Mathematics

Imahn Shekhzadeh

June 11, 2025

Contents

1 Algebra

Definition 1.1 (Action)**.** Let G be a group (where e is the neutral element), and X a set. Then a (left) group action α of G on X is a function

$$\alpha : G \times X \to X \tag{1.1}$$

satisfying the following two axioms $(\forall g, h \in G) \wedge (\forall x \in X)$:

$$\alpha(e, x) = x,$$
$$\alpha(g, \alpha(h, x)) = \alpha(gh, x)$$

The group action is commonly written as $g \cdot x$ s.t. the axioms become

$$e \cdot x = x, \tag{1.2}$$
$$g \cdot (h \cdot x) = (gh) \cdot x \tag{1.3}$$

A right group action of G on X is a function

$$\alpha : X \times G \to X \tag{1.4}$$

satisfying the following two axioms $(\forall g, h \in G) \wedge (\forall x \in X)$:

$$\alpha(x, e) = x, \tag{1.5}$$
$$\alpha(\alpha(x, g), h) = \alpha(x, gh) \tag{1.6}$$

Example 1.2. Consider $G = \mathrm{SO}(2)$, i.e. the group of all rotations in $\mathbb{R}^2$, with the group operation given by the composition of rotations, and let $X = S^1$ be the unit circle in $\mathbb{C}$. Now consider the group action

$$\Phi : \mathrm{SO}(2) \times S^1 \to S^1, (R_\theta, p) \mapsto R_\theta(p) \tag{1.7}$$

This is a group action on $X = S^1$, since

$$\Phi(R_{\theta_1}, \Phi(R_{\theta_2}, p)) = \Phi(R_{\theta_1+\theta_2}, p).$$

In the short notation, one would write this as

$$R_{\theta_1} \cdot (R_{\theta_2} \cdot p) = R_{\theta_1+\theta_2} \cdot p.$$

Definition 1.3 (Signals[1])**.** Let V be a vector space[2] and Ω a set, possibly with additional structure. Then the space of V-valued signals on Ω,

$$\mathcal{X}(\Omega, V) = \{x : \Omega \to V\} \tag{1.8}$$

is a function space with a vector space structure. (Note that each $x \in \mathcal{X}(\Omega, V)$ is a signal.)

Addition and scalar multiplication of signals are defined as

$$(\alpha x + \beta y)(u) := \alpha x(u) + \beta y(u) \quad \forall u \in \Omega\ \forall \alpha, \beta \in \mathbb{R}. \tag{1.9}$$

[1] [18, p. 11]
[2] In computer vision, its dimensions would be called *channels*.

Lemma 1.4. If V is endowed with an inner product $\langle v, w\rangle_V$ and a measure μ on a σ-algebra defined on the set Ω (wrt which an integral can be defined), we have the following induced inner product on $\mathcal{X}(\Omega, V)$

$$\langle x, y\rangle_{\mathcal{X}(\Omega,V)} := \int_\Omega \langle \underbrace{x(u)}_{\in V}, \underbrace{y(u)}_{\in V}\rangle_V d\mu(u) \tag{1.10}$$

Proof. Trivial. ■

Remark 1.5. The domain (set) Ω can be discrete, in which case μ can be chosen to be the counting measure, in which case the integral becomes a sum.

Example 1.6. Let $\Omega = \mathbb{Z}_n \times Z_n$, i.e. a two-dimensional grid, and x an RGB image, i.e. a signal $x : \Omega \to \mathbb{R}^3$, and f a function (such as a single-layer perceptron) operating on $3n^2$-dimensional inputs.

Definition 1.7 (Homomorphism)**.** Let $f : A \to B$ be a map between two sets A and B that are equipped with the same structure. Also assume that $\cdot$ is an operation of the structure. Then f is said to be a homomorphism if $\forall x, y \in A$,

$$f(x \cdot y) = f(x) \cdot f(y). \tag{1.11}$$

Definition 1.8 (Group representation)**.** An n-dimensional real **representation** of a group G is a map $\rho : G \to \mathrm{GL}(n, \mathbb{R})$ (correspondingly, one defines an n-dimensional complex representation) [18, p. 15], where each $g \in G$ is mapped to an invertible matrix, under the *condition* that the map is a homomorphism, i.e. $\rho(gh) = \rho(g)\rho(h)$.

Example 1.9 (Group representation)**.** Let $G = {\mathbb{Z}_n}^3$, where the group operation is addition. The group representation

$$\rho : \mathbb{Z}_n \to \mathrm{GL}(1, \mathbb{C}), g \mapsto \exp\left(\frac{2\pi i g}{n}\right) \tag{1.12}$$

To see that this is indeed a group representation, note that

$$\rho(gh) = \rho(g + h) = \exp\left(\frac{2\pi i gh}{n}\right) = \exp\left(\frac{2\pi i\,(g + h)}{n}\right) = \exp\left(\frac{2\pi i g}{n}\right)\exp\left(\frac{2\pi i h}{n}\right), \tag{1.13}$$

i.e. the group representation is a homomorphism.

Definition 1.10 (Equivariance)**.** Let ρ be a representation of a group G. A function $f : \mathcal{X}(\Omega, V) \to \mathcal{X}(\Omega, V)$ is G-equivariant if

$$f(\rho(g)x) = \rho(g)f(x) \quad \forall g \in G. \tag{1.14}$$

Remark 1.11. Note that $\rho(g) \in \mathrm{GL}(n, \mathbb{K})$, $x \in \mathcal{X}(\Omega, V)$, and hence $\rho(g)x$ is a group action mediated by the representation ρ on the group element g, i.e. $\rho(g)x \in \mathcal{X}(\Omega, V)$. Also note that $f(x) \in \mathcal{X}(\Omega, V)$, i.e. $\rho(g)f(x)$ is a group action as well, where the group element g – mediated by the representation ρ – and hence $\rho(g)f(x) \in \mathcal{X}(\Omega, V)$.

The fact that we have a group action implies according to Def. 1.1 and Eq. (1.14),

$$f\left(\rho(g)(\rho(h)x)\right) \overset{(1.2)}{=} f\left((\rho(g)\rho(h))x\right) = \left(\rho(g)\rho(h)\right) f(x). \tag{1.15}$$

[3] Cyclic group of order n, $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$.

Definition 1.12 (Ring [29])**.** Let R be a set equipped with two binary operations, i.e. maps from $R \times R$ to R that are called *addition* and *multiplication* and that are denoted by $+$ and $\cdot$. Further, let there be two elements that are called *zero* and *unity* and that are denoted by 0 and 1 respectively. Then $(R, +, \cdot)$ is called a *ring* if

1. $(R, +, 0)$ is an abelian group, where 0 is the neutral element for the operation $+$,
2. $(R, \cdot, 1)$ is a monoid, where 1 is the neutral element for the operation $\cdot$,
3. the *distributive laws* hold in R, i.e. for all $a, b, c \in R$, we have

$$((a+b) \cdot c = a \cdot c + b \cdot c) \wedge (a \cdot (b+c) = a \cdot b + a \cdot c)\,,$$

4. we have $0 \cdot a = a \cdot 0 = 0$ for all $a \in R$.

Definition 1.13 (Field [29])**.** Let R be a commutative ring, then we say it is a *field* if $0 \neq 1$ in R and if every non-zero element has a multiplicative inverse.

2 Measure Theory

Definition 2.1 (Generated σ-algebra [19])**.** Let X be a set and $\mathcal{E} \subset \mathcal{P}(X)$ a non-empty collection of subsets of X. The *smallest* σ-algebra containing all the sets of $\mathcal{E}$ is denoted by $\sigma(\mathcal{E})$.

Corollary 2.2. Let $\mathcal{E}_1, \mathcal{E}_2 \subset \mathcal{P}(X)$ be such that $\mathcal{E}_1 \subset \mathcal{E}_2$. Then $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$.

Definition 2.3 (Measurable function)**.** Let $(X, \mathcal{E})$ and $(Y, \mathcal{F})$ be measurable spaces. A map $f : X \to Y$ is said to be $\mathcal{E}$*-measurable* if

$$f^{-1}(\mathcal{F}) := \left\{f^{-1}(A) | A \in \mathcal{F}\right\} := \{\{x \in X | f(x) \in A\} \,| A \in \mathcal{F}\} \subset \mathcal{E}. \tag{2.1}$$

Theorem 2.4 (Generator and measurable function [33])**.** Let $(X, \mathcal{E})$ and $(Y, \mathcal{F})$ be measurable spaces and $\mathcal{F} = \sigma(\mathcal{G})$, i.e. $\mathcal{F}$ is the σ-algebra generated by a family $\mathcal{G} \subset \mathcal{P}(Y)$, where $\mathcal{P}(Y)$ denotes the power set of Y. Then $f : X \to Y$ is measurable if and only if

$$f^{-1}(G) \in \mathcal{E} \; \forall G \in \mathcal{G}. \tag{2.2}$$

Proof. "$\Rightarrow$" Since $\mathcal{F} = \sigma(\mathcal{G})$, it obviously holds that $\mathcal{G} \subset \mathcal{F}$ and therefore $f^{-1}(G) \in \mathcal{E} \; \forall G \in \mathcal{G}$ is ensured by f being measurable.

"$\Leftarrow$" Define the set $\mathcal{M} := \{B \subset Y \mid f^{-1}(B) \in \mathcal{A}\}$. First we want to convince ourselves that $\mathcal{M}$ is a σ-algebra on Y.

1. $\emptyset \in \mathcal{M}$, since $f^{-1}(\emptyset) = \{x \in X \mid f(x) \in \emptyset\} = \emptyset$.
2. Let $B \in \mathcal{M}$, then also $Y \backslash B \in \mathcal{M}$, since $f^{-1}(Y \backslash B) = f^{-1}(Y) \backslash f^{-1}(B)$, as can be easily shown by using the definition of the complement of a set. Since $f^{-1}(Y) = X$, it follows that $f^{-1}(Y \backslash B) = X \backslash f^{-1}(B)$. Since by assumption $B \in \mathcal{M}$ (and therefore $f^{-1}(B) \in \mathcal{A}$) and $\mathcal{A}$ itself is a σ-algebra, it follows that $X \backslash f^{-1}(B) \in \mathcal{A}$.
3. Let $B_i \in \mathcal{M}$ for $i \in \mathbb{N}$, then also $\cup_{i \in \mathbb{N}} B_i \in \mathcal{M}$, since

$$f^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \bigcup_{i \in \mathbb{N}} f^{-1}(B_i).$$

Since $\mathcal{M}$ is a σ-algebra and since $\mathcal{G} \subset \mathcal{M} \Rightarrow \mathcal{F} = \sigma(\mathcal{G}) \subset \mathcal{M} = \sigma(\mathcal{M})$, it follows that f is measurable.
■

Lemma 2.5 (Push-forward measure)**.** Let $(X, \mathcal{E})$ and $(Y, \mathcal{F})$ be measurable spaces. Given a measurable map $f : X \to Y$ and a measure μ on $\mathcal{E}$, let $f_{\#}\mu$ be defined by

$$f_{\#}\mu(A) := \mu(f^{-1}(A)) \quad \forall A \in \mathcal{F}. \tag{2.3}$$

$f_{\#}\mu$ is a measure on $\mathcal{F}$ and called the *push-forward* of μ under f.

Proof. Obviously, $f_{\#}\mu(\emptyset) = \mu(f^{-1}(\emptyset)) \overset{(2.1)}{=} \mu(\emptyset) = 0$. Also, no matter what kind of set $A \in \mathcal{F}$ we take, since $\mu(A) \geq 0$, the same holds for $f_{\#}\mu(A)$. Finally, let $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ be a sequence of mutually disjoint sets, then:

$$
\begin{aligned}
& f_{\#}\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) && (2.4)\\
&= \mu\left(f^{-1}\left(\bigcup_{n\in\mathbb{N}} A_n\right)\right) && (2.5)\\
&\overset{(2.1)}{=} \mu\left(\left\{x \in X \middle| f(x) \in \bigcup_{n\in\mathbb{N}} A_n\right\}\right) && (2.6)\\
&= \mu\left(\bigcup_{n\in\mathbb{N}} \{x \in X | f(x) \in A_n\}\right) && (2.7)\\
&= \mu\left(\bigcup_{n\in\mathbb{N}} f^{-1}(A_n)\right) && (2.8)\\
&= \sum_{n\in\mathbb{N}} \mu\left(f^{-1}(A_n)\right) && (2.9)\\
&= \sum_{n\in\mathbb{N}} f_{\#}\mu(A_n) && (2.10)
\end{aligned}
$$

■

Corollary 2.6 (Push-forward of a probability measure)**.** Let $(X, \mathcal{E})$ and $(Y, \mathcal{F})$ be measurable spaces. Given a measurable map $f : X \to Y$ and a probability measure on $\mathcal{E}$, the push-forward of μ under f, denoted by $f_{\#}\mu$, is also a probability measure.

Proof. Since μ is in particular a measure and thus $f_{\#}\mu$ is also a measure, we only need to show that

$$f_{\#}\mu(Y) = \mu\left(f^{-1}\left(\mu\right)\right) = \mu\left(\{x \in X \mid f(x) \in Y\}\right) = \mu(X) = 1. \tag{2.11}$$

■

Definition 2.7 (σ-finite measure)**.** Let $(X, \mathcal{A})$ be a measurable space and μ a measure on it. If there are sets $A_1, A_2, \ldots \in \mathcal{A}$ with $\mu(A_n) < \infty$ $\forall n \in \mathbb{N}$ that satisfy

$$\bigcup_{n\in\mathbb{N}} A_n = X \tag{2.12}$$

then we say μ is *σ-finite*.

Remark 2.8. Obviously, every finite measure is σ-finite; however, the converse does not necessarily hold [15].

Definition 2.9 (Absolutely continuous measures.)**.** Let μ and ν be two measures on a σ-algebra $\mathcal{A}$. ν is called *absolutely continuous* w.r.t. μ, written as

$$\nu \ll \mu, \tag{2.13}$$

if for each $A \in \mathcal{A}$, $\mu(A) = 0$ implies $\nu(A) = 0$. If μ and ν are both absolutely continuous w.r.t each other μ and ν are called *equivalent*.

Theorem 2.10 (Radon-Nikodym Theorem [39])**.** Let μ be a σ-finite measure on a measurable space $(S, \mathcal{A})$. Then it is equivalent:

1. $\nu \ll \mu$,
2. $d\nu = hd\mu$ for some measurable function $h : S \to \mathbb{R}_+$.

The density h then is μ-a.e. finite and μ-a.e. unique.

Lemma 2.11 (Frechét Inception Distance)**.** For two multivariate Gaussian distributions $\mathcal{G}(\mu_x, \Sigma_x)$, $\mathcal{G}(\mu_y, \Sigma_y)$, the *Frechét Inception Distance* (FID) is defined as [`https://arxiv.org/pdf/1706.08500.pdf`]:

$$d\left(\mathcal{G}(\mu_x, \Sigma_x), \mathcal{G}(\mu_y, \Sigma_y)\right) := \sqrt{||\mu_x - \mu_y||_2^2 + \operatorname{Tr}\left(\Sigma_x + \Sigma_y - 2\left(\Sigma_x \Sigma_y\right)^{\frac{1}{2}}\right)}. \tag{2.14}$$

It is a metric.

Proof. Clearly, $d\left(\mathcal{G}(\mu_x, \Sigma_x), \mathcal{G}(\mu_x, \Sigma_x)\right) = 0$. Also,

$$d\left(\mathcal{G}(\mu_x, \Sigma_x), \mathcal{G}(\mu_y, \Sigma_y)\right) = d\left(\mathcal{G}(\mu_y, \Sigma_y), \mathcal{G}(\mu_x, \Sigma_x)\right),$$

which holds because $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$ for any matrices A and B. To see that

$$d\left(\mathcal{G}(\mu_x, \Sigma_x), \mathcal{G}(\mu_y, \Sigma_y)\right) \geq 0,$$

note that $\operatorname{Tr}(\Sigma_x + \Sigma_y - 2\left(\Sigma_x \Sigma_y\right)^{1/2}) = \operatorname{Tr}\left(\left(\Sigma_x^{1/2} - \Sigma_y^{1/2}\right)^2\right) \geq 0$, since the covariance matrices contain the variances on the diagonal, which are obviously non-negative. It now remains to be shown that also the triangle inequality is fulfilled. For this, note that

$$\begin{aligned}
d\left(\mathcal{G}(\mu_x, \Sigma_x), \mathcal{G}(\mu_y, \Sigma_y)\right) &= \sqrt{||\mu_x - \mu_y||_2^2 + \operatorname{Tr}\left(\Sigma_x + \Sigma_y - 2\left(\Sigma_x \Sigma_y\right)^{\frac{1}{2}}\right)} && (2.15)\\
&= \sqrt{||\mu_x - \mu_y||_2^2 + \operatorname{Tr}\left(\left(\Sigma_x^{1/2} - \Sigma_y^{1/2}\right)^2\right)} && (2.16)\\
&= \sqrt{||\mu_x - \mu_y||_2^2 + ||\sigma_x - \sigma_y||_2^2}, && (2.17)
\end{aligned}$$

where σ_x and σ_y denote vectors containing the standard deviations of the two Gaussian distributions. Clearly,

$$\begin{aligned}
d\left(\mathcal{G}(\mu_x, \Sigma_x), \mathcal{G}(\mu_z, \Sigma_z)\right) &= \sqrt{||\mu_x - \mu_z||_2^2 + ||\sigma_x - \sigma_z||_2^2} && (2.18)\\
&\leq \sqrt{||\mu_x - \mu_y||_2^2 + ||\mu_y - \mu_z||_2^2 + ||\sigma_x - \sigma_y||_2^2 + ||\sigma_y - \sigma_z||_2^2} && (2.19)\\
&\leq \sqrt{||\mu_x - \mu_y||_2^2 + ||\sigma_x - \sigma_y||_2^2} + \sqrt{||\mu_y - \mu_z||_2^2 + ||\sigma_y - \sigma_z||_2^2} && (2.20)\\
&= d\left(\mathcal{G}(\mu_x, \Sigma_x), \mathcal{G}(\mu_y, \Sigma_y)\right) + d\left(\mathcal{G}(\mu_y, \Sigma_y), \mathcal{G}(\mu_z, \Sigma_z)\right), && (2.21)
\end{aligned}$$

since $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$, as one can directly show by squaring for non-negative x, $y \in \mathbb{R}$. ∎

Definition 2.12 (Convergence in probability)**.** Assume we have a sequence of random variables $(X_n)_{n \in \mathbb{N}}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say this sequence converges to another random variable X if

$$\forall \epsilon > 0 : \lim_{n \to \infty} \mathbb{P}\left\{|X_n - X| > \epsilon\right\} = 0. \tag{2.22}$$

3 Inner Product and Normed Spaces

Definition 3.1 (Inner Product Space)**.** Let $\mathbb{K}$ be a field ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$). An inner product space is a vector space V over $\mathbb{K}$ that allows for an **inner product**

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K} \tag{3.1}$$

satisfying the following properties $\forall \alpha, \beta \in \mathbb{K}; x, y, z \in V$:

1. **Conjugate symmetry**:

$$\langle x, y \rangle = \overline{\langle y, x \rangle}, \tag{3.2}$$

   which implies that $\langle x, x \rangle \in \mathbb{R}$, even if $\mathbb{K} = \mathbb{C}$.[4]

   For $\mathbb{K} = \mathbb{R}$, conjugate symmetry is exact symmetry.

2. **Linearity** (in the first argument):

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \tag{3.3}$$

   From the conjugate symmetry property, this means that we have semi-linearity in the second argument:

$$\langle x, \alpha y + \beta z \rangle = \overline{\langle \alpha y + \beta z, x \rangle} = \overline{\alpha \langle y, x \rangle + \beta \langle z, x \rangle} = \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle. \tag{3.4}$$

3. **Positive-definiteness**:

$$\langle x, x \rangle \geq 0 \tag{3.5}$$

   and

$$\langle x, x \rangle = 0 \Leftrightarrow x = 0. \tag{3.6}$$

Definition 3.2 (Normed Space)**.** Let $\mathbb{K}$ be a space ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$). Then a mapping $||\cdot|| : X \to \mathbb{R}$ is called a **norm** if it satisfies the following properties for all φ, $\psi \in X$ and $\alpha \in \mathbb{K}$:

1. **Positivity**:

$$||\varphi|| \geq 0 \tag{3.7}$$

2. **Definiteness**:

$$||\varphi|| = 0 \Leftrightarrow \varphi = 0 \tag{3.8}$$

3. **Homogeneity**:

$$||\alpha \varphi|| = |\alpha| \, ||\varphi|| \tag{3.9}$$

[4] Set $y = x$.

4. **Triangle inequality**:

$$||\varphi + \psi|| \leq ||\varphi|| + ||\psi|| \tag{3.10}$$

A linear space X with a norm $||\cdot||$ is called a **normed linear space** or **normed space** for short. For a normed space, we shall use the notation $(X, ||\cdot||)$.

Example 3.3. Let $X = \mathbb{R}^d$. Then for $1 \leq p < \infty$, the L^p norm of a vector $x \in X$ is defined as:

$$||x||_p := \left(\sum_{j=1}^{d} |x_j|^p \right)^{\frac{1}{p}} \tag{3.11}$$

In the limit $p \to \infty$, we obain the so-called *supremum norm*:

$$||x||_\infty := \max_{1 \leq j \leq d} |x_j|. \tag{3.12}$$

In the special case of $p = 2$, we recover the Euclidean norm.

Proof. First, we show that Eq. (3.12) is indeed the limit of Eq. (3.11):

$$||x||_\infty = \max_{1 \leq j \leq d} |x_j| \leq \sum_{j=1}^{d} |x_j| \leq d \cdot ||x||_\infty \tag{3.13}$$

$$\Rightarrow ||x||_\infty^p = \left(\max_{1 \leq j \leq d} |x_j| \right)^p = \max_{1 \leq j \leq d} |x_j|^p \leq \sum_{j=1}^{d} |x_j|^p \leq d \cdot \max_{1 \leq j \leq d} |x_j|^p = d\, ||x||_\infty^p \tag{3.14}$$

$$\Rightarrow ||x||_\infty \leq \left(\sum_{j=1}^{d} |x_j|^p \right)^{\frac{1}{p}} \leq d^{\frac{1}{p}}\, ||x||_\infty \tag{3.15}$$

$$\Rightarrow \lim_{p \to \infty} ||x||_\infty = ||x||_\infty \leq \lim_{p \to \infty} \left\{ \left(\sum_{j=1}^{d} |x_j|^p \right)^{\frac{1}{p}} \right\} \leq \lim_{p \to \infty} \left\{ d^{\frac{1}{p}}\, ||x||_\infty \right\} = ||x||_\infty \tag{3.16}$$

$$\Rightarrow \lim_{p \to \infty} \left\{ \left(\sum_{j=1}^{d} |x_j|^p \right)^{\frac{1}{p}} \right\} = ||x||_\infty \tag{3.17}$$

To show the norm property of the L^p norm for $1 \leq p < \infty$, we will explicitly show the fulfilling properties of a norm, cf. Defn. (3.2), for all x, y, $z \in X$ and $\alpha \in \mathbb{R}$:

1. Positivity: $||x||_p \geq 0 \; \forall x \in X$,

2. Definiteness: $||x||_p = 0 \Leftrightarrow x = 0$,

3. Homogeneity:

$$||\alpha x||_p = \left(\sum_{j=1}^{d} |\alpha x_j|^p \right)^{\frac{1}{p}} = \left(\sum_{j=1}^{d} |\alpha|^p\, |x_j|^p \right)^{\frac{1}{p}} = |\alpha|\, ||x||_p\,,$$

4. Triangle inequality:

$$||x+y||_p = \left(\sum_{j=1}^{d} |x_j + y_j|^p\right)^{\frac{1}{p}} \leq \left(\sum_{j=1}^{d} |x_j|^p + |y_j|^p\right)^{\frac{1}{p}} = ||x||_p + ||y||_p \,.$$

In case of $p = \infty$, we will only show the triangle inequality, since the other properties are trivial to prove:

$$||x+y||_\infty = \max_{1\leq j\leq d} |x_j + y_j| \leq \max_{1\leq j\leq d} \{|x_j| + |y_j|\} \leq \max_{1\leq j\leq d} |x_j| + \max_{1\leq j\leq d} |y_j| = ||x||_\infty + ||y||_\infty \,,$$

where the last inequality holds since for any $1 \leq j \leq d$, it holds that

$$\begin{aligned} &\left(|x_j| \leq \max_{1\leq k\leq d} |x_k|\right) \wedge \left(|y_j| \leq \max_{1\leq k\leq d} |y_k|\right) \\ &\Rightarrow |x_j| + |y_j| \leq \max_{1\leq k\leq d} |x_k| + \max_{1\leq k\leq d} |y_k| \\ &\Rightarrow \max_{1\leq j\leq d} \{|x_j| + |y_j|\} \leq \max_{1\leq k\leq d} |x_k| + \max_{1\leq k\leq d} |y_k| \end{aligned}$$

■

Theorem 3.4. Let $(X, ||\cdot||)$ be a normed space. Then the "second triangle inequality" holds:

$$|||\varphi|| - ||\psi||| \leq ||\varphi - \psi|| \quad \forall \varphi, \psi \in X. \tag{3.18}$$

Proof. For φ, $\psi \in X$ we have

$$||\varphi|| = ||\varphi - \psi + \psi|| \leq ||\varphi - \psi|| + ||\psi|| \Leftrightarrow ||\varphi|| - ||\psi|| \leq ||\varphi - \psi|| \,. \tag{3.19}$$

By exchanging the roles of φ and ψ we obtain

$$||\psi|| - ||\varphi|| \leq ||\varphi - \psi|| \tag{3.20}$$

and thus

$$|||\varphi|| - ||\psi||| \leq ||\varphi - \psi|| \,. \tag{3.21}$$

■

Theorem 3.5. Let $(X, ||\cdot||)$ be a normed space. Then the addition, scalar multiplication and the norm itself are continuous.

Proof.
- ad continuity of the addition: Let $(\varphi_n)_{n\in\mathbb{N}}$ and $(\psi_n)_{n\in\mathbb{N}}$ be convergent sequences in X with limit elements φ, $\psi \in X$, i.e. $\varphi_n \longrightarrow \varphi$ and $\psi_n \longrightarrow \psi$ for $n \to \infty$. Thus

$$0 \leq ||(\varphi_n + \psi_n) - (\varphi + \psi)|| \leq ||\varphi_n - \varphi|| + ||\psi_n - \psi|| \longrightarrow 0 \quad \text{for } n \to \infty \tag{3.22}$$

and hence $\varphi_n + \psi_n \to \varphi + \psi$ for $n \to \infty$.

- ad continuity of the scalar multiplication: Let $(\alpha_n)_{n\in\mathbb{N}} \in \mathbb{K}$ converge to $\alpha \in \mathbb{K}$ and $(\varphi_n)_{n\in\mathbb{N}} \in X \to \varphi \in X$ for $n \to \infty$. Then

$$0 \leq ||\alpha_n\varphi_n - \alpha\varphi|| = ||\alpha_n(\varphi_n - \varphi) + (\alpha_n - \alpha)\varphi|| \leq ||\alpha_n(\varphi_n - \varphi)|| + ||(\alpha_n - \alpha)\varphi|| \tag{3.23}$$

$$\leq |\alpha_n|\,||\varphi_n - \varphi|| + |\alpha_n - \alpha|\,||\varphi|| \overset{n\to\infty}{\longrightarrow} 0. \tag{3.24}$$

This implies $\alpha_n\varphi_n \to \alpha\varphi$ for $n \to \infty$.

- ad continuity of the norm: Let $\varphi_n \to \varphi$. With Theorem 3.4 we have:

$$0 \leq |\ ||\varphi_n|| - ||\varphi||\ | \leq ||\varphi_n - \varphi|| \overset{n\to\infty}{\longrightarrow} \tag{3.25}$$

and hence $||\varphi_n|| \to ||\varphi||$ for $n \to \infty$.

■

Definition 3.6. Two norms $||\cdot||_a$ and $||\cdot||_b$ on a linear space X are said to be equivalent if and only if there exist positive constants $0 < c \leq C < \infty$ such that

$$c\,||\varphi||_b \leq ||\varphi||_a \leq C\,||\varphi||_b \quad \forall\varphi \in X. \tag{3.26}$$

(It is also possible to write this as $\tilde{c}\,||\varphi||_a \leq ||\varphi||_b \leq \tilde{C}\,||\varphi||_a$ with $\tilde{c} := C^{-1}$ and $\tilde{C} := c^{-1}$, where $0 < \tilde{c} \leq \tilde{C} < \infty$.)

Lemma 3.7. Let X be a linear space and the pairs $(||\cdot||_a, ||\cdot||_c)$ and $(||\cdot||_b, ||\cdot||_c)$ be equivalent. Then also the pair $(||\cdot||_a, ||\cdot||_b)$ is equivalent.

Proof. By assumption, we know that

$$c\,||\varphi||_c \leq ||\varphi||_a \leq C\,||\varphi||_c \quad \forall\varphi \in X \tag{3.27}$$

and

$$d\,||\varphi||_c \leq ||\varphi||_b \leq D\,||\varphi||_c \quad \forall\varphi \in X \tag{3.28}$$

$$\Leftrightarrow ||\varphi||_c \leq \frac{1}{d}\,||\varphi||_b \leq \frac{D}{d}\,||\varphi||_c\,. \tag{3.29}$$

$$\overset{(3.27)}{\Leftrightarrow} \frac{1}{C}\,||\varphi||_a \leq ||\varphi||_c \leq \frac{1}{d}\,||\varphi||_b \leq \frac{D}{d}\,||\varphi||_c \leq \frac{D}{d\cdot c}\,||\varphi||_a \tag{3.30}$$

$$\Leftrightarrow \frac{1}{C}\,||\varphi||_a \leq \frac{1}{d}\,||\varphi||_b \leq \frac{D}{d\cdot c}\,||\varphi||_a \tag{3.31}$$

$$\Leftrightarrow \frac{d}{C}\,||\varphi||_a \leq ||\varphi||_b \leq \frac{D}{c}\,||\varphi||_a\,. \tag{3.32}$$

It is clear that $0 < dC^{-1} \leq Dc^{-1} < \infty$ holds. ■

Theorem 3.8. On a *finite-dimensional* space X over a field $\mathbb{K}$ all norms are equivalent.

