} \quad \text{and} \quad \frac{a}{b} \,\widetilde{\cdot}\, \frac{a'}{b'} \ = \ \frac{aa'}{bb'}.$$

If the context is clear, we briefly write "+" and "·" for addition and multiplication on the natural numbers. ⬡

We need to ensure that the relation in the upper definition is well-defined.

Lemma 3.13. *The relation $\sim$ in the sense of Definition 3.12 is indeed an equivalence relation.*

Again, we leave this proof as an exercise.

3.2. **Orderings.** A second, central class of relations, which we want to recall, are orderings.

Definition 3.14. Let A be a set and R be a elation on A. We call R a *partial ordering*, if R is reflexive, anti-symmetric and transitive. We denote a partial ordering often via $\leq$, $\preceq$ or sometimes with $\subseteq$. A set A joint with a partial ordering is called a *partially ordered set*, or briefly *poset*. Two elements $a, b \in (A, \leq)$ in a partially ordered set are called *comparable*, if $a \leq b$ or $b \leq a$ and *incomparable* otherwise. ⬡

We introduce a couple of properties of posets.

Definition 3.15. Let $(A, \leq)$ be a poset.

(1) The ordering $\leq$ on A is called *total* (or *linear*), if the set of all pairs of elements $a, b \in A$ is comparable via $\leq$. A (via $\leq$) totally ordered subset $A' \subseteq A$ is called a *chain*) (in A w.r.t. $\leq$).
(2) Let $a \in A$ such that for all $b \in A$ it holds that : $b \leq a$. The element a is called *maximal (element)* in A (w.r.t. $\leq$). *Minimal elements* are defined analogously.
(3) Let $A' \subseteq A$ be a subset of A. Then $a \in A$ is an *upper bound* of A', if for all $b \in A'$ it holds that: $b \leq a$. *Lower bounds* are defined respectively. Note that we de **not** require $a \in A'$!
(4) Let a be an upper bound of $A' \subseteq A$ and moreover $a \leq b$ for all other upper bounds $b \in A$ of A'. Then A is called the *least upper bound* or *supremum* or *Join* of A. Analogously, we define the *greatest lower bound* or *infimum* or *Meet*.

⬡

We extend the notion of bounds to define another important structure on sets.

Definition 3.16. Let $(A, \leq)$ be a partially ordered set. Then $(A, \leq)$ is called a *lattice*, if for all $a, b \in A$ Meet and Join of a and b exist. For these, we write briefly $a \sqcap b$ and $a \sqcup b$ respectively.

A lattice $(A, \leq)$ is called *complete* if for every subset $X \subseteq A$ the Join of X, i.e., $\bigsqcup X$, and the Meet X, i.e., $\bigsqcap X$, exist. ⬡

4. Revision of Linear Algebra

4.1. **Systems of Linear Equations and Determinants.** Linear algebra takes its starting point at solving systems of linear equations.

We can solve such systems effectively using the *Gauß-Algorithm*. Let us recall its application in an example

Example 4.1. Let us consider the following system of linear equations:

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 3 \\ 3x_1 - x_2 - 3x_3 &= -1 \\ 2x_1 + 3x_2 + x_3 &= 4 \end{aligned}$$

We translate it into matrix notation and perform row operations to bring it in row-echelon-form:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -1 \\ 2 & 3 & 1 & 4 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & -1 & -1 & -2 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & 0 & -1/7 & -4/7 \end{array}\right]$$

⬡

We recall that every elementary row operation can be performed by multiplying the original matrix with a specific elementary matrix. Let in what follows

$$I_n \ := \begin{bmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Type I: A matrix by exchanging two rows of I_n.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Type II: A matrix obtained by multiplying a row of I_n with a nonzero scalar a (in the field of choice – e.g., $\mathbb{R}, \mathbb{C}$).

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & a & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Type III: A matrix obtained from I_n by adding a multiple of one row to another.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Theorem 4.2. *If E is an elementary matrix, then E is nonsingular and its inverse matrix E^{-1} is an elementary matrix of the same type.*

We recall the definition of a determinant of a matrix.

Definition 4.3. The *determinant* of an $n \times n$ matrix A is the real number

$$\det(A) \ := \ \begin{cases} a_{11} & \text{if } n = 1 \\ \sum_{j=1}^{n} a_{1j} A_{1j} & \text{for } n > 1. \end{cases}$$

where $A_{ij} := (-1)^{i+j} \det(M_{ij})$ for all i, j, and M_{ij} is obtained from A by deleting its i-th row and j-th column. $\det(M_{ij}$ is called a *minor* of A and A_{ij} is called a *cofactor* of a_{ij}. ⬡

Note that a matrix is invertible if and only if its determinant is not zero.

4.2. **Vector Spaces.** Usually, we think in terms of linear algebra about the vector space $\mathbb{R}^2$ or $\mathbb{R}^3$. We can "draw" vectors in this space using the standard basis.

This is, however, not the general axiomatic definition of a vector space, which allows us to speak about vectors and matrices more abstractly. We recall this general definition now.

