Math 330 - Additional Material

Student edition with proofs

Michael Fochler
Department of Mathematics
Binghamton University

This document contains $\boxed{\text{chapter 14-15}}$ of the Math 330 lecture notes.

Last update: February 17, 2026

Contents

14 Compactness **653**
14.1 ε-Nets and Total Boundedness . . . 653
14.2 Sequence Compactness . . . 661
14.3 Open Coverings and the Heine–Borel Theorem . . . 663
14.4 Continuous Functions and Compact Spaces . . . 667
14.5 Exercises for Ch.14 . . . 671

15 Applications of Zorn's Lemma **672**
15.1 More on Partially Ordered Sets . . . 672
15.2 Existence of Bases in Vector Spaces . . . 673
15.3 The Cardinal Numbers are a totally ordered set . . . 674
15.4 Extensions of Linear Functions in Arbitrary Vector Spaces . . . 675
15.5 The Hahn-Banach Extension Theorem $\boxed{\star}$ . . . 679
15.5.1 Sublinear Functionals . . . 679
15.5.2 The Hahn-Banach extension theorem and its Proof . . . 680
15.6 Convexity $\boxed{\star}$ . . . 684
15.7 Exercises for Ch.15 . . . 687

References **688**

List of Symbols **689**

Index **690**

14 Compactness

Introduction 14.1.

Let us say informally that a family $(U_i)_{i\in I}$ covers or is a cover of a set A if $A \subseteq \bigcup [U_i : i \in I]$.

This chapter will show that **(A)**, **(B)** and **(C)** below are equivalent statements for any subspace $(K, d\big|_{K\times K})$ of a metric space (X, d):

(A) Any sequence in K has a convergent subsequence with limit in K.
(B) K is complete and, given any $\varepsilon > 0$, no matter how small, K can be covered by finitely many ε–neighborhoods.
(C) Any open covering $(U_i)_i$ of K has a finite subcovering: one can find finitely many indices $i_1, \dots, i_n$ such that $K \subseteq U_{i_1} \cup \dots \cup U_{i_n}$

Such metric spaces K will be called "compact". Moreover, we will see that the following is true for the metric space $\mathbb{R}^n$ with the Euclidean metric: For $X = \mathbb{R}^n$, each of **(A)**, **(B)** and **(C)** is equivalent to

(D) K is bounded and closed.

Property **(C)** is the only one that makes sense for abstract topological spaces and will be used to define compactness for such spaces. $\square$

14.1 ε-Nets and Total Boundedness

Introduction 14.2.

We start out with a few elementary observations for subsets of $\mathbb{R}^2$.

(a) Let C be a square with side length $\varepsilon > 0$ and "edge points" $\vec{p}_1, \vec{p}_2, \vec{p}_3, \vec{p}_4$. Then each point in C belongs to one or more of the ε–neighborhoods $N_\varepsilon(\vec{p}_1), \dots, N_\varepsilon(\vec{p}_4)$, i.e., $C \subseteq \bigcup_{j=1}^{4} N_\varepsilon(\vec{p}_j)$.

(b) Let A be a bounded subset of $\mathbb{R}^2$, i.e., [1] there exists $\gamma > 0$ and $\vec{x_0} \in \mathbb{R}^2$ such that $A \subseteq N_\gamma(\vec{x_0})$. Then, if $\varepsilon > 0$, this circle of radius γ can be covered by a finite number of squares with side length ε.

(c) Put **(a)** and **(b)** together: Any bounded set A can be covered by finitely many ε–neighborhoods.

(d) Equivalently, for any bounded set $A \subseteq \mathbb{R}^2$ and $\varepsilon > 0$ there exists a finite set $G \subseteq \mathbb{R}^2$ such that $A \subseteq \bigcup_{g\in G} N_\varepsilon(g)$

An exact proof will be given in Proposition 14.1 below that all of the above is true for all bounded subsets of $\mathbb{R}^n$, for any $n \in \mathbb{N}$. On the other hand, there are metric spaces (X, d) with bounded subsets which do not possess property **(d)**. $\square$

[1] See Proposition **??** on p.**??**.

Here is a simple counterexample to **(d)** of the introduction to this chapter.

Example 14.1.

Let X be an infinite set, furnished with the discrete metric d. Then any (nonempty) subset A of X, including X, is bounded: $\mathrm{diam}(A) = 0$ if $|A| = 1$, and $\mathrm{diam}(A) = 1$ otherwise.

