Appendix to Lecture Notes on *Calculus II for Statistics Students*

Peter Philip

May 18, 2015

Contents

A Linear Algebra

A.1 Vector Spaces

In [Phi15a], we encountered the abstract definition of a field in [Phi15a, Def. 4.4], and we studied the fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. Even though we will formulate the following definitions and results using a general field F as defined in [Phi15a, Def. 4.4], for the purposes of the present lecture, you may always think of F as being the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$.

Definition A.1. Let F be a field and let V be a nonempty set with two maps

$$\begin{aligned} +&: V \times V \longrightarrow V, \quad (x, y) \mapsto x + y, \\ \cdot&: F \times V \longrightarrow V, \quad (\lambda, x) \mapsto \lambda \cdot x \end{aligned} \tag{A.1}$$

($+$ is called *(vector) addition* and $\cdot$ is called *scalar multiplication*; often one writes xy instead of $x \cdot y$ – please take care not to confuse the vector addition on V with the addition on F and, likewise, not to confuse the scalar multiplication with the multiplication on F, the symbol $+$ is used for both additions and $\cdot$ is used for both multiplications, but you can always determine from the context which addition or multiplication is meant). Then V is called a *vector space* or a *linear space* over F (sometimes also an F-vector space) if, and only if, the following conditions are satisfied:

(i) V is a commutative group with respect to $+$. The neutral element with respect to $+$ is denoted 0 (do not confuse $0 \in V$ with $0 \in F$ – once again, the same symbol is used for different objects (both objects only coincide for $F = V$)).

(ii) Distributivity:

$$\forall_{\lambda \in F} \quad \forall_{x,y \in V} \quad \lambda(x + y) = \lambda x + \lambda y, \tag{A.2a}$$

$$\forall_{\lambda,\mu \in F} \quad \forall_{x \in V} \quad (\lambda + \mu)x = \lambda x + \mu x. \tag{A.2b}$$

(iii) Compatibility between Multiplication on F and Scalar Multiplication:

$$\forall_{\lambda,\mu \in F} \quad \forall_{x \in V} \quad (\lambda\mu)x = \lambda(\mu x). \tag{A.3}$$

(iv) The neutral element with respect to the multiplication on F is also neutral with respect to the scalar multiplication:

$$\forall_{x \in V} \quad 1x = x. \tag{A.4}$$

If V is a vector space over F, then one calls the elements of V *vectors* and the elements of F *scalars*.

Example A.2. (a) Every field F is a vector space over itself if one uses the field addition in F as the vector addition and the field multiplication in F as the scalar multiplication (as important special cases, we obtain that $\mathbb{R}$ is a vector space over $\mathbb{R}$ and $\mathbb{C}$ is a vector space over $\mathbb{C}$): All the vector space laws are immediate consequences of the corresponding field laws: Def. A.1(i) holds as every field is a commutative group with respect to addition; Def. A.1(ii) follows from the field distributivity [Phi15a, Def. (4.5)] and multiplicative commutativity on F; Def. A.1(iii) is merely the multiplicative associativity on F; and Def. A.1(iv) holds, since scalar multiplication coincides with field multiplication on F.

(b) The reasoning in (a) actually shows that every field F is a vector space over every subfield E of F (over every $E \subseteq F$ that is a field with respect to the addition and multiplication defined on F). In particular, $\mathbb{R}$ is a vector space over $\mathbb{Q}$.

(c) If A is any nonempty set, F is a field, and Y is a vector space over the field F, then we can make $V := \mathcal{F}(A, Y) = Y^A$ (the set of functions from A into Y) into a vector space over F by defining for each $f, g : A \longrightarrow Y$:

$$(f+g) : A \longrightarrow Y, \qquad (f+g)(x) := f(x) + g(x), \tag{A.5a}$$

$$(\lambda \cdot f) : A \longrightarrow Y, \qquad (\lambda \cdot f)(x) := \lambda \cdot f(x) \quad \text{for each } \lambda \in F \tag{A.5b}$$

(note that, for $Y = F = \mathbb{K}$, the above definitions are the same as the ones in [Phi15a, (6.1a)] and [Phi15a, (6.1b)], respectively).

It is an exercise to verify that $(V, +, \cdot)$ is, indeed, a vector space over F.

(d) For each $n \in \mathbb{N}$, $(\mathbb{K}^n, +, \cdot)$, with vector addition and scalar multiplication as defined in (1.1a) and (1.1c), respectively, constitutes a vector space over $\mathbb{K}$. The validity of Def. A.1(i) – Def. A.1(iv) can easily be verified directly, but $(\mathbb{K}^n, +, \cdot)$ can also be seen as a special case of (c) with $A = \{1, \ldots, n\}$ and $Y = F = \mathbb{K}$. To this end, recall that, according to [Phi15a, Ex. 2.15(c)], $\mathbb{K}^n = \mathbb{K}^{\{1,\ldots,n\}} = \mathcal{F}\big(\{1, \ldots, n\}, \mathbb{K}\big)$ is the set of functions from $\{1, \ldots, n\}$ into $\mathbb{K}$. Then $z = (z_1, \ldots, z_n) \in \mathbb{K}^n$ is the same as the function $f : \{1, \ldots, n\} \longrightarrow \mathbb{K}$, $f(j) = z_j$. Thus, (1.1a) is, indeed, the same as (A.5a), and (1.1c) is, indeed, the same as (A.5b).

Definition and Remark A.3. Let $(V, +, \cdot)$ be a vector space over the field F. A subset $U \subseteq V$ is called a *subspace* of V if, and only if, U is a vector space over F with respect to operations $+$ and $\cdot$ it inherits from V. Clearly, every law (commutativity, associativity, etc.) that holds on V must, in particular, hold on U, showing that $\emptyset \neq U \subseteq V$ is a subspace of V if, and only if,

$$\forall_{x,y \in U} \quad x + y \in U, \tag{A.6a}$$

$$\wedge \quad \forall_{\lambda \in F} \quad \forall_{x \in U} \quad \lambda x \in U. \tag{A.6b}$$

which holds if, and only if,

$$\forall_{\lambda,\mu \in F} \quad \forall_{x,y \in U} \quad \lambda x + \mu y \in U. \tag{A.7}$$

Example A.4. (a) $\mathbb{Q}$ is *not* a subspace of $\mathbb{R}$ if $\mathbb{R}$ is considered as a vector space over $\mathbb{R}$ (for example, $\sqrt{2} \cdot 2 \notin \mathbb{Q}$). However, $\mathbb{Q}$ *is* a subspace of $\mathbb{R}$ if $\mathbb{R}$ is considered as a vector space over $\mathbb{Q}$.

(b) From Ex. A.2(c), we know that, for each $\emptyset \neq A$, $\mathcal{F}(A, \mathbb{K})$ constitutes a vector space over $\mathbb{K}$. Thus, as a consequence of Def. and Rem. A.3, a subset of $\mathcal{F}(A, \mathbb{K})$ is a vector space over $\mathbb{K}$ if, and only if, it is closed under addition and scalar multiplication. By using results from [Phi15a], we obtain the following examples:

(i) The set $\mathcal{P}(\mathbb{K})$ of all polynomials mapping from $\mathbb{K}$ into $\mathbb{K}$ is a vector space over $\mathbb{K}$ by [Phi15a, Rem. 6.4]; for each $n \in \mathbb{N}_0$, the set $\mathcal{P}_n(\mathbb{K})$ of all such polynomials of degree at most n is also a vector space over $\mathbb{K}$ by [Phi15a, Rem. (6.4a),(6.4b)].

(ii) If $\emptyset \neq M \subseteq \mathbb{C}$, then the set of continuous functions from M into $\mathbb{K}$, i.e. $C(M, \mathbb{K})$, is a vector space over $\mathbb{K}$ by [Phi15a, Th. 7.38].

(iii) If $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ and $a < b$, then the set of differentiable functions from $I :=]a, b[$ into $\mathbb{K}$ is a vector space over $\mathbb{K}$ by [Phi15a, Th. 9.6(a),(b)]. Moreover, [Phi15a, Th. 9.6(a),(b)] also implies that, for each $k \in \mathbb{N}$, the set of k times differentiable functions from I into $\mathbb{K}$ is a vector space over $\mathbb{K}$, and so is each set $C^k(I, \mathbb{K})$ of k times continuously differentiable functions ([Phi15a, Th. 7.38] is also needed for the last conclusion). The set $C^\infty(I, \mathbb{K})$ is also a vector space over $\mathbb{K}$ by Th. A.5(a) below.

Theorem A.5. *Let V be a vector space over the field F.*

(a) *Let $I \neq \emptyset$ be an index set and $(U_i)_{i \in I}$ a family of subspaces of V. Then the intersection $U := \bigcap_{i \in I} U_i$ is also a subspace of V.*

(b) *In contrast to intersections, unions of subspaces are* almost never *subspaces. More precisely, if U_1 and U_2 are subspaces of V, then*

$$U_1 \cup U_2 \text{ is subspace of } V \quad \Leftrightarrow \quad \Big(U_1 \subseteq U_2 \ \vee \ U_2 \subseteq U_1\Big). \tag{A.8}$$

Proof. See, e.g., [Str08, Th. 8.7]. ■

Definition A.6. Let V be a vector space over the field F.

(a) Let $n \in \mathbb{N}$ and $v_1, \ldots, v_n \in V$. A vector $v \in V$ is called a *linear combination* of $v_1, \ldots, v_n$ if, and only if, there exist $\lambda_1, \ldots, \lambda_n \in F$ such that

$$v = \sum_{i=1} \lambda_i \, v_i. \tag{A.9}$$

Morover, $v \in V$ is called *linearly dependent* of a subset U of V if, and only if, there exists $n \in \mathbb{N}$ and $u_1, \ldots, u_n \in U$ such that v is a linear combination of $u_1, \ldots, u_n$.

(b) A subset U of V is called *linearly independent* if, and only if,

$$\left(n \in \mathbb{N} \quad \wedge \quad u_1, \ldots, u_n \in U \quad \wedge \quad \lambda_1, \ldots, \lambda_n \in F \quad \wedge \quad \sum_{i=1} \lambda_i \, u_i = 0 \right)$$
$$\Rightarrow \quad \lambda_1 = \cdots = \lambda_n = 0. \tag{A.10a}$$

Occasionally, one also wants to have the notion available for families of vectors instead of sets, and one calls a family $(u_i)_{i \in I}$ of vectors in V *linearly independent* if, and only if,

$$\left(n \in \mathbb{N} \quad \wedge \quad i_1, \ldots, i_n \in I \quad \wedge \quad \lambda_1, \ldots, \lambda_n \in F \quad \wedge \quad \sum_{k=1} \lambda_i \, u_{i_k} = 0 \right)$$
$$\Rightarrow \quad \lambda_1 = \cdots = \lambda_n = 0. \tag{A.10b}$$

Sets and families that are not linearly independent are called *linearly dependent*.

Definition A.7. Let V be a vector space over the field F, $A \subseteq V$, and

$$\mathcal{U} := \big\{ U \in \mathcal{P}(V) : \; A \subseteq U \; \wedge \; U \text{ is subspace of } V \big\}. \tag{A.11}$$

Then the set

$$\langle A \rangle := \operatorname{span} A := \bigcap_{U \in \mathcal{U}} U \tag{A.12}$$

is called the *span* of A. Moreover A is called a *spanning set* of V if, and only if, $\langle A \rangle = V$.

Proposition A.8. *Let V be a vector space over the field F and $A \subseteq V$.*

(a) *$\langle A \rangle$ is a subspace of V, namely the smallest subspace of V containing A.*

(b) *If $A = \emptyset$, then $\langle A \rangle = \{0\}$; if $A \neq \emptyset$, then $\langle A \rangle$ is the set of all linear combinations of elements from A, i.e.*

$$\langle A \rangle = \left\{ \sum_{i=1}^{n} \lambda_i \, a_i : \; n \in \mathbb{N} \; \wedge \; \lambda_1, \ldots, \lambda_n \in F \; \wedge \; a_1, \ldots, a_n \in A \right\}. \tag{A.13}$$

Proof. (a) is immediate from Th. A.5(a).

(b): For the case $A = \emptyset$, note that $\{0\}$ is a subspace of V, and that $\{0\}$ is contained in every subspace of V. For $A \neq \emptyset$, let W denote the right-hand side of (A.13), and recall from (A.12) that $\langle A \rangle$ is the intersection of all subspaces U of V that contain A. If U is a subspace of V and $A \subseteq U$, then $W \subseteq U$, since U is closed under vector addition and scalar multiplication, showing $W \subseteq \langle A \rangle$. On the other hand, W is clearly a subspace of V that contains A, showing $\langle A \rangle \subseteq W$, completing the proof of $\langle A \rangle = W$. ∎

Definition A.9. Let V be a vector space over the field F and $B \subseteq V$.

(a) B is called a *generating set* for V if, and only if, $\langle B\rangle = V$. One then also says that V is *generated* or *spanned* by B.

(b) B is called a *basis* for V if, and only if, B is a generating set for V that is also linearly independent (see Def. A.6(b)).

Theorem A.10. *Let V be a vector space over the field F and $B \subseteq V$. Then the following statements* (i) – (iii) *are equivalent:*

(i) *B is a basis for V.*

(ii) *B is a maximal linearly independent subset of V, i.e. B is linearly independent and each set $A \subseteq V$ with $B \subsetneq A$ is linearly dependent.*

(iii) *B is a minimal generating set for V, i.e. $\langle B\rangle = V$ and $\langle A\rangle \subsetneq V$ for each $A \subsetneq B$.*

Proof. See, e.g., [Str08, Th. 9.6]. ■

Theorem A.11 (Coordinates)**.** *Let V be a vector space over the field F and assume $B \subseteq V$ is a basis of V. Then each vector $v \in V$ has unique* coordinates *with respect to the basis B, i.e., for each $v \in V$, there exists a unique finite subset B_v of B and a unique map $c : B_v \longrightarrow F \setminus \{0\}$ such that*

$$v = \sum_{b\in B_v} c(b)\, b. \tag{A.14}$$

Note that, for $v = 0$, one has $B_v = \emptyset$, c is the empty map, and (A.14) *becomes $0 = \sum_{b\in\emptyset} c(b)\, b$, employing the useful convention that sums over the empty set are defined as* 0.

Proof. The existence of B_v and the map c follows from the fact that the basis B is a generating set, $\langle B\rangle = V$. For the uniqueness proof, consider finite sets $B_v, \tilde{B}_v \subseteq B$ and maps $c : B_v \longrightarrow F \setminus \{0\}$, $\tilde{c} : \tilde{B}_v \longrightarrow F \setminus \{0\}$ such that

$$v = \sum_{b\in B_v} c(b)\, b = \sum_{b\in \tilde{B}_v} \tilde{c}(b)\, b. \tag{A.15}$$

Extend both c and $\tilde{c}$ to $A := B_v \cup \tilde{B}_v$ by letting $c(b) := 0$ for $b \in \tilde{B}_v \setminus B_v$ and $\tilde{c}(b) := 0$ for $b \in B_v \setminus \tilde{B}_v$. Then

$$0 = \sum_{b\in A} \big(c(b) - \tilde{c}(b)\big)\, b, \tag{A.16}$$

such that the linear independence of A implies $c(b) = \tilde{c}(b)$ for each $b \in A$, which, in turn, implies $B_v = \tilde{B}_v$ and $c = \tilde{c}$. ■

Remark A.12. If the basis B of V has finitely many elements, then one often enumerates the elements $B = \{b_1, \ldots, b_n\}$, $n = \#B \in \mathbb{N}$, and writes $\lambda_i := c(b_i)$ for $b_i \in B_v$, $\lambda_i := 0$ for $b_i \notin B_v$, such that (A.14) takes the, perhaps more familiar looking, form

$$v = \sum_{i=1}^{n} \lambda_i b_i. \tag{A.17}$$

Theorem A.13. *Every vector space V over a field F has a basis $B \subseteq V$. Moreover, bases of V have a unique cardinality, i.e. if $B \subseteq V$ and $\tilde{B} \subseteq V$ are both bases of V, then $\#B = \#\tilde{B}$.*

Proof. See, e.g., [Str08, Lem. 11.3, Th. 11.5]. ■

Definition A.14. According to Th. A.13, for each vector space V over a field F, the cardinality of its bases is unique. This unique cardinality is called the *dimension* of V and is denoted $\dim V$. If $\dim V < \infty$ (i.e. $\dim V \in \mathbb{N}_0$), then V is called *finite dimensional*, otherwise *infinite dimensional*.

Example A.15. Given a field F and a nonempty set I, let F^I_{fin} denote the set of functions $f : I \longrightarrow F$ such that there exists a finite set $I_f \subseteq I$ satisfying

$$f(i) = 0 \quad \text{for each } i \in I \setminus I_f, \tag{A.18a}$$

$$f(i) \neq 0 \quad \text{for each } i \in I_f. \tag{A.18b}$$

Then $F^I_{\text{fin}} = F^I$ if, and only if, I is finite (for example $F^n_{\text{fin}} = F^n$ for $n \in \mathbb{N}$); in general F^I_{fin} is a strict subset of F^I. However, if $f, g \in F^I_{\text{fin}}$ and $\lambda \in F$, then $I_{\lambda f} = I_f$ for $\lambda \neq 0$, $I_{\lambda f} = \emptyset$ for $\lambda = 0$, and $I_{f+g} \subseteq I_f \cup I_g$, showing F^I_{fin} is always a subspace of F^I. Define

$$e_i : I \longrightarrow F, \quad e_i(j) := \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{if } j \neq i. \end{cases} \tag{A.19}$$

Then $I_{e_i} = \{i\}$ for each $i \in I$, in particular, $e_i \in F^I_{\text{fin}}$ for each $i \in I$. We claim that $B := \{e_i : i \in I\}$ is a basis for F^I_{fin}. Indeed, if $f \in F^I_{\text{fin}}$, then

$$f = \sum_{i \in I_f} f(i) e_i, \tag{A.20}$$

showing $\langle B \rangle = F^I_{\text{fin}}$. If J is a finite subset of I and $(\lambda_j)_{j \in J}$ is a family in F such that

$$\sum_{j \in J} \lambda_j e_j \equiv 0, \tag{A.21}$$

then

$$\underset{j \in J}{\forall} \quad \lambda_j \overset{\text{(A.19)}}{=} \sum_{j \in J} \lambda_j e_j(j) \overset{\text{(A.21)}}{=} 0, \tag{A.22}$$

proving B is linearly independent. Clearly, $\#B = \#I$, so we have shown

$$\dim F^I_{\text{fin}} = \#I. \tag{A.23}$$

In particular, we have shown that, for each $n \in \mathbb{N}$, the set $\{e_j : j = 1, \ldots, n\}$, where

$$e_1 := (1, 0, \ldots, 0), \quad e_2 := (0, 1, \ldots, 0), \quad \ldots, \quad e_n := (0, \ldots, 0, 1), \tag{A.24}$$

forms a basis of F^n (of $\mathbb{R}^n$ if $F = \mathbb{R}$ and of $\mathbb{C}^n$ if $F = \mathbb{C}$),

$$\dim F^n = \dim \mathbb{R}^n = \dim \mathbb{C}^n = n. \tag{A.25}$$

Remark A.16. We will see in Th. A.24 below that, in a certain sense, F^I_{fin} is the only vector space of dimension $\#I$ over F. In particular, for $n \in \mathbb{N}$, you can think of $\mathbb{K}^n$ as the *standard model* of an n-dimensional vector space over $\mathbb{K}$.

A.2 Linear Maps

Definition A.17. Let V and W be vector spaces over the field F.

(a) A map $A : V \longrightarrow W$ is called *F-linear* (or merely *linear* if the field F is understood) if, and only if,

$$\bigvee_{v_1,v_2\in V} \quad A(v_1 + v_2) = A(v_1) + A(v_2), \tag{A.26a}$$

$$\wedge \quad \bigvee_{\lambda\in F} \quad \bigvee_{v\in V} \quad A(\lambda v) = \lambda A(v) \tag{A.26b}$$

or, equivalently, if, and only if,

$$\bigvee_{\lambda,\mu\in F} \quad \bigvee_{v_1,v_2\in V} \quad A(\lambda v_1 + \mu v_2) = \lambda A(v_1) + \mu A(v_2) \tag{A.27}$$

(note that, in general, vector addition and scalar multiplication will be different on the left-hand sides and right-hand sides of the above equations).

One also calls linear maps (vector space) *homomorphisms*. We denote the set of all F-linear maps from V into W by $\mathcal{L}(V, W)$.

(b) A linear map $I : V \longrightarrow W$ is called a (vector space or linear) *isomorphism* if, and only if, it is bijective (i.e. invertible). The vector spaces V and W are called *isomorphic* (denoted $V \cong W$) if, and only if, there exists a vector space isomorphism $I : V \longrightarrow W$.

Theorem A.18. *Let V and W be vector spaces over the field F. If $I : V \longrightarrow W$ is a linear isomorphism, then so is $I^{-1} : W \longrightarrow V$ (i.e. I^{-1} is not only bijective, but also linear).*

Proof. See, e.g., [Str08, Th. 13.5]. ■

Definition A.19. Let V and W be vector spaces over the field F, and $A \in \mathcal{L}(V, W)$. Define the *kernel* and the *image* of A by

$$\ker A := A^{-1}\{0\} = \{v \in V : A(v) = 0\}, \tag{A.28a}$$

$$\operatorname{Im} A := A(V) = \{A(v) : v \in V\}, \tag{A.28b}$$

respectively.