Proof [48]. Let $\dim(X) = n$, $\{e_1, \ldots, e_n\}$ be a basis of X and $||\cdot||$ a norm on X. We can now show that $||\cdot||$ is equivalent to the Euclidean norm $||\sum_{i=1}^{n} \alpha_i e_i||_2 = \left(\sum_{i=1}^{n} |\alpha_i|^2\right)^{1/2}$ as follows:

Set $K := \max\{||e_1||, \ldots, ||e_n||\} > 0$. Then from the triangle inequality for $||\cdot||$ we have:

$$||x|| = \left|\left|\sum_{i=1}^{n} \alpha_i e_i\right|\right| \leq \sum_{i=1}^{n} ||\alpha_i e_i|| = \sum_{i=1}^{n} |\alpha_i| \, ||e_i|| \tag{3.33}$$

Since $(|\alpha_1|, \ldots, |\alpha_n|)^T$, $(||e_1||_, \ldots, ||e_n||)^T \in \mathbb{R}^n$ and

$$\Big\langle (|\alpha_1|, \ldots, |\alpha_n|), (||e_1||_, \ldots, ||e_n||) \Big\rangle = \sum_{i=1}^{n} |\alpha_i| \, ||e_i|| \tag{3.34}$$

we can use the Cauchy-Schwarz inequality:

$$\langle (|\alpha_1|, \ldots, |\alpha_n|), (||e_1||_, \ldots, ||e_n||) \rangle \leq \left|\left|\sum_{i=1}^{n} \alpha_i e_i\right|\right|_2 \cdot \left|\left|\sum_{i=1}^{n} e_i\right|\right|_2 = \sqrt{\sum_{i=1}^{n} |\alpha_i|^2} \cdot \sqrt{\sum_{i=1}^{n} ||e_i||^2} \tag{3.35}$$

$$= ||x||_2 \cdot \sqrt{\sum_{i=1}^{n} K^2} = K\sqrt{n} \, ||x||_2 \quad \forall x \in X, \tag{3.36}$$

where in the last line we used that $K = \max\{||e_1||_, \ldots, ||e_n||\}$ and $x = \sum_{i=1}^{n} \alpha_i e_i$. Putting the last Eq. into Eq. (3.33), we have:

$$||x|| \leq \sum_{i=1}^{n} |\alpha_i| \, ||e_i|| \leq K\sqrt{n} \, ||x||_2 \quad \forall x \in X. \tag{3.37}$$

Now define the set

$$S := \{x \in X \mid ||x||_2 = 1\}. \tag{3.38}$$

This set is closed since it is the preimage of the closed set $\{1\} \subset \mathbb{R}$ under the continuous function $||\cdot||_2$, cf. Theorem 3.5, 4.117. S is also closed since $S \subset B_r(0) = \{\psi \in X \mid ||\psi||_2 < r\}$ for $r > 0$ (here, we take into account that every norm induces a metric). Thus, S is compact according to Heine-Borel (which applies to every finite-dimensional normed vector space). Since every continuous function takes its minimum on a compact set, we know that $||\cdot||$ has a minimum $m > 0$ on S. Since $x \cdot ||x||_2^{-1} \in S$ for $x \neq 0$, we have (m is the minimum of the function $||\cdot||$):

$$m \, ||x||_2 \leq ||x|| \quad \forall x \in X. \tag{3.39}$$

All in all, we proved:

$$m \, ||x||_2 \leq ||x|| \leq K\sqrt{n} \, ||x||_2 \quad \forall x \in X. \tag{3.40}$$

■

Definition 3.9 (Strongly equivalent metrics [21])**.** Let X be a linear space equipped with two metrics d and d'. Then the metrics are *strongly equivalent* if and only if there exist positive constants $0 < \alpha \leq \beta < \infty$ such that

$$\alpha d(x, y) \leq d'(x, y) \leq \beta d(x, y) \quad \forall x, y \in X. \tag{3.41}$$

Remark 3.10. Obviously, Eq. (3.26) can also be written as

$$c \, ||\varphi - \psi||_b \leq ||\varphi - \psi||_a \leq C \, ||\varphi - \psi||_b \quad \forall \varphi, \psi \in X. \tag{3.42}$$

Thus, Theorem 3.8 holds for metric spaces as well if the metric is defined via $d(\varphi, \psi) := ||\varphi - \psi||$ for all $\varphi, \psi \in X$.

4 Topology

Definition 4.1 (Metric Space [35])**.** Let X be a non-empty set. Then the map

$$d : X \times X \to \mathbb{R}$$

is called a *metric* on X if for all φ, ψ, $\chi \in X$ the following properties are given:

- **positivity**: $d(\varphi, \psi) \geq 0$,
- **symmetry**: $d(\varphi, \psi) = d(\psi, \varphi)$,
- **definiteness**: $d(\varphi, \psi) = 0$ iff $\varphi = \psi$,
- **triangle inequality**: $d(\varphi, \psi) \leq d(\varphi, \chi) + d(\chi, \psi)$.

Theorem 4.2. Let X be a linear space and d a metric on X. Then $||x|| := d(x, 0)$ defines a norm on X if $d(tx, 0) = |t|\, d(x, 0)$ ("scaling property") for all $t \in \mathbb{K}$, $x \in X$, and if $d(x + y, y) = d(x, 0)$ for all $x, y \in X$ ("translation invariance").

Proof. Positivity, definiteness and homogeneity of $||\cdot||$ are obvious, for the triangle inequality note that for all $x, y \in X$,

$$\begin{aligned} ||x + y|| &= d(x + y, 0) = d(x, -y) \leq d(x, 0) + d(0, -y) = d(x, 0) + d(-y, 0) \\ &= d(x, 0) + d(y, 0) = ||x|| + ||y|| \,. \end{aligned}$$

■

Definition 4.3 (Topological Space [40])**.** A topology on a set X is a collection τ of subsets of X such that

- the intersection of two members of τ is in τ,
- the union of any collection of members of τ is in τ,
- the empty set $\emptyset$ and X itself are in τ.

A set X endowed with a topological structure τ on it is called a *topological space*. The elements of X are called *points*, and the members of τ are called the *open sets*.

Definition 4.4 (Neighborhood [40])**.** Let (X, τ) be a topological space and $x \in X$, then a set $N \subset X$ is called a *neighborhood* of x in X if there is an open set $U \in \tau$ with $x \in U \subset N$.

Definition 4.5. Let (X, τ) be a topological space, then a sequence $(\varphi_n)_{n \in \mathbb{N}}$ is said to converge to a point $\varphi \in X$ if for each neighborhood U of x, there exists an $N \in \mathbb{N}$ s.t. for all $n \geq N$, we have $\varphi_n \in U$ [9].

Remark 4.6. Unlike metric spaces, the limit point of a converging sequence in topological spaces is not guaranteed to be unique.

Example 4.7. Consider $X = \mathbb{R}$ and the topology $\tau = \{\mathbb{R}, \emptyset\}$. Then a sequence $(\varphi_n)_{n \in \mathbb{N}}$ in $\mathbb{R}$ converges to any point in $\mathbb{R}$, since the neighborhood of any $x \in \mathbb{R}$ is $\mathbb{R}$.

Definition 4.8 (Basis). A *basis* of a topology (M, τ) is a collection of open sets $\mathcal{B}$ such that for all $U \in \tau$ there exists an index set I and corresponding $B_i \in \mathcal{B}$ s. t.

$$\bigcup_{i \in I} B_i = U. \tag{4.1}$$

Example 4.9. Let (X, d) be a metric space, where the metric is the discrete metric from Remark 4.17. Then the collection of all singletons $\{\varphi \in X\}$ is basis of X.

Example 4.10. In a metric space X, the collection

$$\{B_r(x) \mid x \in X, r > 0\}$$

of open balls is a basis for the topology given by the metric on X.

Definition 4.11 (Hausdorff). A topology (M, τ) is called **Hausdorff** if $\forall p \neq q \in M$ $\exists U$, $V \in \tau$ open with $p \in U$, $q \in V$ such that $U \cap V = \emptyset$.

Lemma 4.12. All metric spaces are Hausdorff spaces.

Proof: A visualization is shown in Fig. 1. To prove this more rigorously, define $U := B_r(p)$

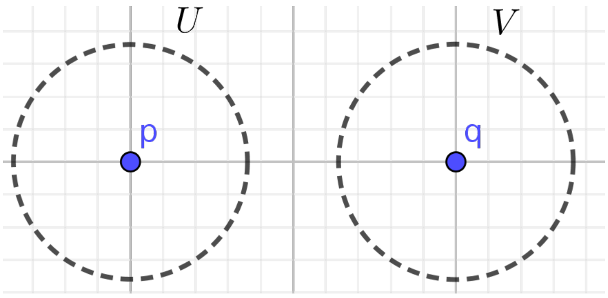


Figure 1: Visualization for why a metric space is a Hausdorff space.

and $V := B_r(q)$ with radius $r := \frac{d(p,q)}{2}$, where from Theorem 4.23 we know that U and V are open. Suppose $U \cap V = \emptyset$ would not hold. Then there exists a $z \in U \cap V$ with

$$d(p, z) < \frac{d(p, q)}{2} \tag{4.2}$$

and

$$d(q, z) < \frac{d(p, q)}{2}. \tag{4.3}$$

Therefore, by the triangle inequality for metric spaces, we have:

$$d(p, q) \leq d(p, z) + d(q, z) < d(p, q), \tag{4.4}$$

which is clearly a contradiction. ■

Definition 4.13. Let (X, τ) be a topological space. Then it is said to be *second countable* if τ has a countable basis.

Example 4.14. For any $n > 0$, the topological spaces $\mathbb{R}^n$ are second countable.

Remark 4.15. Metric spaces are not automatically second countable. For example, take an uncountable set X, endow it with the discrete metric and because of Example 4.9, 4.10, the topological space X is not second countable, since X is uncountable.

Theorem 4.16. Let (X, d) be a metric space. Then we have the inequality

$$|d(\varphi, \psi) - d(\varphi', \psi')| \leq d(\varphi, \varphi') + d(\psi, \psi') \quad \forall \varphi, \varphi', \psi, \psi' \in X. \tag{4.5}$$

Proof. Let $\varphi, \varphi', \psi, \psi' \in X$ be arbitrary. Applying the triangle inequality twice, we have

$$d(\varphi, \psi) \leq d(\varphi, \varphi') + d(\varphi', \psi) \leq d(\varphi, \varphi') + d(\psi, \psi') + d(\psi', \varphi') \tag{4.6}$$

$$\Rightarrow d(\varphi, \psi) - d(\varphi', \psi') \leq d(\varphi, \varphi') + d(\psi, \psi') \tag{4.7}$$

Exchanging the roles of φ and φ' and ψ and ψ', we get the stated inequality. ■

Remark 4.17 (Discrete Metric)**.** Let X be an arbitrary set and $d(\varphi, \psi) = 1 \ \forall \varphi, \ \psi \in X$ and $d(\varphi, \varphi) = 0$. Then d is a metric, called the *discrete metric* on X.

Definition 4.18 (Open Ball)**.** Let (X, d) be a metric space, $\varphi \in X$ and $r > 0$. Then

$$B_r(\varphi) := \{\psi \in X | d(\varphi, \psi) < r\} \subset X \tag{4.8}$$

is called an *open ball* in X with middle point φ and radius r.

Example 4.19. For the discrete metric d according to Remark 4.17, the open balls can be characterized as follows:

$$B_r(\varphi) = \begin{cases} \{\varphi\}, & r \leq 1 \\ X, & r > 1 \end{cases} \tag{4.9}$$

Definition 4.20 (Closed ball)**.** Let (X, d) be a metric space, $\varphi \in X$ and $r > 0$. Then

$$B_r[\varphi] := \{\psi \in X | d(\varphi, \psi) \leq r\} \subset X \tag{4.10}$$

is called a *closed ball* in X with middle point φ and radius r.

Definition 4.21 (Open Set)**.** Let (X, d) be a metric space. Then a subset $U \subset X$ is called *open* in X if for every $\varphi \in U$ there is an open ball $B_r(\varphi)$ that is contained in U, i.e. $B_r(\varphi) \subset U$.

Definition 4.22 (Closed Set)**.** Let (X, d) be a metric space. Then a subset $A \subset X$ is called *closed* in X if the complement $X \backslash A$ is open according to Definition 4.21.

Theorem 4.23. Open balls are open.

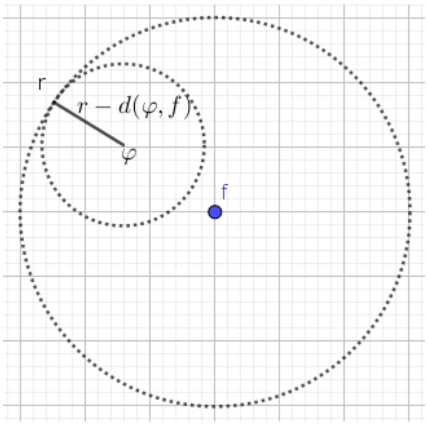
Proof: An illustration is shown in Figure. Let $f \in X$ and for $r > 0$ consider the open ball $U := B_r(f) \subset X$. By definition, for every $\varphi \in U$ it holds that $d(\varphi, f) < r$. Now we show that

$$B_{r-d(\varphi,f)}(\varphi) \subset U. \tag{4.11}$$

For this consider $\psi \in B_{r-d(\varphi,f)}(\varphi)$, i.e. $d(\psi, \varphi) < r - d(\varphi, f)$. With the triangle inequality we obtain:

$$d(f, \psi) \leq d(f, \varphi) + d(\varphi, \psi) < d(f, \varphi) + r - d(\varphi, f) = r, \tag{4.12}$$

i.e. $d(f, \psi) < r$ and thus $\psi \in U = B_r(f)$.



Theorem 4.24. Closed balls are closed.

Proof. Let $f \in X$ and for $r > 0$ consider the closed ball $B_r[f] \subset X$. By definition, for every $\varphi \in X \backslash B_r[f] \subset X$, it holds that $d(\varphi, f) > r$. We will now show that

$$B_{d(\varphi,f)-r}(\varphi) \subset X \backslash B_r[f], \tag{4.13}$$

which would prove that $B_r[f]$ is closed. Let $\psi \in B_{d(\varphi,f)-r}(\varphi)$, i.e. $d(\psi, \varphi) < d(\varphi, f) - r$. With the triangle inequality, we obtain:

$$d(f, \varphi) \leq d(f, \psi) + d(\psi, \varphi) \Leftrightarrow d(f, \psi) \geq d(f, \varphi) - d(\psi, \varphi) > d(f, \phi) - d(\varphi, f) + r = r, \tag{4.14}$$

i.e. $d(f, \psi) > r$, and hence $\psi \notin B_r[f]$, implying that $\psi \in X \backslash B_r[f]$. ■

Theorem 4.25. Let (X, d) be a metric space. The collection of open sets as in Definition 4.21 gives a topology.

Proof. According to Defn. 4.1, we need to prove that X and $\emptyset$ are open, that the union of an arbitrary number of open sets is open, and that the intersection of a finite number of open sets is open.

Clearly, X is open (since for any choice of $r > 0$ and $\varphi \in X$, $B_r(\varphi) \subset X$). $\emptyset$ is also open, since nothing needs to be shown.

Now let $U_1, \ldots, U_n \subset X$ be open, and consider the intersection

$$U := \bigcap_{i=1}^{n} U_i \subset X. \tag{4.15}$$

For any $\varphi \in U$, we know that $\varphi \in U_i$ for all $1 \leq i \leq n$. Since the U_i are open, there is a radius $r_i > 0$ such that $B_{r_i}(\varphi) \subset U_i$ for all $1 \leq i \leq n$. Defining $r := \min_{1 \leq i \leq n}\{r_i\} > 0$, we have that $B_r(\varphi) \subset U_i$ for all $1 \leq i \leq n$ and hence $B_r(\varphi) \subset U$.

Finally, let U_i, $i \in I$, be open, where I is in index set. Consider the union

$$U := \bigcup_{i \in I} U_i \subset X. \tag{4.16}$$

For any $\varphi \in U$ there is an index $i \in I$ such that $\varphi \in U_i$. Since U_i is open, there is an $r > 0$ such that $B_r(\varphi) \subset U_i \subset U$. ■

Definition 4.26 (Convergence of sequence [35])**.** Let (X, d) be a metric space. A sequence $(\varphi_n)_{n\in\mathbb{N}}$ with elements $\varphi_n \in X$, $n \in \mathbb{N}$, is called *convergent* if there is a $\varphi \in X$ s.t.

$$\lim_{n\to\infty} d(\varphi_n, \varphi) = 0, \tag{4.17}$$

i.e. for all $\epsilon > 0$ there exists an $N(\epsilon) \in \mathbb{N}$ s.t.

$$d(\varphi_n, \varphi) < \epsilon \quad \forall n \geq N(\epsilon). \tag{4.18}$$

φ is called the *limit* of the sequence $(\varphi_n)_{n\in\mathbb{N}}$. For this, we write

$$\lim_{n\to\infty} \varphi_n = \varphi \quad \text{or} \quad \varphi_n \overset{n\to\infty}{\longrightarrow} \varphi. \tag{4.19}$$

Non-convergent sequence are called *divergent*.

Theorem 4.27. Let (X, d) be a metric space, and $(\varphi_n)_{n\in\mathbb{N}}$ a convergent sequence in X wrt d. Then the limit is unique.

Proof. Assume the limit is not unique, i.e. $(\varphi_n)_{n\in\mathbb{N}}$ converges to both $\varphi \in X$ and $\psi \in X$. with $\varphi \neq \psi$. Then for $\epsilon := d(\varphi, \psi)/2$ there exists an $N' \in \mathbb{N}$ s.t. for all $n \geq N$, we have

$$d(\varphi_n, \varphi) < \frac{d(\varphi, \psi)}{2} = \epsilon$$

and an $\tilde{N}$ s.t. for all $n \geq \tilde{N}$, we have

$$d(\varphi_n, \psi) < \frac{d(\varphi, \psi)}{2} = \epsilon.$$

Define $N := \max\{N', \tilde{N}\}$, then for all $n \geq N$, we have

$$d(\varphi, \psi) \leq d(\varphi, \varphi_n) + d(\varphi_n, \psi) < d(\varphi, \psi),$$

which is a contradiction. ■

Theorem 4.28. A sequence is convergent iff every subsequence converges to the same limit.

Proof [8]. "$\Longleftarrow$" Every sequence has a trivial subsequence, namely itself.

"$\Longrightarrow$" Let $(\varphi_{n_k})_{k\in\mathbb{N}}$ be a subsequence of $(\varphi_n)_{n\in\mathbb{N}}$. Note that $n_k \geq k$ for all $k \in \mathbb{N}$, which we will prove by induction: Clearly, $n_1 \geq 1$, and $n_k \geq k$ implies $n_{k+1} > n_k \geq k$, i.e. $n_{k+1} \geq k + 1$.

We denote the limit of $(\varphi_n)_{n\in\mathbb{N}}$ by φ, and note that for all $\epsilon > 0$, there exists an $N \in \mathbb{N}$ s.t. for all $n \geq N$, we have $d(\varphi_n, \varphi) < \epsilon$. Thus, for all $k \geq N$ (and hence $n_k \geq N$), we have $d(\varphi_{n_k}, \varphi) < \epsilon$. ■

Definition 4.29 (Accumulation point)**.** Let $U \subset X$. Then $\varphi \in X$ is called an *accumulation point* of U if there is a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U such that $\lim_{n\to\infty} \varphi_n = \varphi$.

Theorem 4.30. A subset $U \subset X$ is closed if and only if it contains all its accumulation points according to Defn. 4.29.

Proof. "$\Longleftarrow$": Let U contain all its accumulation points, and let $\varphi \in X\backslash U$. We will now show that there is an $n \in \mathbb{N}$ such that $B_{\frac{1}{n}}(\varphi) \subset X\backslash U$, proving that $X\backslash U$ is open. Assuming there is no such $n \in \mathbb{N}$, then $B_{\frac{1}{n}}(\varphi) \cap U = \emptyset$ for all $n \in \mathbb{N}$. This means that there is a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U such that $d(\varphi, \varphi_n) < \frac{1}{n}$ for all $n \in \mathbb{N}$. Hence, φ is an accumulation point of U, and since U contains all its accumulation points by assumption, it follows that $\varphi \in U$, contradicting the assumption that $\varphi \in X\backslash U$.

"$\Longrightarrow$": Let U be closed, i.e. $X\backslash U$ is open. Further, let $\varphi \in X$ be an accumulation point of U, i.e. there is a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U such that $\lim\limits_{n\to\infty} \varphi_n = \varphi$. Hence, by definition of convergence, for all $\epsilon > 0$ there exists an $n \in \mathbb{N}$ such that $\varphi_n \in B(\varphi, \epsilon) \cap U$, implying that $\varphi \in U$. Thus, U contains all its accumulation points. ∎

Definition 4.31 (Closure of set)**.** The set

$$\bar{U} := \{\varphi \in X \mid \varphi \text{ is an accumulation point of } U\} \subset X \tag{4.20}$$

is called the closure of the set $U \subset X$.

Remark 4.32. Clearly, $U \subset \bar{U}$, since any $\varphi \in U$ is an accumulation point of U (consider the constant sequence, where for any $n \in \mathbb{N}$, $\varphi_n := \varphi$).

Remark 4.33. According to Theorem 4.30, a subset $U \subset X$ is closed if and only if it contains all its limit points, thus for a closed subset we have $\overline{U} \subset U$. With Remark 4.32, we get U is closed if and only if $U = \overline{U}$.

Theorem 4.34. Let $U \subset X$. Then its closure, $\bar{U}$, is closed.

Proof. According to Theorem 4.30, we need to prove that $\bar{U}$ contains all its limit points. Let $\varphi \in X$ be an accumulation point of $\overline{U}$. Then for all $n \in \mathbb{N}$ there is a $\varphi_n \in \overline{U}$ such that $d(\varphi, \varphi_n) < 1/n$. By Def. 4.31, we also know that since $\varphi_n \in \bar{U}$, it is an accumulation point of U, i.e. for every $\varphi_n \in \overline{U}$ there is a $\psi_n \in U$ such that $d(\varphi_n, \psi_n) < 1/n$. Thus:

$$d(\psi_n, \varphi) \leq d(\psi_n, \varphi_n) + d(\varphi_n, \varphi) < \frac{1}{n} + \frac{1}{n} \overset{n\to\infty}{\longrightarrow} 0, \tag{4.21}$$

i.e. $(\psi_n)_{n\in\mathbb{N}}$ is a sequence in U that converges to φ, i.e. φ is an accumulation point of U, hence $\varphi \in \bar{U}$. ∎

Theorem 4.35. Let (X, d) be a metric space. Then $\varphi \in \overline{U}$ for $U \subset X$ if and only if for every open set $O \subset X$ containing φ, $U \cap O \neq \emptyset$.

Proof. "$\Longrightarrow$" Since $\varphi \in \overline{U}$, there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ converging to φ wrt d, i.e. for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ s.t. for all $n \geq N$, $d(\varphi_n, \varphi) < \epsilon$. Let $O \subset X$ be open, and let $\varphi \in O$. By definition, there exists an open ball $B_r(\varphi) \subset O$ with $r > 0$. Choose $\epsilon = r$ and $\psi = x_N \in U$, then $\psi \in B_r(\varphi) \subset O$.

"$\Longleftarrow$" By assumption, for any non-empty open set $O \subset X$, $U \cap O \neq \emptyset$, where $U \subset X$. Let $\varphi \in X$. Since open balls are open, for any $\epsilon > 0$, $U \cap B_\epsilon(\varphi) \neq \emptyset$. Consider the sequence of open balls $\big(B_{1/n}(\varphi)\big)_{n\in\mathbb{N}}$, for which $U \cap B_{\frac{1}{n}}(\varphi) \neq \emptyset$ holds. Thus for each $n \in \mathbb{N}$, there exists a point $\varphi_n \in X$ s.t. $\varphi_n \in B_{\frac{1}{n}}(\varphi)$ and $\varphi_n \in U$. Hence, by definition, $\varphi \in X$ is an accumulation point of U. ∎

Theorem 4.36. Let $U \subset X$. Then its closure, $\overline{U}$, is the *smallest* closed set in X containing U.

Proof. From Remark 4.32, we know that $U \subset \overline{U}$, and from Theorem 4.34, we know that $\overline{U}$ is closed. To prove that $\overline{U}$ is the smallest *closed* set in X containing U, let $V \subset X$ be another closed set that contains U, i.e. $U \subset V$. Every sequence in U is also a sequence in V, i.e. every accumulation point of U is in V, i.e. $\overline{U} \subset V$. ■

Definition 4.37 (Topological equivalence [40])**.** Two metrics d and d' on a set X are called *equivalent* if they induce the same topology on X, i.e. $\tau_d = \tau_{d'}$.

Theorem 4.38. Two metrics d and d' on a set X are equivalent if and only if for each $x \in X$ and for each $\epsilon > 0$, there exist real numbers δ_1, $\delta_2 > 0$ (depending upon x and ϵ) such that $B^d_{\delta_1}(x) \subset B^{d'}_\epsilon(x)$ and $B^{d'}_{\delta_2}(x) \subset B^d_\epsilon(x)$. [40].

Proof. "$\Longrightarrow$" For each $x \in X$ and for each $\epsilon > 0$, the open ball $B^d_\epsilon(x)$ in the metric space (X, d) is open in the topology $\tau_{d'}$, and the open ball $B^{d'}_\epsilon(x)$ in the metric space (X, d') is open in the topology τ_d. Hence, there exist δ_1, δ_2 such that $B^{d'}_{\delta_1}(x) \subset B^d_\epsilon(x)$ and $B^d_{\delta_2}(x) \subset B^{d'}_\epsilon(x)$.

"$\Longleftarrow$" Suppose that $B^{d'}_{\delta_1}(x) \subset B^d_\epsilon(x)$ and $B^d_{\delta_2}(x) \subset B^{d'}_\epsilon(x)$ hold for each $x \in X$ and $\epsilon > 0$. Given an open set $U \subset X$ in the topology τ_d and a point $x \in U$, there exists an open ball $B^d_r(x) \subset U$ with $r > 0$. By assumption, there exists an open ball $B^{d'}_{\delta_1}(x) \subset B^d_r(x) \subset U$, proving that U is an open set in the topology $\tau_{d'}$. Similarly, one can show that every open set in the topology $\tau_{d'}$ is also open in the topology τ_d. Hence, $\tau_{d'} = \tau_d$. ■

Theorem 4.39. If two metrics d and d' on a set X are strongly equivalent, cf. Def. 3.9, then they are also (topologically) equivalent.

Proof. [5] By definition of strong equivalence, we know that

$$\alpha d(x, y) \leq d'(x, y) \leq \beta d(x, y) \quad \forall x, y \in X, \alpha, \beta > 0. \tag{4.22}$$

According to Theorem 4.38, we need to show that for all $x \in X$ and $\epsilon > 0$, there exist δ_1, $\delta_2 > 0$ such that

$$\left(B^d_{\delta_1}(x) \subset B^{d'}_\epsilon(x)\right) \wedge \left(B^{d'}_{\delta_2}(x) \subset B^d_\epsilon(x)\right) \tag{4.23}$$

To show the first inclusion, let $\delta_1 \leq \epsilon/\beta$, then we have for all x, $y \in X$:

$$\frac{d'(x, y)}{\beta} \leq d(x, y) < \delta_1 \leq \frac{\epsilon}{\beta} \Rightarrow d'(x, y) < \epsilon. \tag{4.24}$$

To show the second inclusion, let $\delta_2 \leq \epsilon\alpha$, then we have for all x, $y \in X$:

$$\alpha d(x, y) \leq d'(x, y) < \delta_2 \leq \epsilon\alpha \Rightarrow d(x, y) < \epsilon. \tag{4.25}$$

■

Remark 4.40. There exist metrics d and d' on a set X that are (topologically) equivalent, but not strongly equivalent. [10]

Example 4.41. Let $X := (0, 1]$, and for any x, $y \in X$, define $d(x, y) := |x - y|$ and $d'(x, y) := |x^{-1} - y^{-1}|$. Clearly, d and d' define metrics on X, and they are not strongly equivalent, since there is no $\beta > 0$ s.t. $d'(x, y) \leq d(x, y)$ for all x, $y \in X$.

However, d and d' are (topologically) equivalent. To see this, remember that acc. to Theorem 4.38, we need to prove that for any $x \in X$ and for any $\epsilon > 0$, there exist $\delta_1(x, \epsilon)$ and $\delta_2(x, \epsilon)$ s.t. $B^d_{\delta_1}(x) \subset B^{d'}_\epsilon(x)$ and $B^{d'}_{\delta_2}(x) \subset B^d_\epsilon(x)$.