On the other hand, if $\varepsilon \leq 1$, then $N_\varepsilon(x) = \{x\}$ for all $x \in X$. thus X is not the union of a finite number of such ε–neighborhoods. $\square$

Considering this counterexample, it makes sense to give a special name for subsets of metric spaces which satisfy **(d)**. We will call such sets totally bounded and refer to G as an ε–net or ε–grid.

Definition 14.1 (ε–nets)**.**

Let $\varepsilon > 0$. Let (X, d) be a metric space and $A \subseteq X$. let $G \subseteq X$ be a subset of X with the following property:

$$\text{For each } x \in A \text{ there exists } g \in G \text{ such that } x \in N_\varepsilon(g), \quad \text{i.e.,} \bigcup_{g \in G} N_\varepsilon(g) \supseteq A. \tag{14.1}$$

In other words, the points of G form a "grid" or "net" fine enough so that no matter what point x of A you choose, you can always find a "grid point" g with distance less than ε to x, because that is precisely the meaning of $x \in N_\varepsilon(g)$.

We call G an ε–**net** or ε–**grid** for A and we call $g \in G$ a **grid point** of the net. $\square$

The relation $\bigcup\limits_{g \in G} N_\varepsilon(g) \supseteq A$ asserts that the family $(N_\varepsilon(g))_{g \in G}$ is a collection of open sets which "covers" all of A We will later call a family of open sets $(U_i)_i$ which satisfies $\bigcup_i U_i \supseteq A$ an open cover of A.

Definition 14.2 (Total boundedness)**.**

Let (X, d) be a metric space and let A be a subset of X. We say that A is **totally bounded** if, for each $\varepsilon > 0$, there exists a finite(!) ε–grid for A. $\square$

Remark 14.1.

(A) $A \subseteq (X, d)$ is totally bounded if and only if for each $\varepsilon > 0$ there is a finite collection $\mathscr{G}_\varepsilon = \{g_1, \dots g_n\}$ of points in X whose ε–balls $N_\varepsilon(g_j)$ cover A: For any $a \in A$ there is $j = j(a)$ such that $d(a, g_j) < \varepsilon$.

(B) Let $\varepsilon > 0$. Since all sets A satisfy $A = \bigcup\limits_{a \in A} \{a\} \subseteq \bigcup\limits_{a \in A} N_\varepsilon\{a\}$, all finite sets are totally bounded

(C) Note that the definition of total boundedness of a set A does not demand that the gridpoints are elements of A.

This allows us, for example, to accept the set $G := \{(i, j) : i, j \in \mathbb{Z}\}$ of all points in the plane with integer coordinates, as a 1–grid for the set $A := \{\, (i + \pi, j + \pi) : i, j \in \mathbb{Z}\}$. $\square$

Proposition 14.1 (ε–nets in $\mathbb{R}^n$)**.**

> ⋆ *Let (X, d) be $\mathbb{R}^n$ with the Euclidean metric.*
>
> ***(A)*** *Let $\varepsilon > 0$. Then the set*
> $$\varepsilon\mathbb{Z}^n \;=\; \{\varepsilon\vec{z} : \vec{z} \in \mathbb{Z}^n\} \;=\; \{(\varepsilon z_1, \dots, \varepsilon z_n) : z_j \in \mathbb{Z} \text{ for } j = 1, \dots, n\}$$
> *is an $(\varepsilon\sqrt{n})$–net for (any subset of) $\mathbb{R}^n$.*
>
> ***(B)*** *Let A be a bounded set in $\mathbb{R}^n$ and $\varepsilon > 0$. Then there is $k \in \mathbb{N}$ and $g_1, \dots g_k \in \varepsilon\mathbb{Z}^n$ such that*
> $$A \;\subseteq\; N_\varepsilon(g_1) \;\cup\; N_\varepsilon(g_2) \;\cup\; \dots, \;\cup\; N_\varepsilon(g_k),$$
> *i.e., A is covered by finitely many ε–neighborhoods of points in the $(\varepsilon/\sqrt{n})$–grid $\varepsilon\mathbb{Z}^n$.*

PROOF of **(A)**

Let $\vec{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. For $j = 1, \dots, n$, let k_j be the integer such that

$$\varepsilon k_j \;\leq\; x_j \;<\; \varepsilon(k_j + 1)\,. \tag{14.2}$$

It is obvious from (14.2) that $\vec{x} \in C$, the n–dimensional cube of side length ε, defined by

$$C \;:=\; C(\vec{x}) \;:=\; \{\vec{y} = (y_1, \dots, y_n) \in \mathbb{R}^n \;:\; \varepsilon k_j \;\leq\; y_j \;<\; \varepsilon(k_j + 1)\}$$