Theorem A.20. *Let V and W be vector spaces over the field F, and $A \in \mathcal{L}(V, W)$.*

(a) $\ker A$ *is a subspace of V and* $\operatorname{Im} A$ *is a subspace of W.*

(b) *A is injective if, and only if,* $\ker A = \{0\}$.

Proof. (a): See, e.g., [Str08, Lem. 12.5].

(b): Since $A(0) = 0$, A being injective implies $\ker A = \{0\}$. Conversely, assume $\ker A = \{0\}$ and $A(v_1) = A(v_2)$ for $v_1, v_2 \in V$. Then $A(-v_1 + v_2) = -A(v_1) + A(v_2) = -A(v_1) + A(v_1) = 0$, i.e. $-v_1 + v_2 \in \ker A$, i.e. $-v_1 + v_2 = 0$, showing $v_1 = v_2$ and the injectivity of A. ■

Theorem A.21 (Dimension Formulas)**.** *Let V and W be vector spaces over the field F, and let $A : V \longrightarrow W$ be linear.*

(a) *If V is finite dimensional, then* $\dim V = \dim \ker A + \dim \operatorname{Im} A$.

(b) *If V is finite dimensional, then* $\dim \operatorname{Im} A \leq \dim V$.

(c) *If W is finite dimensional, then* $\dim \operatorname{Im} A \leq \dim W$.

Proof. See, e.g., [Str08, Th. 12.12]. ■

Proposition A.22. *Let V and W be vector spaces over the field F, and let $A : V \longrightarrow W$ be linear.*

(a) *A is injective if, and only if, for each linearly independent subset S of V, $A(S)$ is a linearly independent subset of W.*

(b) *A is surjective if, and only if, for each generating subset S of V, $A(S)$ is a generating subset of W.*

(c) *A is bijective if, and only if, for each basis B of V, $A(B)$ is a basis of W.*

Proof. (a): If A is not injective, then, according to Th. A.20(b), there exists $0 \neq v \in V$ such that $A(v) = 0$. Then $S := \{v\}$ is linearly independent, whereas $A(S) = \{0\}$ is not. Conversely, if A is injective, $S \subseteq V$ is linearly independent, and $\lambda_1, \ldots, \lambda_n \in F$; $s_1, \ldots, s_n \in S$; $n \in \mathbb{N}$; such that

$$0 = \sum_{i=1}^{n} \lambda_i A(s_i) = A\left(\sum_{i=1}^{n} \lambda_i s_i\right), \tag{A.29}$$

then $\sum_{i=1}^{n} \lambda_i s_i = 0$ by Th. A.20(b), implying $\lambda_1 = \cdots = \lambda_n = 0$, showing that $A(S)$ is also linearly independent.

(b): If A is not surjective, then $\langle A(V)\rangle = \operatorname{Im} A \neq W$, since $\operatorname{Im} A$ is a subspace of W. Conversely, if A is surjective, $S \subseteq V$, $\langle S\rangle = V$, and $w \in W$, then there are $v \in V$; $\lambda_1, \ldots, \lambda_n \in F$; $s_1, \ldots, s_n \in S$; $n \in \mathbb{N}$; such that $A(v) = w$ and $v = \sum_{i=1}^{n} \lambda_i s_i$, i.e. $w = A(v) = \sum_{i=1}^{n} \lambda_i A(s_i)$, proving $w \in \langle A(S)\rangle$. Since $w \in W$ was arbitrary, we have shown $\langle A(S)\rangle = W$.

(c) follows immediately by combining (a) and (b) (recalling that a basis is a linearly independent generating set). ■

Theorem A.23. *Let V and W be vector spaces over the field F. Then each linear map $A : V \longrightarrow W$ is uniquely determined by its values on a basis of V. More precisely, if B is a basis of V, $(w_b)_{b\in B}$ is a family in W, and, for each $v \in V$, B_v and $c_v : B_v \longrightarrow F \setminus \{0\}$ are as in Th.* A.11 *(we now write c_v instead of c to underline the dependence of c on v), then the map*

$$A : V \longrightarrow W, \quad A(v) = A\left(\sum_{b\in B_v} c_v(b)\, b\right) := \sum_{b\in B_v} c_v(b)\, w_b, \tag{A.30}$$

is linear, and $\tilde{A} \in \mathcal{L}(V,W)$ *with*

$$\underset{b\in B}{\forall} \quad \tilde{A}(b) = w_b, \tag{A.31}$$

implies $A = \tilde{A}$.

Proof. We first verify A is linear. Let $v \in V$ and $\lambda \in F$. If $\lambda = 0$, then $A(\lambda v) = A(0) = 0 = \lambda A(v)$. If $\lambda \neq 0$, then $B_{\lambda v} = B_v$, $c_{\lambda v} = \lambda c_v$, and

$$A(\lambda v) = A\left(\sum_{b\in B_{\lambda v}} c_{\lambda v}(b)\, b\right) = \sum_{b\in B_v} \lambda\, c_v(b)\, w_b = \lambda\, A\left(\sum_{b\in B_v} c_v(b)\, b\right) = \lambda A(v). \tag{A.32a}$$

Now let $u, v \in V$. If $u = 0$, then $A(u+v) = A(v) = 0 + A(v) = A(u) + A(v)$, and analogously if $v = 0$. So assume $u, v \neq 0$. If $u + v = 0$, then $v = -u$ and $A(u+v) = A(0) = 0 = A(u)+A(-u) = A(u)+A(v)$. If $u+v \neq 0$, then $B_{u+v} \subseteq B_u \cup B_v$ and c_{u+v} can be extended to $B_u \cup B_v$ by letting

$$c_{u+v}(b) := \begin{cases} c_u(b) + c_v(b) & \text{if } b \in B_u \cap B_v, \\ c_u(b) & \text{if } b \in B_u \setminus B_v, \\ c_v(b) & \text{if } b \in B_v \setminus B_u. \end{cases} \tag{A.32b}$$

One then obtains

$$\begin{aligned} A(u+v) = A\left(\sum_{b\in B_{u+v}} c_{u+v}(b)\, b\right) &= \sum_{b\in B_{u+v}} c_{u+v}(b)\, w_b \\ &= \sum_{b\in B_u} c_u(b)\, w_b + \sum_{b\in B_v} c_v(b)\, w_b = A(u) + A(v). \end{aligned} \tag{A.32c}$$

If $v \in V$ and B_v and c_v are as before, then the linearity of $\tilde{A}$ and (A.31) imply

$$\begin{aligned} \tilde{A}(v) = \tilde{A}\left(\sum_{b\in B_v} c_v(b)\, b\right) &\overset{\tilde{A}\in\mathcal{L}(V,W)}{=} \sum_{b\in B_v} c_v(b)\, \tilde{A}(b) \\ = \sum_{b\in B_v} c_v(b)\, A(b) &\overset{(*)}{=} \sum_{b\in B_v} c_v(b)\, w_b = A(v), \end{aligned} \tag{A.33}$$

where, at $(*)$, it was used that, for each $b \in B$, one has $A(b) = w_b$ (note $B_b = \{b\}$, $c_b(b) = 1$). Since (A.33) establishes $\tilde{A} = A$, the proof is complete. ∎

Theorem A.24. *Let* V *and* W *be vector spaces over the field* F*. Then* $V \cong W$ *(i.e.* V *and* W *are isomorphic) if, and only if,* $\dim V = \dim W$.

Proof. Suppose $\dim V = \dim W$. If B_V is a basis of V and B_W is a basis of W, then there exists a bijective map $i : B_V \longrightarrow B_W$. According to Th. A.23, i defines a unique

linear map $A : V \longrightarrow W$ with $A(b) = i(b)$ for each $b \in B_V$. More precisely, letting, once again, for each $v \in V$, B_v and $c_v : B_v \longrightarrow F \setminus \{0\}$ be as in Th. A.11 (writing c_v instead of c to underline the dependence of c on v),

$$\underset{v\in V}{\forall} \quad A(v) = A\left(\sum_{b\in B_v} c_v(b)\, b\right) = \sum_{b\in B_v} c_v(b)\, i(b). \tag{A.34}$$

It remains to show A is bijective. If $v \neq 0$, then $B_v \neq \emptyset$ and $A(v) = \sum_{b\in B_v} c_v(b)\, i(b) \neq 0$, since $c_v(b) \neq 0$ and $\{i(b) : b \in B_v\} \subseteq B_W$ is linearly independent, showing A is injective by Th. A.20(b). If $w \in W$, then there exists a finite set $\tilde{B}_w \subseteq B_W$ and $c_w : B_w \longrightarrow F$ such that $\sum_{\tilde{b}\in\tilde{B}_w} \tilde{c}_w(\tilde{b})\, \tilde{b}$. Then

$$\begin{aligned} A\left(\sum_{\tilde{b}\in\tilde{B}_w} \tilde{c}_w(\tilde{b})\, i^{-1}(\tilde{b})\right) &\overset{\tilde{A}\in\mathcal{L}(V,W)}{=} \sum_{\tilde{b}\in\tilde{B}_w} \tilde{c}_w(\tilde{b})\, A\left(i^{-1}(\tilde{b})\right) \overset{i^{-1}(\tilde{b})\in B_V}{=} \sum_{\tilde{b}\in\tilde{B}_w} \tilde{c}_w(\tilde{b})\, i\left(i^{-1}(\tilde{b})\right) \\ &= \sum_{\tilde{b}\in\tilde{B}_w} \tilde{c}_w(\tilde{b})\, \tilde{b} = w, \end{aligned} \tag{A.35}$$

showing $\operatorname{Im} A = W$, completing the proof that A is bijective.

If $A : V \longrightarrow W$ is a linear isomorphism and B is a basis for V, then, by Prop. A.22(c), $A(B)$ is a basis for W. As A is bijective, so is $A\restriction_B$, showing $\dim V = \#B = \#A(B) = \dim W$ as claimed. ∎

Definition A.25. Let V and W be vector spaces over the field F. We define an addition and a scalar multiplication on $\mathcal{L}(V, W)$ by

$$(A + B) : V \longrightarrow W, \qquad (A + B)(x) := A(x) + B(x), \tag{A.36a}$$

$$(\lambda \cdot A) : V \longrightarrow W, \qquad (\lambda \cdot A)(x) := \lambda \cdot A(x) \quad \text{for each } \lambda \in F. \tag{A.36b}$$

Theorem A.26. *Let V and W be vector spaces over the field F. The addition and scalar multiplication on $\mathcal{L}(V, W)$ given by* (A.36) *are well-defined in the sense that, if $A, B \in \mathcal{L}(V, W)$ and $\lambda \in F$, then $A + B \in \mathcal{L}(V, W)$ and $\lambda A \in \mathcal{L}(V, W)$. Moreover, with the operations defined in* (A.36)*, $\mathcal{L}(V, W)$ forms a vector space over F.*

Proof. See, e.g., [Str08, Th. 13.2]. ∎

Theorem A.27. *Let V and W be finite dimensional vector spaces over the field F, let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ be bases of V and W, respectively; $m, n \in \mathbb{N}$. Using Th.* A.23*, define maps $A_{ji} \in \mathcal{L}(V, W)$ by letting*

$$\underset{(j,i,k)\in\{1,\ldots,m\}\times\{1,\ldots,n\}^2}{\forall} \quad A_{ji}(v_k) := \begin{cases} w_j & \text{for } k = i, \\ 0 & \text{for } k \neq i. \end{cases} \tag{A.37}$$

Then $\big\{A_{ji} : (j, i) \in \{1, \ldots, m\} \times \{1, \ldots, n\}\big\}$ constitutes a basis for $\mathcal{L}(V, W)$ and, in particular,

$$\dim \mathcal{L}(V, W) = \dim V \cdot \dim W = n \cdot m. \tag{A.38}$$

Proof. See, e.g., [Str08, Th. 13.11]. ■

Theorem A.28. *Let V, W, X be vector spaces over the field F.*

(a) *The composition of linear maps is linear, i.e. if $A \in \mathcal{L}(V,W)$ and $B \in \mathcal{L}(W,X)$, then $B \circ A \in \mathcal{L}(V,X)$.*

(b) *If $A \in \mathcal{L}(V,W)$ and $B, C \in \mathcal{L}(W,X)$, then*

$$A \circ (B + C) = A \circ B + A \circ C. \tag{A.39}$$

(c) *If $A, B \in \mathcal{L}(V,W)$ and $C \in \mathcal{L}(W,X)$, then*

$$(A + B) \circ C = A \circ C + B \circ C. \tag{A.40}$$

Proof. See, e.g., [Str08, Th. 13.3]. ■

A.3 Matrices

Matrices provide a convenient representation for linear maps A between finite dimensional vector spaces V and W. Recall the basis $\{A_{ji} : (j,i) \in \{1,\dots,m\} \times \{1,\dots,n\}\}$ of $\mathcal{L}(V,W)$ that, in Th. A.27, was shown to arise from bases $\{v_1,\dots,v_n\}$ and $\{w_1,\dots,w_m\}$ of V and W, respectively; $m, n \in \mathbb{N}$. Thus, each $A \in \mathcal{L}(V,W)$ can be written in the form

$$A = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ji} A_{ji}, \tag{A.41}$$

with coordinates $(a_{ji})_{(j,i) \in \{1,\dots,m\} \times \{1,\dots,n\}}$ in F. This motivates the following definition of matrices.

Definition A.29. Let F be a field and $m, n \in \mathbb{N}$. A family $(a_{ji})_{(j,i) \in \{1,\dots,m\} \times \{1,\dots,n\}}$ is called an m-by-n or an $m \times n$ *matrix* over F, where $m \times n$ is called the *size*, *dimension* or *type* of the matrix. The a_{ji} are called the *entries* or *elements* of the matrix. One also writes just (a_{ji}) instead of $(a_{ji})_{(j,i) \in \{1,\dots,m\} \times \{1,\dots,n\}}$ if the size of the matrix is understood. One usually thinks of the $m \times n$ matrix (a_{ji}) as the *rectangular array*

$$(a_{ji}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \tag{A.42}$$

with m rows and n columns. One therefore also calls $1 \times n$ matrices *row vectors* and $m \times 1$ matrices *column vectors*, and one calls $n \times n$ matrices *quadratic*. The set of all $m \times n$ matrices over F is denoted by $\mathcal{M}(m,n,F)$, and for the set of all quadratic $n \times n$ matrices, one uses the abbreviation $\mathcal{M}(n,F) := \mathcal{M}(n,n,F)$.

Definition A.30 (Matrix Arithmetic)**.** Let F be a field and $m, n, l \in \mathbb{N}$.

(a) *Matrix Addition:* For $m \times n$ matrices (a_{ji}) and (b_{ji}) over F, define the sum

$$(a_{ji}) + (b_{ji}) := (a_{ji} + b_{ji}). \tag{A.43}$$

(b) *Scalar Multiplication:* For each $m \times n$ matrix (a_{ji}) and each $\lambda \in F$, define

$$\lambda\,(a_{ji}) := (\lambda\, a_{ji}). \tag{A.44}$$

(c) *Matrix Multiplication:* For each $m \times n$ matrix (a_{ji}) and each $n \times l$ matrix (b_{ji}) over F, define the product

$$(a_{ji})(b_{ji}) := \left(\sum_{k=1}^{n} a_{jk} b_{ki} \right)_{(j,i) \in \{1,\ldots,m\} \times \{1,\ldots,l\}}, \tag{A.45}$$

i.e. the product of an $m \times n$ matrix and an $n \times l$ matrix is an $m \times l$ matrix (cf. Th. A.34 below).

Remark A.31. We consider matrices over a field F.

(a) For each $m, n \in \mathbb{N}$, the set $\mathcal{M}(m, n, F)$ of $m \times n$ matrices over F with the operations of Def. A.30(a),(b) constitutes a vector space over F: An $m \times n$ matrix $A = (a_{ji})_{(j,i) \in \{1,\ldots,m\} \times \{1,\ldots,n\}}$ is defined as a family in F, i.e., recalling [Phi15a, Def. 2.14(a)], A is defined as the function $A : \{1, \ldots, m\} \times \{1, \ldots, n\} \longrightarrow F$, $A(j, i) = a_{ji}$; and $\mathcal{M}(m, n, F) = \mathcal{F}\big(\{1, \ldots, m\} \times \{1, \ldots, n\}, F\big)$ (so we notice that matrices are nothing new in terms of objects, but just a new way of thinking about functions from $\{1, \ldots, m\} \times \{1, \ldots, n\}$ into F, that turns out to be convenient in certain contexts). Thus, the operations defined in Def. A.30(a),(b) are precisely the same operations that were defined in (A.5) and $\mathcal{M}(m, n, F)$ is a vector space according to Ex. A.2(c). Clearly, the map

$$\begin{gathered} I : \mathcal{M}(m, n, F) \longrightarrow F^{m \cdot n}, \quad (a_{ji}) \mapsto (\lambda_1, \ldots, \lambda_{m \cdot n}), \\ \text{where } \lambda_k = a_{ji} \text{ if, and only if, } k = (j-1) \cdot n + i, \end{gathered} \tag{A.46}$$

constitutes a linear isomorphism. Other important linear isomorphisms between $\mathcal{M}(m, n, F)$ and vector spaces of linear maps will be provided in Th. A.32 below.

(b) Matrix multiplication is associative whenever all relevant multiplications are defined. More precisely, if A is an $m \times n$ matrix, B is an $n \times l$, and C is an $l \times p$ matrix, then

$$(AB)C = A(BC) : \tag{A.47}$$

Indeed, one has $m \times p$ matrices $(AB)C = (d_{ji})$ and $A(BC) = (e_{ji})$, where

$$d_{ji} = \sum_{\alpha=1}^{l} \left(\sum_{k=1}^{n} a_{jk} b_{k\alpha} \right) c_{\alpha i} = \sum_{\alpha=1}^{l} \sum_{k=1}^{n} a_{jk} b_{k\alpha} c_{\alpha i} = \sum_{k=1}^{n} a_{jk} \left(\sum_{\alpha=1}^{l} b_{k\alpha} c_{\alpha i} \right) = e_{ji}. \tag{A.48}$$

(c) Matrix multiplication is, in general, *not* commutative: If A is an $m \times n$ matrix and B is an $n \times l$ with $m \neq l$, then BA is not even defined. If $m = l$, but $m \neq n$, then AB has dimension $m \times m$, but BA has different dimension, namely $n \times n$. And even if $m = n = l > 1$, then commutativity is, in general not true – for example

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \tag{A.49a}$$

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}. \tag{A.49b}$$

Note that $\lambda = m$ for $F = \mathbb{R}$, but, in general, λ will depend on F, e.g. for $F = \{0, 1\}$, one obtains $\lambda = m \mod 2$.

—

Let us come back to the situation discussed at the beginning of the section above, resulting in (A.41). Let $v = \sum_{i=1}^{n} \lambda_i v_i \in V$ with $\lambda_1, \dots, \lambda_n \in F$. Then

$$\begin{aligned} A(v) &= \sum_{i=1}^{n} \lambda_i A(v_i) = \sum_{i=1}^{n} \lambda_i \sum_{k=1}^{n} \sum_{j=1}^{m} a_{jk} A_{jk}(v_i) \overset{(A.37)}{=} \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{m} a_{ji} w_j \\ &= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_{ji} \lambda_i \right) w_j. \end{aligned} \tag{A.50}$$

Thus, if we represent v by a column vector $\tilde{v}$ (an $n \times 1$ matrix) containing its coordinates $\lambda_1, \dots, \lambda_n$ with respect to the basis $\{v_1, \dots, v_n\}$ and $A(v)$ by a column vector $\tilde{w}$ (an $m \times 1$ matrix) containing its coordinates with respect to the basis $\{w_1, \dots, w_m\}$, then (A.50) shows

$$\tilde{w} = M\tilde{v} = M \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}, \quad \text{where} \quad M := (a_{ji}). \tag{A.51}$$

For finite dimensional vector spaces, the precise relationship between linear maps, bases, and matrices is provided by the following theorem:

Theorem A.32. *Let V and W be finite dimensional vector spaces over the field F, let $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_m\}$ be bases of V and W, respectively; $m, n \in \mathbb{N}$. Then the map*

$$I : \mathcal{L}(V, W) \longrightarrow \mathcal{M}(m, n, F), \quad A \mapsto (a_{ji}), \tag{A.52}$$

where the a_{ji} are given by (A.41) *constitutes a linear isomorphism.*

Proof. According to Th. A.27, $\{A_{ji} : (j,i) \in \{1,\dots,m\} \times \{1,\dots,n\}\}$ forms a basis of $\mathcal{L}(V,W)$. Thus, to every family of coordinates $\{a_{ji} : (j,i) \in \{1,\dots,m\} \times \{1,\dots,n\}\}$ in F, (A.41) defines a unique element of $\mathcal{L}(V,W)$, i.e. I is bijective. It remains to verify that I is linear. To this end, let $\lambda, \mu \in F$ and $A, B \in \mathcal{M}(m,n,F)$ with

$$A = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ji} A_{ji}, \qquad (a_{ji}) = I(A) \in \mathcal{M}(m,n,F), \tag{A.53a}$$

$$B = \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ji} A_{ji}, \qquad (b_{ji}) = I(B) \in \mathcal{M}(m,n,F). \tag{A.53b}$$

Then

$$\lambda A + \mu B = \lambda \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ji} A_{ji} + \mu \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ji} A_{ji} = \sum_{i=1}^{n} \sum_{j=1}^{m} (\lambda a_{ji} + \mu b_{ji}) A_{ji}, \tag{A.54}$$

showing

$$I(\lambda A + \mu B) = (\lambda a_{ji} + \mu b_{ji}) A_{ji}) = \lambda (a_{ji}) + \mu (b_{ji}) = \lambda I(A) + \mu I(B), \tag{A.55}$$

proving the linearity of I. ■

Definition and Remark A.33. In the situation of Th. A.32, for each $A \in \mathcal{L}(V,W)$, one calls the matrix $I(A) = (a_{ji}) \in \mathcal{M}(m,n,F)$ the *(transformation) matrix corresponding* to A *with respect to* the basis $\{v_1,\dots,v_n\}$ of V and the basis $\{w_1,\dots,w_m\}$ of W. If the bases are understood, then one often tends to identify the map with its corresponding matrix.