To show the first inclusion, choose any

$$\delta_1 < \frac{\epsilon x^2}{\epsilon x + 1}. \tag{4.26}$$

Now we have

$$y \in B^d_{\delta_1}(x) \Rightarrow |y - x| < \delta_1 \Rightarrow \frac{1}{xy}|x - y| = \left|\frac{1}{x} - \frac{1}{y}\right| < \frac{\delta_1}{xy} \tag{4.27}$$

Since $z \leq |z|$ for any $z \in \mathbb{R}$,

$$y \in B^d_{\delta_1}(x) \Rightarrow x - y \leq |x - y| < \delta_1 \Rightarrow x - \delta_1 < y \Rightarrow \frac{1}{y} < \frac{1}{x - \delta_1}, \tag{4.28}$$

and hence

$$y \in B^d_{\delta_1}(x) \overset{(4.27)}{\Rightarrow} \left|\frac{1}{x} - \frac{1}{y}\right| < \frac{\delta_1}{xy} \overset{(4.28)}{<} \frac{\delta_1}{x(x - \delta_1)} = \frac{\delta_1}{x^2 - x\delta_1} \overset{(4.26)}{<} \frac{\epsilon x^2}{(\epsilon x + 1)\left(x^2 - \frac{\epsilon x^3}{\epsilon x + 1}\right)} \tag{4.29}$$

$$= \frac{\epsilon x^2}{\epsilon x^3 - \frac{\epsilon^2 x^4}{\epsilon x + 1} + x^2 - \frac{\epsilon x^3}{\epsilon x + 1}} = \frac{\epsilon x^2}{\frac{\epsilon^2 x^4 + \epsilon x^3}{\epsilon x + 1} - \frac{\epsilon^2 x^4}{\epsilon x + 1} + x^2 - \frac{\epsilon x^3}{\epsilon x + 1}} = \epsilon. \tag{4.30}$$

To show the second inclusion, i.e. $B^{d'}_{\delta_2}(x) \subset B^d_\epsilon(x)$, choose $\delta_2 := \epsilon$, then we have:

$$y \in B^{d'}_{\delta_2}(x) \Rightarrow \left|\frac{1}{x} - \frac{1}{y}\right| < \delta_2 \Rightarrow xy\left|\frac{1}{x} - \frac{1}{y}\right| = |y - x| < \delta_2 xy \overset{x,y \in X = (0,1]}{\leq} \delta_2 = \epsilon. \tag{4.31}$$

Theorem 4.42. Two metrics d and d' on a set X are (topologically) equivalent iff each convergent sequence in X wrt d is also convergent in X wrt d'. In that case, their limits coincide.

Proof. "$\Longrightarrow$" Let d and d' be equivalent metrics on X, and let $(\varphi_n)_{n \in \mathbb{N}}$ be a sequence in X, which converges to $\varphi \in X$ wrt d, i.e.

$$d(\varphi, \varphi_n) \overset{n \to \infty}{\longrightarrow} 0. \tag{4.32}$$

According to Theorem 4.38, for any $\epsilon > 0$, there is a $\delta > 0$ s.t. $B^d_\delta(\varphi) \subset B^{d'}_\epsilon(\varphi)$. To this δ, there exists an $N \in \mathbb{N}$ s.t. $d(\varphi, \varphi_n) < \delta$ for all $n \geq N$, cf. Def. 4.26. Hence:

$$\varphi_n \in B^d_\delta(\varphi) \subset B^{d'}_\epsilon(\varphi), \quad \forall n \geq N \tag{4.33}$$

i.e. $d'(\varphi, \varphi_n) < \epsilon$ for any $\epsilon > 0$.

"$\Longleftarrow$" Let $U \subset X$ be d-open in X, i.e. open wrt d. Then $X \backslash U$ is d-closed in X. Let $\varphi \in X$ be a d'-accumulation point of $X \backslash U$, i.e. there exists a sequence $(\varphi_n)_{n \in \mathbb{N}}$ in $X \backslash U$ s.t. $\varphi_n \xrightarrow{d'} \varphi$ for $n \to \infty$. By assumption, $(\varphi_n)_{n \in \mathbb{N}}$ converges to a $\psi \in X$ wrt d. Since $X \backslash U$ is closed, by Theorem 4.30, $\psi \in X \backslash U$ holds.

Let us now show that $\varphi = \psi$. For this, consider the extended sequence

$$(\varphi_1, \varphi, \varphi_2, \varphi, \varphi_3, \varphi, \dots)\,. \tag{4.34}$$

The extended sequence converges to φ wrt d', and hence it also converges wrt d. The subsequence $(\varphi_n)_{n\in\mathbb{N}}$ of the extended sequence converges to ψ wrt d. Hence, $\varphi = \psi$, and the d'-accumulation point $\varphi = \psi$ of $X\backslash U$ lies in $X\backslash U$, i.e. $X\backslash U$ is also d'-closed (by Theorem 4.30). This in turn implies that U is d'-open.

We have shown that d and d' induce the same topology on X, and are hence equivalent. ∎

Definition 4.43 (Bounded set). A subset $U \subset X$ is called *bounded* in X if it is contained in a closed ball in X, i.e. there exists $r > 0$ and $\varphi \in X$ s.t. $U \subset B_r[\varphi]$.

Theorem 4.44. Convergent sequences in X are fully contained in a bounded subset.

Proof. Let $(\varphi_n)_{n\in\mathbb{N}}$ be a convergent sequence in X with limit point $\varphi \in X$. Then there exists an $N \in \mathbb{N}$ s.t. $d(\varphi_n, \varphi) < \epsilon$ for all $n \geq N$ for a fixed $\epsilon > 0$. Defining

$$r := \max\left\{\epsilon, \max_{n\leq N}\{d(\varphi, \varphi_n)\}\right\}, \tag{4.35}$$

we have $d(\varphi, \varphi_n) \leq r$ for all $n \in \mathbb{N}$. Hence, for any $n \in \mathbb{N}$, $\varphi_n \in B_r[\varphi]$.¸ ∎

Definition 4.45 (Isometry). Let (X, d) and (X', d') be metric spaces. Then $f : X \to Y$ is called a distance-preserving map (or *isometry*) if

$$d'(f(\varphi), f(\psi)) = d(\varphi, \psi) \quad \forall \varphi, \psi \in X \tag{4.36}$$

(X, d) and (X', d') are called *isometric* if there is a surjective isometry $f : X \to X'$.

Theorem 4.46. Let (X, d) and (X', d') be metric spaces and $f : X \to X'$ be an isometry. Then it is injective.

Proof. If $f(\varphi) = f(\psi)$ for any φ, $\psi \in X$, then

$$d'(f(\varphi), f(\psi)) = d'(f(\varphi), f(\varphi)) = d(\varphi, \psi) = 0, \tag{4.37}$$

i.e. $\varphi = \psi$. ∎

Remark 4.47. Two isometric spaces (X, d) and (X', d') do not differ in their properties wrt their metrics, but instead in the properties of their individual elements. Any metric property holding in one metric space also holds in the other. Hence, in this sense, we say that (X, d) and (X', d') are *identical*.

Remark 4.48. Surjective isometries are hence bijective.

4.1 Complete metric spaces

Definition 4.49. A sequence $(\varphi_n)_{n\in\mathbb{N}}$ in a metric space (X, d) is a *Cauchy sequence* if

$$\lim_{n,m\to\infty} d(\varphi_n, \varphi_m) = 0, \tag{4.38}$$

i.e. for all $\epsilon > 0$ there exists an $N(\epsilon) \in \mathbb{N}$ s.t.

$$d(\varphi_n, \varphi_m) < \epsilon \quad \forall n, m \geq N. \tag{4.39}$$

Theorem 4.50. Every convergent sequence in X is a cauchy sequence in X.

Proof. Let $(\varphi_n)_{n\in\mathbb{N}}$ be a convergent sequence in X with limit point $\varphi \in X$. Hence, for all $\epsilon > 0$, there is an $N(\epsilon) \in \mathbb{N}$ with $d(\varphi_n, \varphi) < \epsilon$ for all $n \geq N(\epsilon)$. Hence

$$d(\varphi_n, \varphi_m) \leq d(\varphi_n, \varphi) + d(\varphi, \varphi_m) < 2\epsilon \quad \forall n, m \geq N. \tag{4.40}$$

■

Theorem 4.51. A Cauchy sequence in a metric space (X, d) is convergent iff it contains a convergent subsequence.

Proof [11]. "$\Longrightarrow$" Every sequence has a trivial subsequence, namely itself.

"$\Longleftarrow$" Let $(\varphi_{n_k})_{k\in\mathbb{N}}$ be a subsequence of the Cauchy sequence $(\varphi_n)_{n\in\mathbb{N}}$, and assume that $\lim_{k\to\infty} \varphi_{n_k} = \varphi$ wrt d, where $\varphi \in X$, i.e. for all $\epsilon > 0$ there exists an $N' \in \mathbb{N}$ s.t. for all $k \geq N'$, $d(\varphi_{n_k}, \varphi) < \epsilon$ (note that $n_k \geq k$). Since $(\varphi_n)_{n\in\mathbb{N}}$ is Cauchy, there exists an $\tilde{N}$ s.t. for all $m, n \geq \tilde{N}$, we have $d(\varphi_n, \varphi_m) < \epsilon$. Define $N := \max\left\{N', \tilde{N}\right\}$, then for all $k, n \geq N$,

$$d(\varphi_n, \varphi) \leq d(\varphi_n, \varphi_{n_k}) + d(\varphi_{n_k}, \varphi) < 2\epsilon,$$

and thus $(\varphi_n)_{n\in\mathbb{N}}$ converges to φ. ■

Remark 4.52. The converse of Theorem 4.50 is in general not true: Take $X = \mathbb{Q}$ and let $d(x, y) := |x - y|$, for any x, $y \in \mathbb{Q}$. Then we have a metric space $(\mathbb{Q}, d)$, but the sequence $(a_n)_{n\in\mathbb{N}}$ with

$$a_n := \left(1 + \frac{1}{n}\right)^n, \tag{4.41}$$

which is a Cauchy sequence and where all elements are in $\mathbb{Q}$, does not converge in $\mathbb{Q}$ (the limit point is Euler's number e). To prove this, we will make use of the following two Theorems.

Theorem 4.53. Any sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \in \mathbb{R}$ for any $n \in \mathbb{N}$ that is monotonically increasing and upper bounded for any $n \in \mathbb{N}$ is a Cauchy sequence in $(\mathbb{R}, d)$ with $d(x, y) := |x - y|$ for any x, $y \in \mathbb{R}$.

Proof. ([6]) The completeness axiom states that any non-empty subset of $\mathbb{R}$ that has an upper bound has a least upper bound in $\mathbb{R}$. Let M be the least upper bound for the sequence $(a_n)_{n\in\mathbb{N}}$, i.e. $a_n \leq M$ for all $n \in \mathbb{N}$. Since M is the least upper bound, for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ s.t. $M - \epsilon < a_N \leq M$.

By assumption, the sequence is monotonically increasing, i.e. $a_m \geq a_n$ for any $m > n$. Without loss of generality, let $m > n$, then we have

$$M - \epsilon < a_N \leq a_n \leq a_m \leq M, \quad \forall (m, n \geq N) \wedge (m > n) \tag{4.42}$$

from which $a_m - a_n < \epsilon$ follows. Similarly, we can show that $a_n - a_m < \epsilon$ for $n > m$. Hence, $d(a_m, a_n) < \epsilon$ for any m, $n \geq N$. ■

Theorem 4.54. Let $x_1, \dots, x_n \in \mathbb{R}_{\geq 0}$, $n \geq 1$, then the arithmetic mean is bigger than or equal to the geometric mean, i.e.

$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n} \tag{4.43}$$

Proof. ([1]) Via induction. Let $n = 1$, then $x_1 \geq x_1$, which is true.

Assuming that Eq. (4.43) holds for any $n \in \mathbb{N}$, we will make the transition $n \mapsto n+1$ and denote the arithmetic mean of $x_1, \dots, x_{n+1}$ by α:

$$(n+1)\alpha = \sum_{i=1}^{n+1} x_i \Leftrightarrow n\alpha = \sum_{i=1}^{n+1} x_i - \alpha \tag{4.44}$$

If $x_i = \alpha$ for all $1 \leq i \leq n+1$, then we are done, since Eq. (4.43) would be an equality. If all the x_i are not equal to α, then WLOG, let there be exactly two x_i that are not equal to α. Place these two elements at the end, and noting that one needs to be bigger than α and the other smaller than α, WLOG we have

$$(x_n - \alpha > 0) \wedge (\alpha - x_{n+1} > 0) \Rightarrow (x_n - \alpha)(\alpha - x_{n+1}) > 0. \tag{4.45}$$

Now define

$$y := x_n + x_{n+1} - \alpha \geq x_n - \alpha \overset{(4.45)}{>} 0. \tag{4.46}$$

and substitute it into Eq. (4.44):

$$n\alpha = \sum_{i=1}^{n+1} x_i - \alpha = \sum_{i=1}^{n-1} x_i + x_n + x_{n+1} - \alpha = \sum_{i=1}^{n-1} x_i + y \tag{4.47}$$

Thus, α is also the arithmetic mean of the n numbers $x_1, \dots, x_{n-1}, y$. Making use of the induction hypothesis, we get

$$\alpha^{n+1} = \alpha\alpha^n \geq \alpha y \prod_{i=1}^{n-1} x_i. \tag{4.48}$$

From Eq.(4.45) we know that

$$y\alpha - x_n x_{n+1} = (x_n + x_{n+1} - \alpha)\alpha - x_n x_{n+1} = (x_n - \alpha)(\alpha - x_{n+1}) > 0 \Leftrightarrow y\alpha > x_n x_{n+1} \tag{4.49}$$

Substituting Eq. (4.49) into (4.48), we get

$$\alpha^{n+1} \geq \alpha y \prod_{i=1}^{n-1} x_i > x_n x_{n+1} \prod_{i=1}^{n-1} x_i = \prod_{i=1}^{n+1} x_i \Leftrightarrow \alpha \geq \left(\prod_{i=1}^{n+1} x_i\right)^{\frac{1}{n+1}}, \tag{4.50}$$

which proves the induction hypothesis. ■

Proof. (of Remark 4.52) Looking at the proof of Theorem 4.53, we will see that that the statement of Theorem 4.53 also holds for sequences in $\mathbb{Q}$ that are upper bounded and monotonically increasing, since any non-empty subset of $\mathbb{Q} \subset \mathbb{R}$ that is upper bounded has a least upper bound in $\mathbb{R}$ (though not necessarily in $\mathbb{Q}$).

Hence, to show that the sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n = (1+n^{-1})^n$ is a Cauchy sequence in $\mathbb{Q}$, we only need to show that it is monotonically increasing and derive an upper bound.

For the monotonicity, we will make use of Theorem 4.54 ([2]). For this, define the x_i as follows:

$$x_i := \begin{cases} 1, & i = 1 \\ 1 + \frac{1}{n}, & 2 \leq i \leq n+1 \end{cases} \tag{4.51}$$

Then we have

$$\frac{\sum_{i=1}^{n+1} x_i}{n+1} \geq \left(\prod_{i=1}^{n+1} x_i\right)^{\frac{1}{n+1}} \Leftrightarrow \left(\frac{\sum_{i=1}^{n+1} x_i}{n+1}\right)^{n+1} \geq \prod_{i=1}^{n+1} x_i \tag{4.52}$$

$$\Rightarrow \left(\frac{1+n\left(1+\frac{1}{n}\right)}{n+1}\right)^{n+1} = \left(1+\frac{1}{n+1}\right)^{n+1} \geq \left(1+\frac{1}{n}\right)^{n} \quad \forall n \in \mathbb{N}, \tag{4.53}$$

i.e. $a_{n+1} > a_n$.

To prove that the sequence $(a_n)_{n\in\mathbb{N}}$ is upper bounded ([12]), note that

$$a_n = \left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k} = \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \frac{1}{n^k} = 1 + \sum_{k=1}^{n} \frac{n!}{k!\,(n-k)!} \frac{1}{n^k} \tag{4.54}$$

$$= 1 + \sum_{k=1}^{n} \frac{n(n-1)\dots(n-k+1)}{k!} \frac{1}{n^k} \leq 1 + \sum_{k=1}^{n} \frac{1}{k!} \leq 1 + \sum_{k=1}^{n} \frac{1}{2^{k-1}} \tag{4.55}$$

$$\leq 1 + \sum_{k=0}^{\infty} \frac{1}{2^k} = 3. \tag{4.56}$$

■

Example 4.55. We can instruct infinitely many counter-examples to Theorem 4.50 in $\mathbb{Q}$. Let

$$(a_n)_{n\in\mathbb{N}} := \lfloor 10^n \cdot a \rfloor \cdot 10^{-n}, \tag{4.57}$$

where $a \in \mathbb{R}\backslash\mathbb{Q}$. $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{Q}$, which gives a to its n^{th} decimal point, converging to a. To see that it is a Cauchy sequence in $\mathbb{Q}$, note that all sequence elements are rational, that the sequence is bounded by a and that it is monotonically increasing. Thus, by the proof of Theorem 4.53, which also holds for $\mathbb{Q}$, $(a_n)_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{Q}$.

Theorem 4.56. Any sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \in \mathbb{R}$ for any $n \in \mathbb{N}$ that is monotonically increasing and upper bounded for any $n \in \mathbb{N}$ converges in $(\mathbb{R}, d)$ with $d(x,y) := |x-y|$ for any $x, y \in \mathbb{R}$.

Proof. The proof is identical to that of Theorem 4.53. With the same notation as there, we have, cf. Eq. (4.42):

$$M - \epsilon < a_n \le M, \quad \forall n \ge N \tag{4.58}$$
$$\Rightarrow |a_n - M| < \epsilon, \tag{4.59}$$

i.e. the limit point of $(a_n)_{n\in\mathbb{N}}$ is M (the least upper bound). ■

Remark 4.57. Note that Theorem 4.56 does not hold in $\mathbb{Q}$, since the completeness axiom only guarantess the existence of a least upper bound in $\mathbb{R}$.

Theorem 4.58. Let (X, d) be a metric space. Then if $(\varphi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, it is bounded.

Proof. Fix $\epsilon > 0$ and $\psi \in X$. Since $(\varphi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, there exists an $N \in \mathbb{N}$ s.t. $d(\varphi_m, \varphi_n) < \epsilon$ for all $m, n \ge N$. Define

$$M := \max_{1\le i\le N} \{d(\psi, \varphi_n)\}, \tag{4.60}$$

and note that the maximum exists since we take it over a set with finitely many elements. For $n \le N$, we obviously have $d(\psi, \varphi_n) \ge M$. For $n > N$, we have

$$d(\psi, \varphi_n) \le d(\psi, \varphi_N) + d(\varphi_N, \varphi_n) < M + \epsilon.$$

Thus, $(\varphi_n)_{n\in\mathbb{N}}$ is contained in $B_{M+\epsilon+1}[\psi]$. ■

Definition 4.59. Let (X, d) be a metric space. Then (X, d) is called *complete* if every Cauchy sequence in X converges to a $\varphi \in X$ wrt d.

Theorem 4.60. Let (X, d) be a metric space and consider the set Y. If $f : X \to Y$ is a bijection, then we can endow Y with the following metric d':

$$d'(y_1, y_2) := d\left(f^{-1}(y_1), f^{-1}(y_2)\right). \quad \forall y_1, y_2 \in Y \tag{4.61}$$

(X, d) is complete iff (Y, d') is complete [4].

Proof. Let (X, d) be complete, i.e. every Cauchy sequence in X converges in X. Now let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in X with limit point $\varphi \in X$. Consider the sequence $(y_n)_{n\in\mathbb{N}} := (f(x_n))_{n\in\mathbb{N}}$ in Y. Since X is Cauchy, we have $d(x_m, x_n) = d'(y_m, y_n) < \epsilon$ for all $\epsilon > 0$ and m, $n \ge N$. Thus, we have that $(y_n)_{n\in\mathbb{N}}$ is also a Cauchy sequence in Y. The limit point of $(y_n)_{n\in\mathbb{N}}$ is $f(\varphi)$, since $d'(y_n, f(\varphi)) = d(x_n, \varphi) < \epsilon'$ for all $\epsilon' > 0$ and $n \ge N'$. ■

Remark 4.61. To construct incomplete metric spaces, we can thus take an incomplete metric space (X, d), construct a surjection $f : X \to Y$ to another set Y and endow Y with the metric d' from Eq. (4.61), then (Y, d') is also incomplete.

Example 4.62. Let $X = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, and consider the metric $d(x_1, x_2) := |x_1 - x_2|$. Define $Y = \mathbb{R}$, and consider the map $f : X \to Y$, $x \mapsto \tan(x)$, which is surjective. On Y, consider the metric $d'(y_1, y_2) := d(\arctan(y_1), \arctan(y_2)) = |\arctan y_1 - \arctan y_2|$, cf. Eq. (4.61). Since (X, d) is incomplete, (Y, d') is incomplete as well. To see that (X, d) is incomplete, note that $\left(\frac{\pi}{2} - \frac{1}{n}\right)_{n\in\mathbb{N}}$ or $\left(-\frac{\pi}{2} + \frac{1}{n}\right)_{n\in\mathbb{N}}$ are Cauchy in X, yet do not converge in X.

Example 4.63 (Complete metric spaces)**.** Any set X equipped with the discrete metric from Def. 4.17 is complete.

Example 4.64. The metric space $(\mathbb{R}, d)$ with $d(x, y) := |x - y|$ is complete. Also, $(\mathbb{R}^d, ||x||_p)$, cf. Example 3.3, for $1 \leq p \leq \infty$ is complete.

Example 4.65. Let $\mathcal{C}[a, b]$ be the space of all continuous functions $\chi : [a, b] \to \mathbb{R}$ with the metric

$$d_\infty(\varphi, \psi) = ||\varphi - \psi||_\infty := \sup_{x \in [a,b]} \{|\varphi(x) - \psi(y)|\} \quad \forall \varphi, \psi \in \mathcal{C}[a, b]. \tag{4.62}$$

The metric space $(\mathcal{C}[a, b], d_\infty)$ is complete.

Proof. Let $(\varphi_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in $\mathcal{C}[a, b]$, then for any $x \in [a, b]$:

$$|\varphi_n(x) - \varphi_m(x)| \leq \sup_{x \in [a,b]} \{|\varphi_n(x) - \varphi_m(y)|\} < \epsilon. \quad \forall m, n \geq N \tag{4.63}$$

Hence, $(\varphi_n)_{n \in \mathbb{N}}(x)$ is a Cauchy sequence in $\mathbb{R}$ for any $x \in [a, b]$ wrt the metric $d(x, y) := |x - y|$. Since $(\mathbb{R}, d)$ is complete, we know that $\varphi(x) := \lim_{n \to \infty} \varphi_n(x)$ exists.

Finally, let us show that $\varphi \in \mathcal{C}[a, b]$. Since $(\varphi_n(x))_{n \in \mathbb{N}}$ converges in $(\mathbb{R}, d)$, we can choose an $N \in \mathbb{N}$ s.t. $|\varphi_n(x) - \varphi(x)| < \epsilon/3$ for all $n \geq N$ and all $x \in [a, b]$. And since $(\varphi_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in $(\mathcal{C}[a, b], d_\infty)$ (by assumption), i.e. the φ_n are continuous, there exists a $\delta > 0$ s.t. $|\varphi_N(x) - \varphi_N(y)| < \epsilon/3$ with $|x - y| < \delta$. If $|x - y| < \delta$, we get by applying the triangle inequality

$$|\varphi(x) - \varphi(y)| \leq |\varphi(x) - \varphi_N(x)| + |\varphi_N(x) - \varphi_N(y)| + |\varphi_N(y) - \varphi(y)| < \epsilon, \tag{4.64}$$

which shows the continuity of φ, cf. Def. 4.108.

Thus, we have shown that $\varphi \in \mathcal{C}[a, b]$. ∎

Remark 4.66. Let (X, d) be a metric space. Then if $(\varphi_n)_{n \in \mathbb{N}}$ and $(\psi_n)_{n \in \mathbb{N}}$ are Cauchy sequences in (X, d), their sum $(\varphi_n + \psi_n)_{n \in \mathbb{N}}$ is *not* necessarily a Cauchy sequence in X wrt d. (Note that we are assuming here that X is equipped with a binary operation $+$.)

Example 4.67. Consider the metric space $(\mathbb{R}, d)$ with $d(x, y) := |\arctan x - \arctan y|$ for any x, $y \in \mathbb{R}$, cf. Example 4.62. Define the sequences $(a_n)_{n \in \mathbb{N}}$ as $a_n := n$ and $(b_n)_{n \in \mathbb{N}}$ as $b_n := (-1)^n - n$. Note that both sequences are Cauchy, but $(a_n + b_n)_{n \in \mathbb{N}} = ((-1)^n)_{n \in \mathbb{N}}$ is not.

Remark 4.68. Let (X, d) be a metric space. Then if $(\varphi_n)_{n \in \mathbb{N}}$ and $(\psi_n)_{n \in \mathbb{N}}$ are Cauchy sequences in (X, d), their product $(\varphi_n \cdot \psi_n)_{n \in \mathbb{N}}$ is *not* necessarily a Cauchy sequence in X wrt d. (Note that we are assuming here that X is equipped with a binary operation $\cdot$.)

Example 4.69. Consider the metric space $(\mathbb{R}, d)$ with $d(x, y) := |\arctan x - \arctan y|$ for any x, $y \in \mathbb{R}$, cf. Example 4.62. Define the sequences $(a_n)_{n \in \mathbb{N}}$ as $a_n := n$ and $(b_n)_{n \in \mathbb{N}}$ as $b_n := \frac{(-1)^n}{n}$. Note that both sequences are Cauchy, but $(a_n \cdot b_n)_{n \in \mathbb{N}} = ((-1)^n)_{n \in \mathbb{N}}$ is not.

Remark 4.70. In the special case of $X = \mathbb{Q}$ or $X = \mathbb{R}$ with $d(x, y) := |x - y|$ for any x, $y \in X$, then the sum and product of two Cauchy sequences is indeed Cauchy again, cf. Theorem B.3 and B.4.

Proof of Example 4.67. First, we will show that $(a_n)_{n\in\mathbb{N}} = (n)_{n\in\mathbb{N}}$ is Cauchy. For any $\epsilon > 0$, choose $N = N(\epsilon)$ s.t. $N > 1/\epsilon$, and note that for any m, $n \geq N$, we have

$$d(a_m, a_n) = |\arctan m - \arctan n| = \left|\arctan\left(\frac{m-n}{1+mn}\right)\right| \overset{(\star)}{\leq} \left|\frac{m-n}{1+mn}\right| \leq \frac{|m-n|}{mn}, \tag{4.65}$$

where the inequality in $(\star)$ can be obtained by using the integral definition of $\arctan x$. WLOG assume $n \leq m$, then Equation (4.65) can be further simplified to

$$d(a_m, a_n) \overset{(4.65)}{\leq} \frac{|m-n|}{mn} \leq \frac{m}{mn} \overset{n\geq N}{\leq} \frac{m}{mN} = \frac{1}{N} \overset{N>\epsilon^{-1}}{<} \epsilon.$$

To show that $(b_n)_{n\in\mathbb{N}} = ((-1)^n - n)_{n\in\mathbb{N}}$ is Cauchy, choose $N = N(\epsilon)$ s.t. $N > \frac{1}{\epsilon} + 2$, and note that for m, $n \geq N$, we have

$$\begin{aligned}
d(b_m, b_n) &= |\arctan((-1)^m - m)| - |\arctan((-1)^n - n)| && (4.66)\\
&= \left|\arctan\left(\frac{(-1)^m - m - (-1)^n + n}{1 + ((-1)^m - m)\cdot((-1)^n - n)}\right)\right| && (4.67)\\
&\leq \left|\frac{(-1)^m - m - (-1)^n + n}{1 + ((-1)^m - m)\cdot((-1)^n - n)}\right| && (4.68)\\
&= \left|\frac{1}{1 + ((-1)^m - m)\cdot((-1)^n - n)}\right| \cdot |(-1)^m - m - (-1)^n + n| && (4.69)\\
&\leq \left|\frac{1}{1 + ((-1)^m - m)\cdot((-1)^n - n)}\right| |n - m| && (4.70)\\
&= \left|\frac{m-n}{1 + (-1)^{m+n} - m\cdot(-1)^n - n\cdot(-1)^m + mn}\right| && (4.71)\\
&\leq \left|\frac{m-n}{-m\cdot(-1)^n - n\cdot(-1)^m + mn}\right| && (4.72)
\end{aligned}$$

We will distinguish four cases:

1. If m and n are both uneven, then Eq. (4.72) becomes

$$\begin{aligned}
d(b_m, b_n) &\overset{(4.72)}{\leq} \left|\frac{m-n}{-m\cdot(-1)^n - n\cdot(-1)^m + mn}\right| \\
&= \left|\frac{m-n}{m+n+mn}\right| \leq \left|\frac{m-n}{nm}\right|,
\end{aligned}$$

   which reduces to showing that $(a_n)_{n\in\mathbb{N}}$ is Cauchy, cf. Eq. (4.65).

2. If m and n are both even and WLOG we assume $n \leq m$, then Eq. (4.72) becomes

$$\begin{aligned}
d(b_m, b_n) &\overset{(4.72)}{\leq} \left|\frac{m-n}{-m\cdot(-1)^n - n\cdot(-1)^m + mn}\right| \\
&= \left|\frac{m-n}{-m-n+mn}\right| \overset{n\leq m}{\leq} \frac{m}{|-m-n+mn|} = \frac{1}{|-1-\frac{n}{m}+n|} \\
&\overset{(\star)}{=} \frac{1}{-1-\frac{n}{m}+n} \overset{(\diamond)}{\leq} \frac{1}{n-2} \overset{n\geq N}{\leq} \frac{1}{N-2} \overset{N>\epsilon^{-1}+2}{<} \epsilon,
\end{aligned}$$

where the equality in $(\star)$ and the inequality in $(\star)$ follow from $n \geq N > 2$ and $n \leq m$, since

$$n \leq m \Rightarrow \frac{n}{m} \leq 1 \Rightarrow -1 - \frac{n}{m} + n \geq n - 2 > 0.$$

3. If m is uneven and n is even and WLOG we asumme $n \leq m$, then we have

$$\begin{aligned} d(b_m, b_n) &\overset{(4.72)}{\leq} \left| \frac{m-n}{-m \cdot (-1)^n - n \cdot (-1)^m + mn} \right| = \left| \frac{m-n}{-m+n+mn} \right| \\ &\overset{n \leq m}{\leq} \frac{m}{|n-m+mn|} = \frac{1}{\left|\frac{n}{m} - 1 + n\right|} = \frac{1}{\frac{n}{m} - 1 + n} \leq \frac{1}{n-1} \\ &\overset{n \geq N}{\leq} \frac{1}{N-1} \overset{N > \epsilon^{-1}+2}{<} \epsilon. \end{aligned}$$

4. If m is even and n is uneven, proceed similarly to 3.

Finally, note that it is obvious that $(a_n + b_n)_{n\in\mathbb{N}} = ((-1)^n)_{n\in\mathbb{N}}$ is not Cauchy. ■

Proof of Example 4.69. We have already shown in the proof of Example 4.67 that $(a_n)_{n\in\mathbb{N}} = (n)_{n\in\mathbb{N}}$ is Cauchy wrt $d(x, y) := |\arctan x - \arctan y|$ for any $x, y \in \mathbb{R}$. To show that $(b_n)_{n\in\mathbb{N}} = \left(\frac{(-1)^n}{n}\right)_{n\in\mathbb{N}}$ is Cauchy wrt d, choose $N = N(\epsilon) > \frac{2}{\epsilon}$, then we have for any $m, n \geq N$:

$$\begin{aligned} d(b_m, b_n) &= \left| \arctan\left(\frac{(-1)^m}{m}\right) - \arctan\left(\frac{(-1)^n}{n}\right) \right| \\ &\leq \left| \arctan\left(\frac{(-1)^m}{m}\right) \right| + \left| \arctan\left(\frac{(-1)^n}{n}\right) \right| \\ &\leq \left| \frac{(-1)^m}{m} \right| + \left| \frac{(-1)^n}{n} \right| = \frac{1}{m} + \frac{1}{n} \\ &\overset{\text{WLOG } n \leq m}{\leq} \frac{2}{n} \overset{n \geq N}{\leq} \frac{2}{N} \overset{N > 2\epsilon^{-1}}{<} \epsilon. \end{aligned}$$

■

Theorem 4.71. Let (X, d) be a metric space. Then the following statements are true.