Definition 4.4. Let $\mathbb{K}$ be a field[1], V be a set and $+ : V \times V \to V, \cdot : \mathbb{K} \times V \to V$ be operators (i.e., maps). Then we call V joint with the two operators (called *vector addition* and *scalar multiplication*), short $(V, +, \cdot)$, a *vector space* if it satisfies the following conditions:

(1) $(V, +)$ is a commutative group with neutral element 0. That means, it holds for all $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u} \in V$ that

$$\boldsymbol{v} + \boldsymbol{w} \ = \ \boldsymbol{w} + \boldsymbol{v} \qquad (\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{u} \ = \ \boldsymbol{v} + (\boldsymbol{w} + \boldsymbol{u}) \qquad \boldsymbol{0} + \boldsymbol{v} \ = \ \boldsymbol{v} + \boldsymbol{0} \ = \ \boldsymbol{v},$$

[1] If you do not know what this is, just think about the real numbers $\mathbb{R}$ for the moment.

and there exists an element $-\boldsymbol{v} \in V$ with

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

(2) For all $\boldsymbol{v}, \boldsymbol{w} \in V$ and $\lambda, \mu \in \mathbb{K}$ it holds that

$$\lambda \cdot (\boldsymbol{v} + \boldsymbol{w}) \;=\; \lambda \cdot \boldsymbol{v} + \lambda \cdot \boldsymbol{w} \qquad (\lambda + \mu) \cdot \boldsymbol{v} \;=\; \lambda \cdot \boldsymbol{v} + \mu \cdot \boldsymbol{v}$$
$$(\lambda\mu) \cdot \boldsymbol{v} \;=\; \lambda \cdot (\mu \cdot \boldsymbol{v}) \qquad 1 \cdot \boldsymbol{v} \;=\; \boldsymbol{v}.$$

The elements of in V are called *vectors*, the elements in $\mathbb{K}$ are called *scalars*. ⬡

Example 4.5. (1) $\{\mathbf{0}\}$ is a $\mathbb{K}$-vector space with the operations $\mathbf{0}+\mathbf{0} = \mathbf{0}$ and $\lambda \cdot \mathbf{0} = \mathbf{0}$ for alle $\lambda \in \mathbb{K}$.

(2) $\mathbb{R}^n$ is an $\mathbb{R}$-vector space with the usual componentwise addition and scalar multiplication.

(3) The set $\mathrm{Mat}(m \times n; \mathbb{R})$ of (real) $m \times n$ matrices with matrix addition and componentwise scalar multiplication is a vector space.

(4) Let $a, b \in \mathbb{R}$. The space $\mathcal{C}^0[a,b]$ of all real valued continuous functions $f : [a,b] \to \mathbb{R}$ on $[a,b]$ is a vector space with the operators

$$(f+g)(x) \;:=\; f(x) + g(x) \qquad (\lambda \cdot f)(x) \;:=\; \lambda \cdot f(x).$$

(5) The set $\mathbb{R}[x]_d$ of all real univariate polynomials with degree at most d is a vector space (exercise).

⬡

Definition 4.6. Let $(V, +, \cdot)$ be a $\mathbb{K}$-vector space and $H \subseteq V$ a subset. Then $(H, +, \cdot)$ is a *subspace* of V (i.e., $(V, +, \cdot)$), if $(H, +, \cdot)$ is a vector space itself. That means

(1) If $\boldsymbol{v} \in H$ and $\lambda \in \mathbb{K}$, then $\lambda \cdot \boldsymbol{v} \in H$

(2) If $\boldsymbol{v}, \boldsymbol{w} \in H$, then $\boldsymbol{v} + \boldsymbol{w} \in V$.

⬡

Definition 4.7. Let A be a $m \times n$ matrix with entries in $\mathbb{K}$. Let $N(A)$ denote the set of all solutions of the system of linear equations given by $A \cdot \boldsymbol{x} = \mathbf{0}$, i.e.,

$$N(A) \;:=\; \{\boldsymbol{x} \in \mathbb{K}^n \;:\; A \cdot \boldsymbol{x} = \mathbf{0}\}.$$

The set $N(A)$ is called the *nullspace* or the *kernel* of A. ⬡

Proposition 4.8. *Let $A \in \mathrm{Mat}(m \times n; \mathbb{K})$. Then $N(A)$ is a subspace of $\mathbb{K}^n$.*

We leave the proof as an exercise.

Definition 4.9. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ be vectors in a $\mathbb{K}$-vector space V, and let $\lambda_1, \ldots, \lambda \in \mathbb{K}$. A sum of the form

$$\lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_n \boldsymbol{v}_n$$

is called a *linear combination* of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Then set of all possible linear combinations of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is called the *span* of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, i.e.,

$$\mathrm{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) \;:=\; \{\boldsymbol{w} \in V \;:\; \boldsymbol{w} = \lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_n \boldsymbol{v}_n \text{ for some } \lambda_1, \ldots, \lambda \in \mathbb{K}\}.$$

⬡

Theorem 4.10. *Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in V$ be vectors in a $\mathbb{K}$-vector space V. Then $\operatorname{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ is a subspace of V.*

Given this theorem, it makes sense to ask, which vectors are needed to generate an entire vector space.