However, as $I(A)$ depends on the bases, identifying A and $I(A)$ is only admissible as long as one keeps the bases of V and W fixed! Moreover, if one represents matrices as rectangular arrays as in (A.42) (which one usually does), then one actually considers the basis vectors of $\{v_1,\dots,v_n\}$ and $\{w_1,\dots,w_m\}$ as *ordered* from 1 to n (resp. m), i.e. $I(A)$ actually depends on the so-called *ordered bases* $(v_1,\dots,v_n)$ and $(w_1,\dots,w_m)$ (ordered bases are tuples rather than sets and the matrix corresponding to A changes if the order of the basis vectors changes).

Similarly, we had seen in (A.51) that it can be useful to identify a vector $v = \sum_{i=1}^{n} \lambda_i v_i$ with its coordinates $(\lambda_1,\dots,\lambda_n)$, typically represented as an $n \times 1$ matrix (a column vector, as in (A.51)) or a $1 \times n$ matrix (a row vector). Obviously, this identification is also only admissible as long as the basis $\{v_1,\dots,v_n\}$ and its order is kept fixed.

—

The following Th. A.34 is the justification for defining matrix multiplication according to Def. A.30(c).

Theorem A.34. *Let F be a field, let $n,m,l \in \mathbb{N}$, and let V,W,X be finite dimensional vector spaces over F such that V has basis $\{v_1,\dots,v_n\}$, W has basis $\{w_1,\dots,w_m\}$, and*

X has basis $\{x_1, \ldots, x_l\}$. If $A \in \mathcal{L}(V,W)$, $B \in \mathcal{L}(W,X)$, $M = (a_{ji}) \in \mathcal{M}(m,n,F)$ is the matrix corresponding to A with respect to $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$, and $N = (b_{ji}) \in \mathcal{M}(l,m,F)$ is the matrix corresponding to B with respect to $\{w_1, \ldots, w_m\}$ and $\{x_1, \ldots, x_l\}$, then $NM = (\sum_{k=1}^m b_{jk}a_{ki}) \in \mathcal{M}(l,n,F)$ is the matrix corresponding to BA with respect to $\{v_1, \ldots, v_n\}$ and $\{x_1, \ldots, x_l\}$.

Proof. For each $i \in \{1, \ldots, n\}$, one computes

$$\begin{aligned}(BA)(v_i) = B\big(A(v_i)\big) = B\left(\sum_{k=1}^m a_{ki}w_k\right) = \sum_{k=1}^m a_{ki}B(w_k) = \sum_{k=1}^m a_{ki}\sum_{j=1}^l b_{jk}x_j \\ = \sum_{j=1}^l\sum_{k=1}^m b_{jk}a_{ki}x_j = \sum_{j=1}^l\left(\sum_{k=1}^m b_{jk}a_{ki}\right)x_j, \qquad (A.56)\end{aligned}$$

proving $NM = (\sum_{k=1}^m b_{jk}a_{ki})$ is the matrix corresponding to BA with respect to the bases $\{v_1, \ldots, v_n\}$ and $\{x_1, \ldots, x_l\}$. ∎

Definition and Remark A.35. Let F be a field, $A := (a_{ji})_{(j,i)\in\{1,\ldots,m\}\times\{1,\ldots,n\}} \in \mathcal{M}(m,n,F)$, and $m, n \in \mathbb{N}$. Then we define the *transpose* of A, denoted A^{t}, by

$$A^{\mathrm{t}} := (a_{ji})_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,m\}}. \qquad (A.57)$$

Thus, if A is an $m \times n$ matrix, then its transpose is an $n \times m$ matrix, where one obtains A^{t} from A by switching rows and columns. One has to use care when using the notation of (A.57), as one often implicitly assumes that, when writing (a_{ji}), the first index is for rows and the second index for columns. However, this is actually determined by the order of the factors of the cartesian product that determines the domain of the family. Whereas A is the map $f : \{1, \ldots, m\} \times \{1, \ldots, n\} \longrightarrow F$, $f(j,i) = a_{ji}$, its transpose A^{t} is the map $f^{\mathrm{t}} : \{1, \ldots, n\} \times \{1, \ldots, m\} \longrightarrow F$, $f^{\mathrm{t}}(i,j) = f(j,i) = a_{ji}$. To emphasize this in the notation, one can rewrite (A.57) in the form

$$A^{\mathrm{t}} = (b_{ij})_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,m\}}, \quad \text{where} \quad \underset{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,m\}}{\forall} \quad b_{ij} := a_{ji}. \qquad (A.58)$$

For the transpose of A, one also finds the notation A' instead of A^{t}.

Theorem A.36. *Let F be a field and $m, n, l \in \mathbb{N}$.*

(a) *The map*

$$I : \mathcal{M}(m,n,F) \longrightarrow \mathcal{M}(n,m,F), \quad A \mapsto A^{\mathrm{t}}, \qquad (A.59)$$

is a linear isomorphism and

$$\underset{A\in\mathcal{M}(m,n,F)}{\forall} \quad (A^{\mathrm{t}})^{\mathrm{t}} = A. \qquad (A.60)$$

(b) *If $A \in \mathcal{M}(m,n,F)$ and $B \in \mathcal{M}(n,l,F)$, then*

$$(AB)^{\mathrm{t}} = B^{\mathrm{t}}A^{\mathrm{t}}. \qquad (A.61)$$

Proof. (a): It is immediate from (A.57) that (A.60) is valid, showing I has an inverse map and is, hence, bijective. So it just remains to verify I is linear. However, if $A, B \in \mathcal{M}(m,n,F)$, $A = (a_{ji})$, $B = (b_{ji})$, and $\mu, \lambda \in F$, then

$$\begin{aligned}(\lambda A + \mu B)^{\mathrm{t}} &= (\lambda a_{ji} + \mu b_{ji})^{\mathrm{t}}_{(j,i)\in\{1,\ldots,m\}\times\{1,\ldots,n\}} = (\lambda a_{ji} + \mu b_{ji})_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,m\}} \\ &= \lambda (a_{ji})_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,m\}} + \mu (b_{ji})_{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,m\}} \\ &= \lambda (a_{ji})^{\mathrm{t}}_{(j,i)\in\{1,\ldots,m\}\times\{1,\ldots,n\}} + \mu (b_{ji})^{\mathrm{t}}_{(j,i)\in\{1,\ldots,m\}\times\{1,\ldots,n\}} \\ &= \lambda A^{\mathrm{t}} + \mu B^{\mathrm{t}}, \end{aligned} \tag{A.62}$$

thereby establishing the case.

(b): Let $A = (a_{ji})$, $B = (b_{ji})$, $A^{\mathrm{t}} = (a^{\mathrm{t}}_{ji})$, $B^{\mathrm{t}} = (b^{\mathrm{t}}_{ji})$. Then

$$\begin{aligned}(AB)^{\mathrm{t}} &\overset{(A.45)}{=} \left(\sum_{k=1}^{n} a_{jk} b_{ki}\right)^{\mathrm{t}}_{(j,i)\in\{1,\ldots,m\}\times\{1,\ldots,l\}} = \left(\sum_{k=1}^{n} b_{ki} a_{jk}\right)_{(i,j)\in\{1,\ldots,l\}\times\{1,\ldots,m\}} \\ &= \left(\sum_{k=1}^{n} b_{kj} a_{ik}\right)_{(j,i)\in\{1,\ldots,l\}\times\{1,\ldots,m\}} = \left(\sum_{k=1}^{n} b^{\mathrm{t}}_{jk} a^{\mathrm{t}}_{ki}\right)_{(j,i)\in\{1,\ldots,l\}\times\{1,\ldots,m\}} \\ &\overset{(A.45)}{=} B^{\mathrm{t}} A^{\mathrm{t}}, \end{aligned} \tag{A.63}$$

proving (A.61). ■

A.4 Determinants

For each quadratic matrix $A \in \mathcal{M}(n,F)$, one can define its determinant $\det(A) \in F$, resulting in a function $\det : \mathcal{M}(n,F) \longrightarrow F$ that is often useful when studying matrices and linear maps. One can characterize the determinant function axiomatically (see Def. A.38 below), and, with some preparation, one can also provide an explicit formula (see (A.82) below).

One important feature of the determinant is its being nonzero if, and only if, the matrix is invertible (cf. Def. A.46 and Th. A.48(a) below). Another is the fact that the determinant's value only depends on the linear map defined by A and an arbitrary basis of F^n. This allows to define $\det : \mathcal{L}(V,V) \longrightarrow F$ as in Def. and Rem. A.53 below. One can show that, if $A \in \mathcal{L}(V,V) \longrightarrow F$, then $\det(A)$ is a measure of the n-dimensional volume distortion caused by applying A, but, here, we will not pursue this aspect.

Definition A.37. Let F be a field, $n \in \mathbb{N}$. Then the $n \times n$ matrix

$$\mathrm{Id} := \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} = (e_{ji}), \quad \text{where} \quad e_{ji} := \begin{cases} 1 & \text{for } j = i, \\ 0 & \text{for } j \neq i, \end{cases} \tag{A.64}$$

is called *identity matrix* or just *identity* or *unit matrix*. The dependence of Id on n is suppressed in the notation, but n should always be clear from the context. In the literature, one also finds the notation E or I instead of Id.

Definition A.38. Let F be a field, $n \in \mathbb{N}$. A map $\det : \mathcal{M}(n,F) \longrightarrow F$ is called *determinant* if, and only if, it satisfies the following conditions (i) – (iii):

(i) det is *multilinear* with regard to matrix columns, i.e., for each $A \in \mathcal{M}(n,F)$, $b \in \mathcal{M}(n,1,F)$, $i \in \{1,\ldots,n\}$, and $\lambda, \mu \in F$:

$$\begin{aligned} &\det(a_1,\ldots,\lambda\, a_i + \mu\, b,\ldots,a_n) \\ &= \lambda\, \det(A) + \mu\, \det(a_1,\ldots,a_{i-1},b,a_{i+1}\ldots,a_n), \end{aligned} \tag{A.65}$$

where $a_1,\ldots,a_n$ denote the columns of A.

(ii) If the columns $a_1,\ldots,a_n$ of $A = (a_1,\ldots,a_n) \in \mathcal{M}(n,F)$ are linearly dependent, then $\det(A) = 0$.

(iii) $\det(\mathrm{Id}) = 1$.

Notation A.39. If F is a field, $n \in \mathbb{N}$, and $\det : \mathcal{M}(n,F) \longrightarrow F$ is a determinant, then, for $A = (a_{ji}) \in \mathcal{M}(n,F)$, one commonly uses the notation

$$|A| := \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{vmatrix} := \det(A). \tag{A.66}$$

—

In Th. A.45 below, it will be stated that, for each $n \in \mathbb{N}$, there exists a unique determinant. To also state an explicit formula for this determinant, we need to know a few things about permutations.

Definition and Remark A.40. Let $n \in \mathbb{N}$. Each bijective map $\pi : \{1,\ldots,n\} \longrightarrow \{1,\ldots,n\}$ is called a *permutation* of $\{1,\ldots,n\}$. The set of permutations of $\{1,\ldots,n\}$ forms a group with respect to the composition of maps, the so-called *symmetric group* S_n: Indeed, the composition of maps is associative by [Phi15a, Prop. 2.9(a)]; the neutral element is the identity map $e : \{1,\ldots,n\} \longrightarrow \{1,\ldots,n\}$, $e(i) = i$; and, for each $\sigma \in S_n$, its inverse map σ^{-1} is also its inverse element in the group S_n. Caveat: Simple examples show that S_n is *not* commutative.

Definition A.41. Let $k,n \in \mathbb{N}$, $k \leq n$. A permutation $\pi \in S_n$ is called a k-cycle if, and only if, there exist k distinct numbers $i_1,\ldots,i_k \in \{1,\ldots,n\}$ such that

$$\pi(i) = \begin{cases} i_{j+1} & \text{if } j \in \{1,\ldots,k-1\}, \\ i_1 & \text{if } i = i_k, \\ i & \text{if } i \notin \{i_1,\ldots,i_k\}. \end{cases} \tag{A.67}$$

If π is a cycle as in (A.67), then one writes

$$\pi = (i_1\ i_2\ \ldots\ i_k). \tag{A.68}$$

Each 2-cycle is also known as a *transposition*.

Theorem A.42. *Let $n \in \mathbb{N}$.*

(a) *Each permutation can be decomposed into finitely many disjoint cycles: For each $\pi \in S_n$ there exists a decomposition of $\{1, \dots, n\}$ into disjoint sets $A_1, \dots, A_N$, $N \in \mathbb{N}$, i.e.*

$$\{1, \dots, n\} = \bigcup_{i=1}^{N} A_i \quad \text{and} \quad A_i \cap A_j = \emptyset \quad \text{for } i \neq j, \tag{A.69}$$

such that A_i consists of the distinct elements $a_{i1}, \dots, a_{i,N_i}$ and

$$\pi = (a_{N1} \ \dots \ a_{N,N_N}) \cdots (a_{11} \ \dots \ a_{1,N_1}). \tag{A.70}$$

The decomposition (A.70) *is unique up to the order of the cycles.*

(b) *If $n \geq 2$, then every permutation $\pi \in S_n$ is the composition of finitely many transpositions, where each transposition permutes two juxtaposed elements, i.e.*

$$\underset{\pi \in S_n}{\forall} \quad \underset{N \in \mathbb{N}}{\exists} \quad \underset{\tau_1, \dots, \tau_N \in T}{\exists} \quad \pi = \tau_N \circ \cdots \circ \tau_1, \tag{A.71}$$

where $T := \big\{(i \ \ i+1) : i \in \{1, \dots, n-1\}\big\}$.

Proof. (a): We prove the statement by induction on n. For $n = 1$, there is nothing to prove. Let $n > 1$ and choose $i \in \{1, \dots, n\}$. We claim that

$$\underset{k \in \mathbb{N}}{\exists} \quad \left(\pi^k(i) = i \ \wedge \underset{l \in \{1, \dots, k-1\}}{\forall} \quad \pi^l(i) \neq i \right). \tag{A.72}$$

Indeed, since $\{1, \dots, n\}$ is finite, there must be a smallest $k \in \mathbb{N}$ such that $\pi^k(i) \in A_1 := \{i, \pi(i), \dots, \pi^{k-1}(i)\}$. Since π is bijective, it must be $\pi^k(i) = i$ and $(i \ \pi(i) \ \dots, \pi^{k-1}(i))$ is a k-cycle. We are already done in case $k = n$. If $k < n$, then consider $B := \{1, \dots, n\} \setminus A_1$. Then, again using the bijectivity of π, $\pi \restriction_B$ is a permutation on B with $1 \leq \#B < n$. By induction, there are disjoint sets $A_2, \dots, A_N$ such that $B = \bigcup_{j=2}^{N} A_j$, A_j consists of the distinct elements $a_{j1}, \dots, a_{j,N_j}$ and

$$\pi \restriction_B = (a_{N1} \ \dots \ a_{N,N_N}) \cdots (a_{21} \ \dots \ a_{2,N_2}).$$

Since $\pi = (i \ \pi(i) \ \dots, \pi^{k-1}(i)) \circ \pi \restriction_B$, this finishes the proof of (A.70). If there were another, different, decomposition of π into cycles, say, given by disjoint sets $B_1, \dots, B_M$, $\{1, \dots, n\} = \bigcup_{i=1}^{M} B_i$, $M \in \mathbb{N}$, then there were $A_i \neq B_j$ and $k \in A_i \cap B_j$. But then k were in the cycle given by A_i and in the cycle given by B_j, implying $A_i = \{\pi^l(k) : l \in \mathbb{N}\} = B_j$, in contradiction to $A_i \neq B_j$.

(b): We first show that every $\pi \in S_n$ is a composition of finitely many transpositions (not necessarily transpositions from the set T): According to (a), it suffices to show that every cycle is a composition of finitely many transpositions. Since each 1-cycle is

the identity, it is $(i) = \mathrm{Id} = (1\ 2)\,(1\ 2)$ for each $i \in \{1, \ldots, n\}$. If $(i_1\ \ldots\ i_k)$ is a k-cycle, $k \in \{2, \ldots, n\}$, then

$$(i_1\ \ldots\ i_k) = (i_1\ i_2)\,(i_2\ i_3)\cdots(i_{k-1}\ i_k)\,: \tag{A.73}$$

Indeed,

$$\underset{i\in\{1,\ldots,n\}}{\forall} \quad (i_1\ i_2)\,(i_2\ i_3)\cdots(i_{k-1}\ i_k)(i) = \begin{cases} i_1 & \text{for } i = i_k, \\ i_{l+1} & \text{for } i = i_l,\ l \in \{1, \ldots, k-1\}, \\ i & \text{for } i \notin \{i_1, \ldots, i_k\}, \end{cases} \tag{A.74}$$

proving (A.73). To finish the proof of (b), we observe that every transposition is a composition of finitely many elements of T: If $i, j \in \{1, \ldots, n\}$, $i < j$, then

$$(i\ j) = (i\ \ i+1)\cdots(j-2\ j-1)(j-1\ \ j)\cdots(i+1\ \ i+2)(i\ \ i+1)\,: \tag{A.75}$$

Indeed,

$$\begin{aligned} &\underset{k\in\{1,\ldots,n\}}{\forall} \quad (i\ \ i+1)\cdots(j-2\ j-1)(j-1\ \ j)\cdots(i+1\ \ i+2)(i\ \ i+1)(k) \\ &\qquad = \begin{cases} j & \text{for } k = i, \\ i & \text{for } k = j, \\ k & \text{for } i < k < j, \\ k & \text{for } k \notin \{i, i+1, \ldots, j\}, \end{cases} \end{aligned} \tag{A.76}$$

proving (A.75). ■

Definition A.43. Let $n \in \mathbb{N}$. For each permutation $\pi \in S_n$, one defines its *sign*, $\mathrm{sgn}(\pi)$, via the map

$$\mathrm{sgn}:\ S_n \longrightarrow \{-1, 1\}, \quad \mathrm{sgn}(\pi) := \prod_{1\le i<j\le n} \frac{\pi(i) - \pi(j)}{i - j}. \tag{A.77}$$

Note that, for $n = 1$, $\mathrm{sgn}:\ S_1 = \{e\} \longrightarrow \{-1, 1\}$, $\mathrm{sgn}(e) = 1$, as the product in (A.77) is empty.

Proposition A.44. *Let $n \in \mathbb{N}$.*

(a) *The sign is well-defined by* (A.77)*, i.e. the map is, indeed, $\{-1, 1\}$-valued.*

(b) *The function* $\mathrm{sgn}:\ S_n \longrightarrow \{-1, 1\}$ *is a group homomorphism (note that $\{-1, 1\}$ forms a multiplicative subgroup of $\mathbb{R}$), i.e.*

$$\underset{\pi_1,\pi_2\in S_n}{\forall} \quad \mathrm{sgn}(\pi_1 \circ \pi_2) = \mathrm{sgn}(\pi_1)\,\mathrm{sgn}(\pi_2). \tag{A.78}$$

(c) *For $n \geq 2$, if a permutation $\pi \in S_n$ is the composition of k transpositions, then the parity of k is uniquely determined by π (i.e., for a given π, k is either always even or always odd) and*

$$\operatorname{sgn}(\pi) = (-1)^k = \begin{cases} 1 & \text{if } k \text{ is even,} \\ -1 & \text{if } k \text{ is odd.} \end{cases} \tag{A.79}$$

Proof. (a): The map sgn is $\{-1, 1\}$-valued, since the bijectivity of $\pi \in S_n$ implies that the factor $i - j$ appears in the denominator of $\operatorname{sgn}(\pi)$ as defined in (A.77) if, and only if, the factor $i - j$ or the factor $j - i$ appears in the numerator.