(a) Complete subsets of (X, d) are closed.
(b) If (X, d) is complete, then every closed subset of (X, d) is complete.

Proof. (a): Let $U \subset X$ be closed wrt d, and let $\varphi \in X$ be a limit point of $\overline{U}$, i.e. there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U s.t. $\lim_{n\to\infty} \varphi_n = \varphi$. Hence, $(\varphi_n)_{n\in\mathbb{N}}$ converges in X, and is thus a Cauchy sequence in X. Since U is closed, the limit of the sequence lies in U, i.e. U is closed.

(b): Let (X, d) be complete, $U \subset X$ closed and $(\varphi_n)_{n\in\mathbb{N}}$ a Cauchy sequence in U, which converges to a $\varphi \in X$, since X is closed. Since φ is the limit of $(\varphi_n)_{n\in\mathbb{N}}$ and U is closed, it follows that $\varphi \in U$. ■

Definition 4.72. Let (X, d) be a metric space. Then $U \subset X$ is called *dense* in X if $\overline{U} = X$, i.e. to every $\varphi \in X$ there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U that converges to φ wrt d. Alternatively, we can say that to every $\varphi \in X$ and $\epsilon > 0$ there is a $\psi \in U$ s.t. $d(\varphi, \psi) < \epsilon$ (set e.g. $\psi := \varphi_N$, where N is s.t. for all $n \geq N$, $d(\varphi_n, \varphi) < \epsilon$).

Example 4.73. $\mathbb{Q}$ lies dense in $\mathbb{R}$ wrt $d(x, y) := |x - y|$ for all x, $y \in \mathbb{R}$.

Theorem 4.74. If $\varphi \in \mathbb{R}\backslash\mathbb{Q}$ and $\psi \in \mathbb{Q}$, then $\varphi + \psi$ is irrational. This is because any $\psi \in \mathbb{Q}$ can be written as p/q, where p, $q \in \mathbb{Z}$. Thus: If $\varphi + \psi$ was rational, we would have

$$\varphi + \psi = \varphi + \frac{p}{q} = \frac{r}{s} \Rightarrow \varphi = \frac{r}{s} - \frac{p}{q} = \frac{rq - ps}{qs} \in \mathbb{Q}, \tag{4.73}$$

which is a contradiction to the assumption $\varphi \in \mathbb{R}\backslash\mathbb{Q}$. This proves that the sum of a rational and irrational number is always irrational.

Example 4.75 (Dense subsets)**.** $\mathbb{R}\backslash\mathbb{Q}$ lies dense in $\mathbb{R}$ wrt $d(x, y) := |x - y|$ for all x, $y \in \mathbb{R}$.

Proof. We need to show that to every $\varphi \in \mathbb{R}$ there exists a sequence $(\varphi_n)_{n\in\mathbb{N}} \in \mathbb{R}\backslash\mathbb{Q}$ s.t. $\varphi_n \overset{n\to\infty}{\longrightarrow} \varphi$ wrt d. We already know that $\mathbb{Q}$ lies dense in $\mathbb{R}$ wrt d, i.e. there exists a sequence $(\psi_n)_{n\in\mathbb{N}}$ s.t. $\psi_n \overset{n\to\infty}{\longrightarrow} \varphi$. Now define the sequence $(\varphi_n)_{n\in\mathbb{N}} := \left(\psi_n + \frac{a}{n}\right)_{n\in\mathbb{N}}$, where $a \in \mathbb{R}\backslash\mathbb{Q}$. Each sequence element lies in $\mathbb{R}\backslash\mathbb{Q}$, since the sum of a rational and an irrational number is irrational, cf. Remark 4.74, and converges to φ wrt d. ■

Example 4.76 (Weierstrass Approximation Theorem)**.** The algebraic polynomials $\mathcal{P}$ are dense on an interval $[a, b] \subset \mathbb{R}$ with $-\infty < a < b < \infty$ in $\mathcal{C}[a, b]$ wrt the metric d_∞ from Example 4.65. [34, Corollary 6.12]

Proof. We deletegate the proof of this celebrated Theorem to appendix A. ■

Example 4.77. Let (X, d) be a metric space. Then $U \subset X$ lies dense in $\overline{U} \subset X$ wrt d.

Theorem 4.78. Let (X, d) be a metric space. $U \subset X$ dense iff every non-empty open set $O \subset X$ intersects U, i.e. $U \cap O \neq \emptyset$.

Proof. "$\Longrightarrow$" Use Theorem 4.35 and note that $U \subset X$ is dense, i.e. $\overline{U} = X$.

"$\Longleftarrow$" We want to prove $U \subset X$ dense $\Rightarrow$ $U \cap O \neq \emptyset$ for all non-empty open sets $O \subset X$. We will prove this via its contraposition, i.e. if there exists a non-empty open set $O \subset X$ with $U \cap O = \emptyset$, then $U \subset X$ is not complete.

Note that according to Theorem 4.35, if $\varphi \in \overline{U}$ for $U \subset X$, then for every open set $O \subset X$ containing φ, $U \cap O \neq \emptyset$. The contraposition of this statement is what we want to prove. ■

Remark 4.79. Let d and d' be two metrics on X. If $U \subset X$ is dense wrt d, it is not necessarily dense wrt d'.

Example 4.80. Consider $X = \mathbb{R}$ and let $d(x, y) := |x - y|$ for all x, $y \in \mathbb{R}$. While $U = \mathbb{Q} \subset X = \mathbb{R}$ is dense wrt d, it is not dense wrt the discrete metric from Remark 4.17, since for $r \leq 1$, the open balls are given by the singletons $\{\varphi\}$ for any $\varphi \in X$, cf. Example 4.19. Let $\varphi = \pi$, then $B_r(\pi) = \{\pi\}$ for any $r \leq 1$, but clearly, $\mathbb{Q} \cap \{\pi\} = \emptyset$, and thus by Theorem 4.78 $\mathbb{Q}$ is not dense in $\mathbb{R}$ wrt the discrete metric.

Remark 4.81. Let (X, d) be a metric space. Then there are examples of dense subsets $U \subset X$ wrt d that are not complete wrt d, and vice versa.

Example 4.82. Consider the cantor set $\mathcal{C}$ on $[0,1]$, which we consider as a subset of the metric space $([0,1],d)$ with $d(x,y) := |x-y|$ for all $x,\, y \in [0,1]$. Formally, it is defined as follows:

$$C_0 := [0,1], \tag{4.74}$$

$$C_n := \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right) = \frac{1}{3}\left(C_{n-1} \cup (2 + C_{n-1})\right), \tag{4.75}$$

$$\mathcal{C} := \bigcap_{n=0}^{\infty} C_n. \tag{4.76}$$

Intuitively, the Cantor set $\mathcal{C}$ is created by iteratively removing the open middle third of each line segment.

Since for all $n \in \mathbb{N}$, C_n is closed (as the finite union of closed sets) – for a formal proof, use induction – and the arbitrary intersection of closed sets is closed, we have that $\mathcal{C} \subset [0,1]$ is closed wrt d. Thus, by Theorem 4.71 b), $\mathcal{C}$ is complete wrt d. Now consider the open set $O := \left(\frac{1}{2}, \frac{2}{3}\right) \subset [0,1]$, which is open wrt d. Clearly, $O \cap \mathcal{C} = \emptyset$, and thus by Theorem 4.78 $\mathcal{C}$ is not dense wrt d.

Example 4.83. $\mathbb{Q}$ is a dense subset of $X = \mathbb{R}$, where (X,d) with $d(x,y) := |x-y|$ for all $x,\, y \in \mathbb{R}$. However, we have already established that $\mathbb{Q}$ is not complete wrt d, e.g. cf. Remark 4.52.

Definition 4.84 (Completion of metric spaces)**.** Let (X,d) be a metric space. Then a *completion* of X is a triplet $\left(\tilde{X}, \tilde{d}, f\right)$, where $\left(\tilde{X}, \tilde{d}\right)$ is a complete metric space and $f : X \to \tilde{X}$ is an isometry with dense image, i.e. $f(X) \subset \tilde{X}$ is dense wrt $\tilde{d}$ [38, Def. 2.2].

Remark 4.85. Let (X,d) be a metric space. If a completion $\left(\tilde{X}, \tilde{d}, f\right)$ exists, then X is isometric to the dense subset $f(X)$.

Proposition 4.86. To every metric space (X,d), a completion $\left(\tilde{X}, \tilde{d}, f\right)$ exists and up to isometry, it is unique.

Remark 4.87. We can henceforth talk about *the* completion of a metric space.

Example 4.88. Consider the metric space $(\mathbb{Q}, d)$ with $d(r_1, r_2) := |r_1 - r_2|$ for any $r_1, r_2 \in \mathbb{Q}$. In Appendix B, we will construct the completion of $\mathbb{Q}$ (the real numbers $\mathbb{R}$) with the extended metric $d_{\text{ext}}(x,y) := |x-y|$ for any $x,\, y \in \mathbb{R}$ with an isometry $f : \mathbb{Q} \to \mathbb{R}$.

Remark 4.89. The completion of $\mathbb{R}$ wrt $|\cdot|$ is needed for the proof of Proposition 4.86.

Proof of Proposition 4.86. Let (X,d) be a metric space. We will proceed in several steps.

Step 1: We will first construct $\tilde{X}$. Let $(\varphi_n)_{n\in\mathbb{N}}$ and $(\psi_n)_{n\in\mathbb{N}}$ be Cauchy sequences in X. Then by Theorem 4.16,

$$|d(\varphi_n, \psi_n) - d(\varphi_m, \psi_m)| \leq d(\varphi_n, \varphi_m) + d(\psi_n, \psi_m) \overset{n,m\to\infty}{\longrightarrow} 0, \tag{4.77}$$

which implies that $(d(\varphi_n, \psi_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$ with the usual metric. By the completeness $\mathbb{R}$, every Cauchy sequence in $\mathbb{R}$ converges, and thus the limit $\lim_{n\to\infty} d(\varphi, \psi_n)$ exists.

We will now construct an equivalence relation on the set of Cauchy sequences in X as follows: Let $(\varphi_n)_{n\in\mathbb{N}}$ and $(\psi_n)_{n\in\mathbb{N}}$ be Cauchy sequences in X, then we call them equivalent if $d(\varphi_n, \psi_n) \overset{n\to\infty}{\longrightarrow} 0$. By Def. 6.2, we need to show reflexivity (which holds since $d(\varphi_n, \varphi_n) = 0$), symmetry (which holds since $d(\varphi_n, \psi_n) = d(\psi_n, \varphi_n)$) and transitivity (which holds because $d(\varphi_n, \chi_n) \leq d(\varphi_n, \psi_n) + d(\psi_n, \chi_n)$ for any Cauchy sequence $(\chi_n)_{n\in\mathbb{N}}$ in X).

We will denote the set of all equivalent Cauchy sequences in X by $\tilde{X}$,

$$\tilde{X} = \{(\varphi_n)_{n\in\mathbb{N}} \mid (\varphi_n)_{n\in\mathbb{N}} \text{ is a Cauchy sequence in } X\}/\sim . \tag{4.78}$$

Step 2: We will now define a metric $\tilde{d}$ on $\tilde{X}$. Let $\tilde{\varphi}, \tilde{\psi} \in \tilde{X}$ with representations $(\varphi_n)_{n\in\mathbb{N}}$ and $(\psi_n)_{n\in\mathbb{N}}$. We set

$$\tilde{d}\left(\tilde{\varphi}, \tilde{\psi}\right) := \lim_{n\to\infty} d(\varphi_n, \psi_n). \tag{4.79}$$

The map $\tilde{d} : \tilde{X} \times \tilde{X} \to \mathbb{R}$ is well-defined, since for other representations $(\varphi'_n)_{n\in\mathbb{N}} \sim (\varphi_n)_{n\in\mathbb{N}}$ and $(\psi'_n)_{n\in\mathbb{N}} \sim (\psi_n)_{n\in\mathbb{N}}$, we have by Theorem 4.16 the inequality

$$|d(\varphi'_n, \psi'_n) - d(\varphi_n, \psi_n)| \leq d(\varphi'_n, \varphi_n) + d(\psi'_n, \psi_n) \overset{n\to\infty}{\longrightarrow} 0. \tag{4.80}$$

And we have already established in Step 1 that the Cauchy sequence $(d(\varphi_n, \psi_n))_{n\in\mathbb{N}} \in \mathbb{R}$ is convergent wrt the usual metric on $\mathbb{R}$.

To show that $\tilde{d}$ is indeed a metric on $\tilde{X}$, note that for any $\tilde{\varphi}, \tilde{\psi}, \tilde{\chi} \in \tilde{X}$ (with representations $(\varphi_n)_{n\in\mathbb{N}}$, $(\psi_n)_{n\in\mathbb{N}}$ and $(\chi_n)_{n\in\mathbb{N}}$), we have:

1. Positivity: $\tilde{d}\left(\tilde{\varphi}, \tilde{\varphi}\right) = \lim_{n\to\infty} d(\varphi_n, \varphi_n) = 0$.
2. Symmetry: $\tilde{d}\left(\tilde{\varphi}, \tilde{\psi}\right) = \lim_{n\to\infty} d(\varphi_n, \psi_n) = \lim_{n\to\infty} d(\psi_n, \varphi_n) = \tilde{d}\left(\tilde{\psi}, \tilde{\varphi}\right)$.
3. Definiteness: $\tilde{d}\left(\tilde{\varphi}, \tilde{\psi}\right) = \lim_{n\to\infty} d(\varphi_n, \psi_n) = 0$ iff $(\varphi_n)_{n\in\mathbb{N}} \sim (\psi_n)_{n\in\mathbb{N}}$, i.e. $\tilde{\varphi} = \tilde{\psi}$, cf. Theorem 6.4.
4. Triangle inequality:

$$\tilde{d}\left(\tilde{\varphi}, \tilde{\psi}\right) = \lim_{n\to\infty} d(\varphi_n, \psi_n) \leq \lim_{n\to\infty} \{d(\varphi_n, \chi_n) + d(\chi_n, \psi_n)\}$$

Since $(d(\varphi_n, \chi_n))_{n\in\mathbb{N}}$ and $(d(\chi_n, \psi_n))_{n\in\mathbb{N}}$ are Cauchy sequences in $\mathbb{R}$, also cf. Step 1, we have

$$\begin{aligned}\tilde{d}\left(\tilde{\varphi}, \tilde{\psi}\right) &\leq \lim_{n\to\infty} \{d(\varphi_n, \chi_n) + d(\chi_n, \psi_n)\} = \lim_{n\to\infty} d(\varphi_n, \chi_n) + \lim_{n\to\infty} d(\chi_n, \psi_n) \\ &= \tilde{d}\left(\tilde{\varphi}, \tilde{\chi}\right) + \tilde{d}\left(\tilde{\chi}, \tilde{\psi}\right),\end{aligned}$$

which shows the triangle inequality.

Step 3: We now embed X isometrically as a dense subset of $\tilde{X}$. For this, consider the map $f : X \to \tilde{X}, \varphi \mapsto \left[(\varphi_n)_{n\in\mathbb{N}}\right]$ with $\varphi_n = \varphi$ for any $n \in \mathbb{N}$. f is an isometry, since

$$\tilde{d}\left(f(\varphi), f(\psi)\right) = \lim_{n\to\infty} d\left(\varphi_n, \psi_n\right) = \lim_{n\to\infty} d(\varphi, \psi) = d(\varphi, \psi). \tag{4.81}$$

The image of f under X, i.e. $f(X) \subset \tilde{X}$, is dense in $\tilde{X}$ wrt $\tilde{d}$, since for any $\tilde{\varphi} \in \tilde{X}$ with representation $(\varphi_n)_{n\in\mathbb{N}}$ the sequence $(f(\varphi_n))_{n\in\mathbb{N}} \in f(X)$ converges to $\tilde{\varphi}$ wrt $\tilde{d}$:

$$\lim_{n\to\infty} \tilde{d}\left(\tilde{\varphi}, f(\varphi_n)\right) = \lim_{n\to\infty}\lim_{m\to\infty} d(\varphi_m, \varphi_n) = 0, \tag{4.82}$$

since $(\varphi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X wrt d.

Step 4: We will now show that $\left(\tilde{X}, \tilde{d}\right)$ is a complete metric space.

Let $(\tilde{\varphi}_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\tilde{X}$. Since $f(X) \subset \tilde{X}$ is dense wrt $\tilde{d}$, to every $\tilde{\varphi}_n$ there is a $\varphi_n \in X$ s.t. $\tilde{d}\left(\tilde{\varphi}_n, f(\varphi_n)\right) < 1/n$. Hence, by applying the triangle equality twice, we have

$$\begin{aligned} d(\varphi_n, \varphi_m) = \tilde{d}\left(f(\varphi_n), f(\varphi_m)\right) &\le \tilde{d}\left(f(\varphi_n), \tilde{\varphi}_n\right) + \tilde{d}\left(\tilde{\varphi}_n, \tilde{\varphi}_m\right) + \tilde{d}\left(\tilde{\varphi}_m, f(\varphi_m)\right) &\text{(4.83)}\\ &< \frac{1}{n} + \tilde{d}\left(\tilde{\varphi}_n, \tilde{\varphi}_m\right) + \frac{1}{m} \overset{n,m\to\infty}{\longrightarrow} 0, &\text{(4.84)} \end{aligned}$$

since by assumption, $(\tilde{\varphi}_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Hence, $(\varphi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X, and hence a representation for some $\tilde{\varphi} \in \tilde{X}$. We will now show that $(\tilde{\varphi}_n)_{n\in\mathbb{N}}$ converges to $\tilde{\varphi}$:

$$\begin{aligned} \tilde{d}\left(\tilde{\varphi}_n, \tilde{\varphi}\right) &\le \tilde{d}\left(\tilde{\varphi}, f(\varphi_n)\right) + \tilde{d}\left(f(\varphi_n), \tilde{\varphi}_n\right) < \tilde{d}\left(\tilde{\varphi}, f(\varphi_n)\right) + \frac{1}{n} &\text{(4.85)}\\ &= \lim_{m\to\infty} d\left(f(\varphi_n)_m, \varphi_m\right) + \frac{1}{n} = \lim_{m\to\infty} d\left(\varphi_n, \varphi_m\right) + \frac{1}{n} &\text{(4.86)}\\ \Rightarrow \lim_{n\to\infty} \tilde{d}\left(\tilde{\varphi}_n, \tilde{\varphi}\right) &\le \lim_{n,m\to\infty} d(\varphi_n, \varphi_m) = 0. &\text{(4.87)} \end{aligned}$$

Hence, $\left(\tilde{X}, \tilde{d}\right)$ is complete.

Step 5: We will now show $\tilde{X}$, a completion of X, is unique up to isometry.

Let (Y, d_Y, f) and (Z, d_Z, g) be completions of the metric space (X, d), i.e. (Y, d_Y) and (Z, d_Z) are complete metric spaces, and $f : X \to Y$ and $g : X \to Z$ are isometries with dense subsets $f(X) \subset Y$ wrt d_Y and $g(X) \subset Z$ wrt d_Z.

We will now show that (Y, d_Y) and (Z, d_Z) are isometric. For this, let $\psi \in Y$. Since $f(X) \subset Y$ is dense wrt d_Y, there is a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in X s.t.

$$\lim_{n\to\infty} d_Y(\psi, f(\varphi_n)) = 0.$$

Now define $h : Y \to Z$ via

$$h(\psi) := \lim_{n\to\infty} g(\varphi_n). \tag{4.88}$$

Note the following properties of h:

- The limit in Eq. (4.88) exists, since

  $$\begin{aligned} d_Z(g(\varphi_n), g(\varphi_m)) &= d_X(\varphi_n, \varphi_m) = d_Y(f(\varphi_n), f(\varphi_m)) &\text{(4.89)}\\ \Rightarrow \lim_{n,m\to\infty} d_Z(g(\varphi_n), g(\varphi_m)) &= \lim_{n,m\to\infty} d_Y(f(\varphi_n), f(\varphi_m)) = d_Y(\psi, \psi) = 0, &\text{(4.90)} \end{aligned}$$

  since by Theorem 4.16 we have that

  $$|d_Y(f(\varphi_n), f(\varphi_m)) - d_Y(\psi, \psi)| \le d_Y(f(\varphi_n), \psi) + d_Y(f(\varphi_m), f(\psi)).$$

  Eq. (4.90) shows that $(g(\varphi_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in Z, which also converges in Z, since by assumption, (Z, d_Z) is complete.

- Let $(\varphi'_n)_{n\in\mathbb{N}}$ be another sequence in X converging to ψ wrt d_Y. Then

$$\begin{aligned} d_Z(g(\varphi_n), g(\varphi'_n)) &= d_X(\varphi_n, \varphi'_n) = d_Y(f(\varphi_n), f(\varphi'_n)) \overset{n\to\infty}{\longrightarrow} d_Y(\psi,\psi) = 0 \\ &\Rightarrow \lim_{n\to\infty} g(\varphi_n) = \lim_{n\to\infty} g(\varphi'_n), \end{aligned}$$

  which makes the limit in Eq. (4.88) well-defined.

- The map h is an isometry, since from the triangle inequality and the inequality in Eq. (4.16), we have for arbitrary ψ, $\psi' \in Y$

$$\begin{aligned} &|d_Z(h(\psi), h(\psi')) - d_Y(\psi,\psi')| \\ =&\left| d_Z(h(\psi), h(\psi')) + \underbrace{d_Y(f(\varphi_n), f(\varphi'_n)) - d_Y(f(\varphi_n), f(\varphi'_n))}_{=0} - d_Y(\psi,\psi') \right| \\ =&\left| d_Z(h(\psi), h(\psi')) + d_Y(f(\varphi_n), f(\varphi'_n)) - \underbrace{d_Y(f(\varphi_n), f(\varphi'_n))}_{=d_X(\varphi_n,\varphi'_n) = d_Z(g(\varphi_n), g(\varphi'_n))} - d_Y(\psi,\psi') \right| \\ \le& |d_Z(h(\psi), h(\psi')) - d_Z(g(\varphi_n), g(\varphi'_n))| + |d_Y(f(\varphi_n), f(\varphi'_n)) - d_Y(\psi,\psi')| \\ \le& d_Z(h(\psi), g(\varphi_n)) + d_Z(h(\psi'), g(\varphi'_n)) + d_Y(f(\varphi_n),\psi) + d_Y(f(\varphi'_n),\psi') \\ &\overset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

  since we choose the sequences $(\varphi_n)_{n\in\mathbb{N}}$ and $(\varphi'_n)_{n\in\mathbb{N}}$ in X such that $f(\varphi_n) \overset{n\to\infty}{\longrightarrow} \psi$ and $f(\varphi'_n) \overset{n\to\infty}{\longrightarrow} \psi'$ (wrt d_Y each). By definition, cf. Eq. (4.88), we then have $h(\psi) = \lim_{n\to\infty} g(\varphi_n)$ and $h(\psi') = \lim_{n\to\infty} g(\varphi'_n)$.

  The above calculation shows that $d_Z(h(\psi), h(\psi')) = d_Y(\psi,\psi')$ for arbitrary $\psi, \psi' \in Y$.

- Finally, h is surjective, since for every $\chi \in Z$ there is a sequence $(\varphi_n)_{n\in\mathbb{N}} \in X$ s.t. $\lim_{n\to\infty} g(\varphi_n) = \chi$, since $g(X)$ is dense in Z wrt d_Z. Now consider the limit $\psi := \lim_{n\to\infty} f(\varphi_n) \in Y$ and note that $\chi = h(\psi) = \lim_{n\to\infty} g(\varphi_n)$.

  Note that the existence of the limit $\psi = \lim_{n\to\infty} f(\varphi_n) \in Y$ follows from

$$d_Y(f(\varphi_n), f(\varphi_m)) = d_X(\varphi_n, \varphi_m) = d_Z(g(\varphi_n), g(\varphi_m)) \overset{n\to\infty}{\longrightarrow} 0, \tag{4.91}$$

  i.e. $(f(\varphi_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in Y, and hence by the completeness of (Y, d_Y), the sequence also converges in Y.

  This proof of the existence of the limit $\psi = \lim_{n\to\infty} f(\varphi_n)$ is very similar to how we proved the existence of the limit in Eq. (4.88) for the definition of h.

■

4.2 Compactness

Definition 4.90 (Compact)**.** Let U be a subset of a metric space. Then it is called

- *compact* if every open covering of U contains a finite sub-covering, i.e. if to every family $\{V_j\}_{j\in J}$ of open sets with $U \subset \bigcup_{j\in\mathbb{N}} V_j$ there are finitely many V_{j_1}, ..., V_{j_n} s.t.

$$U \subset \bigcup_{k=1}^{n} V_{j_k},$$

- *sequentially compact* if every sequence in U contains a convergent subsequence in U.

Example 4.91. Consider $U := (0, 1]$, which is a subset of the metric space $(\mathbb{R}, |\cdot|)$. Then $V_j := (1/j, 1 + 1/j)$, $j \in \mathbb{N}$, are an open covering of U, but U is not compact, since there is no *finite* covering of U with open sets from from $\{V_j\}_{j\in\mathbb{N}}$.

U is also not sequentially compact, since the sequence $(1/n)_{n\in\mathbb{N}}$ does not contain a convergent subsequence in U $(0 \notin U)$.

Example 4.92. $(\mathbb{R}, |\cdot|)$ is obviously not compact. Also, it is not sequentially compact, since $(n)_{n\in\mathbb{N}}$ does not contain a convergent subsequence in $\mathbb{R}$.

Definition 4.93 (Totally bounded set)**.** A subset U of a metric space is called *totally bounded* if for every $\epsilon > 0$ there are finitely many elements $\varphi_1, \dots, \varphi_n \in U$ s.t.

$$U \subset \bigcup_{i=1}^{n} B_\epsilon(\varphi_i),$$

i.e. to every $\varphi \in U$ there is a $\varphi_k \in U$, $1 \leq k \leq n$, s.t. $d(\varphi, \varphi_k) < \epsilon$.

Theorem 4.94. Every totally bounded set is bounded.

Proof. Let $\varphi_1, \dots, \varphi_n \in U$ be s.t.

$$U \subset \bigcup_{i=1}^{n} B_\epsilon(\varphi_i)$$

for all $\epsilon > 0$. Now note that

$$U \subset \bigcup_{i=1}^{n} B_\epsilon(\varphi_i) \subset \bigcup_{i=1}^{n} B_\epsilon[\varphi_i],$$

and that by Theorem 4.25, the finite union of closed sets is also closed. Thus, U is bounded. ∎

Remark 4.95. The converse of Theorem 4.94 is not true in general: There are bounded sets that are not totally bounded.

Example 4.96. Consider $\mathbb{R}$ with the metric $d(x, y) := \min\{1, |x - y|\}$. Then $\mathbb{R}$ (as a subset of $\mathbb{R}$) is bounded, but not totally bounded.

Proof. $\mathbb{R}$ is bounded wrt d, since $\mathbb{R} \subset B_r[0]$, where $r \geq 1$. However, $\mathbb{R}$ is not totally bounded. For this, let $\epsilon \leq 1$ be arbitrary, then it is clear that we need infinitely many open balls to cover $\mathbb{Z}$, and thus also $\mathbb{R}$. ∎

Theorem 4.97. Every sequentially compact set is totally bounded.

Proof. We will proof this by contradiction. So let U be sequentially compact, but not totally bounded. The fact that U is totally bounded can be written with quantifiers:

$$\forall \epsilon > 0 \ \exists \psi_1, \dots, \psi_n \in U : U \subset \bigcup_{i=1}^{n} B_\epsilon(\psi_i)$$

So if U is not totally bounded, we can write this as

$$\exists \epsilon > 0 \ \forall \psi_1, \dots, \psi_n \in U : \exists \psi \in U, \psi \notin \bigcup_{i=1}^{n} B_\epsilon(\psi_i)$$

So there is an $\epsilon > 0$ s.t. for arbitrary $\psi_1, \dots, \psi_n \in U$, $n \in \mathbb{N}$, there is a $\psi_{n+1} \in U$ s.t. $d(\psi_{n+1}, \psi_j) \geq \epsilon$ for all $j \in \{1, \dots, n\}$. Thus, we can also find a $\psi_{n+2} \in U$ s.t. $d(\psi_{n+2}, \psi_j) \geq \epsilon$ for all $j \in \{1, \dots, n+1\}$. Hence, we can recursively construct a sequence $(\psi_n)_{n\in\mathbb{N}}$ in U s.t.

$$d(\psi_l, \psi_k) \geq \epsilon \quad \text{for } n < k < l.$$

However, such a sequence cannot contain a convergent subsequence, which is in contradiction to U being sequentially compact. ■

Definition 4.98. A set U of a metric space is called separable, if U contains a countable dense subset.