Definition 4.11. Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in V$ be vectors in a $\mathbb{K}$-vector space V. Then the set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is called a *spanning set* or *generating set* of V, if $\operatorname{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n) = V$. ⬡

Generating sets of vector spaces are not unique. This is easy to see, since if G is a generating set of a vector space V and $\boldsymbol{v} \in V \setminus G$, then $G \cup \{\boldsymbol{v}\}$ still is a generating set. Thus, a natural question is how many vectors we need in a spanning set for V.

Proposition 4.12. *Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in V$ be vectors in a $\mathbb{K}$-vector space V such that $\operatorname{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n) = V$. Let $j \in [n]$. Then the set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \setminus \{\boldsymbol{v}_j\}$ is still a spanning set of V if one of the following conditions is satisfied:*

1. *One of the $\boldsymbol{v}_i \in \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is a linear combination of $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \setminus \{\boldsymbol{v}_i\}$.*
2. *We have $\mathbf{0} \in \operatorname{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$.*

Following this proposition, we give the following definition.

Definition 4.13. Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in V$ be vectors in a $\mathbb{K}$-vector space V. Then $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are called *linear independent* if the only possible choice of λ_i with

$$\lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_n \boldsymbol{v}_n \;=\; \mathbf{0}$$

is $\lambda_1, \dots, \lambda_n = 0$. $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are called *linear dependent* otherwise. ⬡

Now, we see that the two definitions of spanning sets and linear independence come together in a natural way.

Theorem 4.14. *Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in V$ be vectors in a $\mathbb{K}$-vector space V. Let $\boldsymbol{w} \in \operatorname{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$. Then $\boldsymbol{w}$ can be written as a linear combination of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ **uniquely** if and only if $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are linear independent.*

This theorem motivates the following definition.

Definition 4.15. Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in V$ be vectors in a $\mathbb{K}$-vector space V. Then $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is called a *basis*, if

1. $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is a spanning set of V, and
2. $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are linear independent.

⬡

A quite straightforward consequence based on this definition is the following theorem.

Theorem 4.16. *Let $S := \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \subseteq V$ be a spanning set in a $\mathbb{K}$-vector space V. Then every collection of vectors $\boldsymbol{w}_1, \dots, \boldsymbol{w}_m$ with $m > n$ is linear dependent. In particular, if S is a basis, then **every** basis in V consists of exactly n vectors.*

This theorem motivates the definition of the dimension of a vector space.

Definition 4.17. Let V be a vector space and let B be a basis of V such that B consists of n elements. Then we say that V has *dimension* n. We say that the vector space $\{\mathbf{0}\}$ is of dimension 0. If there is an infinite set that spans V, then we call V *infinite dimensional*. ⬡

We now apply the notion of dimensions to matrices.

Definition 4.18. Let A be an $m \times n$ matrix. The subspace of $\mathbb{R}^m$ spaned by the rows of A is called the *rowspace* of A. Analogously, the subspace of $\mathbb{R}^n$ spaned by the columns of A is called the *columnspace* of A.

The dimension of the rowspace of A is called the *rank* of A, short: $\operatorname{rank}(A)$. ⬡

Theorem 4.19 (Rank-Nullity-Theorem)**.** *Let A be an $m \times n$ matrix. Then we have*

$$\operatorname{rank}(A) + \dim(N(A)) \;=\; n.$$

Indeed, we could equivalently work with the columnspace due to the following theorem.

Theorem 4.20. *Let A be an $m \times n$ matrix. Then the dimension of its rowspace equals the dimension of its column space.*

Finally, we take a brief look at linear maps between vector spaces.

Definition 4.21. Let $(V, +, \cdot)$ and $(W, \oplus, \odot)$ be $\mathbb{K}$-vector spaces and let $L : V \to W$ be a map. Then L is called a *linear transformation* or just *linear*, if for all $\boldsymbol{v}, \boldsymbol{w} \in V$ and $\lambda, \mu \in \mathbb{K}$ it holds that

$$L(\lambda \boldsymbol{v} + \mu \boldsymbol{w}) \;=\; \lambda \odot L(\boldsymbol{v}) \oplus \mu \odot L(\boldsymbol{w}).$$

The *kernel* of L is defined as

$$\ker(L) \;:=\; \{\boldsymbol{v} \in V \;:\; L(\boldsymbol{v}) = \mathbf{0}\}.$$

Let S be a subspace of V. Then the *image* of S w.r.t. L is defined as

$$L(S) \;:=\; \{\boldsymbol{w} \in W \;:\; \exists \boldsymbol{v} \in V : L(\boldsymbol{v}) = \boldsymbol{w}\}.$$

The image of V itself, i.e., $L(V)$ is also called the *range* of L. ⬡

Timo de Wolff, Technische Universität Braunschweig, Institut für Analysis und Algebra, AG Algebra, Universitätsplatz 2, 38106 Braunschweig, Germany

Email address: `t.de-wolff@tu-braunschweig.de`