(b): Let $\pi_1, \pi_2 \in S_n$. One computes

$$\begin{aligned} \operatorname{sgn}(\pi_1 \circ \pi_2) &= \prod_{1 \leq i < j \leq n} \frac{\pi_1(\pi_2(i)) - \pi_1(\pi_2(j))}{i - j} \\ &= \prod_{1 \leq i < j \leq n} \left(\frac{\pi_1(\pi_2(i)) - \pi_1(\pi_2(j))}{\pi_2(i) - \pi_2(j)} \cdot \frac{\pi_2(i) - \pi_2(j)}{i - j} \right) \\ &\overset{\pi_2 \text{ bij.}}{=} \operatorname{sgn}(\pi_1) \operatorname{sgn}(\pi_2). \end{aligned} \tag{A.80}$$

(c): If $\tau \in S_n$ is a transposition, then there are elements $i, j \in \{1, \ldots, n\}$ such that $i < j$ and $\tau = (i\ j)$. Thus,

$$\operatorname{sgn}(\tau) = \frac{\tau(i) - \tau(j)}{i - j} = \frac{j - i}{i - j} = -1 \tag{A.81}$$

holds for every transposition τ. In consequence, if $\pi \in S_n$ is the composition of k transpositions, $k \in \mathbb{N}$, then (A.79) must hold and, in particular, k is always even if $\operatorname{sgn}(\pi) = 1$ and k is always odd if if $\operatorname{sgn}(\pi) = -1$. ∎

Theorem A.45. *Let F be a field. For each $n \in \mathbb{N}$, there exists a unique determinant, i.e. there is a unique map* $\det : \mathcal{M}(n, F) \longrightarrow F$, *satisfying* (i) – (iii) *of Def.* A.38. *Moreover, this map is given by*

$$\det : \mathcal{M}(n, F) \longrightarrow F, \quad \det\big((a_{ji})\big) := \sum_{\pi \in S_n} \operatorname{sgn}(\pi)\, a_{1\pi(1)} \cdots a_{n\pi(n)}. \tag{A.82}$$

Proof. See, e.g., [Str08, Th. 17.5, Th. 17.11(a)]. ∎

Definition A.46. Let F be a field, $n \in \mathbb{N}$. A quadratic matrix $A \in \mathcal{M}(n, F)$ is called *invertible* or *regular* if, and only if,

$$\exists_{B \in \mathcal{M}(n,F)} \quad AB = \mathrm{Id}\,. \tag{A.83}$$

One then usually writes A^{-1} instead of B and calls A^{-1} the *inverse matrix* of A. If A is not regular, then it is called *singular*.

Remark A.47. If V is a finite dimensional vector space over a field F, and $\{v_1, \ldots, v_n\}$ is a basis of V, $n \in \mathbb{N}$, then, due to Th. A.32, a linear map $A \in \mathcal{L}(V, V)$ is bijective if, and only if, its transformation matrix $I(V)$ with respect to the given basis is invertible.

—

Important properties of the determinant are compiled in the following Th. A.48.

Theorem A.48. *Let F be a field, $n \in \mathbb{N}$, let $A \in \mathcal{M}(n, F)$, and let $c_1, \ldots, c_n$ denote the columns of A, whereas $r_1, \ldots, r_n$ denote the rows of A, i.e.*

$$A = (c_1, \ldots, c_n) = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}. \tag{A.84}$$

(a) $\det(A) = 0$ *if, and only if, A is singular. If A is invertible, then* $\det(A^{-1}) = (\det(A))^{-1}$.

(b) *If $B \in \mathcal{M}(n, F)$, then* $\det(AB) = \det(A)\det(B)$.

(c) $\det(A^{\mathrm{t}}) = \det(A)$.

(d) *If $\lambda \in F$, then* $\det(\lambda A) = \lambda^n \det(A)$.

(e) *The value of the determinant remains the same if one column of a matrix is replaced by the sum of that column and a scalar multiple of another column. More generally, the determinant remains the same if one column of a matrix is replaced by the sum of that column and a linear combination of the other columns, i.e., if $\lambda_1, \ldots, \lambda_n \in F$ and $i \in \{1, \ldots, n\}$, then*

$$\det(A) = \det(c_1, \ldots, c_n) = \det\left(c_1, \ldots, c_{i-1}, c_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} \lambda_j\, c_j, c_{i+1}, \ldots, c_n\right). \tag{A.85}$$

(f) *Switching columns i and j, where $i, j \in \{1, \ldots, n\}$, $i \neq j$, changes the sign of the determinant, i.e.*

$$\det(c_1, \ldots, c_i, \ldots, c_j, \ldots, c_n) = -\det(c_1, \ldots, c_j, \ldots, c_i, \ldots, c_n). \tag{A.86}$$

(g) det *is* multilinear *with regard to matrix rows, i.e., for each $b \in \mathcal{M}(1, n, F)$, $i \in \{1, \ldots, n\}$, and $\lambda, \mu \in F$:*

$$\det\begin{pmatrix} r_1 \\ \vdots \\ r_{i-1} \\ \lambda\, r_i + \mu\, b \\ r_{i+1} \\ \vdots \\ r_n \end{pmatrix} = \lambda\, \det(A) + \mu\, \det\begin{pmatrix} r_1 \\ \vdots \\ r_{i-1} \\ b \\ r_{i+1} \\ \vdots \\ r_n \end{pmatrix}. \tag{A.87}$$

(h) *The value of the determinant remains the same if one row of a matrix is replaced by the sum of that row and a scalar multiple of another row. More generally, the determinant remains the same if one row of a matrix is replaced by the sum of that row and a linear combination of the other rows, i.e., if* $\lambda_1, \dots, \lambda_n \in F$ *and* $i \in \{1, \dots, n\}$*, then*

$$\det(A) = \det \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \det \begin{pmatrix} r_1 \\ \vdots \\ r_{i-1} \\ r_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} \lambda_j \, r_j \\ r_{i+1} \\ \vdots \\ r_n \end{pmatrix}. \tag{A.88}$$

(i) *Switching rows* i *and* j*, where* $i, j \in \{1, \dots, n\}$*,* $i \neq j$*, changes the sign of the determinant, i.e.*

$$\det \begin{pmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_j \\ \vdots \\ r_n \end{pmatrix} = -\det \begin{pmatrix} r_1 \\ \vdots \\ r_j \\ \vdots \\ r_i \\ \vdots \\ r_n \end{pmatrix}. \tag{A.89}$$

Proof. (a): See, e.g., [Str08, Th. 17.7(b), Th. 17.11(a)].

(b): See, e.g., [Str08, Th. 17.11(b)].

(c): See, e.g., [Str08, Lem. 18.1].

(d) is an immediate consequence of Def. A.38(i).

(e): One computes, for $i < j$,

$$\begin{aligned} &\det\Big(c_1, \dots, c_{i-1}, c_i + \lambda_j \, c_j, c_{i+1}, \dots, c_n\Big) \\ &\overset{\text{Def. A.38(i)}}{=} \lambda_j^{-1} \det\Big(c_1, \dots, c_{i-1}, c_i + \lambda_j \, c_j, c_{i+1}, \dots, \lambda_j \, c_j, \dots, c_n\Big) \\ &\overset{\text{Def. A.38(i)}}{=} \lambda_j^{-1} \det\Big(c_1, \dots, c_{i-1}, c_i, c_{i+1}, \dots, \lambda_j \, c_j, \dots, c_n\Big) \\ &\qquad + \lambda_j^{-1} \det\Big(c_1, \dots, c_{i-1}, \lambda_j \, c_j, c_{i+1}, \dots, \lambda_j \, c_j, \dots, c_n\Big) \\ &\overset{\text{Def. A.38(ii)}}{=} \lambda_j^{-1} \det\Big(c_1, \dots, c_{i-1}, c_i, c_{i+1}, \dots, \lambda_j \, c_j, \dots, c_n\Big) + 0 \\ &\overset{\text{Def. A.38(i)}}{=} \det\Big(c_1, \dots, c_n\Big) = \det(A). \end{aligned} \tag{A.90}$$

The general case of (A.85) then follows by induction.

(f): We compute

$$\begin{aligned}
&\det(c_1,\dots,c_i,\dots,c_j,\dots,c_n)+\det(c_1,\dots,c_j,\dots,c_i,\dots,c_n)\\
&\overset{\text{(e)}}{=} \det(c_1,\dots,c_i+c_j,\dots,c_j,\dots,c_n)+\det(c_1,\dots,c_j+c_i,\dots,c_i,\dots,c_n)\\
&\overset{\text{Def. A.38(i)}}{=} \det(c_1,\dots,c_i+c_j,\dots,c_i+c_j,\dots,c_n) \overset{\text{Def. A.38(ii)}}{=} 0, \qquad \text{(A.91)}
\end{aligned}$$

proving (f).

(g) is inferred by combining Def. A.38(i) with (c).

(h) is inferred by combining (e) with (c).

(i) is inferred by combining (f) with (c). ■

Theorem A.49 (Block Matrices)**.** *The determinant of so-called block matrices, where one block is a zero matrix (all entries* 0*), can be computed as the product of the determinants of the corresponding blocks. More precisely, if* $n, m \in \mathbb{N}$*, then*

$$\begin{vmatrix} a_{11} & \dots & a_{1n} & & & \\ \vdots & \vdots & \vdots & & * & \\ a_{n1} & \dots & a_{nn} & & & \\ 0 & \dots & 0 & b_{11} & \dots & b_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & b_{m1} & \dots & b_{mm} \end{vmatrix} = \det(a_{ji}) \det(b_{ji}). \qquad \text{(A.92)}$$

Proof. See, e.g., [Str08, Th. 18.3]. ■

Definition A.50. Let F be a field, $n \in \mathbb{N}$, $n \geq 2$, $A = (a_{ji}) \in \mathcal{M}(n, F)$. For each $j, i \in \{1, \dots, n\}$, let M_{ji} the $n \times n$ submatrix of A obtained by deleting the jth row and the ith column of A – the M_{ji} are sometimes called the *minor matrices* of A; define

$$A_{ji} := (-1)^{i+j} \det(M_{ij}), \qquad \text{(A.93)}$$

where the A_{ji} are called *cofactors* of A and the $\det(M_{ij})$ are called the *minors* of A. Let $\tilde{A} := (A_{ji})$ denote the matrix of cofactors.

Theorem A.51. *Let* F *be a field,* $n \in \mathbb{N}$*,* $n \geq 2$*,* $A = (a_{ji}) \in \mathcal{M}(n, F)$*. Moreover, let* $\tilde{A} := (A_{ji})$ *be the matrix of cofactors according to Def.* A.50*.*

(a) $A\tilde{A} = (\det A)\,\mathrm{Id}$*.*

(b) *If* $\det A \neq 0$*, then* $\det \tilde{A} = (\det A)^{n-1}$*.*

(c) *If* $\det A \neq 0$*, then* $A^{-1} = (\det A)^{-1}\,\tilde{A}$*.*

(d) Laplace Expansion by Rows: $\det A = \sum_{j=1}^{n} a_{ij} A_{ji}$ *(expansion with respect to the ith row).*

(e) Laplace Expansion by Columns: $\det A = \sum_{j=1}^{n} A_{ij} a_{ji}$ *(expansion with respect to the ith column).*

Proof. See, e.g., [Str08, Th. 18.6]. ■

Theorem A.52. *Let F be a field, $n \in \mathbb{N}$, let V be an n-dimensional vector space over F, and $A \in \mathcal{L}(V,V)$. Moreover, let $B_1 = \{v_1, \ldots, v_n\}$ and $B_2 = \{w_1, \ldots, w_n\}$ be bases of V. If $M = (m_{ji})$ is the transformation matrix corresponding to A with respect to B_1 and $N = (n_{ji})$ is the transformation matrix corresponding to A with respect to B_2 (i.e., for each $i \in \{1, \ldots, n\}$, $A(v_i) = \sum_{j=1}^{n} m_{ji} v_j$ and $A(w_i) = \sum_{j=1}^{n} n_{ji} w_j$, cf. Def. and Rem.* A.33*), then* $\det(M) = \det(N)$.

Proof. See, e.g., [Str08, Th. 17.11(a)]. ■

Definition and Remark A.53. Let F be a field, $n \in \mathbb{N}$, and let V be an n-dimensional vector space over F. Then Th. A.52 allows to define a determinant function for linear maps by

$$\det : \mathcal{L}(V,V) \longrightarrow F, \quad \det(A) := \det(M), \tag{A.94}$$

where M is a transformation matrix for A with respect to an arbitrary basis of V. Then Th. A.32 shows that Th. A.48(a),(b),(d) yield the following properties of the new determinant function defined in (A.94):

(a) If $A \in \mathcal{L}(V,V)$, then $\det(A) = 0$ if, and only if, A is not bijective. If A is bijective, then $\det(A^{-1}) = (\det(A))^{-1}$.

(b) If $A, B \in \mathcal{M}(n,F)$, then $\det(AB) = \det(A)\det(B)$.

(c) If $A \in \mathcal{L}(V,V)$ and $\lambda \in F$, then $\det(\lambda A) = \lambda^n \det(A)$. In particular, det is *not* linear for $n > 1$.

In [Str08, §17], the author actually defines the determinant first for linear maps $A \in \mathcal{L}(V,V)$, establishes properties including the above properties, and only then defines the determinant function for square matrices. Several alternative, but equivalent, approaches are possible and can be found in the literature.

B Metric Spaces

B.1 Metric Subspaces

Definition B.1. If (X,d) is a metric space, $M \subseteq X$, then (M,d) is called a *metric subspace* of (X,d) (if d is understood, one also speaks of M as a metric subspace of X). Thus, the metric on the subspace M is just the metric on X restricted to M.

Remark B.2. One sees immediately that a metric subspace (M,d) of a metric space (X,d) is, indeed, a metric space: Since d satisfies the laws (i) – (iii) from Def. 1.17 for all $x, y, z \in X$, in particular, d satisfies the same laws for all $x, y, z \in M \subseteq X$.

Definition B.3. Let (X, d) be a metric space, and let (M, d) be a metric subspace of (X, d). Is $A \subseteq M$ open with respect to (M, d), then one says that A is *open* in M or *M-open* or *relatively open*. For $A \subseteq M$ closed with respect to (M, d), one introduces analogous terms. Moreover, for $x \in M$, $r > 0$, call

$$B_{r,M}(x) := M \cap B_r(x) = \{y \in M : d(x, y) < r\}, \tag{B.1}$$

the *open M-ball* with radius r and center x.

Caveat B.4. One has to use care when working with a subspace (M, d) of a metric space (X, d): As will be seen in Ex. B.5, the notions and properties with respect to M are in general very different from the corresponding notions and properties with respect to X. For example, a set that is M-open might not be X-open and a set that is M-closed might not be X-closed!

Example B.5. (a) If (M, d) is a metric subspace of a metric space (X, d), then, according to Lem. 1.27(b), M is always both M-open and M-closed (irrespective of M being X-open or X-closed).

(b) Let $X = \mathbb{R}$ with the usual metric, i.e. $d(x, y) = |x - y|$ for each $x, y \in \mathbb{R}$. Let $M = [0, 1]$. According to (a), is both M-closed and M-open, even though $[0, 1]$ is not open in X. When noting before that $\mathbb{Q}$ and $]0, 1]$ are metric spaces that are not complete, we already considered metric subspaces of $\mathbb{R}$ without making use of the term subspace. If $M =]0, 1]$, then, again, M is both M-closed and M-open, even though $]0, 1]$ is neither closed nor open in X. Moreover, $]0, \frac{1}{2}]$ is M-closed (but not X-closed) and $]\frac{1}{2}, 1]$ is M-open (but not X-open).

Proposition B.6. *Let (M, d) be a metric subspace of a metric space (X, d).*

(a) *A subset A of M is M-open if, and only if, there is a set $O \subseteq X$ which is X-open and $A = O \cap M$.*

(b) *A subset A of M is M-closed if, and only if, there is a set $C \subseteq X$ which is X-closed and $A = C \cap M$.*

Proof. (a): Suppose A is M-open. Then, for each $a \in A$, there is $\epsilon_a > 0$ such that the open M-ball $B_{\epsilon_a,M}(a)$ is contained in A, i.e.

$$B_{\epsilon_a,M}(a) = M \cap B_{\epsilon_a}(a) \subseteq A. \tag{B.2}$$

Let $O := \bigcup_{a \in A} B_{\epsilon_a}(a)$. Then O is X-open by Th. 1.29(a). Moreover,

$$O \cap M = \bigcup_{a \in A} \big(M \cap B_{\epsilon_a}(a)\big) = \bigcup_{a \in A} B_{\epsilon_a,M}(a) = A, \tag{B.3}$$

where the last equality is due to (B.2) and the fact that $a \in A$ implies $a \in B_{\epsilon_a,M}(a)$.

Conversely, if $O \subseteq X$ is X-open and $A = O \cap M$, then each $a \in A$ is an X-interior point of O, i.e. there is $\epsilon > 0$ such that the open X-ball $B_\epsilon(a)$ is contained in O, i.e.

$B_\epsilon(a) \subseteq O$. Intersecting with M yields $B_{\epsilon,M}(a) = M \cap B_\epsilon(a) \subseteq M \cap O = A$, i.e. the open M-ball $B_{\epsilon,M}(a)$ is contained in A, showing that a is an M-interior point of A. As a was an arbitrary point of A, A is M-open.

(b): If A is M-closed, then $M \setminus A$ is M-open. According to (a), there is an X-open set $O \subseteq X$ such that $M \setminus A = M \cap O$. Then $C := X \setminus O$ is an X-closed set and $M \cap C = M \cap (X \setminus O) = M \setminus (M \cap O) = M \setminus (M \setminus A) = A$.

Conversely, if there is an X-closed set $C \subseteq X$ with $A = C \cap M$, then $O := X \setminus C$ is an X-open set satisfying $O \cap M = M \cap (X \setminus C) = M \setminus (C \cap M) = M \setminus A$. Thus, according to (a), $M \setminus A$ is M-open, i.e. A is M-closed. ■

B.2 Norm-Preserving and Isometric Maps

Definition B.7. (a) Given normed vector spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ over $\mathbb{K}$, a function $f : X \longrightarrow Y$ is called *norm-preserving* if, and only if,

$$\big\|f(x)\big\|_Y = \|x\|_X \quad \text{for each } x \in X. \tag{B.4}$$

(b) Given metric spaces (X, d_X) and (Y, d_Y), a function $f : X \longrightarrow Y$ is called *distance-preserving* or *isometric* if, and only if,

$$d_Y\big(f(x), f(y)\big) = d_X(x, y) \quad \text{for each } x, y \in X. \tag{B.5}$$

Lemma B.8. *Given normed vector spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ over $\mathbb{K}$, a $\mathbb{K}$-linear function $f : X \longrightarrow Y$ is norm-preserving if, and only if, f is isometric with respect to the induced metrics.*

Proof. The function f is norm-preserving if, and only if, $\big\|f(x)\big\|_Y = \|x\|_X$ for each $x \in X$. This, in turn is the case if, and only if,

$$\big\|f(x) - f(y)\big\|_Y = \big\|f(x - y)\big\|_Y = \|x - y\|_X \quad \text{for each } x, y \in X, \tag{B.6}$$

where it was used that f is linear. As (B.6) states that f is isometric with respect to the induced metrics, the proof is complete. ■

The following examples show that the assertion of Lem. B.8 becomes false if the word "linear" is omitted.

Example B.9. (a) Let $(X, \|\cdot\|_X)$ be a normed vector space over $\mathbb{K}$, and $f : X \longrightarrow \mathbb{K}$, $f(x) := \|x\|_X$. If we take $\|\cdot\|_Y$ to be the usual norm on $\mathbb{K}$, i.e. $\|y\|_Y := |y|$, then, for each $x \in X$, $\big\|f(x)\big\|_Y = \big|\|x\|_X\big| = \|x\|_X$, i.e. f is norm-preserving. However, if $\dim X > 0$ (i.e. if $X \neq \{0\}$), then f is *not* isometric with respect to the induced metrics: Take any $0 \neq x \in X$. One computes

$$\big\|f(x) - f(-x)\big\|_Y = \big|\|x\|_X - \|x\|_X\big| = 0 \neq \|x - (-x)\|_X = 2\|x\|_X. \tag{B.7}$$

(b) Consider $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$, where $X = Y = \mathbb{K}$ and $\|x\|_X = \|x\|_Y = |x|$ for each $x \in \mathbb{K}$. Then $f : X \longrightarrow Y$, $f(x) := 1 + x$, is isometric due to $|f(x) - f(y)| = |1 + x - (1 + y)| = |x - y|$, but f is *not* norm-preserving, since $0 = |0| \neq |f(0)| = 1$.