Example 4.99. Consider the metric space $(\mathbb{R}, d)$, where d is the usual metric. Then $\mathbb{Q} \subset \mathbb{R}$ is separable wrt d.

Theorem 4.100. Every totally bounded set is separable.

Proof. Let U be totally bounded. There for every $\epsilon_m = 1/m > 0$ there is a finite sequence of elements $\varphi_{1,m}, \dots, \varphi_{n_m,m} \in U$ s.t.

$$U \subset \bigcup_{j=1}^{n_m} B_{\epsilon_m}(\varphi_{j,m}).$$

Consider the set

$$V := \{\varphi_{1,1}, \dots, \varphi_{n_1,1}, \varphi_{1,2}, \dots, \varphi_{n_2,2}, \varphi_{1,3}, \dots, \varphi_{n_3,3}, \dots\dots\},$$

which is the set of elements of the sequence $(\varphi_{1,m}, \dots, \varphi_{n_m,m})_{m\in\mathbb{N}}$. Obviously, V is countable. Also, for every $\varphi \in U$ and $\epsilon > 0$, there is an $\epsilon_m = 1/m < \epsilon$ and a $\varphi_{j,m} \in V$ s.t. $d(\varphi_{j,m}, \varphi) < \epsilon_m = 1/m < \epsilon$. Thus, V is dense in U. ■

Theorem 4.101. Let U be a subset of a metric space, then the following statements are equivalent:

(a) U is compact,

(b) U is sequentially compact,

(c) U is totally bounded and complete.

Proof. "(a) $\implies$ (b)": Proof by contradiction: Let U be compact, but not sequentially compact. Then there is a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U that does not have a convergent subsequence. Thus, to every $\varphi \in U$ there exists an open ball $B_\varphi = B_{r_\varphi}(\varphi)$, where B_φ only contains finitely many elements of $(\varphi_n)_{n\in\mathbb{N}}$.

However, since U is compact, there are finitely many elements $\psi_1, \dots, \psi_n \in U$ s.t.

$$U \subset \bigcup_{k=1}^{n} B_{\psi_k},$$

where B_{ψ_k} only contains finitely many points of $(\varphi_n)_{n\in\mathbb{N}}$. But then, U contains only finitely many elements of $(\varphi_n)_{n\in\mathbb{N}}$, which is a contradiction.

"(b) $\implies$ (c)": Let U be sequentially compact and thus totally bounded by Theorem (4.97). It now remains to be shown that U is complete. Let $(\varphi_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in U, then it contains a convergent subsequence in U, since U is sequentially compact, and thus by Theorem 4.51, $(\varphi_n)_{n\in\mathbb{N}}$ converges.

"(c) $\implies$ (a)": Let U be complete and totally bounded. We will show that U is compact via contradiction.

Let $\{V_j\}_{j\in J}$ be an open cover of U s.t.

$$U \subset \bigcup_{j\in J} V_j. \tag{4.92}$$

We now define

$$\mathcal{B} := \left\{ B \subset U \mid B \subset \bigcup_{i\in I} V_i \Rightarrow I \text{ infinite, where } I \subset J \right\}$$

and we will show that $U \notin \mathcal{B}$.

Since U is totally bounded, for all $\epsilon > 0$ there is an open cover

$$U \subset \bigcup_{i=1}^{n_\epsilon} B_\epsilon(\varphi_i) \tag{4.93}$$

with points $\varphi_1, \dots, \varphi_n \in U$. Further, to *each* $B \in \mathcal{B}$, there is an $i(\epsilon)$ with $B_\epsilon(\varphi_i) \cap B \in \mathcal{B}$. This holds because according to Eq. (4.93),

$$B \cap U = B \subset \left(\bigcup_{i=1}^{n_\epsilon} B_\epsilon(\varphi_i) \right) \cap B = \bigcup_{i=1}^{n_\epsilon} \left(B_\epsilon(\varphi_i) \cap B \right),$$

where the last equality follows from the distributive law for sets. Since $B_\epsilon(\varphi_i) \cap B \subset B$, we have equality, i.e.

$$B = \bigcup_{i=1}^{n_\epsilon} \left(B_\epsilon(\varphi_i) \cap B \right).$$

Since we assumed $B \in \mathcal{B}$, this implies that there is an $i(\epsilon)$ s.t. $B_\epsilon(\varphi_i) \cap B \in \mathcal{B}$, since otherwise, B could be covered by a finite collection of $\{V_i\}_{i\in I}$.

If $U \in \mathcal{B}$, define $B_1 := U \in \mathcal{B}$, and then recursively construct (for $\epsilon_k = 1/k$)

$$B_k := B_{\epsilon_k}(\varphi_k) \cap B_{k-1} \in \mathcal{B} \quad \text{for } k \geq 2. \tag{4.94}$$

By construction, we have

$$B_k \subset B_{k-1} \subset \cdots \subset B_2 \subset B_1 = U.$$

Now, for any n, choose $\psi_n \in B_n$, then for all $l, k \in \mathbb{N}$ we have (WLOG, let $l \geq k$),

$$d(\psi_k, \psi_l) \leq d(\psi_k, \varphi_k) + d(\varphi_k, \psi_l) < 2\epsilon_k, \tag{4.95}$$

and hence $(\psi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in U. Since U is complete, $(\psi_n)_{n\in\mathbb{N}}$ converges to a $\psi \in U$ wrt d. Since $\psi \in V_{j_0}$ for a $j_0 \in J$, cf. Eq. (4.92), for sufficiently large k, we have

$$B_k \subset B_{\epsilon_k}(\varphi_k) \subset B_{2\epsilon_k}(\psi_k) \subset B_{2\epsilon_k + d(\psi_k,\psi)}(\psi) \subset V_{j_0}.$$

For the second inclusion, note that for any $\varphi \in B_{\epsilon_k}(\varphi_k)$,

$$d(\varphi, \psi_k) \leq d(\psi_k, \varphi_k) + d(\varphi_k, \varphi) < 2\epsilon_k.$$

For the third inclusion, note that for any $\chi \in B_{2\epsilon_k}(\psi_k)$,

$$d(\psi, \chi) \leq d(\psi, \psi_k) + d(\psi_k, \chi) < d(\psi, \psi_k) + 2\epsilon_k.$$

For the final inclusion, note that V_{j_0} is open, i.e. for $\psi \in U$ there is an $r > 0$ s.t. $B_r(\psi) \subset V_{j_0}$. Since $2\epsilon_k + d(\psi_k, \psi) = 2/k + d(\psi_k, \psi)$ becomes arbitrarily small, for sufficiently large k, the last inclusion holds.

However, $B_k \in V_{j_0}$ is a contradiction to $B_k \in \mathcal{B}$, cf. the construction of the B_k in Eq. (4.94). ■

Corollary 4.102. Compact sets in metric spaces are closed and bounded.

Proof. Let (X, d) be a metric space, and let $U \subset X$ be compact.

Let $\varphi \in \overline{U}$, i.e. φ is an accumulation point of U, and thus there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U converging to φ wrt d. Thus, by Theorem 4.28, every subsequence of $(\varphi_n)_{n\in\mathbb{N}}$ also converges to φ in $\overline{U}$. Since U is sequentially compact, cf. Theorem 4.101, there is a subsequence $(\varphi_{n_k})_{k\in\mathbb{N}}$ of $(\varphi_n)_{n\in\mathbb{N}}$ converging to $\psi \in U \subset \overline{U}$, and since limits of convergent sequences are unique, cf. Theorem 4.27, it follows that $\varphi = \psi$, i.e. $\varphi \in U$, which shows that U is closed.

To see that U is bounded, note that by Theorem 4.101, U is totally bounded. Hence, for any fixed $\epsilon > 0$, there exist points $\varphi_1, \ldots, \varphi_{n_\epsilon} \in U$ with

$$U \subset \bigcup_{i=1}^{n_\epsilon} B_\epsilon(\varphi_1),$$

i.e. for $\varphi \in U$ we have $\varphi \in B_\epsilon(\varphi_j)$ for $1 \leq j \leq n$. Now define

$$r := \epsilon + \max_{1\leq j\leq n} \{d(\varphi_1, \varphi_j)\},$$

then we have

$$d(\varphi, \varphi_1) \leq d(\varphi, \varphi_j) + d(\varphi_j, \varphi_1) < \epsilon + r - \epsilon = r.$$

Thus, $U \subset B_r[\varphi_1]$. ■

Remark 4.103. For topological spaces, this does not hold in general.

Definition 4.104 (Relatively compact)**.** A subset U of a metric space (X, d) is called *relatively compact* in X if its closure $\overline{U} \subset X$ is compact.

Theorem 4.105. $U \subset X$ is sequentially compact iff U is relatively compact and closed.

Proof. "$\Longrightarrow$" Since U is sequentially compact, it is compact, and thus closed. Since it is closed, we have that $U = \overline{U}$, and thus $\overline{U}$ is compact.

"$\Longleftarrow$" Since U is closed, we have that $U = \overline{U}$, and since U is relatively compact in X, $U = \overline{U}$ is compact, and hence also sequentially compact. ∎

Theorem 4.106. A subset $U \subset X$ of a metric space (X, d) is relatively compact iff every sequence in U contains a convergent subsequence in X.

Proof. "$\Longrightarrow$" Let U be relatively compact, i.e. $\overline{U}$ is compact, and thus also sequentially compact. Since every sequence $(\varphi_n)_{n\in\mathbb{N}}$ in U also lies in $\overline{U}$, $U \subset \overline{U}$, $(\varphi_n)_{n\in\mathbb{N}}$ contains a convergent subsequence in $\overline{U} \subset X$.

"$\Longleftarrow$" Let every sequence in U contain a convergent subsequence in X. We now show that $\overline{U}$ is sequentially compact, which implies its compactness.

Let $(\varphi_n)_{n\in\mathbb{N}}$ be a sequence in $\overline{U}$. Since U is dense in $\overline{U}$, for every φ_n, $n \in \mathbb{N}$, there is a $\psi_n \in U$ s.t. $d(\varphi_n, \psi_n) < 1/n$. That way we can construct a sequence $(\psi_n)_{n\in\mathbb{N}}$ in U, and by assumption it contains a convergent subsequence $(\psi_{n_k})_{k\in\mathbb{N}}$ in X; denote the limit by $\psi \in X$. Now note that

$$d(\varphi_{n_k}, \psi) \leq d(\varphi_{n_k}, \psi_{n_k}) + d(\psi_{n_k}, \psi) < \frac{2}{k} \overset{k\to\infty}{\longrightarrow} 0,$$

i.e. every sequence $(\varphi_n)_{n\in\mathbb{N}}$ in $\overline{U}$ contains a convergent subsequence $(\varphi_{n_k})_{k\in\mathbb{N}}$ in X. ∎

Theorem 4.107. Relationship between totally bounded and relatively compact subsets [49, Theorem 4.27]:

(a) Any relatively compact subset of a metric space is totally bounded.

(b) Any totally bounded subset of a complete metric space is relatively compact.

Proof. (a): Let U be relatively compact, i.e. $\overline{U}$ is compact, and thus by Theorem 4.101 also totally bounded. Since $U \subset \overline{U}$, U is also totally bounded.

(b): Let $U \subset X$ be totally bounded, where (X, d) is a complete metric space. This means that for any fixed $\epsilon > 0$, there exists an $n_\epsilon \in \mathbb{N}$ s.t.

$$U \subset \bigcup_{i=1}^{n_\epsilon} B_\epsilon(\varphi_i),$$

where $\varphi_1, \ldots, \varphi_{n_\epsilon} \in U$. Since U lies dense in $\overline{U}$, for any $\psi \in \overline{U}$ there is a $\varphi \in U$ s.t. $d(\varphi, \psi) < \epsilon$. Since $\varphi \in U$, there is a $j \in \{1, \ldots, n_\epsilon\}$ s.t. $\varphi \in B_\epsilon(\varphi_j)$. Then

$$d(\psi, \varphi_j) \leq d(\psi, \varphi) + d(\varphi, \varphi_j) < 2\epsilon,$$

i.e. $\overline{U}$ can be covered as follows:

$$\overline{U} \subset \bigcup_{i=1}^{n_\epsilon} B_{2\epsilon}(\varphi_i),$$

implying that $\overline{U}$ is totally bounded. Since $\overline{U}$ is closed and by Theorem 4.71, closed subsets of complete metric spaces are also complete, by Theorem 4.101, $\overline{U}$ is compact. ∎

4.3 Continuity

Definition 4.108 (ϵ-δ definition of continuity [30])**.** Let (X, d) and (Y, d') be metric spaces. A function $f : X \to Y$ is called *continuous* at $p \in X$ if

$$\forall \epsilon > 0 \ \exists \delta(\epsilon, p) > 0 \ \forall z \in X : d(z, p) < \delta \Rightarrow d'(f(z), f(p)) < \epsilon. \tag{4.96}$$

f is called continuous on X if it is continuous for every $p \in X$.

Remark 4.109. The notion of continuity depends on the metrics. For example [3], let $X = \mathbb{R}$, and let d be the usual metric on $\mathbb{R}$, and d' the discrete metric. Then the function $\mathrm{id} : (\mathbb{R}, d') \to (\mathbb{R}, d), x \mapsto x$ is continuous for all $p \in \mathbb{R}$, since for $\delta = 1$, if we have $d'(z, p) < \delta = 1$ for all $z \in \mathbb{R}$, then $z = p$, and thus $d(z, p) = 0 < \epsilon$ for any $\epsilon > 0$.

Conversely, $\mathrm{id} : (\mathbb{R}, d) \to (\mathbb{R}, d'), x \mapsto x$ is not continuous on $\mathbb{R}$. To see this, let $\epsilon = 1$, consider arbitrary $\delta > 0$ and define $p := \min\{1, \delta\} \in \mathbb{R}$ and $z := p/2 \neq p$. Then $d(z, p) = |p|/2 < \delta$, but $d'(z, p) = 1 \geq \epsilon$.

Definition 4.110 (Uniform continuity [44])**.** Let (X, d) and (Y, d') be metric spaces. A function $f : X \to Y$ is called *uniformly continuous* on X

$$\forall \epsilon > 0 \ \exists \delta(\epsilon) > 0 \ \forall p, z \in X : d(z, p) < \delta \Rightarrow d'(f(z), f(p)) < \epsilon. \tag{4.97}$$

Remark 4.111. While uniform continuity implies continuity, the opposite is in general not true.

Example 4.112. Consider the metric space $(\mathbb{R}_{>0}, d)$, where $d(x, y) := |x - y|$ for any $x, y \in \mathbb{R}_{>0}$. The function $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is continuous on $\mathbb{R}_{>0}$, but not uniformly continuous on $\mathbb{R}_{>0}$.

Proof. To see that f is not uniformly continous on $\mathbb{R}_{>0}$, choose $\epsilon = 1$, then for all $\delta > 0$ take $x := \min\{\delta, 1\} \in \mathbb{R}_{>0}$ and $y := x/2 \in \mathbb{R}_{>0}$, then $|x - y| = \frac{x}{2} < \delta$, but

$$|f(x) - f(y)| = \left| \frac{1}{x} - \frac{1}{y} \right| = \frac{1}{x} \geq 1 = \epsilon.$$

∎

Theorem 4.113. Let (X, d) and (Y, d') be metric spaces. A function $f : X \to Y$ is continuous at $p \in X$ iff for every sequence $(p_n)_{n\in\mathbb{N}}$ in X that converges to p wrt d, the sequence $(f(p_n))_{n\in\mathbb{N}}$ converges to $f(p)$ wrt d'.

Proof. "$\Longrightarrow$" Let f be continuous at $p \in X$, and $(p_n)_{n\in\mathbb{N}}$ a sequence converging to p. Then for all $\epsilon > 0$ there exists a $\delta(\epsilon, p)$ s.t. for all $z \in X$, $d(z, p) < \delta$ implies $d'(f(z), f(p)) < \epsilon$. Since $(p_n)_{n\in\mathbb{N}}$ converges to p wrt d, for the same δ, there exists an $N(\delta)$ s.t. for all $n \geq N$, $d(p_n, p) < \delta$, and thus $d'(f(p_n), f(p)) < \epsilon$.

"$\Longleftarrow$" Let $(p_n)_{n\in\mathbb{N}}$ be a sequence in X converging to $p \in X$ wrt d, and $(f(p_n))_{n\in\mathbb{N}}$ converging to $f(p)$ wrt d'. Assume that f is not continuous at p, i.e. there exists an $\epsilon > 0$ s.t. for all $\delta > 0$ there exists a $z \in X$ s.t. $d(z, p) < \delta$, but $d'(f(z), f(p)) \geq \epsilon$. Fix this $\epsilon > 0$, and consider an arbitrary $\delta > 0$. Since $(p_n)_{n\in\mathbb{N}}$ converges to p, there exists an $N(\delta) \in \mathbb{N}$ s.t. for all $n \geq N$, $d(p_n, p) < \delta$, but $d'(f(p_n), f(p)) \geq \epsilon$. But this is a contradiction to $(f(p_n))_{n\in\mathbb{N}}$ converging to $f(p)$ wrt d'. ∎

Theorem 4.114. Any function $f : [a, b] \to \mathbb{R}$ defined on the compact interval $[a, b] \subset \mathbb{R}$, where on $\mathbb{R}$ we consider the usual metric, that is continuous on $[a, b]$, is uniformly continuous on $[a, b]$.

Proof. [30, p. 285]. ■

Theorem 4.115. Any function $f : [a, b] \to \mathbb{R}$ defined on the compact interval $[a, b] \subset \mathbb{R}$ (where on $\mathbb{R}$ we consider the usual metric) that is continuous is bounded, in the sense that $\{f(x) \mid x \in [a, b]\} \subset \mathbb{R}$ is bounded, cf. Defn. 4.43,

Proof. [30, p. 161]. ■

Lemma 4.116. Consider the map $f : \mathbb{R}^m \to \mathbb{R}^n$. Then f is continuous if all the $f_i : \mathbb{R}^m \to \mathbb{R}$, $i \in \{1, \dots, n\}$ are continuous.

Proof. Since all metrics are strongly equivalent, it does not matter which metric we equip $\mathbb{R}$, $\mathbb{R}^m$ and $\mathbb{R}^n$ with, cf. Remark 3.10; for the following, consider the Euclidean metric. Suppose each f_i is continuous for all $i \in \{1, \dots, n\}$, i.e.:

$$\forall p \in \mathbb{R}^m : \forall \epsilon_i > 0 \ \exists \delta_i > 0 : d_{\mathbb{R}}(f_i(p), f_i(z)) < \frac{\epsilon_i}{\sqrt{n}} \ \forall z \in D \quad \text{with } d_{\mathbb{R}^m}(p, z) < \delta_i. \tag{4.98}$$

Now define $\epsilon := \max\{\epsilon_1, \dots, \epsilon_n\}$. Thus:

$$d_{\mathbb{R}^n}(f(p), f(z)) = \sqrt{\sum_{i=1}^{n} (f_i(p) - f_i(z))^2} = \sqrt{\sum_{i=1}^{n} d_{\mathbb{R}}^2 (f_i(p), f_i(z))} < \sqrt{\sum_{i=1}^{n} \left(\frac{\epsilon}{\sqrt{n}}\right)^2} = \epsilon. \tag{4.99}$$

■

Theorem 4.117. For a map $f : X \to Y$ between two metric spaces (X, d_X) and (Y, d_Y) the following statements are equivalent:

(i) f is continuous,

(ii) preimages $f^{-1}(V) := \{x \in X \mid f(x) \in V\}$ of <u>all</u> open sets $V \subset Y$ are open,

(iii) preimages $f^{-1}(A)$ of <u>all</u> closed sets $A \subset Y$ are closed.

Proof. (i) ⇒ (ii) Assume that f is continuous and that $V \subset Y$ is open and let $a \in f^{-1}(V)$. Since V is an open set, $\exists \epsilon > 0$ s.t. $B_\epsilon(f(a)) \subset V$, cf. Def. 4.21. By assumption, f is continuous at $a \in f^{-1}(V) \subset X$ and therefore $\forall \epsilon > 0 \ \exists \delta > 0$ s.t.

$$d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \epsilon \ \forall x \in X.$$

Put differently, $x \in B_\delta(a)$ implies $f(x) \in B_\epsilon(f(a)) \subset V$. Thus, $B_\delta(a) \subset f^{-1}(V) \Rightarrow f^{-1}(V) \subset X$ is an open set, cf. Def. 4.21.

(ii) ⇒ (i) Assume that $f^{-1}(Y) \subset X$ is open for $V \subset Y$ open and let $a \in X$, $\epsilon > 0$. From Theorem 4.23 we know that $B_\epsilon(f(a)) \subset Y$ is open. Thus, by assumption, $f^{-1}(B_\epsilon(f(a))) = \{x \in X \mid f(x) \in B_\epsilon(f(a))\} = \{x \in X \mid f(x) \in \{y \in Y \mid d_Y(y, f(a)) < \epsilon\}\}$ is open as well. Clearly, $a \in f^{-1}(B_\epsilon(f(a)))$. Therefore, it follows from Def. 4.21 that $\exists \delta > 0$ s.t. $B_\delta(a) \subset f^{-1}(B_\epsilon(f(a)))$. Thus, $\forall x \in B_\delta(a) : d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \epsilon$.

This proves that $f : X \to Y$ is continuous in every point $a \in X$. [24]

(ii) $\Rightarrow$ (iii) Assume that the preimages $f^{-1}(V)$ of all open sets $V \subset Y$ are open. Since $f^{-1}(Y \backslash V) = f^{-1}(Y) \backslash f^{-1}(V)$, as can be easily shown by using the definition of the complement of a set, we have for all open sets $V \subset Y$:

$$f^{-1}(Y \backslash V) = f^{-1}(Y) \backslash f^{-1}(V) = X \backslash f^{-1}(V). \tag{4.100}$$

Since $f^{-1}(V)$ is open by assumption, $f^{-1}(Y \backslash V) = X \backslash f^{-1}(V)$ is closed.

(iii) $\Rightarrow$ (ii) Assume that the preimages $f^{-1}(A)$ of all closed sets $A \subset Y$ are closed. Then:

$$f^{-1}(Y \backslash A) = f^{-1}(Y) \backslash f^{-1}(A) = X \backslash f^{-1}(A). \tag{4.101}$$

Since $f^{-1}(A)$ is closed by assumption, $f^{-1}(Y \backslash A) = X \backslash f^{-1}(A)$ is open. [7] ■

Corollary 4.118. Let (X, τ_1) and (Y, τ_2) be topological spaces coming from metric spaces. Then $f : X \to Y$ is continuous if and only if $f^{-1}(G)$ for every open set $G \in Y$.

Definition 4.119 (Homeomorphism)**.** A *homeomorphism* between two topological spaces X and Y is an invertible function $f : X \to Y$ such that f and f^{-1} are continuous [40, p. 33].

Example 4.120. The Euclidean space $\mathbb{R}^n$, equipped with the usual topology, is homeomorphic to the open ball $B_r(\varphi) = \{x \in \mathbb{R}^n \mid ||x - \varphi|| < r\} \subset \mathbb{R}^n$ (consider the open ball as a metric space with the *induced metric* from the whole space of $\mathbb{R}^n$ and then equip it with the usual topology on a metric space given by the open sets).

Proof: Consider the map

$$f : \mathbb{R}^n \to B_r(\varphi),\ x \mapsto \frac{r \cdot (x - \varphi)}{1 + ||x - \varphi||}.$$

Obviously, f is continuous with inverse

$$f^{-1} : B_r(\varphi) \to \mathbb{R}^n,\ x \mapsto \frac{x}{r - ||x||} + \varphi,$$

which is also continuous. One can easily show that f and f^{-1} are inverses to each other. Thus, f describes a homeomorphism.

Example 4.121 (Stereographic projection)**.** The unit sphere S^n, embedded in $\mathbb{R}^{n+1}$, without the north pole, i.e. $S^n \backslash \{p\} \subset \mathbb{R}^{n+1}$, where $S^n := \{x \in \mathbb{R}^{n+1} \mid ||x||_2 = 1\}$ and $p := \{x \in \mathbb{R}^{n+1} \mid x_i = 0\ \forall i \in [1, n], x_{n+1} = 1\}$, is homeomorphic to $\mathbb{R}^n$ (consider the unit sphere as a metric space with the *induced metric* from the whole space of $\mathbb{R}^{n+1}$ and then equip it with the usual topology on a metric space given by the open sets).

Proof. Consider the map

$$f : S^n \backslash \{p\} \to \mathbb{R}^n,\ x \mapsto \frac{1}{1 - x_{n+1}} \left(x_1, \ldots, x_n\right)^T. \tag{4.102}$$

Obviously, f is continuous. Its inverse is given by

$$f^{-1} : \mathbb{R}^n \to S^n \backslash \{p\}, \ x \mapsto \begin{pmatrix} 2x_1/\big(1 + ||x||_2^2\big) \\ \vdots \\ 2x_n/\big(1 + ||x||_2^2\big) \\ 1 - 2/\big(1 + ||x||_2^2\big) \end{pmatrix}. \tag{4.103}$$

It is trivial to show that f and f^{-1} are inverses to each other. One can also easily show that $||f^{-1}(x)||_2 = 1$; thus, f^{-1} does indeed bring us to the unit ball S^n. To see that f^{-1} is continuous, note that all components are continuous and thus, according to Lemma 4.116, the function itself is continuous. ■

5 Linear Functional Analysis

5.1 Bounded Linear Operators

Let us start by defining the notions of *linearity* and *boundedness* of operators (functionals).

Definition 5.1. Let X, Y be normed linear spaces. Then an operator $A : X \to Y$ is called *linear* if

$$A(\alpha\varphi + \beta\psi) = \alpha A(\varphi) + \beta A(\psi) \quad \forall \varphi, \psi \in X, \forall \alpha, \beta \in \mathbb{K}. \tag{5.1}$$

Theorem 5.2. A linear operator $A : X \to Y$ from a normed linear space $(X, ||\cdot||_X)$ into another normed linear space $(Y, ||\cdot||_Y)$ is continuous on X iff it is continuous at a single point in X.

Proof. "$\Longrightarrow$" By definition.

"$\Longleftarrow$" Let A be continuous at $\varphi_0 \in X$, and let $(\varphi_n)_{n\in\mathbb{N}} \subset X$ converge to $\varphi \in X$ wrt $||\cdot||_X$, then

$$A(\varphi_n) = A(\varphi_n + \varphi_0 - \varphi) - A(\varphi_0 - \varphi) \overset{n\to\infty}{\longrightarrow} A(\varphi) \tag{5.2}$$

wrt $||\cdot||_Y$. ■

Definition 5.3. Let $(X, ||\cdot||_X)$ and $(Y, ||\cdot||_Y)$ be normed linear spaces. Then an operator $A : X \to Y$ is called *bounded* if there is a constant $C(A) > 0$ s.t.

$$||A\varphi||_Y \leq C\, ||\varphi||_X \quad \forall \varphi \in X. \tag{5.3}$$

Any such constant C is called an *upper bound* for A.

Theorem 5.4. A linear operator $A : X \to Y$ between normed spaces $(X, ||\cdot||_X)$ and $(Y, ||\cdot||_Y)$ is bounded iff the operator norm

$$||A|| := \sup_{||\varphi||_X = 1} ||A\varphi||_Y = \sup_{\varphi \in X\setminus\{0\}} \frac{||A\varphi||_Y}{||\varphi||_X} \tag{5.4}$$

on the linear space $\mathcal{L}(X, Y)$ of all bounded linear operators between the two normed spaces $(X, ||\cdot||_X)$ and $(Y, ||\cdot||_Y)$ is finite, i.e.

$$||A|| < \infty. \tag{5.5}$$

The operator norm is the smallest upper bound for A.

Proof. First, let us prove that the operator norm does indeed define a norm on $\mathcal{L}(X, Y)$. Note that the sum of two operators $A, B \in \mathcal{L}(X, Y)$ is defined pointwise, i.e. for $\varphi \in X$: $(A + B)(\varphi) := A\varphi + B\varphi$. Similarly, scalar multiplication is defined for $\alpha \in \mathbb{K}$ as $(\alpha A)(\varphi) := \alpha A\varphi$.

The positivity, definiteness and homogeneity of $||A||$ are trivial to show. For the triangle equality, note that for any $A, B \in \mathcal{L}(X, Y)$, we have

$$||A + B|| = \sup_{||\varphi||_X = 1} ||(A + B)(\varphi)||_Y = \sup_{||\varphi||_X = 1} ||A\varphi + B\varphi||_Y \tag{5.6}$$

$$\leq \sup_{||\varphi||_X = 1} \{||A\varphi||_Y + ||B\varphi||_Y\} \tag{5.7}$$

$$\leq \sup_{||\varphi||_X = 1} \{||A\varphi||_Y\} + \sup_{||\varphi||_X = 1} \{||B\varphi||_Y\} \tag{5.8}$$

$$= ||A|| + ||B||\,. \tag{5.9}$$

Now to the statement A bounded iff $||A|| < \infty$.

"$\Longrightarrow$" Let A be upper bounded with bound $C > 0$, then

$$||A|| = \sup_{||\varphi||_X=1} ||A\varphi||_Y \leq C < \infty. \tag{5.10}$$

This also shows that $||A||$ is the smallest upper bound for A.

"$\Longleftarrow$" We have

$$||A|| \, ||\varphi||_X = \sup_{\varphi \in X\backslash\{0\}} \frac{||A\varphi||_Y}{||\varphi||_X} ||\varphi||_X = \sup_{\varphi \in X\backslash\{0\}} ||A\varphi||_Y \geq ||A\varphi||_Y \,, \tag{5.11}$$

which shows that A is bounded with upper bound $C = ||A|| < \infty$ (note that for $v = 0$, there is nothing to show). ■

Theorem 5.5. A linear operator $A : X \to Y$ for normed linear spaces $(X, ||\cdot||_X)$ and $(Y, ||\cdot||_Y)$ is continuous iff it is bounded.