Note that this is the following cube: [2] Its 2^n edgepoints are the vectors $(e_1, \dots, e_n)$ for which

$$\text{the } j\text{–th coordinate is either } e_j \;=\; \varepsilon k_j\,, \text{ or } e_j \;=\; \varepsilon(k_j + 1)\,. \tag{14.3}$$

Let $\vec{x}^\star = (x_1^\star, \dots, x_n^\star) \in \mathbb{R}^n$ be defined as follows. For $j = 1, \dots, n$, let

$$x_j^\star \;:=\; \begin{cases} \varepsilon k_j & \text{if } x_j \leq \varepsilon(k_j + 1/2)\,, \\ \varepsilon(k_j + 1) & \text{else}\,. \end{cases}$$

Thus $\vec{x}^\star$ is an edge point of the cube $C(\vec{x})$ since each $x_j^\star$ is of the form (14.3). Moreover, since each $x_j^\star$ satisfies $|x_j - x_j^\star| \leq \varepsilon/2$,

$$d(\vec{x}, \vec{x}^\star) \;=\; \sqrt{\sum_{j=1}^n (x_j - x_j^\star)^2} \;\leq\; \sqrt{n \cdot \left(\frac{\varepsilon}{2}\right)^2} \;=\; \frac{\varepsilon\sqrt{n}}{2} \;<\; \varepsilon\sqrt{n}\,. \tag{14.4}$$

We have found for arbitrary $\vec{x} \in \mathbb{R}^n$ a vector $\vec{x}^\star \in \varepsilon\mathbb{R}^n$ such that $\vec{x} \in N_{\varepsilon\sqrt{n}}(\vec{x}^\star)$. [3] This proves that $\varepsilon\mathbb{Z}^n$ is an $\varepsilon\sqrt{n}$–net in $\mathbb{R}^n$.

PROOF of **(B)**

[2] Draw pictures for dimensions $1, 2, 3$ with $\varepsilon = 1$!

[3] Here is an example. If $n = 5, \varepsilon = 1$, and $\vec{x} = (12.85, -12.35, \frac{1}{3}, 9, -\pi)$, then the associated grid point is $\vec{x}^\star = (13, -12, 0, 9, -3)$. The distance is:

$$d(\vec{x}, \vec{x}^\star) \;=\; \sqrt{.15^2 + .35^2 + (1/3)^2 + 0 + (\pi - 3)^2} \;\leq\; \sqrt{1/2 + 1/2 + 1/2 + 0 + 1/2} \;\leq\; \varepsilon \cdot \sqrt{n}.$$

We see that part ***A*** of the lemma is true for this specific example.

Intuitively clear but very messy. Here is an outline.

For convenience, let $\varepsilon' := \varepsilon/\sqrt{n}$.

First we choose a radius R so big that $A \subseteq N_R(\vec{0})$. This is possible according to Proposition **??** on p.**??**. Next we choose $M \in \mathbb{N}$ which is so big that $M\varepsilon' > R$. Let $\vec{c}^{(1)}, \vec{c}^{(2)}, \dots, \vec{c}^{(2^n)}$ be the 2^n points in $\mathbb{R}^n$ for which each coordinate is either $M\varepsilon'$ or $-M\varepsilon'$. Those are the edge points of the n–dimensional cube

$$C \ := \ \{\vec{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \,:\, |x_j| \leq M\varepsilon \text{ for all } j = 1, \dots, n\}\,.$$

Since there are only $2M+1$ integers m such that $|m| \leq M$, there are only finitely many points $\vec{y} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ which belong to both C and $\varepsilon\mathbb{Z}^n$, i.e., such that each coordinate y_j is of the form $y_j = \varepsilon' m$ for some integer m which satisfies

$$-\varepsilon' M \ \leq \ y_j \ \leq \ \varepsilon' M\,, \quad \text{i.e.,} \quad -M \ \leq \ m \ \leq \ M\,.$$

Those points $\vec{y}$ are the edge points of $(2M)^n$ n–dimensional cubes $C_1, C_2, \dots, C_{(2M)^n}$ of side length ε' whose union equals C.

We have seen in the proof of **(A)** that each C_i is covered by the $\sqrt{n}\cdot\varepsilon'$–neighborhoods of its 2^n edge points. Since $\sqrt{n}\cdot\varepsilon' = \varepsilon$, each C_i is covered by a finite number of ε–neighborhoods. Since

$$C \ = \ \bigcup \left[\, C_j \,:\, j = 1, \dots, (2M)^n \right]\,,$$

C also is covered by a finite number of ε–neighborhoods. Since $A \subseteq C$, it follows that A is covered by a finite number of ε–neighborhoods. ■

Note for part **(A)** that$\varepsilon\mathbb{Z}^n$ is as intuitive a grid as you can think of, especially if you look at the 2–dimensional plane or 3–dimensional space and consider $\varepsilon = 1$.