Lemma B.10. *Isometric functions between metric spaces are one-to-one (in particular, isometric functions between normed spaces are one-to-one).*

Proof. Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \longrightarrow Y$ be an isometric function, i.e. $d_Y\big(f(x), f(y)\big) = d_X(x, y)$ for each $x, y \in X$. If $x \neq y$, then $0 \neq d_X(x, y) = d_Y\big(f(x), f(y)\big)$. Thus, $f(x) \neq f(y)$, showing that f is one-to-one. ■

Example B.11. If a function between normed spaces is just norm-preserving, but not isometric, then this function is not necessarily one-to-one: To see this, we reemploy the function f from Ex. B.9(a), i.e. let $(X, \|\cdot\|_X)$ be a normed vector space over $\mathbb{K}$, $\dim X > 0$, and $f : X \longrightarrow \mathbb{K}$, $f(x) := \|x\|_X$. In Ex. B.9(a), we saw that f is norm-preserving, but not isometric. Since, for $x \neq 0$, one has $x \neq -x$, but $f(x) = \|x\|_X = f(-x)$, f is not one-to-one.

Remark B.12. If $(X, \|\cdot\|)$ is a normed space, d is the induced metric, and $M \subseteq X$, then (M, d) can be considered as the metric subspace of (X, d) according to Def. B.3. Thus, every subset of a normed space is turned into a metric space in a natural way. It is quite remarkable that actually *every* metric space arises in this way. That means, given any metric space (M, d), there exists a normed space $(X, \|\cdot\|)$ and an isometric (in particular, one-to-one) function $f : M \longrightarrow X$: One can choose X as the $\mathbb{R}$-vector space of bounded functions from M into $\mathbb{R}$ with the sup-norm (for $F \in X$, define $\|F\| := \sup\{|F(x)| : x \in M\}$) and $f : M \longrightarrow X$, $f(x) = f_x$, where $f_x : M \longrightarrow \mathbb{R}$, $f_x(y) = d(x, y) - d(x_0, y)$ with some fixed $x_0 \in M$. However, the normed space X can be very large (i.e. much larger than M), and, thus, in practice, it is not always useful to study X in order to learn more about the metric space M.

B.3 Uniform Continuity and Lipschitz Continuity

This section provides some additional important results regarding uniformly continuous functions (see Def. 1.49(b)) and Lipschitz continuous functions (see Def. 1.49(c)). We start with an auxiliary result:

Lemma B.13. *If f, g are real-valued functions on a set X, i.e. if $f, g : X \longrightarrow \mathbb{R}$, then, for each $x, y \in X$,*

$$\big|\max(f, g)(x) - \max(f, g)(y)\big| \leq \max\big\{|f(x) - f(y)|, |g(x) - g(y)|\big\}, \tag{B.8a}$$

$$\big|\min(f, g)(x) - \min(f, g)(y)\big| \leq \max\big\{|f(x) - f(y)|, |g(x) - g(y)|\big\}. \tag{B.8b}$$

Proof. By possibly switching the names of f and g, one can assume, without loss of generality, that $\max(f, g)(x) = f(x)$, i.e. $g(x) \leq f(x)$. If $g(y) \leq f(y)$ as well, then

$\big|\max(f,g)(x) - \max(f,g)(y)\big| = |f(x) - f(y)|$ and $\big|\min(f,g)(x) - \min(f,g)(y)\big| = |g(x) - g(y)|$, i.e. (B.8) is true. If $g(y) > f(y)$, then

$$\big|\max(f,g)(x) - \max(f,g)(y)\big| = |f(x) - g(y)| \leq \begin{cases} \leq |g(x) - g(y)| & \text{for } f(x) \leq g(y), \\ < f(x) - f(y) & \text{for } f(x) > g(y), \end{cases} \tag{B.9a}$$

$$\big|\min(f,g)(x) - \min(f,g)(y)\big| = |g(x) - f(y)| \leq \begin{cases} < |g(x) - g(y)| & \text{for } g(x) \leq f(y), \\ \leq f(x) - f(y) & \text{for } g(x) > f(y), \end{cases} \tag{B.9b}$$

showing that (B.8) holds in all cases. ■

Theorem B.14. *Let (X,d) be a metric space (e.g. a normed space), $(Y, \|\cdot\|)$ a normed vector space over $\mathbb{K}$, and assume that $f, g : X \longrightarrow Y$ are uniformly continuous. Then $f+g$ and λf are uniformly continuous for each $\lambda \in \mathbb{K}$, i.e. the set of all uniformly continuous functions from X into Y constitutes a subspace of the vector space $\mathcal{F}(X,Y)$ over $\mathbb{K}$. Moreover, if $Y = \mathbb{K} = \mathbb{R}$, then $\max(f,g)$, $\min(f,g)$, f^+, f^-, $|f|$ are all uniformly continuous.*

Proof. As f and g are uniformly continuous, given $\epsilon > 0$, there exist $\delta_f > 0$ and $\delta_g > 0$ such that, for each $x, y \in X$,

$$d(x,y) < \delta_f \;\Rightarrow\; \big\|f(x) - f(y)\big\| < \epsilon/2, \tag{B.10a}$$

$$d(x,y) < \delta_g \;\Rightarrow\; \big\|g(x) - g(y)\big\| < \epsilon/2. \tag{B.10b}$$

Thus, if $d(x,y) < \min\{\delta_f, \delta_g\}$, then

$$\big\|(f+g)(x) - (f+g)(y)\big\| \leq \big\|f(x) - f(y)\big\| + \big\|g(x) - g(y)\big\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \tag{B.10c}$$

showing that $f+g$ is uniformly continuous. Next, if $\lambda = 0$, then $\lambda f \equiv 0$, and obviously uniformly continuous. For $\lambda \neq 0$, choose $\delta > 0$ such that $d(x,y) < \delta$ implies $\|f(x) - f(y)\| < \epsilon/|\lambda|$. Then

$$\big\|(\lambda f)(x) - (\lambda f)(y)\big\| = |\lambda|\big\|f(x) - f(y)\big\| < |\lambda|\,\frac{\epsilon}{|\lambda|} = \epsilon, \tag{B.10d}$$

showing that λf is uniformly continuous. If $Y = \mathbb{K} = \mathbb{R}$, then $d(x,y) < \min\{\delta_f, \delta_g\}$ together with Lem. B.13 implies

$$\big|\max(f,g)(x) - \max(f,g)(y)\big| < \epsilon/2 < \epsilon, \tag{B.10e}$$

$$\big|\min(f,g)(x) - \min(f,g)(y)\big| < \epsilon/2 < \epsilon, \tag{B.10f}$$

showing the uniform continuity of $\max(f,g)$ and $\min(f,g)$ and, in turn, also of f^+, f^-, and $|f|$. ■

Theorem B.15. *Let (X,d) be a metric space (e.g. a normed space), $(Y,\|\cdot\|)$ a normed vector space over $\mathbb{K}$, and assume that $f,g : X \longrightarrow Y$ are Lipschitz continuous. Then $f+g$ and λf are Lipschitz continuous for each $\lambda \in \mathbb{K}$, i.e. the set $\operatorname{Lip}(X,Y)$ constitutes a subspace of the vector space $\mathcal{F}(X,Y)$ over $\mathbb{K}$. Moreover, if $Y = \mathbb{K} = \mathbb{R}$, then $\max(f,g)$, $\min(f,g)$, f^+, f^-, $|f|$ are all Lipschitz continuous.*

Proof. As f and g are Lipschitz continuous, there exist $L_f \geq 0$ and $L_g \geq 0$ such that, for each $x, y \in X$,

$$\big\|f(x) - f(y)\big\| \leq L_f\, d(x,y), \tag{B.11a}$$
$$\big\|g(x) - g(y)\big\| \leq L_g\, d(x,y). \tag{B.11b}$$

Thus,

$$\begin{aligned}\big\|(f+g)(x) - (f+g)(y)\big\| &\leq \big\|f(x) - f(y)\big\| + \big\|g(x) - g(y)\big\| \\ &\leq L_f\, d(x,y) + L_g\, d(x,y) = (L_f + L_g) d(x,y),\end{aligned} \tag{B.11c}$$

showing that $f+g$ is Lipschitz continuous with Lipschitz constant $L_f + L_g$. Next, for $\lambda \in \mathbb{K}$,

$$\big\|(\lambda f)(x) - (\lambda f)(y)\big\| = |\lambda| \big\|f(x) - f(y)\big\| \leq |\lambda| L_f\, d(x,y), \tag{B.11d}$$

showing that λf is Lipschitz continuous with Lipschitz constant $|\lambda| L_f$. For $Y = \mathbb{K} = \mathbb{R}$, Lem. B.13 shows $\max(f,g)$ and $\min(f,g)$ are Lipschitz continuous with Lipschitz constant $\max\{L_f, L_g\}$, f^+ and f^- are Lipschitz continuous with Lipschitz constant L_f, and $|f|$ is Lipschitz continuous with Lipschitz constant $2L_f$. ∎

Caveat B.16. Products and quotients of uniformly continuous functions are not necessarily uniformly continuous; products and quotients of Lipschitz continuous functions are not necessarily Lipschitz continuous: Even though $f \equiv 1$ and $g(x) = x$ are Lipschitz continuous, it was shown in Examples 1.52(a),(b), respectively, that f/g and g^2 are not even uniformly continuous on $\mathbb{R}^+$.

B.4 Viewing $\mathbb{C}^n$ as $\mathbb{R}^{2n}$

Remark B.17. Recall that the set of complex numbers $\mathbb{C}$ is *defined* to be $\mathbb{R}^2$, where the imaginary unit is $i := (0,1) \in \mathbb{R}^2$, which allows to write each $z = (x,y) \in \mathbb{C} = \mathbb{R}^2$ as $z = x + iy$, where $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$. This, for each $n \in \mathbb{N}$, gives rise to the $\mathbb{R}$-linear bijective map

$$I : \mathbb{C}^n \longrightarrow \mathbb{R}^{2n}, \quad I\big((x_1,y_1),\ldots,(x_n,y_n)\big) := (x_1,y_1,\ldots,x_n,y_n), \tag{B.12}$$

allowing to canonically identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$.

—

The identification (B.12) allows the identification of metric structures on $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ due to the following general result:

Proposition B.18. *Let X, Y be sets, let $d : X \times X \longrightarrow \mathbb{R}_0^+$ be a metric on X, and let $I : X \longrightarrow Y$ be bijective. Then*

$$d_Y : Y \times Y \longrightarrow \mathbb{R}_0^+, \quad d_Y(x,y) := d\big(I^{-1}(x), I^{-1}(y)\big), \tag{B.13}$$

defines a metric on Y such that (X,d) and (Y,d_Y) are isometric (with the map I providing the isometry).

Proof. Let $x, y, z \in Y$. Then

$$d_Y(x,y) = 0 \quad \Leftrightarrow \quad d\big(I^{-1}(x), I^{-1}(y)\big) = 0 \quad \Leftrightarrow \quad I^{-1}(x) = I^{-1}(y) \quad \Leftrightarrow \quad x = y, \tag{B.14}$$

showing that d_Y is positive definite. Moreover,

$$d_Y(x,y) = d\big(I^{-1}(x), I^{-1}(y)\big) = d\big(I^{-1}(y), I^{-1}(x)\big) = d_Y(y,x), \tag{B.15}$$

showing d_Y is symmetric. Finally,

$$\begin{aligned} d_Y(x,z) &= d\big(I^{-1}(x), I^{-1}(z)\big) \leq d\big(I^{-1}(x), I^{-1}(y)\big) + d\big(I^{-1}(y), I^{-1}(z)\big) \\ &= d_Y(x,y) + d_Y(y,z), \end{aligned} \tag{B.16}$$

proving the triangle inequality for d_Y and completing the proof that d_Y constitutes a metric. That I provides an isometry between (X,d) and (Y,d_Y) is immediate from (B.13). ■

Corollary B.19. *Let $n \in \mathbb{N}$, let $d : \mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{R}_0^+$ be a metric, and let I be the map from* (B.12)*. Then*

$$d_{\mathrm{r}} : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \longrightarrow \mathbb{R}_0^+, \quad d_{\mathrm{r}}(x,y) := d\big(I^{-1}(x), I^{-1}(y)\big), \tag{B.17}$$

defines a metric on $\mathbb{R}^{2n}$ such that $(\mathbb{C}^n, d)$ and $(\mathbb{R}^{2n}, d_{\mathrm{r}})$ are isometric (with the map I providing the isometry). Moreover, the map $d \mapsto d_{\mathrm{r}}$ is bijective between the set of metrics on $\mathbb{C}^n$ and the set of metrics on $\mathbb{R}^{2n}$. ■

Proposition B.20. *Let $n \in \mathbb{N}$. If $\|\cdot\|$ constitutes a norm on the vector space $\mathbb{C}^n$ over $\mathbb{C}$ in the sense of Def.* 1.19*, then*

$$\|\cdot\|_{\mathrm{r}} : \mathbb{R}^{2n} \longrightarrow \mathbb{R}_0^+, \quad \|(x_1, y_1, \ldots, x_n, y_n)\|_{\mathrm{r}} := \big\|\big((x_1, y_1), \ldots, (x_n, y_n)\big)\big\| \tag{B.18}$$

defines a norm on the vector space $\mathbb{R}^{2n}$ over $\mathbb{R}$ such that $(\mathbb{C}^n, \|\cdot\|)$ and $(\mathbb{R}^{2n}, \|\cdot\|_{\mathrm{r}})$ are isometric (with the map I from (B.12) *providing the isometry – even more precisely, if d and d_{r} denote the respective induced metrics, then the relation between d and d_{r} is given by* (B.17)*).*

Proof. Exercise. ■

Example B.21. Let $n \in \mathbb{N}$, $p \in [1, \infty]$, and let $\|\cdot\|$ denote the p-norm on the vector space $\mathbb{R}^n$ over $\mathbb{R}$, i.e. $\|x\| := (\sum_{j=1}^n |x_j|^p)^{1/p}$ for $p < \infty$ and $\|x\| = \max\{|x_j| : j = 1, \ldots, n\}$ for $p = \infty$. Then it is an exercise to show

$$\|\cdot\|_{\mathrm{c}} : \mathbb{C}^n \longrightarrow \mathbb{R}_0^+, \quad \|(z_1, \ldots, z_n)\|_{\mathrm{c}} := \|(|z_1|, \ldots, |z_n|)\| \tag{B.19}$$

defines a norm on the vector space $\mathbb{C}^n$ over $\mathbb{C}$.

Remark B.22. As a consequence of Th. 1.95, every norm on the normed vector space $\mathbb{C}^n$ over $\mathbb{C}$ generates precisely the same open subsets of $\mathbb{C}^n$ – in other words, there is only one *norm topology* on $\mathbb{C}^n$. Analogously, there is only one norm topology on $\mathbb{R}^n$ as every norm on the normed vector space $\mathbb{R}^n$ over $\mathbb{R}$ generates precisely the same open subsets of $\mathbb{R}^n$ Moreover, Prop. B.20 shows that the open sets of the norm topology on $\mathbb{C}^n$ are actually precisely the same as the open sets of the norm topology on $\mathbb{R}^{2n}$.

Theorem B.23. *Let $n \in \mathbb{N}$, $A \subseteq \mathbb{C}^n$. Then A is bounded in the normed vector space $\mathbb{C}^n$ over $\mathbb{C}$ if, and only if, A is bounded in the normed vector space $\mathbb{R}^{2n}$ over $\mathbb{R}$.*

Proof. Exercise. ■

B.5 Banach Fixed Point Theorem a.k.a. Contraction Mapping Principle

Definition B.24. Let $\emptyset \neq A$ be a subset of a metric space (X, d). A map $\varphi : A \longrightarrow A$ is called a *contraction* if, and only if, there exists $0 \leq L < 1$ satisfying

$$d\big(\varphi(x), \varphi(y)\big) \leq L\, d(x, y) \quad \text{for each } x, y \in A. \tag{B.20}$$

Remark B.25. According to Def. B.24, $\varphi : A \longrightarrow A$ is a contraction if, and only if, φ is Lipschitz continuous with Lipschitz constant $L < 1$.

The following Th. B.26 constitutes the Banach fixed point theorem. It is also known as the contraction mapping principle. Its proof is surprisingly simple, e.g. about an order of magnitude easier than the proof of the Brouwer fixed point theorem.

Theorem B.26 (Banach Fixed Point Theorem)**.** *Let $\emptyset \neq A$ be a closed subset of a complete metric space (X, d) (for example, a Banach space). If $\varphi : A \longrightarrow A$ is a contraction with Lipschitz constant $0 \leq L < 1$, then φ has a unique fixed point $x_* \in A$. Moreover, for each initial value $x_0 \in A$, the sequence $(x_n)_{n \in \mathbb{N}_0}$, defined by*

$$x_{n+1} := \varphi(x_n) \quad \textit{for each } n \in \mathbb{N}_0, \tag{B.21}$$

converges to x_*:

$$\lim_{n \to \infty} \varphi^n(x_0) = x_*. \tag{B.22}$$

Furthermore, for each such sequence, we have the error estimate

$$d(x_n, x_*) \leq \frac{L}{1 - L}\, d(x_n, x_{n-1}) \leq \frac{L^n}{1 - L}\, d(x_1, x_0) \tag{B.23}$$

for each $n \in \mathbb{N}$.

Proof. We start with uniqueness: Let $x_*, x_{**} \in A$ be fixed points of φ. Then

$$d(x_*, x_{**}) = d\big(\varphi(x_*), \varphi(x_{**})\big) \leq L\, d(x_*, x_{**}), \tag{B.24}$$

which implies $1 \leq L$ for $d(x_*, x_{**}) > 0$. Thus, $L < 1$ implies $d(x_*, x_{**}) = 0$ and $x_* = x_{**}$. Next, we turn to existence. A simple induction on $m - n$ shows

$$\begin{gathered} d(x_{m+1}, x_m) \leq L\, d(x_m, x_{m-1}) \leq L^{m-n}\, d(x_{n+1}, x_n) \\ \text{for each } m, n \in \mathbb{N}_0,\ m > n. \end{gathered} \tag{B.25}$$

This, in turn, allows us to estimate, for each $n, k \in \mathbb{N}_0$:

$$\begin{aligned} d(x_{n+k}, x_n) &\leq \sum_{m=n}^{n+k-1} d(x_{m+1}, x_m) \overset{\text{(B.25)}}{\leq} \sum_{m=n}^{n+k-1} L^{m-n}\, d(x_{n+1}, x_n) \\ &\leq \frac{1}{1-L}\, d(x_{n+1}, x_n) \overset{\text{(B.25)}}{\leq} \frac{L^n}{1-L}\, d(x_1, x_0) \to 0 \quad \text{for } n \to \infty, \end{aligned} \tag{B.26}$$

establishing that $(x_n)_{n\in\mathbb{N}_0}$ constitutes a Cauchy sequence. Since X is complete, this Cauchy sequence must have a limit $x_* \in X$, and since the sequence is in A and A is closed, $x_* \in A$. The continuity of φ allows to take limits in (B.21), resulting in $x_* = \varphi(x_*)$, showing that x_* is a fixed point and proving existence.

Finally, the error estimate (B.23) follows from (B.26) by fixing n and taking the limit for $k \to \infty$. ∎

Example B.27. Suppose, we are looking for a fixed point of the map $\varphi(x) = \cos x$ (or, equivalently, for a zero of $f(x) = \cos x - x$). To apply the Banach fixed point theorem, we need to restrict φ to a set A such that $\varphi(A) \subseteq A$. This is the case for $A := [0, 1]$. Moreover, $\varphi : A \longrightarrow A$ is a contraction, due to $\sin 1 < 1$ and the mean value theorem providing $\tau \in]0, 1[$, satisfying

$$\big|\varphi(x) - \varphi(y)\big| = \big|\varphi'(\tau)\big|\, |x - y| < (\sin 1)|x - y| \tag{B.27}$$

for each $x, y \in A$. Since $\mathbb{R}$ is complete and A is closed in $\mathbb{R}$, the Banach fixed point theorem yields the existence of a unique fixed point $x_* \in [0, 1]$ and $\lim \varphi^n(x_0) = x_*$ for each $x_0 \in [0, 1]$.

B.6 Unit Balls in Normed Spaces

The goal of this section is to prove that a normed vector space is finite-dimensional if, and only if, its closed unit ball is compact (see Th. B.29). In preparation, we show that finite-dimensional normed vector spaces are always closed:

Theorem B.28. *Let $(X, \|\cdot\|)$ be a normed vector space over $\mathbb{K}$. If $U \subseteq X$ is a subspace such that $\dim U = n \in \mathbb{N}$, then U is closed.*

Proof. Let $(b_1, \dots, b_n)$ be a basis of U. Then

$$A : U \longrightarrow \mathbb{K}^n, \quad A\left(\sum_{k=1}^{n} \alpha_k\, b_k\right) := (\alpha_1, \dots, \alpha_n), \tag{B.28}$$

defines a linear isomorphism (cf. Th. A.24). We define a norm on $\mathbb{K}^n$ by letting

$$\|\cdot\| : \mathbb{K}^n \longrightarrow \mathbb{R}_0^+, \quad \|z\| := \left\|A^{-1}(z)\right\|. \tag{B.29}$$

Indeed, (B.29) defines a norm: $\|0\| = \left\|A^{-1}(0)\right\| = \|0\| = 0$; if $z \in \mathbb{K}^n$ and $\|z\| = \left\|A^{-1}(z)\right\| = 0$, then $A^{-1}(z) = 0$, i.e. $z = 0$, showing $\|\cdot\|$ to be positive definite. Moreover

$$\forall_{z\in\mathbb{K}^n} \quad \forall_{\lambda\in\mathbb{K}} \quad \|\lambda z\| = \left\|A^{-1}(\lambda z)\right\| = |\lambda| \left\|A^{-1}(z)\right\| = |\lambda|\, \|z\|, \tag{B.30}$$

showing $\|\cdot\|$ to be homogeneous of degree 1. Finally,

$$\forall_{z,w\in\mathbb{K}^n} \quad \|z + w\| = \left\|A^{-1}(z + w)\right\| \leq \left\|A^{-1}(z)\right\| + \left\|A^{-1}(w)\right\| = \|z\| + \|w\|, \tag{B.31}$$

showing the triangle inequality to hold for $\|\cdot\|$.