Proof. "$\Longleftarrow$" Since A is bounded, by definition, we know that for all $\varphi \in X$, there is an upper bound $C > 0$ s.t.

$$||A\varphi||_Y \leq C \cdot ||\varphi||_X \,. \tag{5.12}$$

For an arbitrary $\epsilon > 0$, choose $\delta := \epsilon/C$, then for $||\varphi - \psi||_X < \delta = \epsilon/C$ (where $\psi \in X$ is arbitary) we have because of the linearity of A

$$||A\varphi - A\psi||_Y = ||A(\varphi - \psi)||_Y \leq C \, ||\varphi - \psi||_X < \epsilon, \tag{5.13}$$

which proves the continuity of A, cf. Def. 4.108 [13].

"$\Longrightarrow$" Let A be continuous on X. Then by Theorem 4.117, we know that $A^{-1}\left(B_1^Y(0)\right)$ is open, since $B_1^Y(0)$ is open. By the linearity of A, $0 \in A^{-1}\left(B_1^Y(0)\right)$, since $A(0) = 0 \in B_1^Y(0)$. This means that there is an $r > 0$ s.t. $B_r^V(0) \subset A^{-1}\left(B_1^Y(0)\right)$, i.e. the image of $B_r^V(0)$ is contained in $B_1^Y(0)$.

Now, for any $a > 1$, choose $C := a/r$, and we will show that C is an upper bound for A. For $\varphi \in X\backslash\{0\}$ (for $\varphi = 0$, there is nothing to show), note that

$$\left|\left|\frac{r}{a\,||\varphi||_X}\varphi\right|\right|_X = \frac{r}{a} < r,$$

i.e.

$$\frac{r}{a\,||\varphi||_X}\varphi \in B_r^V(0) \Rightarrow A\left(\frac{r}{a\,||\varphi||_X}\varphi\right) = \frac{r}{a\,||\varphi||_X}A\varphi \in B_1^Y(0) \tag{5.14}$$

$$\Rightarrow \left|\left|\frac{r}{a\,||\varphi||_X}A\varphi\right|\right|_Y < 1 \Rightarrow ||A\varphi||_Y \leq \frac{a}{r}\,||\varphi||_X = C\,||\varphi||_X \,, \tag{5.15}$$

which completes the proof [28, p. 2]. ■

6 Set Theory

Definition 6.1 (Binary Relation [45]). A *binary relation* over a set X is some relation R where for all x, $y \in X$ the statement xRy is either true or false.

Definition 6.2 (Equivalence Relation and Class [22]). An *equivalence relation* on a set X is a binary relation $\sim$ with the following properties $\forall x$, y, $z \in X$:

- **reflexivity**: $x \sim x$,
- **symmetry**: $x \sim y \Leftrightarrow y \sim x$,
- **transitivity**: $(x \sim y) \wedge (y \sim z) \Rightarrow x \sim z$.

Definition 6.3 (Equivalence Class). Let $\sim$ be an equivalence relation on X. Then the *equivalence class* of an element $x \in X$ is defined as

$$[x] := \{y \in X \mid x \sim y\} = \{y \in X \mid y \sim x\} \subset X. \tag{6.1}$$

The set of all equivalences classes in X wrt an equivalence relation $\sim$ is denoted by $X/\sim$. Note that $X/\sim \subset P(X)$, where $P(X)$ denotes the power set of X [27]. The surjective map $f : X \to X/\sim, x \mapsto [x]$ is called the *canoninal surjection*.

Theorem 6.4. Let $\sim$ be an equivalence relation on X. For any $x, y \in X$, it holds that $[x] = [y]$ iff $x \sim y$.

Proof. "$\Longrightarrow$" Since $x \in [x] = [y]$ we have $x \sim y$.

"$\Longleftarrow$" Let $x \sim y$ for arbitrary $x, y \in X$ hold, and consider $x' \in [x]$, i.e. $x' \sim x$. By transitivity, we have $x' \sim y$, i.e. $[x] \subset [y]$. Similarly, we show $[y] \subset [x]$, and hence $[x] = [y]$. ■

Theorem 6.5. Let $\sim$ be an equivalence relation on X. Then for $x, y \in X$, we either have $[x] \cap [y] = \emptyset$ or $[x] = [y]$.

Proof. We know that the statement $x \sim y$ is either true or false. If it is true, then by Theorem 6.4, $[x] = [y]$. If $x \not\sim y$, then no $x' \in [x] = \{z \in X \mid x \sim z\}$ can lie in $[y] = \{z' \in X \mid x \sim z'\}$ (and vice versa), because otherwise, by transitivity, $x \sim y$, which we said is false. Hence, $[x] \cap [y] = \emptyset$. ■

Definition 6.6. Every element of an equivalence class characterizes it, and can be used to *represent* it. Such a chosen element is called a *representative*.

Definition 6.7 (Partially Ordered Set [42]). A *partially ordered set* $(X, \leq)$ is a set X, equipped with a binary relation $\leq$, that satisfies the following properties $\forall x$, y, $z \in X$:

- **reflexivity**: $x \leq x$,
- **antisymmetry**: $(x \leq y) \wedge (y \leq x) \Rightarrow x = y$,
- **transitivity**: $(x \leq y) \wedge (y \leq z) \Rightarrow x \leq z$.

Definition 6.8 (Incomparability). In Definition 6.7, the phrasing "partially ordered" is used to emphasize that there might exist elements x, $y \in X$ s.t. both $x \leq y$ and $y \leq x$ are wrong. These pairs are called *incomparable*. If either $x \leq y$ or $y \leq x$ is true, then we say that the pair is *comparable*.

Example 6.9. Consider $X := \{\{1\}, \{2\}, \{1,2\}\}$ with $\subset$ as partial ordering. Obviously, the elements $\{1\}$ and $\{2\}$ are incomparable.

Definition 6.10 (Chain, Upper Bound, Maximal Element)**.** For preparing the Kuratowski-Zorn lemma, the following definitions come in handy:

a) A *chain* C is a partially ordered set where every pair of elements in C is comparable. One might also say that C is a *totally ordered set*.

b) An *upper bound* (if existent) of a subset $S \subset X$, where X is a partially ordered set, is an element $u \in X$ such that

$$s \leq u \;\forall s \in S. \tag{6.2}$$

Since $S \subset X$, S itself is a partially ordered set.

c) A *maximal element* (if existent) of a partially ordered set X is an element $m \in X$ such that

$$\text{if } m \leq x \text{ for some } x \in X, \text{ then } x = m. \tag{6.3}$$

This is equivalent to saying that there is no $x \in X$ such that $m \leq x$ and $x \neq m$.

Remark 6.11. For an arbitrary partially ordered set X, a maximal element (if existent) does not have to be unique. For example, consider $X := \{\{1\}, \{2\}, \{3\}, \{1,2\}\}$ with $\subset$ as partial ordering. Both $\{3\}$ and $\{1,2\}$ are maximal elements. However, if we consider chains, then maximal elements are indeed unique by definition.

Theorem 6.12 (Kuratowski-Zorn Lemma)**.** Let $(M, \leq)$ be a non-empty partially ordered set. If every chain $C \subset M$ has an upper bound, then M has a maximal element.

Remark 6.13. The upper bound of every chain $C \subset M$ need not be in C, by definition of a chain, but it must be in M.

7 Differential Geometry

Definition 7.1 (Smooth Atlas [37]). Let M be a second countable Hausdorff topological space. An *n-dimensional smooth atlas on M* is a collection of maps

$$\mathcal{A} = \{(\varphi_i, U_i) \mid i \in I\}, \quad \varphi_i : U_i \to \varphi_i(U_i) \subset \mathbb{R}^n$$

such that all $U_i \subset M$ are open, all φ_i are homeomorphisms, I is an index set and

- $\{U_i \mid i \in I\}$ is an open covering of M,
- $\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$ are smooth $\forall i,\, j \in I$.

The tuples (φ_i, U_i), $i \in I$, are so-called *charts* on M, the maps $\varphi_i \circ \varphi_j^{-1}$ are called *transition maps* or *changes of coordinates* and n is the *dimension* of M.

Remark 7.2. To see why the domain of the transition maps $\varphi_i \circ \varphi_j^{-1}$ is $\varphi_j\left(U_i \cap U_j\right)$, note that the expression $\varphi_i\left(\varphi_j^{-1}\left(x\right)\right)$ only makes sense if

$$(x \in \varphi_j(U_j)) \wedge \left(\varphi_j^{-1}\left(x\right) \in U_i\right) \Rightarrow (x \in \varphi_j(U_j)) \wedge \left(x \in \varphi_j\left(U_i\right)\right) \Rightarrow x \in \varphi_j(U_j \cap U_i).$$

Similarly, we can convince ourselves that the codomain of the transition maps $\varphi_i \circ \varphi_j^{-1}$ is given by $\varphi_i(U_i \cap U_j)$. Since $x \in \varphi_j(U_j)$, it follows that $\varphi_j^{-1}(x) \in U_j$. In addition, due to the domain of the homeomorphism φ_i, it must hold that $\varphi_j^{-1}(x) \in U_i$. Thus:

$$\begin{aligned}
&\left(\varphi_j^{-1}(x) \in U_j\right) \wedge \left(\varphi_j^{-1}(x) \in U_i\right)\\
\Rightarrow &\left(\varphi_i(\varphi_j^{-1}(x)) \in \varphi_i(U_j)\right) \wedge \left(\varphi_i(\varphi_j^{-1}(x)) \in \varphi_i(U_i)\right)\\
\Rightarrow &\, \varphi_i\left(\varphi_j^{-1}(x)\right) \in \varphi_i\left(U_i \cap U_j\right).
\end{aligned}$$

Definition 7.3 (Equivalence of Atlases). Let M be a second countable Hausdorff topological space. Two atlases $\mathcal{A}$ and $\mathcal{B}$ on M are called *equivalent* if $\mathcal{A} \cup \mathcal{B}$ is an atlas on M.

Remark 7.4. To see that not all atlases are equivalent to each other, consider $M = \mathbb{R}$ (which is a second countable Hausdorff topological space). Consider the atlases $\mathcal{A} = \{(\varphi, M)\}$ with $\varphi : M \to M$, $x \mapsto x$ and $\mathcal{B} = \{(\psi, M)\}$ with $\psi : M \to M$, $x \mapsto x^3$. The atlases are not equivalent, since $\varphi \circ \psi^{-1} : M \to M$, $x \mapsto \sqrt[3]{x}$ is not smooth (the derivative is not continuous).

8 Normalizing Flows

Definition 8.1. Let M, N be two manifolds, $g : M \to N$ a differentiable map. If g is a bijection and its inverse $g^{-1} : N \to M$ is differentiable as well, then we call f a *diffeomorphism*. We talk of a C^k *diffeomorphism* if both g and g^{-1} are k-times continuously differentiable.

Theorem 8.2 (Change of variables [31])**.** Let $U, V \subset \mathbb{R}^n$ be open subsets and $T : U \to V$ a diffeomorphism, cf. Def. 8.1. Then the function $f : V \to \mathbb{C} \cup \{\infty\}$ is integrable over V if and only if the function

$$(f \circ T) \cdot \left| \det \left(\frac{\partial T_\mu}{\partial x_\nu} \right)_{\mu\nu} \right| \tag{8.1}$$

is integrable over U. In this case, it holds that

$$\int_U (f \circ T)(x) \cdot \left| \det \left(\frac{\partial T_\mu}{\partial x_\nu}(x) \right)_{\mu,\nu} \right| dx = \int_V f(y) dy. \tag{8.2}$$

Remark 8.3. If T is a diffeomorphism, then also T^{-1} is a diffeomorphism, thus we could also have chosen T^{-1} in the formulation of Theorem 8.2.

Theorem 8.4 (Inverse function theorem [46])**.** Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on some open set $V \subset \mathbb{R}^n$ containing $\mathbf{a}$ and suppose $\det Jg(\mathbf{a}) \neq 0$, where J shall be the Jacobi matrix of g. Then there is some open set containing $\mathbf{a}$ and an open set $W \subset \mathbb{R}^n$ containing $g(\mathbf{a})$ such that $g : V \to W$ has a continuous inverse $g^{-1} : W \to V$ which is differentiable for all $\mathbf{y} \in W$.

As matrices, we can write this as

$$J(g^{-1})(\mathbf{y}) = \left[(Jg) \left(g^{-1}(\mathbf{y}) \right) \right]^{-1} \tag{8.3}$$

Remark 8.5. An example for a function that is invertible and continuously differentiable but not a diffeomorphism is $g : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^3$. Its inverse is obviously $g^{-1} : \mathbb{R} \to \mathbb{R}, x \mapsto \sqrt[3]{x}$, cf. [16] for a nice plot. However, $\frac{dg^{-1}}{dx}|_{x=0}$ does not exist. The reason is that $\det Jg(0) = 0$ and hence the inverse function theorem does not apply.

- Let $\mathbf{U}$ be a random variable and let $p(\mathbf{U})$ describe the probability distribution of it, e.g. a uniform distribution between 0 and 1. We now make a simple transformation and obtain a new random variable $\mathbf{X}$, where we again denote by $p(\mathbf{X})$ the probability distribution of $\mathbf{X}$. We obtain $\mathbf{X}$ in the following way:

  $$p(\mathbf{X}) = p(\mathbf{U}) \left| \det \left(\frac{\partial \mathbf{f}}{\partial \mathbf{U}} \right) \right|^{-1}, \tag{8.4}$$

  where $\mathbf{f}$ denotes an invertible (and hence bijective) mapping.

- Without proof, it holds that

  $$\left| \det \left(\frac{\partial \mathbf{f}}{\partial \mathbf{U}} \right) \right|^{-1} = \left| \det \left(\frac{\partial \mathbf{f}^{-1}}{\partial \mathbf{U}} \right) \right| \tag{8.5}$$

and thus we can rewrite Eq. (8.4) as

$$p(\mathbf{U}) = p(\mathbf{X}) \left| \det \left(\frac{\partial \mathbf{f}^{-1}}{\partial \mathbf{U}} \right) \right|^{-1}. \tag{8.6}$$

Since we assumed $\mathbf{f}$ to be invertible, $\mathbf{f}^{-1}$ is well-defined.

- In practice, we will want $\mathbf{f}$ to be both invertible and to have a **tractable** Jacobian, i.e. a Jacobian that we can easily calculate. For $\mathbf{f}$ to have a Jacobian at all, each of its first-order partial derivatives must exist [23]. So-called *autoregressive flows* have the property that their Jacobian is an upper triangular matrix. For an upper triangular matrix, it holds that its determinant is given by the product of its diagonal elements [17].

Definition 8.6 (Determinant)**.** Let D be an $n \times n$ matrix and let S_n denote the symmetric group over n. Then the determinant of D is defined as:

$$\det(D) := \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \right), \tag{8.7}$$

cf. [20].

Lemma 8.7. Let A be a $k \times k$, 0 an $k \times n$, C an $n \times k$ and D an $n \times n$ matrix; then

$$\det \left(\begin{pmatrix} A & 0 \\ C & D \end{pmatrix} \right) = \det(A) \det(D). \tag{8.8}$$

Proof. Define

$$B := \begin{pmatrix} A & 0 \\ C & D \end{pmatrix}. \tag{8.9}$$

Clearly,

$$b_{i,j} = \begin{cases} a_{i,j} & i, j \leq k, \\ 0 & i \leq k, j \geq k+1, \\ c_{i-k,j} & i \geq k+1, j \leq k, \\ d_{i-k,j-k} & i, j \geq k+1. \end{cases} \tag{8.10}$$

We can write the determinant of B as

$$\det(B) = \sum_{\sigma \in S_{n+k}} \operatorname{sgn}(\sigma) \prod_{i=1}^{n+k} b_{i,\sigma(i)}. \tag{8.11}$$

From Eq. (8.10) we know that all summands of the form $\sigma(i) = j$ with $i \leq k$, $j \geq k+1$ are 0. Therefore, we can consider all permutations of the form $\sigma(i) = j$ with $i, j \leq k$ or $\sigma(i) = j$ with $i \geq k+1$, $j \leq k$. We can also write this in the form $\sigma(i) = \pi(i)$ for $i \leq k$

and $\sigma(k+i) = k + \tau(i)$ for $1 \leq i \leq n$, where $\pi \in S_k$ and $\tau \in S_n$. Denote the set of all such permutations by $\tilde{S}_{k+n}$. Thus:

$$
\begin{aligned}
\det(B) &= \sum_{\sigma \in \tilde{S}_{k+n}} \operatorname{sgn}(\sigma) \prod_{i=1}^{n+k} b_{i,\sigma(i)} && (8.12) \\
&= \sum_{\sigma \in \tilde{S}_{k+n}} \operatorname{sgn}(\sigma) \prod_{i=1}^{k} b_{i,\sigma(i)} \prod_{i=k+1}^{n+k} b_{i,\sigma(i)} && (8.13) \\
&\overset{(8.10)}{=} \sum_{\sigma \in \tilde{S}_{k+n}} \operatorname{sgn}(\sigma) \prod_{i=1}^{k} a_{i,\sigma(i)} \prod_{i=k+1}^{n+k} d_{i-k,\sigma(i)-k} && (8.14) \\
&= \sum_{\sigma \in \tilde{S}_{k+n}} \operatorname{sgn}(\sigma) \prod_{i=1}^{k} a_{i,\sigma(i)} \prod_{i=1}^{n} d_{i,\sigma(i+k)-k} && (8.15) \\
&= \sum_{\pi \in S_k, \tau \in S_n} \operatorname{sgn}(\pi)\operatorname{sgn}(\tau) \prod_{i=1}^{k} a_{i,\pi(i)} \prod_{i=1}^{n} d_{i,\tau(i)} && (8.16) \\
&= \sum_{\pi \in S_k} \operatorname{sgn}(\pi) \prod_{i=1}^{k} a_{i,\pi(i)} \sum_{\tau \in S_n} \operatorname{sgn}(\tau) \prod_{i=1}^{n} d_{i,\tau(i)} && (8.17) \\
&= \det(A)\det(D) && (8.18)
\end{aligned}
$$

[14]

Definition 8.8. Let $h(\cdot\ ;\theta) : \mathbb{R} \to \mathbb{R}$ be a bijection parametrized by θ. Then an *autoregressive model* is a function

$$
g : \mathbb{R}^D \to \mathbb{R}^D, \begin{pmatrix} x_1 \\ \dots \\ x_D \end{pmatrix} \mapsto \begin{pmatrix} h\left(x_1; \Theta_1\right) \\ h\left(x_2; \Theta_2\left(x_1\right)\right) \\ \dots \\ h\left(x_D; \Theta_D\left(x_1, \dots, x_{D-1}\right)\right) \end{pmatrix} \tag{8.19}
$$

The functions Θ_t for $t = 2, \dots, D$ are arbitrary functions whose domain is $\mathbb{R}^{t-1}$, Θ_1 is a constant.

Remark 8.9. The Jacobian matrix of an autoregressive flow is given as follows:

$$
Dg = \begin{pmatrix} \partial g_1/\partial x_1 & \partial g_1/\partial x_2 & \dots & \partial g_1/\partial x_D \\ \partial g_2/\partial x_1 & \partial g_2/\partial x_2 & \dots & \partial g_2/\partial x_D \\ \vdots & \vdots & \ddots & \vdots \\ \partial g_D/\partial x_1 & \partial g_D/\partial x_2 & \dots & \partial g_D/\partial x_D \end{pmatrix} \tag{8.20}
$$

One can easily convince oneself that Dg is a lower triangular matrix.

Theorem 8.10. Normalizing flows in $\mathbb{R}^D$ come from the push-forward of a measure.

Proof. Let $\mathcal{Y}$, $\mathcal{Z} \subset \mathbb{R}^D$ be open, $\Sigma_{\mathcal{Y}} = \mathcal{B}(\mathcal{Y})$, $\Sigma_{\mathcal{Z}} = \mathcal{B}(\mathcal{Z})$ and $g : \mathcal{Z} \to \mathcal{Y}$ be a diffeomorphism. The function $g : \mathcal{Z} \to \mathcal{Y}$ is measurable if and only if $g^{-1}(G) \in \Sigma_{\mathcal{Z}}$ for every set G that is open in $\mathcal{Y}$, cf. Theorem 2.4 (Borel σ-algebras are generated by the open sets). Since for functions between two topological spaces it holds that they are continuous if and only if the inverse image of an open set is again open, we have that g is indeed measurable.

Now define the probability measure μ on the measurable space $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$ as

$$\mu(I) := \int_I p_{\mathbf{Z}}(z) d\lambda(z) \; \forall I \in \Sigma_{\mathcal{Z}}, \tag{8.21}$$

where $p_{\mathbf{Z}} : \mathcal{Z} \to \mathbb{R}$ shall be a PDF and λ the Lebesgue measure. (The Lebesgue measure is defined on the completion of $\mathcal{B}(\mathbb{R}^D)$, which is "larger" than $\mathcal{B}(\mathbb{R}^D)$.) By considering the push-forward of μ under the measurable map g, we have $\forall J \in \Sigma_{\mathcal{Y}}$:

$$g_\star\mu\,(J) = \mu(g^{-1}(J)) \overset{(8.21)}{=} \int_{g^{-1}(J)} p_{\mathbf{Z}}(z) d\lambda(z) \tag{8.22}$$

Since by assumption $g : \mathcal{Z} \to \mathcal{Y}$ and therefore also the inverse $g^{-1} : \mathcal{Y} \to \mathcal{Z}$ are a diffeomorphism, we can use the change of variables formula from Theorem 8.2, since we assume $p_{\mathbf{Z}}$ to be integrable over $\mathcal{Z}$. We apply Theorem 8.2 to g^{-1} instead of g, cf. Remark 8.3:

$$g_\star\mu(J) = \int_{g^{-1}(J)} p_{\mathbf{Z}}(z) d\lambda(z) = \int_J \underbrace{\left(p_{\mathbf{Z}} \circ g^{-1}\right)(y) \cdot \left|\det Dg^{-1}(y)\right|}_{=p_{\mathbf{Y}}(y)} d\lambda(y) \tag{8.23}$$

■

Definition 8.11. A rational-quadratic function takes the form of a quotient of two quadratic polynomials.

$$\frac{\alpha^{(k)}(\xi)}{\beta^{(k)}(\xi)} = \frac{a_0\xi^2 + a_1\xi + a_2}{b_0\xi^2 + b_1\xi + b_2} \tag{8.24}$$

9 Diffusion-Based Models

Remark 9.1 (ELBO VAEs)**.** One can easily show that the (marginal) log-likelihood of the data is given as [36]

$$\log p(x) = \text{ELBO} + D_{\text{KL}}\left[q_\phi(z|x)||p(z|x)\right], \tag{9.1}$$

where $p(z|x)$ is the true posterior and ELBO is the usual evidence lower bound, i.e.

$$\text{ELBO} = \mathbb{E}_{q_\phi(z|x)}\left[\log p_\psi(x|z)\right] - D_{\text{KL}}\left[q_\phi(z|x)||p(z)\right] \tag{9.2}$$

Eq. (9.1) shows that since the (marginal) log-likelihood is not parameterized by any NN parameters, maximizing the ELBO necessary leads to a lower $D_{\text{KL}}\left[q_\phi(z|x)||p(z|x)\right]$, which is the reverse KLD.

Lemma 9.2 (ELBO)**.** Let $q(x_0)$ denote the true (unknown) distribution of a real image x_0, and let $p_\theta(x_0)$ be the model's approximation to $q(x_0)$, then we have the following ELBO-like loss:

$$\mathbb{E}_{q(x_0)}\left[\log p_\theta(x_0)\right] \geq -\mathbb{E}_{q(x_0,\ldots,x_T)}\left[\log \frac{q(x_1,\ldots,x_T \mid x_0)}{p_\theta(x_0,\ldots,x_T)}\right]. \tag{9.3}$$

Proof. [47]

$$\begin{aligned}
\log p_\theta(x_0) &\geq \log p_\theta(x_0) - D_{\text{KL}}\left[q(x_1,\ldots,x_T \mid x_0) \;||\; p_\theta(x_1,\ldots,x_T \mid x_0)\right] && (9.4)\\
&= \log p_\theta(x_0) - \mathbb{E}_{q(x_1,\ldots,x_T|x_0)}\left[\log \frac{q(x_1,\ldots,x_T \mid x_0)}{p_\theta(x_1,\ldots,x_T \mid x_0)}\right] && (9.5)\\
&= \log p_\theta(x_0) - \mathbb{E}_{q(x_1,\ldots,x_T|x_0)}\left[\log \frac{q(x_1,\ldots,x_T \mid x_0)}{p_\theta(x_0,\ldots,x_T)/p_\theta(x_0)}\right] && (9.6)\\
&= -\mathbb{E}_{q(x_1,\ldots,x_T|x_0)}\left[\log \frac{q(x_1,\ldots,x_T \mid x_0)}{p_\theta(x_0,\ldots,x_T)}\right] && (9.7)\\
\Rightarrow \mathbb{E}_{q(x_0)}\left[\log p_\theta(x_0)\right] &\geq -\mathbb{E}_{q(x_0)}\mathbb{E}_{q(x_1,\ldots,x_T|x_0)}\left[\log \frac{q(x_1,\ldots,x_T \mid x_0)}{p_\theta(x_0,\ldots,x_T)}\right] && (9.8)
\end{aligned}$$

Assuming that the assumptions of Fubini's theorem hold and from the monotonicity of the expectation, we have:

$$\begin{aligned}
\mathbb{E}_{q(x_0)}\left[\log p_\theta(x_0)\right] &\geq \mathbb{E}_{q(x_1,\ldots,x_T|x_0)q(x_0)}\left[\log \frac{q(x_1,\ldots,x_T \mid x_0)}{p_\theta(x_0,\ldots,x_T)}\right] && (9.9)\\
&= E_{q(x_0,\ldots,x_T|x_0)}\left[\log \frac{q(x_1,\ldots,x_T \mid x_0)}{p_\theta(x_0,\ldots,x_T)}\right]. && (9.10)
\end{aligned}$$

■

Definition 9.3 (Wiener process)**.** Let W_t be a real-valued continuous-time stochastic process. It is said to be a *Wiener process* if the following properties hold:

- $W_0 = 0$,
- W has independent increments, i.e. $\forall t > 0$:, the terms $W_{t+u} - W_t$, $u \geq 0$ are independent of past values W_s, $s \leq t$,

- W has Gaussian increments: $W_{t+u} - W_t \sim \mathcal{N}(0, u)$,
- W has continuous paths, i.e. $\forall t$, W_t is continuous in t.

source: `https://en.wikipedia.org/wiki/Wiener_process`

Remark 9.4. If ξ_1, ξ_2, $\ldots$ be i.i.d random variables with a mean of 0 and standard deviation of 1. For every n, define a continuous time stochastic process

$$W_n(t) := \frac{1}{\sqrt{n}} \sum_{1 \le k \le \lfloor nt \rfloor} \xi_k \tag{9.11}$$

This is what makes Wiener processes so powerful (and explains the ubiquity of Brownian motion). According to Donsker's theorem, the above expression becomes a Wiener process.

10 Miscellaneous

Lemma 10.1 (Chain rule for KL-divergences)**.** Let $p(x, y)$ and $q(x, y)$ be two arbitrary PDF's. Then the following hods:

$$D_{\mathrm{KL}}(p(x,y)||q(x,y)) = D_{\mathrm{KL}}(p(x)||q(x)) + D_{\mathrm{KL}}(p(y|x)||q(y|x)) \tag{10.1}$$

Proof. Brute-force calculation yields:

$$\begin{aligned}
D_{\mathrm{KL}}(p(x,y)||q(x,y)) &= \int\int p(x,y) \log \frac{p(x,y)}{q(x,y)} d\lambda(x) d\lambda(y) && (10.2)\\
&= \int\int p(x,y) \log \frac{p(x)p(y|x)}{q(x)q(y|x)} d\lambda(x) d\lambda(y) && (10.3)\\
&= \int\int p(x,y) \log \frac{p(x)}{q(x)} d\lambda(x) d\lambda(y) + \int\int p(x,y) \log \frac{p(y|x)}{q(y|x)} d\lambda(x) d\lambda(y) && (10.4)\\
&= D_{\mathrm{KL}}(p(x)||q(x)) + \int p(x) d\lambda(x) \int p(y|x) \log \frac{p(y|x)}{q(y|x)} d\lambda(y) && (10.5)\\
&= D_{\mathrm{KL}}(p(x)||q(x)) + D_{\mathrm{KL}}(p(y|x)||q(y|x)) && (10.6)
\end{aligned}$$

■

Definition 10.2 (Mutual information)**.** Let $(X, Y) \sim P_{(X,Y)} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, where $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ is the space of all probability measures over the space $\mathcal{X} \times \mathcal{Y}$. The mutual information between the random variables X and Y is now defined as [25, Def. 10.1]:

$$I(X;Y) := D_{\mathrm{KL}}\left(P_{(X,Y)}||P_X \otimes P_Y\right), \tag{10.7}$$

where P_X and P_Y are the marginal measures of the coupling measure $P_{(X,Y)}$ and $P_X \otimes P_Y$ is the induced **product measure**.

Definition 10.3 (Divergence)**.** Let p and q be two probability distributions, then a divergence D must satisfy: $D(p||q) \geq 0\ \forall p, q$ and $D(p||q) = 0 \Leftrightarrow p = q$ a.e. Note that the triangle inequality and the symmetry property need not be satisfied in general.

Example 10.4 (f-Divergence)**.** Let P and Q be two probability measures defined on the σ-algebra over a space Ω such that $P \ll Q$, i.e. P is absolutely continuous wrt Q. Then, for a convex function $f : [0, \infty) \to \mathbb{R}$ s.t.