Let $(u^k)_{k\in\mathbb{N}}$ be a sequence in U such that $\lim_{k\to\infty} u^k = x \in X$. Then $(u^k)_{k\in\mathbb{N}}$ is a Cauchy sequence and, as A is norm-preserving in consequence of (B.29), $(Au^k)_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{K}^n$. Since $\mathbb{K}^n$ is complete, there is $z \in \mathbb{K}^n$ such that $\lim_{k\to\infty} Au^k = z$ and $\lim_{k\to\infty} u^k = A^{-1}z$, showing $x = A^{-1}z \in U$, i.e. U is closed. ■

Theorem B.29. *A normed vector space $(X, \|\cdot\|)$ over $\mathbb{K}$ is finite-dimensional if, and only if, its closed unit ball $\overline{B}_1(0)$ is compact.*

Proof. Let X be finite-dimensional. If $(b_1, \dots, b_n)$ denotes a basis of X, then (B.28) defines a linear isomorphism $A : X \longrightarrow \mathbb{K}^n$. If we define a norm on $\mathbb{K}^n$ via (B.29), then A^{-1} becomes norm-preserving and, in particular, continuous. Then $\overline{B}_1(0)$ in X must be compact as the countinuous image (under A^{-1}) of $\overline{B}_1(0)$ in $\mathbb{K}^n$.

Conversely, let X be infinite-dimensional. To show that $\overline{B}_1(0)$ is not compact, we construct, via recursion, a sequence $(x^k)_{k\in\mathbb{N}}$ in $\overline{B}_1(0)$ (actually in the sphere $S_1(0)$) that does not have a convergent subsequence: Fix $n \in \mathbb{N}$ and assume $(x^1, \dots, x^n)$ to be already constructed such that

$$\forall_{k\in\{1,\dots,n\}} \quad \|x^k\| = 1, \tag{B.32a}$$

$$\forall_{\substack{k,l\in\{1,\dots,n\},\\ k\neq l}} \quad \|x^k - x^l\| \geq \frac{1}{2}. \tag{B.32b}$$

Let $U := \operatorname{span}\{x^1, \dots, x^n\}$. Since X is infinite-dimensional, we have $U \neq X$. Let $x \in X \setminus U$. Since U is closed by Th. B.28, it is

$$d := \inf\left\{\|x - u\| : u \in U\right\} > 0. \tag{B.33}$$

Moreover, there exists $u_0 \in U$ such that $\|x - u_0\| \leq 2d$. Set

$$x^{n+1} := \frac{x - u_0}{\|x - u_0\|}. \tag{B.34}$$

Then $\|x^{n+1}\| = 1$ and, for each $u \in U$ is $\|x - u_0\|u + u_0 \in U$, implying

$$\|u - x^{n+1}\| = \frac{\big\|\|x - u_0\|u - x + u_0\big\|}{\|x - u_0\|} \geq \frac{d}{\|x - u_0\|} \geq \frac{1}{2}. \tag{B.35}$$

Thus, (B.32) holds with n replaced by $n + 1$, where (B.32b) means that $(x^k)_{k\in\mathbb{N}}$ can not have a convergent subsequence. ■

C Differential Calculus in $\mathbb{R}^n$

C.1 Proof of the Chain Rule

Proof of Th. 2.28. As usual, we first consider the case $\mathbb{K} = \mathbb{R}$. Since f is differentiable at ξ and g is differentiable at $f(\xi)$, according to Lem. 2.21, there are functions $r_f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $r_g : \mathbb{R}^m \longrightarrow \mathbb{R}^p$ satisfying

$$r_f(h) = f(\xi + h) - f(\xi) - Df(\xi)(h), \tag{C.1a}$$

$$r_g(h) = g\big(f(\xi) + h\big) - g\big(f(\xi)\big) - Dg\big(f(\xi)\big)(h) \tag{C.1b}$$

for each h such that $\|h\|_2$ is sufficiently small, as well as

$$\lim_{h\to 0} \frac{r_f(h)}{\|h\|_2} = 0, \quad \lim_{h\to 0} \frac{r_g(h)}{\|h\|_2} = 0. \tag{C.2}$$

Defining $r_{g\circ f} : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ by

$$r_{g\circ f}(h) := \begin{cases} (g \circ f)(\xi + h) - (g \circ f)(\xi) - \Big(Dg\big(f(\xi)\big) \circ Df(\xi)\Big)(h) & \text{for } \xi + h \in G_f, \\ 0 & \text{otherwise,} \end{cases} \tag{C.3}$$

it remains to show

$$\lim_{h\to 0} \frac{r_{g\circ f}(h)}{\|h\|_2} = 0. \tag{C.4}$$

For each $h \in \mathbb{R}^n$ with $\|h\|_2$ sufficiently small, we use (C.1) to compute

$$\begin{aligned} (g \circ f)(\xi + h) &= g\Big(f(\xi) + Df(\xi)(h) + r_f(h)\Big) \\ &= g\big(f(\xi)\big) + Dg\big(f(\xi)\big)\Big(Df(\xi)(h) + r_f(h)\Big) + r_g\Big(Df(\xi)(h) + r_f(h)\Big), \end{aligned} \tag{C.5a}$$

implying

$$r_{g\circ f}(h) = Dg\big(f(\xi)\big)\Big(r_f(h)\Big) + r_g\Big(Df(\xi)(h) + r_f(h)\Big). \tag{C.5b}$$

From Th. 1.67, we know that $Dg(f(\xi))$ is Lipschitz continuous with some Lipschitz constant $L_g \in \mathbb{R}_0^+$. Thus, for each $0 \neq h \in \mathbb{R}^n$,

$$0 \leq \frac{\Big\|Dg\big(f(\xi)\big)\Big(r_f(h)\Big)\Big\|_2}{\|h\|_2} \leq \frac{L_g \,\|r_f(h)\|_2}{\|h\|_2}, \tag{C.6a}$$

implying

$$\lim_{h\to 0} \frac{\Big\|Dg\big(f(\xi)\big)\Big(r_f(h)\Big)\Big\|_2}{\|h\|_2} = 0 \tag{C.6b}$$

due to (C.2). Thus, to prove (C.4), it merely remains to show

$$\lim_{h\to 0} \frac{\Big\|r_g\Big(Df(\xi)(h) + r_f(h)\Big)\Big\|_2}{\|h\|_2} = 0. \tag{C.7}$$

To that end, we rewrite, for $Df(\xi)(h) + r_f(h) \neq 0$,

$$\frac{\Big\|r_g\Big(Df(\xi)(h) + r_f(h)\Big)\Big\|_2}{\|h\|_2} = \frac{\big\|Df(\xi)(h) + r_f(h)\big\|_2 \Big\|r_g\Big(Df(\xi)(h) + r_f(h)\Big)\Big\|_2}{\|h\|_2 \big\|Df(\xi)(h) + r_f(h)\big\|_2}. \tag{C.8a}$$

Next, note

$$\lim_{h\to 0} \big\|Df(\xi)(h) + r_f(h)\big\|_2 = 0 \quad \overset{\text{(C.2)}}{\Rightarrow} \quad \lim_{h\to 0} \frac{\Big\|r_g\Big(Df(\xi)(h) + r_f(h)\Big)\Big\|_2}{\big\|Df(\xi)(h) + r_f(h)\big\|_2} = 0. \tag{C.8b}$$

Once again, from Th. 1.67, we know that $Df(\xi)$ is Lipschitz continuous with some Lipschitz constant $L_f \in \mathbb{R}_0^+$, implying

$$\frac{\big\|Df(\xi)(h) + r_f(h)\big\|_2}{\|h\|_2} \leq \frac{\big\|Df(\xi)(h)\big\|_2 + \big\|r_f(h)\big\|_2}{\|h\|_2} \leq L_f + 1 \tag{C.8c}$$

for $0 \neq \|h\|_2$ sufficiently small. Combining (C.8a) – (C.8c) proves (C.7) and, thus, (C.4). Together with (C.3) and Lem. 2.21, this shows that $g \circ f$ is differentiable at ξ with $D(g \circ f)(\xi) = Dg\big(f(\xi)\big) \circ Df(\xi)$.

In the case $\mathbb{K} = \mathbb{C}$, we can apply the case $\mathbb{K} = \mathbb{R}$ to obtain the differentiability of $\operatorname{Re}(g \circ f) = (\operatorname{Re} g) \circ f$ and of $\operatorname{Im}(g \circ f) = (\operatorname{Im} g) \circ f$ at ξ, and, in consequence, the differentiability of $g \circ f$ at ξ. Moreover, to verify the chain rule, we use the chain rule of the case $\mathbb{K} = \mathbb{R}$ to compute

$$\begin{aligned} D(g \circ f)(\xi) &= D\operatorname{Re}(g \circ f)(\xi) + i\, D\operatorname{Im}(g \circ f)(\xi) \\ &= D\big((\operatorname{Re} g) \circ f\big)(\xi) + i\, D\big((\operatorname{Im} g) \circ f\big)(\xi) \\ &= D\operatorname{Re} g\big(f(\xi)\big) \circ Df(\xi) + i\, D\operatorname{Im} g\big(f(\xi)\big) \circ Df(\xi) \\ &= Dg\big(f(\xi)\big) \circ Df(\xi), \end{aligned} \tag{C.9}$$

thereby completing the proof. ■

C.2 Bounded Derivatives Imply Lipschitz Continuity

First, we provide an $\mathbb{R}^m$-valued variant of Th. 2.35:

Theorem C.1. *Let $m, n \in \mathbb{N}$, let $G \subseteq \mathbb{R}^n$ be open, and let $f : G \longrightarrow \mathbb{R}^m$ be differentiable. Suppose there exists $M \in \mathbb{R}_0^+$ such that $|\partial_k f_l(\xi)| \leq M$ for each $k \in \{1, \ldots, n\}$, each $l \in \{1, \ldots, m\}$, and each $\xi \in G$. If G is convex, then f is Lipschitz continuous with Lipschitz constant $L := mM$ with respect to the 1-norms on $\mathbb{R}^n$ and $\mathbb{R}^m$ and with Lipschitz constant cL, $c > 0$, with respect to arbitrary norms on $\mathbb{R}^n$ and $\mathbb{R}^m$.*

Proof. According to Th. 2.35, each f_l is M-Lipschitz with respect to the 1-norm on $\mathbb{R}^n$. Thus, we obtain, for each $x, y \in G$,

$$\|f(y) - f(x)\|_1 = \sum_{l=1}^{m} |f_l(y) - f_l(x)| \leq mM\|y - x\|_1, \tag{C.10}$$

showing that, with respect to the 1-norms on $\mathbb{R}^n$ and $\mathbb{R}^m$, f is Lipschitz continuous with Lipschitz constant mM. Since all norms on $\mathbb{R}^n$ and $\mathbb{R}^m$ are equivalent, we also get that f is Lipschitz continuous with Lipschitz constant cL, $c > 0$, with respect to all other norms on $\mathbb{R}^n$ and $\mathbb{R}^m$. ■

It is sometimes useful if the bound on the derivatives is the same as the resulting Lipschitz constant (which, for $m > 1$, is not the case in the above Th. C.1). The following Th. C.3 provides a variant, where the constants are the same, formulated for functions $f : I \longrightarrow \mathbb{R}^n$, defined on open intervals $I \subseteq \mathbb{R}$, and making use of the Euclidean norm $\|\cdot\|_2$ on $\mathbb{R}^n$. We will start with some auxiliary results regarding the Euclidean norm and the Euclidean inner product:

Proposition C.2. *Let $I \subseteq \mathbb{R}$ be an open interval, and let $g, h : I \longrightarrow \mathbb{R}^n$ be differentiable, $n \in \mathbb{N}$.*

(a) *The function*

$$f : I \longrightarrow \mathbb{R},\ f(x) := g(x) \bullet h(x) = \sum_{j=1}^{n} g_j(x) h_j(x), \tag{C.11}$$

is differentiable and

$$f' : I \longrightarrow \mathbb{R},\ f'(x) = g'(x) \bullet h(x) + h(x) \bullet h'(x). \tag{C.12}$$

(b) *The function*

$$\alpha : I \longrightarrow \mathbb{R}, \quad \alpha(x) := \|g(x)\|_2 = \sqrt{g(x) \bullet g(x)}, \tag{C.13}$$

is differentiable at each $x \in I$ such that $g(x) \neq 0$. Moreover,

$$\underset{\substack{x \in I,\\ g(x) \neq 0}}{\forall} \quad \alpha'(x) = \frac{g(x) \bullet g'(x)}{\alpha(x)} = \frac{g(x) \bullet g'(x)}{\|g(x)\|_2}. \tag{C.14}$$

Proof. (a) is immediate from the product rule.

(b) is an easy consequence of (a), as (a) implies α to be differentiable at each $x \in I$ such that $g(x) \neq 0$, and

$$\underset{\substack{x\in I,\\ g(x)\neq 0}}{\forall} \quad \alpha'(x) = \frac{2\, g(x) \bullet g'(x)}{2\,\sqrt{g(x) \bullet g(x)}} = \frac{g(x) \bullet g'(x)}{\alpha(x)}, \tag{C.15}$$

completing the proof. ■

Theorem C.3. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f :]a, b[\longrightarrow \mathbb{R}^n$ be differentiable with uniformly bounded derivative, i.e. with*

$$\underset{M\in\mathbb{R}_0^+}{\exists} \quad \underset{x\in]a,b[}{\forall} \quad \|f'(x)\|_2 = \sqrt{\sum_{j=1}^{n} |f_j'(x)|^2} \leq M. \tag{C.16}$$

Then f is M-Lipschitz, i.e.

$$\underset{x_1,x_2\in]a,b[}{\forall} \quad \|f(x_1) - f(x_2)\|_2 \leq M\,|x_1 - x_2|. \tag{C.17}$$

Proof. For $x_1 = x_2$, there is nothing to prove. Thus, assume $x_1 \neq x_2$ and define the auxiliary function

$$g : [0,1] \longrightarrow \mathbb{R}^n, \quad g(t) := f\big(x_1 + t(x_2 - x_1)\big) - f(x_1). \tag{C.18}$$

According to the chain rule of Th. 2.28, g is differentiable on $]0,1[$ and

$$\underset{t\in]0,1[}{\forall} \quad g'(t) = (x_2 - x_1) f'\big(x_1 + t(x_2 - x_1)\big), \tag{C.19}$$

implying

$$\underset{t\in]0,1[}{\forall} \quad \|g'(t)\|_2 \leq M\,|x_1 - x_2|. \tag{C.20}$$

We now introduce another auxiliary function, namely

$$\alpha : [0,1] \longrightarrow \mathbb{R}, \quad \alpha(t) := \|g(t)\|_2. \tag{C.21}$$

Then α is continuous (as f and the norm are both continuous), satisfying $\alpha(0) = \|g(0)\|_2 = 0$ and $\alpha(1) = \|g(1)\|_2 = \|f(x_2) - f(x_1)\|_2$. If $\alpha(1) = 0$, then (C.17) is trivially true, and, thus, we proceed to assume $\alpha(1) > 0$. Then the continuity of α implies

$$s := \sup\big\{t \in [0,1] : \alpha(t) = 0\big\} < 1, \quad \alpha(s) = 0. \tag{C.22}$$

In consequence, α is positive on $]s,1[$ and, thus, differentiable on $]s,1[$ by Prop. C.2(b). The mean value theorem [Phi15a, Th. 9.17] implies the existence of $\sigma \in]s,1[$ such that

$$\begin{aligned} \alpha(1) \quad &= \quad \alpha(1) - \alpha(s) = (1-s)\,\alpha'(\sigma) \overset{\text{(C.14)}}{=} (1-s)\,\frac{g(\sigma) \bullet g'(\sigma)}{\alpha(\sigma)} \\ &\overset{(1.81)}{\leq} \quad (1-s)\,\frac{\|g(\sigma)\|_2 \|g'(\sigma)\|_2}{\|g(\sigma)\|_2} \overset{\text{(C.20)}}{\leq} (1-s)\,M\,|x_1 - x_2| \\ &\leq \quad M\,|x_1 - x_2|, \end{aligned} \tag{C.23}$$

which establishes the case. ■

C.3 Surjectivity of Directional Derivatives

We finish the proof of Th. 2.38 by showing that, for $n \geq 2$, the map

$$D : S_1(0) \longrightarrow [-\alpha, \alpha], \quad D(e) := \nabla f(\xi) \cdot e = \sum_{j=1}^{n} \epsilon_j \partial_j f(\xi), \quad \alpha = \| \nabla f(\xi) \|_2, \tag{C.24}$$

is surjective (we already know from (2.65) that $D(e) \in [-\alpha, \alpha]$ for each $e \in S_1(0)$). We also recall $e_{\max} = \nabla f(\xi)/\alpha$, $e_{\min} = -e_{\max}$, $D(e_{\max}) = \alpha$, $D(e_{\min}) = -\alpha$.

The idea is to rotate $e_{\max}$ into $e_{\min}$. This can be achieved using a suitable function

$$\rho : [0, \pi] \longrightarrow S_1(0) \subseteq \mathbb{R}^n, \quad \rho = (\rho_1, \ldots, \rho_n).$$

We have to define ρ differently, depending on $n \geq 2$ being even or odd. To this end, let $(\epsilon_1, \ldots, \epsilon_n) := e_{\max}$. If n is even, then define

$$\forall_{j \in \{1,\ldots,n\}} \quad \rho_j : [0, \pi] \longrightarrow [-1, 1], \quad \rho_j(\theta) := \begin{cases} \epsilon_j \cos\theta + \epsilon_{j+1} \sin\theta & \text{if } j \text{ is odd,} \\ -\epsilon_{j-1} \sin\theta + \epsilon_j \cos\theta & \text{if } j \text{ is even;} \end{cases} \tag{C.25a}$$

if n is odd (note $n \geq 3$ in this case), then define

$$\rho_j : [0, \pi] \longrightarrow [-1, 1],$$

$$\forall_{j \in \{1,\ldots,n\}} \quad \rho_j(\theta) := \begin{cases} \epsilon_j \cos\theta + \epsilon_{j+1} \sin\theta & \text{if } j < n-2 \text{ is odd,} \\ -\epsilon_{j-1} \sin\theta + \epsilon_j \cos\theta & \text{if } j < n-2 \text{ is even,} \\ \epsilon_{n-2} \cos\theta + \sqrt{\epsilon_{n-1}^2 + \epsilon_n^2} \sin\theta & \text{if } j = n-2, \\ \epsilon_{n-1} \cos\theta - \frac{\epsilon_{n-2}\,\epsilon_{n-1}}{\sqrt{\epsilon_{n-1}^2 + \epsilon_n^2}} \sin\theta & \text{if } j = n-1, \\ \epsilon_n \cos\theta - \frac{\epsilon_{n-2}\,\epsilon_n}{\sqrt{\epsilon_{n-1}^2 + \epsilon_n^2}} \sin\theta & \text{if } j = n. \end{cases} \tag{C.25b}$$

For the sake of readability, we assumed $\epsilon_{n-1} \neq 0$ or $\epsilon_n \neq 0$ in (C.25b). There is always at least one $j_0 \in \{1, \ldots, n\}$ such that $\epsilon_{j_0} \neq 0$. If $j_0 \notin \{n-1, n\}$, then one merely needs to interchange the roles of j_0 and n in (C.25b).

Clearly, for every $n \geq 2$, each ρ_j is continuous, i.e. ρ is continuous.