(i) $f(1) = 0$,

(ii) f is strictly convex at $x = 1$,

the f-divergence is now defined as [26],

$$D_f(P||Q) := \int_\Omega f\left(\frac{dP}{dQ}\right) dQ, \tag{10.8}$$

where dP/dQ is the Radon-Nikodym derivative.

Proof. 1. **Non-negativity**: We know that

$$\mathbb{E}_Q\left[\frac{dP}{dQ}\right] = \int_\Omega \frac{dP}{dQ} dQ = 1. \tag{10.9}$$

Since f is a convex function, by Jensen's inequality,

$$0 = f(1) = f\left(\mathbb{E}_Q\left[\frac{dP}{dQ}\right]\right) \leq \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right] = \int_\Omega f\left(\frac{dP}{dQ}\right) dQ \overset{(10.8)}{=} D_f(P||Q). \tag{10.10}$$

2. **Zero iff equal**: If $P = Q$ a.e., then $dP/dQ = 1$ a.e., and hence

$$D_f(P||Q) = \int_\Omega f(1) dQ = \int_\Omega 0 dQ = 0. \tag{10.11}$$

Conversely, if $D_f(P||Q) = 0$, since f is strictly convex at 1 and convex everywhere else, this implies that the only minimum point of f is at $f(1) = 0$. Therefore, the integrand of Eq. (10.8), i.e. $f\left(\frac{dP}{dQ}\right)$, must be 0 a.e., implying that

$$\frac{dP}{dQ} = 1 \Rightarrow P = Q. \tag{10.12}$$

■

A Weierstrass Approximation Theorem

In the following, we will prove the Weierstrass approximation theorem, as stated in Example 4.76. The proof will use *Korovkin*[5] *sequences*. The approach is taken from Chapter 6 of [34].

Definition A.1 (Positivity of Operator)**.** A linear operator $K : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$ is *positive* on $\mathcal{C}[a,b]$ if $Kf \geq 0$ for all $f \in \mathcal{C}[a,b]$ satisfying $f \geq 0$, where all inequalities are taken pointwise on $[a,b]$.

Definition A.2 (Monotonicity of Operator)**.** A linear operator $K\colon \mathcal{C}[a,b] \to \mathcal{C}[a,b]$ is *monotone* on $\mathcal{C}[a,b]$ if for any $f,\, g \in \mathcal{C}[a,b]$ satisfying $f \leq g$, we have $Kf \leq Kg$, where all inequalities are taken pointwise on $[a,b]$.

Remark A.3. A linear operator is positive iff it is monotone.

Proof. "$\Longrightarrow$" Consider $h := g - f \in \mathcal{C}[a,b]$, where $g,\, f \in \mathcal{C}[a,b]$, with $h = g - f \geq 0$, i.e. $f \leq g$, then by the positivity and linearity of K we have $Kh = Kg - Kf \geq 0$, or equivalently $Kf \leq Kg$.

"$\Longleftarrow$" Assume that for any $f,\, g \in \mathcal{C}[a,b]$ satisfying $f \leq g$, we have $Kf \leq Kg$. Take $f = 0$, then $g \geq 0$ implies $Kg \geq 0$ for any $g \in \mathcal{C}[a,b]$. ∎

Definition A.4 (Korovkin sequence)**.** A sequence $(K_n)_{n\in\mathbb{N}}$ of linear and positive operators $K_n : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$ is called a *Korovkin sequence* on $\mathcal{C}[a,b]$ if $K_n p \overset{n\to\infty}{\longrightarrow} p$ wrt d_∞ for all $p \in \mathcal{P}_2$, where $\mathcal{P}_n$ is the linear space of all univariate polynomials of degree at most n defined on $[a,b]$.

Theorem A.5 (Korovkin, 1953)**.** For a compact interval $[a,b] \subset \mathbb{R}$, let $(K_n)_{n\in\mathbb{N}}$ be a Korovkin sequence on $\mathcal{C}[a,b]$. Then for any $f \in \mathcal{C}[a,b]$, we have

$$K_n f \overset{n\to\infty}{\longrightarrow} f \quad \text{wrt } d_\infty. \tag{A.1}$$

Proof. Let $f \in \mathcal{C}[a,b]$ be arbitrary. Then, f is bounded on $[a,b]$, cf. Theorem 4.115, i.e. there is an $M > 0$ s.t. $||f||_\infty \leq M$. Also, f is uniformly continuous on $[a,b]$, cf. Theorem 4.114, i.e.

$$\forall \epsilon > 0\ \exists \delta > 0\ \forall x, y \in [a,b] : |x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon. \tag{A.2}$$

Now let $t \in [a,b]$ be fixed, $x \in [a,b]$ and $\epsilon > 0$ be arbitrary, then if $|x - t| < \delta$, we have $f(x) - f(t) < \epsilon$. Applying the linear and monotone operator K_n, for $n \in \mathbb{N}$, to both sides of the inequality (wrt x), we have

$$(K_n f)(x) - f(t)(K_n 1)(x) < \epsilon (K_n 1)(x) \tag{A.3}$$

$$\Rightarrow |(K_n f)(x) - f(t)(K_n 1)(x)| < \epsilon \left|(K_n 1)(x)\right|. \tag{A.4}$$

By assumption, for any $\tilde{\epsilon} > 0$ there is an $N(\tilde{\epsilon}) \in \mathbb{N}$ satisfying

$$d_\infty\left(K_n x^k, x^k\right) = \left|\left|K_n x^k - x^k\right|\right|_\infty < \tilde{\epsilon} \qquad \text{for } k \in \{0,1,2\}\ \forall n \geq N. \tag{A.5}$$

This in particular implies

$$|(K_n 1)(x)| \leq ||K_n 1||_\infty = ||K_n 1 - 1 + 1||_\infty \leq ||K_n 1 - 1||_\infty + 1 < \tilde{\epsilon} + 1 \tag{A.6}$$

[5] Pavel Petrovich Korovkin (1913 – 1985), Russian mathematician.

for all $n \geq N$. Using the triangle inequality and putting Eq. (A.6) into (A.4), we get the following inequality:

$$
\begin{aligned}
|(K_n f)(x) - f(t)| &\leq |(K_n f)(x) - f(t)(K_n 1)(x)| + |f(t)(K_n 1)(x) - f(t)| && \text{(A.7)}\\
&\overset{\text{(A.4)}}{<} \epsilon\,|(K_n 1)(x)| + |f(t)|\,|(K_n 1)(x) - 1| && \text{(A.8)}\\
&\overset{\text{(A.6)}}{<} \epsilon(\tilde{\epsilon} + 1) + M\,||K_n 1 - 1||_\infty \quad \forall n \geq N && \text{(A.9)}\\
&< \epsilon(\epsilon \tilde{+} 1) + M\tilde{\epsilon} \quad \forall n \geq N. && \text{(A.10)}
\end{aligned}
$$

For $x = t$, we have the inequality

$$|(K_n f)(t) - f(t)| < \epsilon\,(\tilde{\epsilon} + 1) \quad \forall n \geq N. \tag{A.11}$$

The RHS can be bounded from above by an arbitrarily small $\hat{\epsilon} > 0$, so that for some $N' = N(\hat{\epsilon}) \in \mathbb{N}$ we have

$$d_\infty(K_n f, f) < \hat{\epsilon} \quad \forall n \geq N'. \tag{A.12}$$

■

In the following, we will explicitly construct a Korovkin sequence on $\mathcal{C}[a, b]$.

Remark A.6. In the following, we might restrict ourselves to the continuous functions $\mathcal{C}[0, 1]$ on $[0, 1]$. This is WLOG, since for any $x \in [a, b]$, we can transform it into $[0, 1]$ via a simple affine-linear transformation.

Definition A.7. Let $x \in [0, 1]$, then the *Bernstein*[6] *polynomials* are defined as

$$\beta_j^{(n)}(x) := \binom{n}{j} x^j (1 - x)^{n-j} \in \mathcal{P}_n \quad \text{with } 0 \leq j \leq n, n \in \mathbb{N}_0 \tag{A.13}$$

Example A.8. Here are some Bernstein polynomials:

$$\boldsymbol{n = 0}: \quad \beta_0^{(0)}(x) = 1$$

$$\boldsymbol{n = 1}: \quad \beta_0^{(1)}(x) = 1 - x \quad \beta_1^{(1)}(x) = x$$

$$\boldsymbol{n = 2}: \quad \beta_0^{(2)}(x) = (1 - x)^2 \quad \beta_1^{(2)}(x) = 2x(1 - x) \quad \beta_2^{(2)}(x) = x^2$$

$$\boldsymbol{n = 3}: \quad \beta_0^{(3)}(x) = (1 - x)^3 \quad \beta_1^{(3)}(x) = 3x(1 - x)^2 \quad \beta_2^{(3)}(x) = 3x^2(1 - x) \quad \beta_3^{(3)}(x) = x^3$$

Remark A.9. The Bernstein polynomials $\beta_0^{(n)}, \ldots, \beta_n^{(n)} \in \mathcal{P}_n$, for $n \in \mathbb{N}_0$, ...

(a) ...form a basis for the polynomial space $\mathcal{P}_n$;

(b) ...are positive on $[0, 1]$, i.e. $\beta_j^{(n)}(x) \geq 0$ for all $x \in [0, 1]$;

(c) ...are symmetrical around $x = 1/2$, i.e. $\beta_j^{(n)}(x) = \beta_{n-j}^{(n)}(1 - x)$;

(d) ...are a *partition of unity* on $[0, 1]$, i.e.

$$\sum_{j=0}^{n} \beta_j^{(n)}(x) = 1 \; \forall x \in [0, 1]. \tag{A.14}$$

[6] Sergei Natanovich Bernstein (1880 – 1968)

Proof. (a) From linear algebra, we know that a set is a basis of a finite-dimensional vector space iff it has the same dimension and all vectors in the set are linearly independent. Since $\dim(\mathcal{P}_n) = n+1$ and $\left\{\beta_0^{(n)}(x), \ldots, \beta_n^{(n)}(x)\right\}$ contains $n+1$ vectors, it only remains to show that the $\beta_0^{(n)}(x), \ldots, \beta_n^{(n)}(x)$ are linearly independent.

For this, note that [32, proof Lemma 2.5 ix)]

$$\beta_j^{(n)}(x) = \binom{n}{j} x^j (1-x)^{n-j} \in \mathcal{P}_n \quad \text{with } 0 \le j \le n \tag{A.15}$$

$$= \binom{n}{j} x^j \sum_{i=0}^{n-j} \binom{n-j}{i} (-x)^i \tag{A.16}$$

$$= \sum_{i=0}^{n-j} \binom{n}{j} \binom{n-j}{i} (-1)^i x^{j+i} \tag{A.17}$$

$$= \sum_{i=j}^{n} \binom{n}{j} \binom{n-j}{i-j} (-1)^{i-j} x^i \tag{A.18}$$

$$= \sum_{i=j}^{n} \binom{n}{i} \binom{i}{j} (-1)^{i-j} x^i \tag{A.19}$$

Now consider the equation (with $\lambda_k \in \mathbb{R}$) for all $0 \le k \le n$:

$$0 = \sum_{k=0}^{n} \lambda_k \beta_k^{(n)} \overset{\text{(A.19)}}{=} \sum_{k=0}^{n} \lambda_k \sum_{i=k}^{n} \binom{n}{i} \binom{i}{k} (-1)^{i-k} x^i \tag{A.20}$$

For the double sum, we will use the following identity with $a_{ik} \in \mathbb{R}$ [43, p. 2]:

$$\sum_{k=0}^{n} \sum_{i=k}^{n} a_{ik} = \sum_{0 \le k \le i \le n} a_{ik} = \sum_{i=0}^{n} \sum_{k=0}^{i} a_{ik} = \sum_{k=0}^{n} \sum_{i=0}^{k} a_{ki}, \tag{A.21}$$

since $0 \le i \le n$, and for a fixed i, $0 \le k \le i$, and where we exchanged the indices $(i \mapsto k,\ k \mapsto i)$. Thus, Eq. (A.20) becomes

$$0 = \sum_{k=0}^{n} \lambda_k \beta_k^{(n)} = \sum_{k=0}^{n} x^k \sum_{i=0}^{k} \lambda_i \binom{n}{k} \binom{k}{i} (-1)^{k-i} \tag{A.22}$$

$$= \lambda_0 + (-n\lambda_0 + n\lambda_1)\, x + \left(\lambda_0 n\,(n-1) - \lambda_1 n(n-1) + \frac{\lambda_2 n(n-1)}{2}\right) x^2 + \ldots \tag{A.23}$$

Thus, every polynomial $p \in \mathcal{P}_n$ can be written as the linear combination of $\left\{\beta_0^{(n)}(x), \ldots, \beta_n^{(n)}(x)\right\}$. Also, note that we already know that the monomials $\{x^0, x^1, \ldots, x^n\}$ form a basis for $\mathcal{P}_n$, thus

$$\lambda_0 = 0 \tag{A.24}$$

$$-n\lambda_0 + n\lambda_1 = 0 \tag{A.25}$$

$$\left(\lambda_0 n\,(n-1) - \lambda_1 n(n-1) + \frac{\lambda_2 n(n-1)}{2}\right) = 0 \tag{A.26}$$

$$\vdots$$

By putting $\lambda_0 = 0$ into Eq. (A.25), we get $\lambda_1 = 0$. Putting $\lambda_0 = \lambda_1 = 0$ into Eq. (A.26), we have $\lambda_2 = 0$. Thus, we can recursively show that $\lambda_k = 0$ for all $0 \leq k \leq n$, and hence the set $\left\{\beta_j^{(n)}(x)\right\}_{0\leq j\leq n}$ is linearly independent.

(b) This is straightforward, since $x^j \geq 0$, $(1-x)^{n-j} \geq 0$ and $\binom{n}{j} \geq 0$ for all $x \in [0,1]$, hence $\beta_j^{(n)}(x) \geq 0$.

(c) We have

$$\beta_{n-j}^{(n)}(1-x) = \binom{n}{n-j}(1-x)^{n-j}x^j = \binom{n}{j}(1-x)^{n-j}x^j = \beta_j^{(n)}(x). \tag{A.27}$$

(d) By the binomial theorem we have

$$\sum_{j=0}^{n} \beta_j^{(n)}(x) = \sum_{j=0}^{n} \binom{n}{j} x^j (1-x)^{n-j} = (x + (1-x))^n = 1. \tag{A.28}$$

■

Definition A.10. For $n \in \mathbb{N}$, define the *Bernstein operator* $B_n : \mathcal{C}[0,1] \to \mathcal{P}_n$ as

$$(B_n f)(x) := \sum_{j=0}^{n} f\left(\frac{j}{n}\right) \beta_j^{(n)}(x) \quad \forall f \in \mathcal{C}[0,1], \tag{A.29}$$

where the $\beta_j^{(n)}$ are the Bernstein polynomials from Def. A.7.

Remark A.11. The Bernstein operators B_n are linear operators on $\mathcal{C}[0,1]$. They are also positive (and hence monotone) on $\mathcal{C}[0,1]$, since the $\beta_j^{(n)}$ are positive, cf. Remark A.9 b).

Remark A.12. The Bernstein operators $B_n : \mathcal{C}[0,1] \to \mathcal{P}_n$ are also bounded (in the sense of Def. 5.3) wrt $||\cdot||_\infty$ ($C = 1$ is one upper bound). By Theorem 5.5, they are thus also continuous on $\mathcal{C}[0,1]$.

Proof. We show the boundedness of the Bernstein operators $B_n : \mathcal{C}[0,1] \to \mathcal{P}_n$ directly. Let $f \in \mathcal{C}[0,1]$ and $n \in \mathbb{N}$ be arbitrary, then

$$||B_n f||_\infty = \left|\left|\sum_{j=0}^{n} f\left(\frac{j}{n}\right)\beta_j^{(n)}(x)\right|\right|_\infty = \sup_{x\in[0,1]}\left\{\left|\sum_{j=0}^{n} f\left(\frac{j}{n}\right)\beta_j^{(n)}(x)\right|\right\} \tag{A.30}$$

$$\leq \sup_{x\in[0,1]}\left\{\left|\sup_{x\in[0,1]}\{|f(x)|\}\sum_{j=0}^{n}\beta_j^{(n)}(x)\right|\right\} = ||f||_\infty \sup_{x\in[0,1]}\left\{\sum_{j=0}^{n}\beta_j^{(n)}(x)\right\} = ||f||_\infty\,, \tag{A.31}$$

since the Bernstein polynomials form a partition of unity, cf. Remark A.9 d).

Hence, we have shown that $||B_n f||_\infty \leq ||f||_\infty$, which proves the boundedness of B_n and that $C = 1$ is one upper bound for B_n. ■

Theorem A.13. The sequence $(B_n)_{n\in\mathbb{N}}$ with $B_n : \mathcal{C}[0,1] \to \mathcal{P}_n$ according to Def. A.7 forms a Korovkin sequence on $\mathcal{C}[0,1]$, cf. Def. A.4.

Proof. The Bernstein operators B_n for $n \in \mathbb{N}$ reproduce linear polynomials: Let $1 \in \mathcal{C}[0,1]$ be the constant function that assigns each $x \in [0,1]$ the value 1, then

$$B_n 1 = 1, \tag{A.32}$$

and by the linearity of B_n we have $B_n(c) = c$ for any $c \in \mathbb{R}$.

We also have $B_n p_1 = p_1$, where $p_1 \in \mathcal{C}[0,1]$ is the identity function:

$$B_n p_1 = \sum_{j=0}^{n} p_1\left(\frac{j}{n}\right)\beta_j^{(n)}(x) = \sum_{j=0}^{n} \frac{j}{n}\beta_j^{(n)}(x) = \sum_{j=1}^{n} \frac{j}{n}\beta_j^{(n)}(x) \tag{A.33}$$

$$= \sum_{j=1}^{n} \binom{n-1}{j-1} x^j (1-x)^{n-j} = \sum_{j=0}^{n-1} \binom{n-1}{j} x^{j+1}(1-x)^{n-j-1} \tag{A.34}$$

$$= x\sum_{j=0}^{n-1} \beta_j^{(n-1)}(x) = x = p_1(x), \tag{A.35}$$

since the $\beta_j^{(n-1)}$ form a partition of unity, cf. Remark A.9 d). This implies that $B_n(\alpha p_1 + c) = \alpha p_1 + c$ for any real α, c.

Now consider $p_2 \in \mathcal{P}_2$, where $p_2(x) := x^2$, and consider the following sequence of polynmials of degree 2 for $n \geq 2$:

$$f_n(x) := \frac{x(nx-1)}{n-1} = \frac{n}{n-1}x^2 - \frac{1}{n-1}x \in \mathcal{P}_2. \tag{A.36}$$

For $n \geq 2$, we have

$$\begin{aligned}
(B_n f_n)(x) &= \sum_{j=0}^{n} f\left(\frac{j}{n}\right)\beta_j^{(n)}(x) = \sum_{j=0}^{n} f\left(\frac{j}{n}\right)\binom{n}{j} x^j(1-x)^{n-j} \\
&= \sum_{j=0}^{n} \frac{j(j-1)}{n(n-1)}\binom{n}{j} x^j(1-x)^{n-j} = \sum_{j=2}^{n} \frac{j(j-1)}{n(n-1)}\binom{n}{j} x^j(1-x)^{n-j} \\
&= \sum_{j=2}^{n} \frac{(n-2)!}{(n-j)!\,(j-2)!} x^j(1-x)^{n-j} = x^2 \sum_{j=0}^{n-2} \binom{n-2}{j} x^j (1-x)^{n-j-2} \\
&= x^2 \sum_{j=0}^{n-2} \beta_j^{(n-2)} = x^2 = p_2(x),
\end{aligned} \tag{A.37}$$

since the Bernstein polynomials $\left\{\beta_j^{(n-2)}\right\}_{0\leq j\leq n-2}$ form a partition of unity, cf. Remark A.9 d).

Now note that any polynomial $p \in \mathcal{P}_2$ can be written as $p = \alpha p_2 + \beta p_1 + \gamma$ for some $\alpha, \beta, \gamma \in \mathbb{R}$: Using the linearity and boundedness of the Bernstein operators (where $C = 1$ is one upper bound), we have:

$$B_n p = \alpha B_n p_2 + \beta B_n p_1 + B_n \gamma = \alpha B_n p_2 + \beta p_1 + \gamma, \tag{A.38}$$

since $B_n p_1 = p_1$ and $B_n \gamma = \gamma$, as we have previously shown. Further:

$$\begin{aligned}
||B_n p - p||_\infty &= ||\alpha B_n p_2 - \alpha p_2||_\infty = |\alpha|\, ||B_n p_2 - p_2||_\infty \overset{(A.37)}{=} |\alpha|\, ||B_n(p_2 - f_n)||_\infty \\
&\leq |\alpha|\, ||p_2 - f_n||_\infty = |\alpha| \sup_{x\in[0,1]} \{|p_2(x) - f_n(x)|\} \\
&= |\alpha| \sup_{x\in[0,1]} \left\{ \left| x^2 - \frac{nx^2}{n-1} - \frac{x}{n-1} \right| \right\} = |\alpha| \sup_{x\in[0,1]} \left\{ \left| -\frac{x^2}{n-1} - \frac{x}{n-1} \right| \right\} \\
&= |\alpha| \sup_{x\in[0,1]} \left\{ \frac{x^2 + x}{n-1} \right\} \overset{x=1}{=} \frac{2\,|\alpha|}{n-1} \xrightarrow{n\to\infty} 0.
\end{aligned} \tag{A.39}$$

■

Proof. (Weierstrass Approximation Theorem) Consider the sequence $(B_n)_{n\in\mathbb{N}}$ of Bernstein operators, which are a Korovkin sequence on $\mathcal{C}[0,1]$, cf. Theorem A.13. Let $f \in \mathcal{C}[0,1]$ and $\epsilon > 0$, then by Theorem A.5 there exists an $N(\epsilon) \in \mathbb{N}$ s.t. for all $n \geq N(\epsilon)$ we have $d_\infty(B_n f, f) < \epsilon$. Noting that $B_n f \in \mathcal{P}_n \subset \mathcal{P}$ completes the proof. ■

Example A.14. Let $f : \mathcal{C}[0,\pi] \to \mathcal{C}[0,\pi], x \mapsto \sin x$. To approximate this function arbitrarily well, we will consider $(B_n f)_{n\in\mathbb{N}}$, where $(B_n)_{n\in\mathbb{N}}$ is the sequence of Bernstein operators. Since in Def. A.10 we considered the B_n to be defined on $\mathcal{C}[0,1]$, we need an affine-linear transformation to map $[0,\pi]$ to $[0,1]$. For this, consider $x \mapsto x/\pi$. And since we considered f to be defined on $[0,1]$ in Def. A.10, consider the mapping $x \mapsto \pi x$ to get to $[0,\pi]$. With this mapping, the Bernstein operators B_n are defined on $\mathcal{C}[0,\pi]$ as follows:

$$(B_n f)(x) = \sum_{j=0}^{n} \sin\left(\pi \frac{j}{n}\right) \beta_j^{(n)}(x/\pi) = \frac{1}{\pi^j} \sum_{j=0}^{n} \sin\left(\pi \frac{j}{n}\right) \binom{n}{j} x^j (1 - x/\pi)^{n-j}. \tag{A.40}$$

For different $n \in \mathbb{N}$, Fig. 2 shows some approximating polynomials $B_n f \in \mathcal{P}_n$.

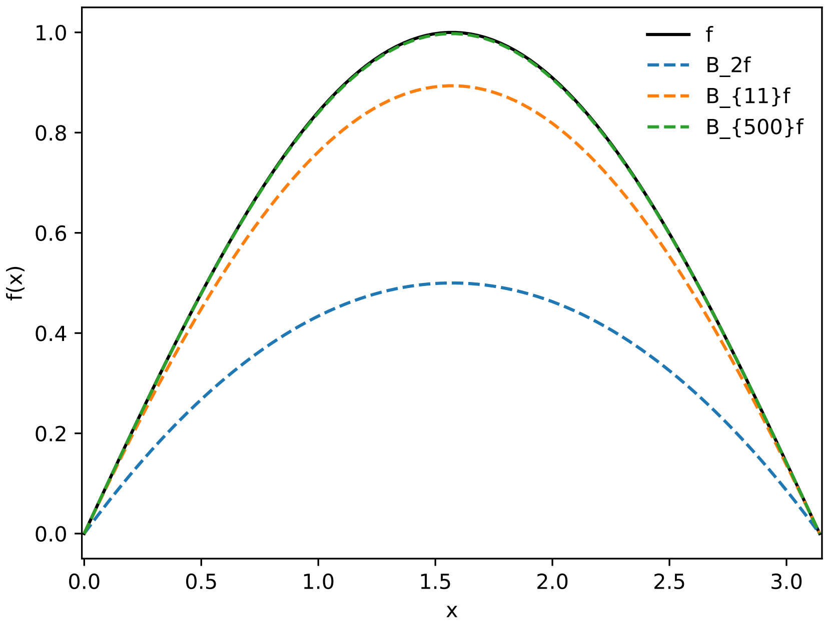


Figure 2: Approximation of $sin(x)$ by linear combination of Bernstein polynomials on $[a, b]$.

B Construction of $\mathbb{R}$

We will follow Ref. [41] to construct the real numbers $\mathbb{R}$. There are two possibilities to do so:

1. Dedekind completion: Every non-empty subset has a least upper bound (wrt $\leq$).
2. Cauchy completion: Every Cauchy sequence converges wrt $|\cdot|$.

We will use the second definition of completion.

Definition B.1. The (archimedean) absolute value of $\mathbb{Q}$ is the function

$$|\cdot| : \mathbb{Q} \to \mathbb{Q}_{\geq 0}, |x| := \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases} \tag{B.1}$$

Definition B.2. A *Cauchy sequence* of rational numbers is a sequence $(x_n)_{n\in\mathbb{N}}$ s.t. for every $\epsilon \in \mathbb{Q}_{>0}$ there exists an $N = N(\epsilon) \in \mathbb{N}$ s.t. for all m, $n \geq N$, $|x_m - x_n| < \epsilon$.

Theorem B.3. Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be two Cauchy sequences in $\mathbb{Q}$, then their sum $(x_n + y_n)_{n\in\mathbb{N}}$ is also Cauchy.

Proof. First, we note that the sum of two rational numbers is also rational. Since both $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are Cauchy sequences, for all $\epsilon \in \mathbb{Q}_{>0}$ there is an $N' \in \mathbb{N}$ s.t. for all $m', n' \geq N$, we have $|x_{m'} - x_{n'}| < \epsilon$ and an $\tilde{N} \in \mathbb{N}$ s.t. for all $\tilde{m}, \tilde{n} \geq \tilde{N}$, we have $|y_{\tilde{m}} - y_{\tilde{n}}| < \epsilon$. Let $N := \max\{N', \tilde{N}\}$, then for all $m, n \geq N$, we have

$$|x_m + y_m - (x_n + y_n)| = |x_m - x_n + y_m - y_n| \leq |x_m - x_n| + |y_m - y_n| < 2\epsilon,$$

which proves that $(x_n + y_n)_{n\in\mathbb{N}}$ is also a Cauchy sequence. ■

Theorem B.4. Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be two Cauchy sequences in $\mathbb{Q}$, then their product $(x_n \cdot y_n)_{n\in\mathbb{N}}$ is also Cauchy.

Proof. First, we note that the multiplication and division of two rational numbers is rational. By Theorem 4.58, Cauchy sequences are bounded, hence choose $M \in \mathbb{Q}$ s.t. $|x_n| \leq M$ and $|y_n| \leq M$ for all $n \in \mathbb{N}$, and note that for any $\epsilon \in \mathbb{Q}_{>0}$ there exists $N' \in \mathbb{N}$ s.t. for all $m', n' \geq N'$, we have $|x_{m'} - x_{n'}| < \epsilon/M$ and an $\tilde{N} \in \mathbb{N}$ s.t. for all $\tilde{m}, \tilde{n} \geq \tilde{N}$, $|y_{\tilde{m}} - y_{\tilde{n}}| < \epsilon/M$. Let $N := \max\{N', \tilde{N}\}$, then for all $m, n \geq N$, we have

$$\begin{aligned} |x_m y_m - x_n y_n| &= |x_m y_m + x_m y_n - x_m y_n - x_n y_n| = |x_m (y_m - y_n) + y_n (x_m - x_n)| \\ &\leq |x_m(y_m - y_n)| + |y_n(x_m - x_n)| = |x_m| \, |y_m - y_n| + |y_n| \, |x_m - x_n| \\ &\leq M \cdot \epsilon/M + M \cdot \epsilon/M = 2\epsilon, \end{aligned}$$

where we used that $|xy| = |x| \, |y|$. ■

Remark B.5. The set of Cauchy sequences in $\mathbb{Q}$ forms a ring, cf. Def. 1.12, where $0 = (0)_{n\in\mathbb{N}}$ and $1 = (1)_{n\in\mathbb{N}}$ are the neutral elements for the additive and multiplicative operation respectively. Since multiplication is commutative, we have a commutative ring.

Remark B.6. The set of Cauchy sequences in $\mathbb{Q}$ does not form a field, cf. Def. 1.13, though, since many Cauchy sequences do not have a multiplicative inverse, such as $(1, 0, 0, 0, \dots)$. Thus, the set of Cauchy sequences in $\mathbb{Q}$ does not form a field.

Definition B.7. We say that a rational Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ is *equivalent to zero* if $\lim_{n\to\infty}|x_n| = 0$. We say that two rational Cauchy sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are rational, denoted by $(x_n)_{n\in\mathbb{N}} \sim (y_n)_{n\in\mathbb{N}}$, if $(x_n - y_n)_{n\in\mathbb{N}}$ is equivalent to 0.

Theorem B.8. The operation $\sim$ is an equivalence relation on the set of all Cauchy sequences in $\mathbb{Q}$.

Proof. Note that any rational Cauchy sequence is equivalent to itself, since $(x_n - x_n)_{n\in\mathbb{N}} = (0)_{n\in\mathbb{N}}$, and $\lim_{n\to\infty}|0| = 0$. We also have symmetry, since $|x_n - y_n| = |y_n - x_n|$. Finally, if $(x_n)_{n\in\mathbb{N}} \sim (y_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}} \sim (z_n)_{n\in\mathbb{N}}$, then $(x_n)_{n\in\mathbb{N}} \sim (z_n)_{n\in\mathbb{N}}$, since $|x_n - z_n| = |x_n - y_n + y_n - z_n| \le |x_n - y_n| + |y_n - z_n|$, thus the operation $\sim$ is also transitive.

Thus, according to Def. 6.2, $\sim$ forms an equivalence relation. ■

Definition B.9 (Real numbers)**.** We define $\mathbb{R} := \mathbb{Q}/{\sim}$. We extend the absolute value of $\mathbb{Q}$ to $\mathbb{R}$ via

$$\left|\left[(x_n)_{n\in\mathbb{N}}\right]\right| := \left[(|x_n|)_{n\in\mathbb{N}}\right], \tag{B.2}$$

where $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{Q}$.

Remark B.10. We embed $\mathbb{Q}$ in $\mathbb{R}$ via the map $x \mapsto \left[(x)_{n\in\mathbb{N}}\right]$, i.e. the equivalence class of a constant sequence, whose elements are always x. This map is injective. It is standard to abuse notation and write x for $\left[(x)_{n\in\mathbb{N}}\right]$. That way, we can see $\mathbb{Q}$ as a subset of $\mathbb{R}$.

Before proving that $\mathbb{Q}$ is dense in $\mathbb{R}$ and that $\mathbb{R}$ is a field, we need to look at the algebraic structure of $\mathbb{R}$, and at what it means to say $\left[(a_n)_{n\in\mathbb{N}}\right] < \left[(b_n)_{n\in\mathbb{N}}\right]$.

Definition B.11. Let s, $t \in \mathbb{R}$, i.e. there are rational Cauchy sequences $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ s.t. $s = \left[(s_n)_{n\in\mathbb{N}}\right]$ and $t = \left[(t_n)_{n\in\mathbb{N}}\right]$.

a) Define the map $+ : \mathbb{R}\times\mathbb{R} \to \mathbb{R}, s + t := \left[(s_n + t_n)_{n\in\mathbb{N}}\right]$,

b) and define $\cdot : \mathbb{R}\times\mathbb{R} \to \mathbb{R}, s \cdot t := \left[(s_n \cdot t_n)_{n\in\mathbb{N}}\right]$.

Remark B.12. Since we are dealing with equivalence classes, and $(s_n)_{n\in\mathbb{N}}$ is merely a representation of $\left[(s_n)_{n\in\mathbb{N}}\right]$, we need to verify that the above definitions are well-defined, i.e. representation-independent.

Theorem B.13. The addition and multiplication of equivalence classes as defined in Def. B.11 are well-defined.

Proof. Let $s, t \in \mathbb{R}$, i.e. there exist rational Cauchy sequences $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ s.t. $s = \left[(s_n)_{n\in\mathbb{N}}\right]$ and $t = \left[(t_n)_{n\in\mathbb{N}}\right]$. Now let $(s'_n)_{n\in\mathbb{N}}, (t'_n)_{n\in\mathbb{N}}$ be rational Cauchy sequences s.t. $(s_n)_{n\in\mathbb{N}} \sim (s'_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}} \sim (t'_n)_{n\in\mathbb{N}}$, i.e. $\left[(s_n)_{n\in\mathbb{N}}\right] = \left[(s'_n)_{n\in\mathbb{N}}\right]$ and $\left[(t_n)_{n\in\mathbb{N}}\right] = \left[(t'_n)_{n\in\mathbb{N}}\right]$.

Since

$$|s'_n + t'_n - (s_n + t_n)| \le |s'_n - s_n| + |t'_n - t_n| \overset{n\to\infty}{\longrightarrow} 0,$$

we have $(s'_n + t_n)_{n\in\mathbb{N}} \sim (s_n + t_n)_{n\in\mathbb{N}}$, and thus $\left[(s'_n + t'_n)_{n\in\mathbb{N}}\right] = \left[(s_n + t_n)_{n\in\mathbb{N}}\right]$, which shows the well-definedness of $+$.

For the multiplication $\cdot$, note that

$$\begin{aligned}|s_n' t_n' - s_n t_n| &= |s_n' t_n' + t_n' s_n - t_n' s_n - s_n t_n| = |t_n' (s_n' - s_n) + s_n (t_n' - t_n)| \\ &\leq |t_n'| \, |s_n' - s_n| + |s_n| \, |t_n' - t_n| \\ &\leq M \, |s_n' - s_n| + M \, |t_n' - t_n| \xrightarrow{n\to\infty} 0,\end{aligned}$$

where we proceeded as in the proof of Theorem B.4 and chose an $M \in \mathbb{Q}$ s.t. $|t_n'| \leq M$ and $|s_n| \leq M$ for all $n \in \mathbb{N}$, cf. Theorem 4.58. Thus, we have shown that $(s_n' t_n')_{n\in\mathbb{N}} \sim (s_n t_n)_{n\in\mathbb{N}}$, and hence $\left[(s_n' t_n')_{n\in\mathbb{N}}\right] = \left[(s_n t_n)_{n\in\mathbb{N}}\right]$, showing the well-definedness of $\cdot$. ∎

Theorem B.14. $\mathbb{R}$, equipped with the two operations $+$ and $\cdot$, is a field.

Proof. Here, we will only prove that every $s \in \mathbb{R}\backslash\{0\}$ has a multiplicative inverse (the rest is genereally easier to prove than this one).

We shall first understand what the existence of the multiplicative inverse means. Let $(s_n)_{n\in\mathbb{N}}$ be a rational Cauchy sequence s.t. $s = \left[(s_n)_{n\in\mathbb{N}}\right]$. Since $s \neq 0$, we know that $(s_n)_{n\in\mathbb{N}} \not\sim (0)_{n\in\mathbb{N}}$, i.e. $(s_n)_{n\in\mathbb{N}}$ does not converge to 0. We now need to find a multiplicative inverse t s.t. $s \cdot t = 1$. While this might seem trivial (since every $s_n \in \mathbb{Q}$ has a multiplicative inverse, i.e. we could choose $t_n = s_n^{-1}$), there is a subtle difficulty. The fact that s is non-zero does *not* mean that all sequence elements of $(s_n)_{n\in\mathbb{N}}$ are unequal to zero. For example, let $s = 1$, then $(s_n)_{n\in\mathbb{N}} = (0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, \dots)$ is in the equivalence class of $[(1)_{n\in\mathbb{N}}]$. However, the point is that *eventually*, the sequence elements must be non-zero (since s is non-zero), and thus there is an $N \in \mathbb{N}$ s.t. for $n \geq N$, we have $s_n \neq 0$. Thus, define a sequence $(t_n)_{n\in\mathbb{N}}$ that for $n < N$ is zero and for $n \geq N$ let $t_n := s_n^{-1}$. Now the multiplicative inverse of s is $t = \left[(t_n)_{n\in\mathbb{N}}\right]$. ∎

Definition B.15 (Order of $\mathbb{R}$)**.** Let $s \in \mathbb{R}$, i.e. there is a rational Cauchy sequence $(s_n)_{n\in\mathbb{N}}$ s.t. $\left[(s_n)_{n\in\mathbb{N}}\right] = s$. We say that s is *positive*, denoted by $s > 0$, if $s \neq 0$ and if there is an $N \in \mathbb{N}$ s.t. $s_n > 0$ for all $n \geq N$.

Let $t \in \mathbb{R}$, i.e. there is a rational Cauchy sequence $(t_n)_{n\in\mathbb{N}}$ s.t. $\left[(t_n)_{n\in\mathbb{N}}\right] = t$. Then we write $s > t$ if $s - t$ is positive.

Theorem B.16. The order of the real numbers is well-defined.

Proof. Fix $s \in \mathbb{R}$, i.e. $s = \left[(s_n)_{n\in\mathbb{N}}\right]$ for a rational Cauchy sequence $(s_n)_{n\in\mathbb{N}}$. Also, let s be positive, i.e. there is an $N \in \mathbb{N}$ s.t. $s_n > 0$ for all $n \geq N$. We need to show that for any other representative of $\left[(s_n)_{n\in\mathbb{N}}\right]$, i.e. $(s_n')_{n\in\mathbb{N}} \in \left[(s_n)_{n\in\mathbb{N}}\right]$ with $(s_n')_{n\in\mathbb{N}} \sim (s_n)_{n\in\mathbb{N}}$, there is an $N' \in \mathbb{N}$ s.t. $s_n' > 0$ for all $n \geq N'$.

We first show that there exists a rational number $r \in \mathbb{Q}_{>0}$ and an $\tilde{N} \in \mathbb{N}$ s.t. $s_n > r$ for all $n \geq \tilde{N}$ by contradiction. Suppose that for all $r \in \mathbb{Q}_{>0}$ and all $\tilde{N} \in \mathbb{N}$ there exists an $n = n(r, \tilde{N}) \geq \tilde{N}$ s.t. $s_n \leq r$. Then there is a subsequence $(s_{n_k})_{k\in\mathbb{N}}$ of $(s_n)_{n\in\mathbb{N}}$ s.t. $0 < s_{n_k} \leq 1/k$. We will prove this via induction. Set $n_1 := n(1, N)$. Assume we have constructed $n_1, n_2, \dots, n_k$ s.t. $n_1 < n_2 < \cdots < n_k$ and $0 < s_{n_j} \leq 1/j$ for $j \in \{1, \dots, k\}$. Then set $n_{k+1} := n(1/(k+1), n_k + 1)$. We have $n_k < n_{k+1}$ and $0 < s_{n_{k+1}} \leq 1/(k+1)$.

Since $(s_n)_{n\in\mathbb{N}}$ is Cauchy, there exists an $M_k \in \mathbb{N}$ s.t. $|s_n - s_m| < 1/k$ for all $n, m \geq M_k$. Choose $l \geq k$ s.t. $n_l \geq M_k$. Then for $n \geq M_k$, we have

$$|s_n| = |s_n - s_{n_l} + s_{n_l}| \leq |s_n - s_{n_l}| + |s_{n_l}| < \frac{1}{k} + \frac{1}{l} \leq \frac{2}{k} \xrightarrow{k\to\infty} 0. \tag{B.3}$$

This means that $(s_n)_{n\in\mathbb{N}}$ converges to 0, and thus $s = 0$, which is a contradiction. Hence, there exists a rational number $r \in \mathbb{Q}_{>0}$ and an $\tilde{N} \in \mathbb{N}$ s.t. $s_n > r$ for all $n \geq \tilde{N}$.

Let $(s'_n)_{n\in\mathbb{N}}$ be another representative of $\left[(s_n)_{n\in\mathbb{N}}\right]$, i.e. $(s_n)_{n\in\mathbb{N}} \sim (s'_n)_{n\in\mathbb{N}}$, and thus $s'_n - s_n \overset{n\to\infty}{\longrightarrow} 0$. This in turn implies that $|s'_n - s_n| < r$ for $n \geq N'$. Let $n \geq \max\{\tilde{N}, N'\}$, then we have

$$s'_n = s_n - (s_n - s'_n) \geq s_n - |s_n - s'_n| > r - r = 0,$$

since $s_n - s'_n \leq |s_n - s'_n|$, which implies $-(s_n - s'_n) \geq -|s_n - s'_n|$. ∎

Theorem B.17. Let $s, t, r \in \mathbb{R}$, then if $s > t$, we also have $s + r > t + r$.

Proof. Let $(s_n)_{n\in\mathbb{N}}, (t_n)_{n\in\mathbb{N}}, (r_n)_{n\in\mathbb{N}}$ be rational Cauchy sequences s.t. $s = \left[(s_n)_{n\in\mathbb{N}}\right]$, $t = \left[(t_n)_{n\in\mathbb{N}}\right]$ and $r = \left[(r_n)_{n\in\mathbb{N}}\right]$. Since $s > t$, we know that there is an $N \in \mathbb{N}$ s.t. for all $n \geq N$, we have

$$s_n - t_n > 0 \Leftrightarrow s_n > t_n \Leftrightarrow s_n + r_n > t_n + r_n \Leftrightarrow \underbrace{(s_n + r_n) - (t_n + r_n)}_{= s_n - t_n} > 0$$

Since $s_n - t_n$ does not go to 0 for $n \to \infty$ (since $s > t$), we have that $(s_n + r_n) - (t_n + r_n)$ does not either, and thus $s + r > t + r$. ∎

The density of $\mathbb{Q}$ in $\mathbb{R}$ follows almost immediately from the construction of $\mathbb{R}$ from $\mathbb{Q}$.

Theorem B.18. $\mathbb{Q}$ is dense in $\mathbb{R}$, i.e. for any $\epsilon \in \mathbb{Q}_{>0}$ and $s \in \mathbb{R}$ there is a rational number $r \in \mathbb{Q}$ s.t. $|s - r| < \epsilon$.

Proof. Since $s \in \mathbb{R}$, there is a rational Cauchy sequence $(s_n)_{n\in\mathbb{N}}$ s.t. $s = \left[(s_n)_{n\in\mathbb{N}}\right]$. Since $(s_n)_{n\in\mathbb{N}}$ is Cauchy, there is an $N \in \mathbb{N}$ s.t. for all $m, n \geq N$, we have $|s_n - s_m| < \epsilon$. Let $r := s_N \in \mathbb{R}$, and note that we can embed it into $\mathbb{R}$ as $\left[(a_N)_{n\in\mathbb{N}}\right]$. Thus, $|s - r| = \left[(|s_n - s_N|)_{n\in\mathbb{N}}\right]$. For $n \geq N$, we have $|s_n - s_N| < \epsilon$ and thus $s_n - s_N < \epsilon$ and $s_N - s_n < \epsilon$, since for any rational number a, $a \leq |a|$. Thus, $(s_n - s_N) - \epsilon$ and $(s_N - s_n) - \epsilon$ are negative, i.e. $s - r < \epsilon$ and $r - s < \epsilon$, which can be summarized to $|s - r_{\text{¸}}| < \epsilon$. ∎

Remark B.19. The density of $\mathbb{Q}$ in $\mathbb{R}$ implies that we could replace $\epsilon \in \mathbb{Q}_{>0}$ with $\epsilon \in \mathbb{R}_{>0}$ throughout.

Theorem B.20. $\mathbb{R}$ has the Archimedean property, i.e. for all $s, t \in \mathbb{R}$ there exists a natural number $m \in \mathbb{N}$ s.t. $m \cdot s > t$.

Proof. Since $s, t \in \mathbb{R}$, there exist rational Cauchy sequences $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ s.t. $s = \left[(s_n)_{n\in\mathbb{N}}\right]$ and $t = \left[(t_n)_{n\in\mathbb{N}}\right]$.

Recall that by $m \in \mathbb{N} \subset \mathbb{Q}$, we mean the m that is embedded into $\mathbb{R}$, i.e. $\left[(m)_{n\in\mathbb{N}}\right]$. Thus, we need to show that there exists an $m, N \in \mathbb{N}$ s.t. $ms_n - t_n > 0$ for all $n \geq N$ and that $(ms_n - t_n)_{n\in\mathbb{N}}$ does not converge to 0, cf. Definition B.15.

We will prove the first part by contradiction. Assume that for all $m, N \in \mathbb{N}$, there exists an $n \geq N$ s.t. $ms_n - t_n \leq 0$. For this, note that $(t_n)_{n\in\mathbb{N}}$ is bounded, cf. Theorem 4.58, i.e. there is an $M \in \mathbb{Q}$ s.t. $t_n \leq M$ for all $n \in \mathbb{N}$. By the Archimedean property of $\mathbb{Q}$, for any $\epsilon \in \mathbb{Q}_{>0}$ there is an $m \in \mathbb{N}$ s.t. $m\epsilon/2 > M$, or equivalently $M/m < \epsilon/2$. Thus,

$$ms_n - t_n \leq 0 \Leftrightarrow s_n \leq \frac{t_n}{m} \leq \frac{M}{m} < \frac{\epsilon}{2}. \tag{B.4}$$

Since $(s_n)_{n\in\mathbb{N}}$ is Cauchy, there exists an $N \in \mathbb{N}$ s.t. for all $n', k \geq N$, $|s_k - s_{n'}| < \epsilon/2$, i.e. $s_k - s_{n'} < \epsilon/2$. Now choose an $n' \geq n \geq N$, then for all $k \geq N$, we have

$$s_k < s_{n'} + \frac{\epsilon}{2} \overset{\text{(B.4)}}{<} \epsilon.$$

Thus, s cannot be positive, contradicting the assumption that $s > 0$. Hence, we have shown that there exists an $m, N \in \mathbb{N}$ s.t. $ms_n - t_n > 0$ for all $n \geq N$.

If $(ms_n - t_n)_{n\in\mathbb{N}}$ converges to 0, e.g. for $(s_n)_{n\in\mathbb{N}} = (1)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}} = (m)_{n\in\mathbb{N}}$, then we can choose $m+1$ instead of m, since $(m+1)s_n - t_n = ms_n - t_n + s_n \overset{n\to\infty}{\longrightarrow} \lim_{n\to\infty} s_n > 0$, since $s \neq 0$. ■

We can now prove that $\mathbb{R}$ is complete.

Proposition B.21. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence of *real numbers*, then it converges to an $x \in \mathbb{R}$, i.e. $\mathbb{R}$ is complete.

Proof. Fix an x_n of the sequence. By the density of $\mathbb{Q}$ in $\mathbb{R}$, cf. Theorem B.18, we know there exists a $q_n \in \mathbb{Q}$ s.t. $|x_n - q_n| < 1/n$. We will now show that $(q_n)_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{Q}$. Let $\epsilon \in \mathbb{Q}_{>0}$, then by the Archimedean property of $\mathbb{Q}$, there exists an $N \in \mathbb{N}$ s.t. $N\epsilon/3 > 1$, or equivalently $1/N < \epsilon/3$. Since $(x_n)_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{R}$, there exists an $M \in \mathbb{R}$ s.t. for all $n, m \geq M$, we have $|x_n - x_m| < \epsilon/3$. Now, if $n, m \geq \max\{N, M\}$, we have

$$\begin{aligned}|q_n - q_m| &= |q_n - x_n + x_n - x_m + x_m - q_m| \leq |q_n - x_n| + |x_n - x_m| + |x_m - q_m| \\ &< \frac{1}{n} + \frac{\epsilon}{3} + \frac{1}{m} \leq \frac{2}{N} + \frac{\epsilon}{3} < \epsilon,\end{aligned}$$

which proves that $(q_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, and thus it represents a real number, which we shall denote by x. We now want to show that $(x_n)_{n\in\mathbb{N}}$ converges to x. First of all, by definition of x, $q_n - x \overset{n\to\infty}{\longrightarrow} 0$. We also know that $|x_n - q_n| < 1/n$, cf. above. Thus,

$$|x_n - x| = |x_n - q_n + q_n - x| \leq |x_n - q_n| + |q_n - x| < \frac{1}{n} + |q_n - x| \overset{n\to\infty}{\longrightarrow} 0.$$

■

Remark B.22. For the proof of the uniqueness of the completion of $\mathbb{Q}$, we refer to step 5 of the proof of Proposition 4.86.

References

[1] Wikipedia. accessed on 16.05.2025. URL: https://en.wikipedia.org/wiki/AM%E2%80%93GM_inequality#Proof_by_induction_#1.

[2] Srivatsan (https://math.stackexchange.com/users/13425/srivatsan). *Proving :* $(1+\frac{1}{n+1})^{n+1} \geq (1 + \frac{1}{n})^n$. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/64864 (version: 2011-12-10). eprint: https://math.stackexchange.com/q/64864. URL: https://math.stackexchange.com/q/64864.

[3] George Law (https://math.stackexchange.com/users/141584/george-law). *A function which is continuous with respect to some metric but not continuous with respect to some other metric.* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/208 (version: 2017-01-07). eprint: https://math.stackexchange.com/q/2087322. URL: https://math.stackexchange.com/q/2087322.

[4] Kavi Rama Murthy (https://math.stackexchange.com/users/142385/kavi-rama-murthy). *Is there any metric on R with which it is incomplete.* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/3789234 (version: 2020-08-13). eprint: https://math.stackexchange.com/q/3789234. URL: https://math.stackexchange.com/q/3789234.

[5] Conrado Costa (https://math.stackexchange.com/users/226425/conrado-costa). *Strongly equivalent metrics are equivalent.* Mathematics Stack Exchange. URL:https://math.stackexchange (version: 2015-07-30). eprint: https://math.stackexchange.com/q/1379634. URL: https://math.stackexchange.com/q/1379634.

[6] Mark Bennet (https://math.stackexchange.com/users/2906/mark-bennet). *Show that a bounded, monotone increasing sequence is a Cauchy sequence.* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/2169936 (version: 2017-03-03). eprint: https://math.stackexchange.com/q/2169936. URL: https://math.stackexchange.com/q/2169936.

[7] Ink (https://math.stackexchange.com/users/34881/ink). *The preimage of continuous function on a closed set is closed.* Mathematics Stack Exchange. URL:https://math.stackexcha (version: 2013-07-07). eprint: https://math.stackexchange.com/q/437837. URL: https://math.stackexchange.com/q/437837.

[8] CodeKingPlusPlus (https://math.stackexchange.com/users/38879/codekingplusplus). *Prove: If a sequence converges, then every subsequence converges to the same limit.* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/213285 (version: 2012-10-13). eprint: https://math.stackexchange.com/q/213285. URL: https://math.stackexchange.com/q/213285.

[9] Learner (https://math.stackexchange.com/users/48763/learner). *Convergence of sequences in topological spaces.* Mathematics Stack Exchange. URL:https://math.stackexchange.com (version: 2022-06-30). eprint: https://math.stackexchange.com/q/289740. URL: https://math.stackexchange.com/q/289740.

[10] Bruno Andrades (https://math.stackexchange.com/users/578498/bruno-andrades). *Topological Equivalence of Metrics Does Not Imply Strong Equivalence.* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/5065146 (version: 2025-05-14). eprint: https://math.stackexchange.com/q/5065146. URL: https://math.stackexchange.com/q/5065146.

[11] Dome (https://math.stackexchange.com/users/70644/dome). *Cauchy sequence is convergent iff it has a convergent subsequence*. Mathematics Stack Exchange. URL:https://math.st (version: 2014-03-24). eprint: `https://math.stackexchange.com/q/354965`. URL: `https://math.stackexchange.com/q/354965`.

[12] Thomas Andrews (https://math.stackexchange.com/users/7933/thomas-andrews). *Proof that* $(1+1/n)^n$ *is Cauchy*. Mathematics Stack Exchange. URL:https://math.stackexchange.c (version: 2025-05-15). eprint: `https://math.stackexchange.com/q/5065619`. URL: `https://math.stackexchange.com/q/5065619`.

[13] Haha (https://math.stackexchange.com/users/94689/haha). *is bounded linear operator necessarily continuous?* Mathematics Stack Exchange. URL:https://math.stackexchange.com (version: 2024-03-16). eprint: `https://math.stackexchange.com/q/556667`. URL: `https://math.stackexchange.com/q/556667`.

[14] Zilin J. (https://math.stackexchange.com/users/94898/zilin-j). *Determinant of a block lower triangular matrix*. Mathematics Stack Exchange. URL:https://math.stackexchange.con (version: 2015-08-04). eprint: `https://math.stackexchange.com/q/1221066`. URL: `https://math.stackexchange.com/q/1221066`.

[15] Davide Giraudo (https://math.stackexchange.com/users/9849/davide-giraudo). *Finite measures are* σ*-finite*. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q (version: 2015-11-23). eprint: `https://math.stackexchange.com/q/144687`. URL: `https://math.stackexchange.com/q/144687`.

[16] *5.1 Roots and Radicals*. URL: `https://saylordotorg.github.io/text_intermediate-algebra/s08-01-roots-and-radicals.html`.

[17] Khan Academy. *Upper triangular determinant*. URL: `https://www.khanacademy.org/math/linear-algebra/matrix-transformations/determinant-depth/v/linear-algebra-upper-triangular-determinant`.

[18] Michael M. Bronstein et al. *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*. 2021. arXiv: `2104.13478 [cs.LG]`.

[19] *Chapter 1. Sigma-Algebras*. URL: `https://www.math.lsu.edu/~sengupta/7312s02/sigmaalg.pdf`.

[20] Wikipedia contributors. *Determinant*. 2021. URL: `https://en.wikipedia.org/wiki/Determinant`.

[21] Wikipedia contributors. *Equivalence of metrics*. 2020. URL: `https://en.wikipedia.org/wiki/Equivalence_of_metrics`.

[22] Wikipedia contributors. *Equivalence relation*. 2021. URL: `https://en.wikipedia.org/wiki/Equivalence_relation`.

[23] Wikipedia contributors. *Jacobian matrix and determinant*. 2021. URL: `https://en.wikipedia.org/wiki/Jacobian_matrix_and_determinant`.

[24] Jeff Diller. *Continuous Functions and Open Sets*. 2014. URL: `https://www3.nd.edu/~jdiller/teaching/archive/fall14_20850/continuity.pdf`.

[25] *ECE 5630 Lecture Notes*. `https://people.ece.cornell.edu/zivg/ECE_5630_Lectures10.pdf`. Accessed on 11/26/23.

[26] *ECE 5630 Lecture Notes 6*. `https://people.ece.cornell.edu/zivg/ECE_5630_Lectures6.pdf`. Accessed on 11/26/23.

[27] *Equivalence class*. online. accessed on 05/28/25. URL: https://en.wikipedia.org/wiki/Equivalence_class.

[28] *February 18, 2021*. online. accessed on 05/23/25. URL: https://ocw.mit.edu/courses/18-102-introduction-to-functional-analysis-spring-2021/22bf2f774fb16177603249l568148707_MIT18_102s21_lec2.pdf.

[29] Darij Grinberg. *An introduction to the algebra of rings and fields. Text for Math 332 Winter 2023 and Winter 2025 at Drexel University*. accessed on 06/02/25. 2025. URL: https://www.cip.ifi.lmu.de/~grinberg/t/23wa/23wa.pdf.

[30] Ralf Holtkamp. *Mathematik I fur Studierende der Geophysik/Ozeanographie, Meteorologie und Physik. Vorlesungsskript*. 2021. URL: https://www.math.uni-hamburg.de/home/holtkamp/mfp1/mfp1rh.pdf.

[31] Ralf Holtkamp. *Mathematik III für Studierende der Physik Vorlesungsskript*. 2021. URL: https://www.math.uni-hamburg.de/home/holtkamp/mfp3/mfp3rh_druckskript.pdf.

[32] Tianyi Hu. "Bernstein Polynomials and their Application in the Finite Element Method". Bachelor's Thesis.

[33] John K. Hunter. *CHAPTER 3. Measurable functions*. URL: https://www.math.ucdavis.edu/~hunter/measure_theory/measure_notes_ch3.pdf.

[34] Armin Iske. *Approximation Theory and Algorithms for Data Analysis*. 1st ed. Springer Cham, 2018, p. 358. URL: https://link.springer.com/book/10.1007/978-3-030-05228-7.

[35] Armin Iske. "Funktionalanalysis. Universität Hamburg, Sommersemester 2019". unpublished lecture notes. 2019.

[36] Fei-Fei Li, Jiajun Wu, and Ruohan Gao. *Lecture 13: Generative Models*. University lecture. 2022. URL: http://cs231n.stanford.edu/slides/2022/lecture_13_jiajun.pdf (visited on 08/09/2022).

[37] David Lindemann. *Differential geometry. Lecture 1: Smooth manifolds*. 2020. URL: https://www.math.uni-hamburg.de/home/lindemann/material/DG2020L1_slides.pdf.

[38] *Math 396. Completions*. online. accessed on 05/29/25. URL: https://virtualmath1.stanford.edu/~conrad/diffgeomPage/handouts/completion.pdf.

[39] Harkrishan Lal Vasudeva Satish Shirali. *Measure and Integration*. 1st ed., p. 598. DOI: 10.1007/978-3-030-18747-7. URL: https://katalogplus.sub.uni-hamburg.de/vufind/Record/1678681113?rank=1.

[40] Tej Bahadur Singh. *Introduction to Topology*. Springer, Singapore, 2019, p. 452. ISBN: 978-3-662-55406-7. DOI: https://doi.org/10.1007/978-981-13-6954-4.

[41] Andrew Sutherland. *18.095 Lecture Series in Mathematics*. online. accessed on 05/29/25. 2015. URL: https://math.mit.edu/classes/18.095/2015IAP/lecture1/padic.pdf.

[42] Chee Han Tan. *Zorn's Lemma*. URL: https://www.math.utah.edu/~tan/6710_FA/Hahn%20Banach%20Theorem.pdf.

[43] *The Fubini Principle in Discrete Math*. online. accessed on 05/22/25. URL: https://web.math.ucsb.edu/~cmart07/fubini.pdf.

[44] *Uniform continuity*. online. accessed on 05/23/25. URL: https://en.wikipedia.org/wiki/Uniform_continuity#Definition_for_functions_on_metric_spaces.

[45] Stanford University. *Binary Relations*. 2007. URL: https://web.stanford.edu/class/archive/cs/cs103/cs103.1142/lectures/07/Small07.pdf.

[46] Nolan Wallach. *MATH 23b, SPRING 2005. THEORETICAL LINEAR ALGEBRA AND MULTIVARIABLE CALCULUS*. 2005. URL: http://www.math.ucsd.edu/~nwallach/inverse%5B1%5D.pdf.

[47] Lilian Weng. *What are Diffusion Models?* 2021. URL: https://lilianweng.github.io/posts/2021-07-11-diffusion-models/.

[48] Dirk Werner. *Funktionalanalysis*. Springer Spektrum, Berlin, Heidelberg, 2018, p. 586. ISBN: 978-3-662-55406-7. DOI: https://link.springer.com/book/10.1007%2F978-3-662-55407-4.

[49] Miles Wheeler. *4.4 Relative Compactness*. online. accessed on 06/11/2025. URL: https://www.mileshwheeler.com/ma30252/sec-rel.html.