Next, we verify that ρ, indeed, maps into $S_1(0)$ (which, in particular, implies each ρ_j maps into $[-1, 1]$): If $n \geq 2$ is even, then, for each odd $j \leq n-1$, one has

$$\begin{aligned} &\forall_{\theta \in [0,\pi]} \quad (\rho_j(\theta))^2 + (\rho_{j+1}(\theta))^2 \\ &= (\epsilon_j \cos\theta + \epsilon_{j+1} \sin\theta)^2 + (-\epsilon_j \sin\theta + \epsilon_{j+1} \cos\theta)^2 \\ &= \epsilon_j^2 \cos^2\theta + 2\epsilon_j\epsilon_{j+1} \cos\theta \sin\theta + \epsilon_{j+1}^2 \sin^2\theta \\ &\quad + \epsilon_j^2 \sin\theta - 2\epsilon_j\epsilon_{j+1} \cos\theta \sin\theta + \epsilon_{j+1}^2 \cos^2\theta \\ &= \epsilon_j^2(\cos^2\theta + \sin^2\theta) + \epsilon_{j+1}^2(\cos^2\theta + \sin^2\theta) = \epsilon_j^2 + \epsilon_{j+1}^2, \end{aligned} \tag{C.26}$$

implying

$$\forall_{\theta \in [0,\pi]} \quad \|\rho(\theta)\|_2^2 = \sum_{j=1}^{n} (\rho_j(\theta))^2 = \sum_{j=1}^{n} \epsilon_j^2 = 1. \tag{C.27}$$

If $n \geq 3$ is odd, then (C.26) still holds for each odd $j \leq n-4$. Additionally,

$$\begin{aligned}
&(\rho_{n-2}(\theta))^2 + (\rho_{n-1}(\theta))^2 + (\rho_n(\theta))^2 \\
&= \epsilon_1^2 \cos^2\theta + 2\epsilon_1\sqrt{\epsilon_2^2+\epsilon_3^2}\,\sin\theta\cos\theta + (\epsilon_2^2+\epsilon_3^2)\,\sin^2\theta \\
&\quad + \epsilon_2^2\cos^2\theta - 2\frac{\epsilon_1\,\epsilon_2^2}{\sqrt{\epsilon_2^2+\epsilon_3^2}}\,\sin\theta\cos\theta + \frac{\epsilon_1^2\,\epsilon_2^2}{\epsilon_2^2+\epsilon_3^2}\,\sin^2\theta \\
&\quad + \epsilon_3^2\cos^2\theta - 2\frac{\epsilon_1\,\epsilon_3^2}{\sqrt{\epsilon_2^2+\epsilon_3^2}}\,\sin\theta\cos\theta + \frac{\epsilon_1^2\,\epsilon_3^2}{\epsilon_2^2+\epsilon_3^2}\,\sin^2\theta \\
&\underset{\theta\in[0,\pi]}{\forall} \quad = (\epsilon_1^2+\epsilon_2^2+\epsilon_3^2)\,\cos^2\theta \\
&\quad + \frac{2\epsilon_1\,(\epsilon_2^2+\epsilon_3^2-\epsilon_2^2-\epsilon_3^2)}{\sqrt{\epsilon_2^2+\epsilon_3^2}}\,\sin\theta\cos\theta \\
&\quad + \left(\epsilon_2^2+\epsilon_3^2+\frac{\epsilon_1^2\,(\epsilon_2^2+\epsilon_3^2)}{\epsilon_2^2+\epsilon_3^2}\right)\sin^2\theta \\
&= (\epsilon_1^2+\epsilon_2^2+\epsilon_3^2)\,(\cos^2\theta+\sin^2\theta) = \epsilon_{n-2}^2+\epsilon_{n-1}^2+\epsilon_n^2,
\end{aligned} \tag{C.28}$$

i.e. (C.27) is true also for each $n \geq 3$ odd.

Clearly, D is also continuous and, thus, so is $D \circ \rho : [0,\pi] \longrightarrow [-\alpha,\alpha]$. Moreover, as $\sin(0) = \sin(\pi) = 0$, $\cos(0) = 1$, $\cos(\pi) = -1$, we obtain

$$\underset{n\geq 2}{\forall} \quad \underset{j\in\{1,\ldots,n\}}{\forall} \quad \Big(\rho_j(0) = \epsilon_j \quad \wedge \quad \rho_j(\pi) = -\epsilon_j\Big), \tag{C.29}$$

implying

$$\underset{n\geq 2}{\forall} \quad \Big(\rho(0) = e_{\max} \quad \wedge \quad (D\circ\rho)(0) = \alpha \quad \wedge \quad \rho(\pi) = e_{\min} \quad \wedge \quad (D\circ\rho)(\pi) = -\alpha\Big). \tag{C.30}$$

The continuity of $D \circ \rho$ and the intermediate value theorem [Phi15a, Th. 7.57] imply $D \circ \rho$ to be surjective, i.e. D must be surjective as well.

C.4 Implicit Function Theorem

We start with a preparatory proposition.

Proposition C.4. *Let $\|\cdot\|$ be some norm on $\mathbb{R}^n$, $n \in \mathbb{N}$. Moreover, let $a \in \mathbb{R}^n$, $r > 0$, and let $f : B_r(a) \longrightarrow \mathbb{R}^n$ be defined on the open r-ball with center a with respect to $\|\cdot\|$. If A is an invertible $n \times n$ matrix over $\mathbb{R}$ such that*

$$\left\|A^{-1}f(a)\right\| < \frac{r}{2} \tag{C.31}$$

and such that the map

$$F : B_r(a) \longrightarrow \mathbb{R}^n, \quad F(x) := x - A^{-1}f(x), \tag{C.32}$$

is Lipschitz continuous with Lipschitz constant $L = 1/2$, then f has a unique zero $\xi \in B_r(a)$. Moreover, for each $x_0 \in B_r(a)$, ξ is the limit of the sequence $(x_k)_{k\in\mathbb{N}_0}$, recursively defined by

$$\underset{k\in\mathbb{N}_0}{\forall} \quad x_{k+1} := F(x_k). \tag{C.33}$$

Proof. Set

$$s_0 := \max\left\{2\left\|A^{-1}f(a)\right\|, \|x_0 - a\|\right\} \overset{(C.31)}{\in} [0, r[. \tag{C.34}$$

The idea is to show that, for each $s_0 < s < r$, the Banach fixed point Th. B.26 applies to the contraction

$$F_s : \overline{B}_s(a) \longrightarrow \overline{B}_s(a), \quad F_s(x) := F(x). \tag{C.35}$$

We verify that F_s, indeed, maps $\overline{B}_s(a)$ into $\overline{B}_s(a)$: If $x \in \overline{B}_s(a)$, then

$$\begin{aligned}\left\|F(x) - a\right\| &\leq \left\|F(x) - F(a)\right\| + \left\|F(a) - a\right\| \leq \frac{1}{2}\|x - a\| + \left\|A^{-1}f(a)\right\| \\ &\leq \frac{s}{2} + \frac{s_0}{2} < s,\end{aligned} \tag{C.36}$$

showing $F_s(x) \in \overline{B}_s(a)$ (in particular, this shows the x_k are well-defined by (C.33)). As F is Lipschitz continuous with Lipschitz constant $L = 1/2$, so is F_s, i.e. F_s is, indeed, a contraction. As $\overline{B}_s(a)$ is closed, the Banach fixed point Th. B.26 yields that F_s has a unique fixed point ξ and, moreover, $\xi = \lim_{k\to\infty} x_k$. Since this holds for each $s \in]s_0, r[$, ξ must also be the unique fixed point of F. The proof is concluded by noting

$$\underset{y\in B_r(a)}{\forall} \quad f(y) = 0 \quad \Leftrightarrow \quad F(y) = y - A^{-1}f(y) = y, \tag{C.37}$$

that means y is a zero of f if, and only if, y is a fixed point of F. ∎

Remark C.5. If the map f in Prop. C.4 is differentiable with invertible derivatives $Df(x)$, and if, instead of using a constant matrix A in the definition of (C.33), one uses $(Df(x_k))^{-1}$, then the iteration defined by (C.33) is known as *Newton's method* (in n dimensions, cf. [Phi15b, Sec. 6.3]). In consequence, if $A \approx (Df(x_k))^{-1}$ in (C.33), then the defined iteration is sometimes referred to as a *simplified Newton's method*.

Notation C.6. Let $k, m, n \in \mathbb{N}$, let $G \subseteq \mathbb{R}^n \times \mathbb{R}^m$ be open, and consider a map $f : G \longrightarrow \mathbb{R}^k$. If $(\xi, \eta) \in G$ and f is differentiable at (ξ, η), then let $D_y f(\xi, \eta)$ and $D_x f(\xi, \eta)$ denote the linear maps

$$D_y f(\xi,\eta) : \mathbb{R}^m \longrightarrow \mathbb{R}^k, \qquad \big(D_y f(\xi,\eta)\big)(h) := \big(Df(\xi,\eta)\big)(0,h), \tag{C.38a}$$

$$D_x f(\xi,\eta) : \mathbb{R}^n \longrightarrow \mathbb{R}^k, \qquad \big(D_x f(\xi,\eta)\big)(h) := \big(Df(\xi,\eta)\big)(h,0), \tag{C.38b}$$

respectively.

Theorem C.7 (Implicit Function Theorem)**.** *Let $m, n \in \mathbb{N}$, let $G \subseteq \mathbb{R}^n \times \mathbb{R}^m$ be open, and let $f : G \longrightarrow \mathbb{R}^m$ be continuously differentiable, i.e. $f \in C^1(G, \mathbb{R}^m)$. If $(\xi, \eta) \in G$ is such that*

$$f(\xi,\eta) = 0 \quad \textit{and} \quad A := D_y f(\xi,\eta) \textit{ is invertible}, \tag{C.39}$$

then there exist open neighborhoods $U_\xi \subseteq \mathbb{R}^n$ of ξ and $V_\eta \subseteq \mathbb{R}^m$ of η, and a continuously differentiable map $g : U_\xi \longrightarrow V_\eta$ such that the zeros of f in $U_\xi \times V_\eta$ are given precisely by the graph of g, i.e.

$$(U_\xi \times V_\eta) \cap f^{-1}\{0\} = \{(x, g(x)) : x \in U_\xi\}, \tag{C.40a}$$

which can be restated as

$$\forall_{(x,y)\in U_\xi \times V_\eta} \quad \Big(f(x,y) = 0 \quad \Leftrightarrow \quad y = g(x)\Big). \tag{C.40b}$$

Moreover,

$$\forall_{x\in U_\xi} \quad Dg(x) = -\big(D_y f(x, g(x))\big)^{-1} D_x f(x, g(x)) \tag{C.41}$$

and, if $f \in C^\alpha(G, \mathbb{R}^m)$, $\alpha \in \mathbb{N} \cup \{\infty\}$, then $g \in C^\alpha(U_\xi, \mathbb{R}^m)$.

Proof. Fix some arbitrary norms on $\mathbb{R}^n$ and on the set $\mathcal{M}(m, \mathbb{R})$ of real $m \times m$ matrices (for readability's sake, we will denote both norms by $\|\cdot\|$). On $\mathbb{R}^m$, we will use the 1-norm $\|\cdot\|_1$ to apply Th. C.1. According to the hypothesis, A is invertible. Thus, $\det(A) > 0$. Since the map $B \mapsto \det(B)$ is continuous (cf. Ex. 1.66(a)), and the map $D_y f : G \longrightarrow \mathcal{M}(m, \mathbb{R})$ is continuous due to the assumed continuous differentiability of f, the set

$$G_0 := \big\{(x, y) \in G : \det\big(D_y f(x, y)\big) > 0\big\} \subseteq G \tag{C.42}$$

is an open neighborhood of (ξ, η). Next, we consider the map

$$F : G_0 \longrightarrow \mathbb{R}^m, \quad F(x, y) := y - A^{-1} f(x, y). \tag{C.43}$$

Then F is continuously differentiable with

$$D_y F : G_0 \longrightarrow \mathcal{M}(m, \mathbb{R}), \quad D_y F(x, y) = \mathrm{Id} - A^{-1} D_y f(x, y), \tag{C.44}$$

being continuous as well. Thus, since $D_y F(\xi, \eta) = \mathrm{Id} - A^{-1}A = 0$, there exists $r > 0$ such that the open r-balls $B_r(\xi) \subseteq \mathbb{R}^n$ and $B_r(\eta) \subseteq \mathbb{R}^m$ satisfy

$$(\xi, \eta) \in B_r(\xi) \times B_r(\eta) \subseteq \left\{(x, y) \in G_0 : \forall_{k,l=1,\ldots,m} \big|\partial_{y_k} F_l(x, y)\big| < \frac{1}{2m}\right\} \subseteq G_0. \tag{C.45}$$

As we assume f to be continuous with $f(\xi, \eta) = 0$, there exists $s \in]0, r]$ such that

$$B_s(\xi) \subseteq \left\{x \in \mathbb{R}^n : \big\|A^{-1} f(x, \eta)\big\|_1 < \frac{r}{2}\right\} \subseteq B_r(\xi). \tag{C.46}$$

To construct the map $g : B_s(\xi) \longrightarrow B_r(\eta)$, we fix $x \in B_s(\xi)$ and apply Prop. C.4 to the function

$$f_x : B_r(\eta) \longrightarrow \mathbb{R}^m, \quad f_x(y) := f(x, y). \tag{C.47}$$

To verify that the hypotheses of Prop. C.4 are satisfied, we observe $\|A^{-1} f_x(\eta)\|_1 < \frac{r}{2}$ holds due to $x \in B_s(\xi)$ and (C.46), the map $F_x : B_r(\eta) \longrightarrow \mathbb{R}^m$, $F_x(y) := y - A^{-1} f_x(y) = F(x, y)$, is Lipschitz continuous with Lipschitz constant $L = m\frac{1}{2m} = \frac{1}{2}$ due

to (C.45) and Th. C.1. Thus, according to Prop. C.4, the function f_x has a unique zero $g(x)$ in $B_r(\eta)$, which defines the function g.

Note that, in the above argument, for each $0 < \rho < r$, one can choose $s(\rho) < s$ such that (C.46) holds with s replaced by $s(\rho)$ and r replaced by ρ, then showing that g maps $B_{s(\rho)}$ into B_ρ. We now choose some arbitrary $\rho \in]0, r[$ and set

$$U_\xi := B_{s(\rho)}(\xi), \quad V_\eta := B_\rho(\eta) \tag{C.48}$$

for the desired neighborhoods of the theorem. We verify g to be continuous on U_ξ: Let $x \in U_\xi$ and let $(x_k)_{k\in\mathbb{N}}$ be a sequence in U_ξ with $\lim_{k\to\infty} x_k = x$. We have to show $\lim_{k\to\infty} g(x_k) = g(x)$. If $\lim_{k\to\infty} g(x_k) = g(x)$ does not hold, then, without loss of generality, we may assume that there exists $\epsilon > 0$ such that $\|g(x_k) - g(x)\| > \epsilon$ for each $k \in \mathbb{N}$ (after having replaced $(x_k)_{k\in\mathbb{N}}$ with a suitable subsequence). Moreover, we may replace $(x_k)_{k\in\mathbb{N}}$ with another subsequence such that there exists $y \in \overline{B}_\rho(\eta) \subseteq B_r(\eta)$ satisfying $y = \lim_{k\to\infty} g(x_k)$ (this is due to the Bolzano-Weierstrass Th. 1.16(b), as $g(x_k) \in B_\rho(\eta)$ for each $k \in \mathbb{N}$). Then the continuity of f implies

$$f(x, y) = \lim_{k\to\infty} f\big(x_k, g(x_k)\big) = 0, \tag{C.49}$$

showing $g(x) = y = \lim_{k\to\infty} g(x_k)$ (due to (C.40b) – here we need $y \in B_r(\eta)$, which was the reason for choosing $\rho < r$), which is in contradiction to the choice of the x_k, and proves the continuity of g.

Next, we show that g is differentiable at each $x \in U_\xi$, where the derivative is given by (C.41): To this end, let $x \in U_\xi$ and note the existence of $\big(D_y f(x, g(x))\big)^{-1}$ due to $(x, g(x)) \in U_\xi \times V_\eta \subseteq G_0$. According to Def. 2.19, we have to show

$$\lim_{h\to 0} \frac{g(x+h) - g(x) + \big(D_y f(x, g(x))\big)^{-1} D_x f(x, g(x))\, h}{\|h\|} = 0. \tag{C.50}$$

Let $0 \neq h \in \mathbb{R}^n$ be sufficiently small such that $x+h \in U_\xi$. Using the notation of the mean value Th. 2.32, for each $l \in \{1, \ldots, m\}$, there exist $x_{h,l} \in S_{x,x+h}$ and $y_{h,l} \in S_{g(x),g(x+h)}$ such that

$$\begin{aligned} 0 &= f_l\big(x+h, g(x+h)\big) - f_l\big(x, g(x)\big) \\ &= f_l\big(x+h, g(x+h)\big) - f_l\big(x, g(x+h)\big) + f_l\big(x, g(x+h)\big) - f_l\big(x, g(x)\big) \\ &= D_x f_l\big(x_{h,l}, g(x+h)\big)(h) + D_y f_l\big(x, y_{h,l}\big)\big(g(x+h) - g(x)\big). \end{aligned} \tag{C.51}$$

Note that the two derivatives occurring in (C.51) have the form of gradients, which, according to our usual convention, we can interpret as row vectors. Joining m row vectors into a matrix, we can write the m equations of (C.51) in matrix form as

$$0 = X_h h + Y_h\big(g(x+h) - g(x)\big), \tag{C.52}$$

where

$$X_h := \begin{pmatrix} D_x f_1\big(x_{h,1}, g(x+h)\big) \\ \vdots \\ D_x f_m\big(x_{h,m}, g(x+h)\big) \end{pmatrix}, \quad Y_h := \begin{pmatrix} D_y f_1\big(x, y_{h,1}\big) \\ \vdots \\ D_y f_m\big(x, y_{h,m}\big) \end{pmatrix}. \tag{C.53}$$

As we already know g to be continuous, $h \to 0$ implies $g(x+h) \to g(x)$. Thus, since $y_{h,l} \in S_{g(x),g(x+h)}$, $h \to 0$ implies $y_{h,l} \to g(x)$ for each $l \in \{1,\dots,m\}$, and, as all partials of f are continuous as well, $Y_h \to D_y f\big(x, g(x)\big)$. Since the maps $B \mapsto \det(B)$ and $B \mapsto \|B^{-1}\|$ are continuous (cf. Ex. 1.53 and Ex. 1.66(a),(b)), $h \to 0$ implies $\det(Y_h) \to \det\big(D_y f\big(x, g(x)\big)\big) \neq 0$ and Y_h is invertible for sufficiently small h with $(Y_h)^{-1} \to \big(D_y f\big(x, g(x)\big)\big)^{-1}$. For such sufficiently small h, we can rewrite (C.52) as

$$g(x+h) - g(x) = -(Y_h)^{-1} X_h h. \tag{C.54}$$

Also, since $x_{h,l} \in S_{x,x+h}$, $h \to 0$ implies $x_{h,l} \to x$ and, then, the continuity of g together with the continuity of the partials of f implies $X_h \to D_x f(x, g(x))$. Thus, we can finish the proof of (C.41) by noting

$$\begin{aligned} &\lim_{h\to 0} \frac{\big\| g(x+h) - g(x) + \big(D_y f(x,g(x))\big)^{-1} D_x f(x,g(x))\, h \big\|_1}{\|h\|} \\ &= \lim_{h\to 0} \frac{\big\| -(Y_h)^{-1} X_h h + \big(D_y f(x,g(x))\big)^{-1} D_x f(x,g(x))\, h \big\|_1}{\|h\|} = 0. \end{aligned} \tag{C.55}$$

It remains to prove that g is C^α if f is C^α, $\alpha \in \mathbb{N} \cup \{\infty\}$. To this end, for $\alpha \in \mathbb{N}$, we will show by induction on $\beta = 1, \dots, \alpha$ that each partial derivative of g at $x \in U_\xi$ of order β is a rational function of partials of f of order $\leq \beta$, all taken at $(x, g(x))$, and of partials of g of order $\leq \beta - 1$, all taken at x (in particular, the denominator of this rational function does not have any zeros in U_ξ): For $\beta = 1$, the claim follows from (C.41): The entries of $D_x f(x, g(x))$ are polynomials of first partials of f taken at $(x, g(x))$; the entries of $\big(D_y f(x, g(x))\big)^{-1}$ are, according to Th. A.51(c), rational functions, where both the numerator and the denominator polynomial are polynomials of first partials of f taken at $(x, g(x))$ (in particular, the entries of the right-hand side of (C.41) do not involve any first partials of g). For the induction step, let $1 < \beta \leq \alpha$. By induction, we know the partials of g of order $\beta - 1$ are rational functions of partials of f of order $\leq \beta - 1$, all taken at $(x, g(x))$, and of partials of g of order $\leq \beta - 2$, all taken at x. Taking the derivative of partials of g of order $\leq \beta - 2$ evaluated at x, yields partials of g of order $\leq \beta - 1$ still evaluated at x; according to the chain rule of Th. 2.28, taking the derivative of partials of f of order $\leq \beta - 1$ evaluated at $(x, g(x))$, yields polynomials of partials of f of order $\leq \beta$ evaluated at $(x, g(x))$ and of first partials of g evaluated at x. Thus, applying the product and the quotient rule establishes the case. ■

Theorem C.8 (Inverse Function Theorem)**.** *Let $n \in \mathbb{N}$, let $G \subseteq \mathbb{R}^n$ be open, and let $f : G \longrightarrow \mathbb{R}^n$ be continuously differentiable, i.e. $f \in C^1(G, \mathbb{R}^n)$. If $\xi \in G$ is such that*

$$Df(\xi) \text{ is invertible}, \tag{C.56}$$

then there exists an open neighborhood $U \subseteq G$ of ξ such that $V := f(U)$ is open and the restriction $f : U \longrightarrow V$ is bijective with continuously differentiable inverse function $f^{-1} : V \longrightarrow U$. Moreover,

$$\forall_{y \in V} \quad D(f^{-1})(y) = \Big(Df\big(f^{-1}(y)\big)\Big)^{-1} \tag{C.57}$$

and, if $f \in C^\alpha(U, \mathbb{R}^n)$, $\alpha \in \mathbb{N} \cup \{\infty\}$, then $f^{-1} \in C^\alpha(V, \mathbb{R}^n)$.

Proof. The idea is to apply the implicit function Th. C.7 to the continuously differentiable map

$$F : G \times \mathbb{R}^n \longrightarrow \mathbb{R}^n, \quad F(x,y) := f(x) - y. \tag{C.58}$$

Here, as compared to Th. C.7, the roles of the variables x and y are switched. Letting $\eta := f(\xi)$, we have

$$F(\xi,\eta) = f(\xi) - \eta = 0, \quad \text{and} \quad D_x F(\xi,\eta) = Df(\xi) \text{ is invertible.} \tag{C.59}$$

Thus, Th. C.7 applies and yields an open neighborhood $\tilde{U} \subseteq G$ of ξ, an open neighborhood $V \subseteq \mathbb{R}^n$ of η, and a C^1 map $g : V \longrightarrow \tilde{U}$ such that

$$\forall_{(x,y)\in\tilde{U}\times V} \quad \Big(F(x,y) = f(x) - y = 0 \quad \Leftrightarrow \quad x = g(y)\Big). \tag{C.60}$$

If we let $U := g(V)$, then $U \subseteq \tilde{U}$ is a neighborhood of $\xi = g(f(\xi))$, and (C.60) implies that $f : U \longrightarrow V$ and $g : V \longrightarrow U$ are inverse to each other, in particular, they are both bijective with $f^{-1} = g$. To verify that U is open, consider the (still continuous) map $f : \tilde{U} \longrightarrow \mathbb{R}^n$ and observe $U = f^{-1}(V)$. As V is open, Th. 1.54(ii) implies the existence of $O \subseteq \mathbb{R}^n$ open with $U = O \cap \tilde{U}$. Since both O and $\tilde{U}$ are open, U must be open as well.

Using (C.41), we obtain, for each $y \in V$,

$$\begin{aligned} Dg(y) &= -\big(D_x F(g(y),y)\big)^{-1} D_y F(g(y),y) \\ &= -\Big(Df\big(g(y)\big)\Big)^{-1} (-\operatorname{Id}) = \Big(Df\big(g(y)\big)\Big)^{-1}, \end{aligned} \tag{C.61}$$

proving (C.57). Finally, if f is C^α on U, then F is C^α on $U \times \mathbb{R}^n$, such that Th. C.7 implies $g = f^{-1}$ to be C^α as well. ∎

Corollary C.9. *Let $n \in \mathbb{N}$, let $G \subseteq \mathbb{R}^n$ be open, and let $f : G \longrightarrow \mathbb{R}^n$ be continuously differentiable, i.e. $f \in C^1(G,\mathbb{R}^n)$. If $Df(x)$ is invertible for each $x \in G$, then f maps open sets to open sets, i.e. if $O \subseteq G$ is open, then $f(O)$ is open as well.*

Proof. Let $O \subseteq G$ be open. We have to show that each point $\eta \in f(O)$ is an interior point of $f(O)$. To this end, let $\eta \in f(O)$ and let $\xi \in O$ be such that $f(\xi) = \eta$. Since $Df(\xi)$ is invertible by hypothesis, we can apply the inverse function Th. C.8 to the restriction of f to O, obtaining open neighborhoods $U \subseteq O$ of ξ and $V \subseteq f(O)$ of η such that $f : U \longrightarrow V$ is bijective. In particular, η is an interior point of $f(O)$, proving $f(O)$ to be open. ∎

D Riemann Integral for $\mathbb{C}$-Valued Functions

D.1 Riemann Integrability

Notation D.1. Let $I := [a,b] \subseteq \mathbb{R}^n$ be an interval, $a, b \in \mathbb{R}^n$, $a < b$. By $\mathcal{R}(I,\mathbb{R}) := \mathcal{R}(I)$ we denote the set of all Riemann integrable functions $f : I \longrightarrow \mathbb{R}$ (cf. Def. 4.5(b)).

Definition D.2. Let $I := [a,b] \subseteq \mathbb{R}^n$ be an interval, $a,b \in \mathbb{R}^n$, $a < b$. We call a function $f : I \longrightarrow \mathbb{C}$ *Riemann integrable* if, and only if, both $\operatorname{Re} f$ and $\operatorname{Im} f$ are Riemann integrable. The set of all Riemann integrable functions $f : I \longrightarrow \mathbb{C}$ is denoted by $\mathcal{R}(I,\mathbb{C})$. If $f \in \mathcal{R}(I,\mathbb{C})$, then

$$\int_I f := \left(\int_I \operatorname{Re} f, \int_I \operatorname{Im} f \right) = \int_I \operatorname{Re} f + i \int_I \operatorname{Im} f \in \mathbb{C} \tag{D.1}$$

is called the Riemann integral of f over I.

Theorem D.3. *Let $I := [a,b] \subseteq \mathbb{R}^n$, $a,b \in \mathbb{R}^n$, $n \in \mathbb{N}$, $a < b$, $f : I \longrightarrow \mathbb{C}$. If f is continuous, then f is Riemann integrable over I.*

Proof. If f is continuous, then $\operatorname{Re} f$ and $\operatorname{Im} f$ are both continuous, and, thus, the statement follows from the real-valued case of Th. 4.14. ■

Theorem D.4. *Let $I := [a,b] \subseteq \mathbb{R}^n$, $a,b \in \mathbb{R}^n$, $n \in \mathbb{N}$, $a < b$.*

(a) *If $f,g \in \mathcal{R}(I,\mathbb{C})$, then $\bar{f}, fg, \in \mathcal{R}(I,\mathbb{C})$. If, in addition, there exists $\delta > 0$ such that $|g(x)| \geq \delta$ for each $x \in I$, then $f/g \in \mathcal{R}(I,\mathbb{C})$.*

(b) *If $f \in \mathcal{R}(I,\mathbb{R})$ and $\phi : f(I) \longrightarrow \mathbb{C}$ is Lipschitz continuous, then $\phi \circ f \in \mathcal{R}(I,\mathbb{C})$.*

(c) *If $f \in \mathcal{R}(I,\mathbb{C})$ and $\phi : f(I) \longrightarrow \mathbb{R}$ is Lipschitz continuous, then $\phi \circ f \in \mathcal{R}(I,\mathbb{R})$.*

Proof. All the following proofs are completely analogous to the respective 1-dimensional case in [Phi15a, Th. G.4].

(a): Since

$$\bar{f} = (\operatorname{Re} f, -\operatorname{Im} f), \tag{D.2a}$$
$$fg = (\operatorname{Re} f \operatorname{Re} g - \operatorname{Im} f \operatorname{Im} g,\ \operatorname{Re} f \operatorname{Im} g + \operatorname{Im} f \operatorname{Re} g), \tag{D.2b}$$
$$1/g = (\operatorname{Re} g/|g|^2, -\operatorname{Im} g/|g|^2), \tag{D.2c}$$

everything follows from the real-valued case of Th. 4.12(a) and of Th. 4.15(b),(c), where $|g| \geq \delta > 0$ guarantees $|g|^2 \geq \delta^2 > 0$).

(b): Assume ϕ to be L-Lipschitz, $L \geq 0$. For each $x,y \in f(I)$, one has

$$|\operatorname{Re}\phi(x) - \operatorname{Re}\phi(y)| \overset{\text{[Phi15a, Th. 5.11(d)]}}{\leq} |\phi(x) - \phi(y)| \leq L|x-y|, \tag{D.3a}$$
$$|\operatorname{Im}\phi(x) - \operatorname{Im}\phi(y)| \overset{\text{[Phi15a, Th. 5.11(d)]}}{\leq} |\phi(x) - \phi(y)| \leq L|x-y|, \tag{D.3b}$$

showing $\operatorname{Re}\phi$ and $\operatorname{Im}\phi$ are L-Lipschitz, such that $\operatorname{Re}(\phi \circ f)$ and $\operatorname{Im}(\phi \circ f)$ are Riemann integrable by Th. 4.15(a).

(c): Assume ϕ to be L-Lipschitz, $L \geq 0$. If $f \in \mathcal{R}(I,\mathbb{C})$, then $\operatorname{Re} f, \operatorname{Im} f \in \mathcal{R}(I,\mathbb{R})$, and, given $\epsilon > 0$, Riemann's integrability criterion of Th. 4.13 provides partitions Δ_1, Δ_2 of I such that $R(\Delta_1, \operatorname{Re} f) - r(\Delta_1, \operatorname{Re} f) < \epsilon/2L$, $R(\Delta_2, \operatorname{Im} f) - r(\Delta_2, \operatorname{Im} f) < \epsilon/2L$, where

R and r denote upper and lower Riemann sums, respectively (cf. (4.11)). Letting Δ be a joint refinement of Δ_1 and Δ_2, we have (cf. Def. 4.8(a),(b) and Th. 4.10(a))

$$R(\Delta, \operatorname{Re} f) - r(\Delta, \operatorname{Re} f) < \epsilon/2L, \quad R(\Delta, \operatorname{Im} f) - r(\Delta, \operatorname{Im} f) < \epsilon/2L. \tag{D.4}$$

Recalling that, for each $g : I \longrightarrow \mathbb{R}$, it is

$$r(\Delta, g) = \sum_{p \in P(\Delta)} m_p(g)|I_p|, \tag{D.5a}$$

$$R(\Delta, g) = \sum_{p \in P(\Delta)} M_p(g)|I_p|, \tag{D.5b}$$

where $P(\Delta)$ is according to Def. 4.2,

$$m_p(g) := \inf\{g(x) : x \in I_p\}, \quad M_p(g) := \sup\{g(x) : x \in I_p\}, \tag{D.5c}$$

we obtain, for each $\xi_p, \eta_p \in I_p$,

$$\begin{aligned}
&\big|(\phi \circ f)(\xi_p) - (\phi \circ f)(\eta_p)\big| \\
&\leq L\,\big|f(\xi_p) - f(\eta_p)\big| \overset{\text{[Phi15a, Th. 5.11(d)]}}{\leq} L\,\big|\operatorname{Re} f(\xi_p) - \operatorname{Re} f(\eta_p)\big| + L\,\big|\operatorname{Im} f(\xi_p) - \operatorname{Im} f(\eta_p)\big| \\
&\leq L\,\big(M_p(\operatorname{Re} f) - m_p(\operatorname{Re} f)\big) + L\,\big(M_p(\operatorname{Im} f) - m_p(\operatorname{Im} f)\big),
\end{aligned} \tag{D.6}$$

and, thus,

$$\begin{aligned}
R(\Delta, \phi \circ f) - r(\Delta, \phi \circ f) &= \sum_{p \in P(\Delta)} \big(M_p(\phi \circ f) - m_p(\phi \circ f)\big)|I_p| \\
&\overset{\text{(D.6)}}{\leq} \sum_{p \in P(\Delta)} L\,\big(M_p(\operatorname{Re} f) - m_p(\operatorname{Re} f)\big)|I_p| + \sum_{p \in P(\Delta)} L\,\big(M_p(\operatorname{Im} f) - m_p(\operatorname{Im} f)\big)|I_p| \\
&= L\,\big(R(\Delta, \operatorname{Re} f) - r(\Delta, \operatorname{Re} f)\big) + L\,\big(R(\Delta, \operatorname{Im} f) - r(\Delta, \operatorname{Im} f)\big) \overset{\text{(D.4)}}{<} \epsilon.
\end{aligned} \tag{D.7}$$

Thus, $\phi \circ f \in \mathcal{R}(I, \mathbb{R})$ by Th. 4.13. ■

Theorem D.5. *Let $n \in \mathbb{N}$, $a, b \in \mathbb{R}^n$, $a < b$, $I := [a, b]$.*

(a) *The integral is linear: More precisely, if $f, g \in \mathcal{R}(I, \mathbb{C})$ and $\lambda, \mu \in \mathbb{C}$, then $\lambda f + \mu g \in \mathcal{R}(I, \mathbb{C})$ and*

$$\int_I (\lambda f + \mu g) = \lambda \int_I f + \mu \int_I g. \tag{D.8}$$

(b) *For each $f \in \mathcal{R}(I, \mathbb{C})$, one has $|f| \in \mathcal{R}(I, \mathbb{R})$ and*

$$\left|\int_I f\right| \leq \int_I |f|. \tag{D.9}$$

Proof. (a): One computes, using the real-valued case of Th. 4.12(a),

$$\begin{aligned}\int_I(\lambda f) &= \left(\int_I(\operatorname{Re}\lambda\operatorname{Re}f-\operatorname{Im}\lambda\operatorname{Im}f),\ \int_I(\operatorname{Re}\lambda\operatorname{Im}f+\operatorname{Im}\lambda\operatorname{Re}f)\right)\\ &= \left(\operatorname{Re}\lambda\int_I\operatorname{Re}f-\operatorname{Im}\lambda\int_I\operatorname{Im}f,\ \operatorname{Re}\lambda\int_I\operatorname{Im}f+\operatorname{Im}\lambda\int_I\operatorname{Re}f\right)\\ &= \lambda\int_I f \end{aligned} \tag{D.10a}$$

and

$$\begin{aligned}\int_I(f+g) &= \left(\int_I\operatorname{Re}(f+g),\ \int_I\operatorname{Im}(f+g)\right)=\left(\int_I\operatorname{Re}f+\int_I\operatorname{Re}g,\ \int_I\operatorname{Im}g+\int_I\operatorname{Im}g\right)\\ &= \left(\int_I\operatorname{Re}f,\ \int_I\operatorname{Im}f\right)+\left(\int_I\operatorname{Re}g,\ \int_I\operatorname{Im}g\right)=\int_I f+\int_I g.\end{aligned} \tag{D.10b}$$

(b): As the modulus is 1-Lipschitz by the inverse triangle inequality, $|f|\in\mathcal{R}(I,\mathbb{R})$ by Th. D.4(c). Let Δ be an arbitrary partition of I. Then, using the notation from the proof of Th. D.4(c) above, we obtain the following estimate of intermediate Riemann sums (cf. (4.11c)):

$$\begin{aligned}\Big|\big(\rho(\Delta,\operatorname{Re}f),\,\rho(\Delta,\operatorname{Im}f)\big)\Big| &:= \left|\left(\sum_{p\in P(\Delta)}\operatorname{Re}f(\xi_p)\,|I_p|,\ \sum_{p\in P(\Delta)}\operatorname{Im}f(\xi_p)\,|I_p|\right)\right|\\ &\le \sum_{p\in P(\Delta)}\Big|\big(\operatorname{Re}f(\xi_p),\,\operatorname{Im}f(\xi_p)\big)\Big|\,|I_p|\\ &= \sum_{p\in P(\Delta)}|f(\xi_p)|\,|I_p|=\rho(\Delta,|f|).\end{aligned} \tag{D.11}$$

Since the intermediate Riemann sums in (D.11) converge to the respective integrals by (4.30b), one obtains

$$\left|\int_I f\right| = \lim_{|\Delta|\to0}\Big|\big(\rho(\Delta,\operatorname{Re}f),\,\rho(\Delta,\operatorname{Im}f)\big)\Big| \overset{\text{(D.11)}}{\le} \lim_{|\Delta|\to0}\rho(\Delta,|f|)=\int_I|f|, \tag{D.12}$$

proving (D.9). ∎

D.2 Fubini Theorem

Definition D.6. Let $I=[a,b]\subseteq\mathbb{R}^n$ be an interval, $a,b\in\mathbb{R}^n$, $n\in\mathbb{N}$, $a<b$, and suppose $f: I\longrightarrow\mathbb{C}$ is bounded (i.e. both $\operatorname{Re}f$ and $\operatorname{Im}f$ are bounded). Define

$$J_*(f,I) := J_*(\operatorname{Re}f,I)+i\,J_*(\operatorname{Im}f,I), \tag{D.13a}$$

$$J^*(f,I) := J^*(\operatorname{Re}f,I)+i\,J^*(\operatorname{Im}f,I). \tag{D.13b}$$

As in the $\mathbb{R}$-valued case, we call $J_*(f,I)$ the *lower Riemann integral* of f over I and $J^*(f,I)$ the *upper Riemann integral* of f over I.

Theorem D.7. *Let $a, b, c, d, e, f \in \mathbb{R}^n$, $n \in \mathbb{N}$, $a < b$, $c < d$, $e < f$, $I = [a, b]$, $J = [c, d]$, $K = [e, f]$. If $I = J \times K$ and $f \in \mathcal{R}(I, \mathbb{C})$, then*

$$\int_I f = \int_I f(x, y)\, \mathrm{d}(x, y) = \int_K \int_J f(x, y)\, \mathrm{d}x\ \mathrm{d}y = \int_J \int_K f(x, y)\, \mathrm{d}y\ \mathrm{d}x\,. \tag{D.14}$$

As in the real-valued Th. 4.16*, there is a slight abuse of notation in* (D.14)*, as it can happen that a function $x \mapsto f(x, y)$ is not Riemann integrable over J and that a function $y \mapsto f(x, y)$ is not Riemann integrable over K. As in Th.* 4.16*, in that case, one can choose either the lower or the upper Riemann integral for the inner integrals in* (D.14)*. Independently of the choice, the resulting function $y \mapsto \int_J f(x, y)\, \mathrm{d}x$ is Riemann integrable over K, $x \mapsto \int_K f(x, y)\, \mathrm{d}y$ is Riemann integrable over J, and the validity of* (D.14) *is unaffected. By applying* (D.14) *inductively, one obtains*

$$\int_I f = \int_I f(x)\, \mathrm{d}x = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n)\, \mathrm{d}x_n \cdots \mathrm{d}x_1, \tag{D.15}$$

where, for the inner integrals, one can choose the upper or lower Riemann integral, and one can also permute their order without changing the overall value.

Proof. We show how to obtain the present $\mathbb{C}$-valued case from the $\mathbb{R}$-valued case of Th. 4.16. One computes, using lower Riemann integrals for the inner integrals,

$$\begin{aligned} \int_I f &= \int_I \operatorname{Re} f + i \int_I \operatorname{Im} f \\ &\overset{\text{Th. 4.16}}{=} \int_K J_*\big(\operatorname{Re} f(\cdot, y), J\big)\, \mathrm{d}y + i \int_K J_*\big(\operatorname{Im} f(\cdot, y), J\big)\, \mathrm{d}y \\ &\overset{\text{(D.13a)}}{=} \int_K J_*\big(f(\cdot, y), J\big)\, \mathrm{d}y\,, \end{aligned} \tag{D.16}$$

proving $\int_I f = \int_K \int_J f(x, y)\, \mathrm{d}x\ \mathrm{d}y$ with the inner integral interpreted as lower Riemann integral. Clearly, the same calculation works if the inner integral is interpreted as upper Riemann integral, and it also still works if J and K are switched, completing the proof of (D.14). As mentioned in the statement, (D.15) follows from (D.14) by induction. ■

D.3 Change of Variables

Theorem D.8. *Let $a, b, c, d \in \mathbb{R}^n$, $n \in \mathbb{N}$, $a < b$, $c < d$, $I := [a, b]$, $J := [c, d]$, $\phi : I \longrightarrow \mathbb{R}^n$, $f : J \longrightarrow \mathbb{C}$. If, on the interior of I, ϕ is one-to-one, Lipschitz continuous, and has continuous first partials, $\phi(I) \subseteq J$, and $(f \circ \phi)|\det J_\phi| \in \mathcal{R}(I)$, then $f\chi_{\phi(I)} \in \mathcal{R}(J)$ and the following* change of variables *formula holds:*

$$\int_J f\chi_{\phi(I)} = \int_I (f \circ \phi)|\det J_\phi|. \tag{D.17}$$

Proof. We show how to obtain the present $\mathbb{C}$-valued case from the $\mathbb{R}$-valued case of Th. 4.18. One computes

$$\begin{aligned}\int_J f\chi_{\phi(I)} &= \int_J \operatorname{Re} f\chi_{\phi(I)} + i\int_J \operatorname{Im} f\chi_{\phi(I)} \\ &\overset{(4.61)}{=} \int_I \big((\operatorname{Re} f)\circ\phi\big)|\det J_\phi| + i\int_I \big((\operatorname{Im} f)\circ\phi\big)|\det J_\phi| \\ &= \int_I (f\circ\phi)|\det J_\phi|, \end{aligned} \tag{D.18}$$

thereby establishing the case. ■

References

[Phi15a] P. Philip. *Calculus I for Computer Science and Statistics Students*. Lecture Notes, Ludwig-Maximilians-Universität, Germany, 2014/2015, available in PDF format at `http://www.math.lmu.de/~philip/publications/lectureNotes/calc1_forInfAndStatStudents.pdf`.

[Phi15b] P. Philip. *Numerical Analysis I*. Lecture Notes, Ludwig-Maximilians-Universität, Germany, 2014/2015, available in PDF format at `http://www.math.lmu.de/~philip/publications/lectureNotes/numericalAnalysis.pdf`.

[Str08] Gernot Stroth. *Lineare Algebra*, 2nd ed. Berliner Studienreihe zur Mathematik, Vol. 7, Heldermann Verlag, Lemgo, Germany, 2008 (German).