Theorem 14.1.

> *Bounded subsets of $\mathbb{R}^n$ are totally bounded.*

PROOF: This is immediate from Proposition14.1**(B)**. ■

We have seen that all bounded subsets in $\mathbb{R}^n$ are totally bounded. In the remainder of this subchapter we will see that

- the reverse is true in any metric space: totally bounded subsets are always bounded,
- a subset A of a metric space is totally bounded if and only if all sequences in A possess subsequences which are Cauchy.

We start by proving the following:

(a) All sequences in a totally bounded set possess a subsequence which is Cauchy.
(b) All sequences in a bounded subset of $\mathbb{R}^n$ possess a subsequence which is Cauchy.

We saw earlier in this subchapter that all bounded subsets in $\mathbb{R}^n$ are totally bounded. Thus a proof of **(a)** also is a proof of **(b)**. We will prove **(b)** anyway and do so first, since this proof is easier to visualize than that of **(a)**

Proposition 14.2.

> *Let $A \subseteq \mathbb{R}^n$ be bounded and let $(x_n)_n$ be a sequence such that $x_n \in A$ for all n. Then there exists a subsequence x_{n_j} which is Cauchy.*

PROOF:

We will construct a sequence of sets $A = A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_j \downarrow 0$ and such that each A_j contains infinitely members of $(x_n)_n$. This allows us to find indices $n_0 < n_1 < \dots$ such that $x_{n_j} \in A_j$ for all j. It then follows from Proposition **??** in the chapter on completeness in metric spaces that $\left(x_{n_j}\right)_j$ is Cauchy.

Step 0: Since A_0 is bounded, there exists an n–dimensional cube C_0 such that $A_0 \subseteq C_0$. Let γ be the side length, δ_0 the diameter, and let $\vec{a}^{(0)} = (a_1^{(0)}, \dots, a_n^{(0)})$ be the center of that cube.

The following is a very generous estimate of δ_0. Let $\vec{x} = (x_1, \dots, x_n) \in C_0$. Since $\vec{a}^{(0)}$ is the center of C_0, $|x_j - a_j^{(0)}| \leq \gamma/2$ for each j. Thus

$$d(\vec{x}, \vec{a}^{(0)}) \;=\; \sqrt{\sum_{j=1}^{n} (x_j - a_j^{(0)})^2} \;\leq\; \sqrt{n \cdot \left(\frac{\gamma}{2}\right)^2} \;=\; \frac{\gamma\sqrt{n}}{2} \;<\; \gamma\sqrt{n}\,. \tag{14.5}$$

It follows from $A_0 \subseteq C_0$ that

$$\delta_0 \;\leq\; \operatorname{diam}(C_0) \;\leq\; \gamma\sqrt{n}\,. \tag{14.6}$$

Step 1: We subdivide C_0 into 2^n cubes $C_{1,1}, \dots C_{1,2^n}$ of side length $\gamma/2 = 2^{-1}\gamma$.

It follows from $A_0 \subseteq C_0$ that $A_0 = (C_{1,1} \cap A_0) \cup \dots \cup (C_{1,2^n} \cap A_0)$. Since $(x_n)_n$ is an infinite sequence, there is at least one index $m_1 \in [1, 2^n]_{\mathbb{Z}}$ such that $C_{1,m_1} \cap A_0$ possesses infinitely many members

$$x_{1,1} := x_{n_1},\; x_{1,2} := x_{n_2},\; x_{1,3} := x_{n_3},\; \dots$$

of that sequence. Let $A_1 := C_{1,m_1} \cap A_0$ and $\delta_1 := \operatorname{diam}(A_1)$. Then (14.6) yields

$$\delta_1 \;\leq\; \operatorname{diam}(C_{1,m_1}) \;=\; \frac{1}{2} \cdot \operatorname{diam}(C_0) \;=\; \frac{\gamma\sqrt{n}}{2^1}\,.$$

Since $x_n \in A$ for all n, we also have $x_{1,n} \in A$ for all n. Thus the infinite subsequence $(x_{1,n})_n$ of the original sequence $(x_n)_n$ lives in the subset A_1 of A with diameter $\leq (\gamma\sqrt{n}) \cdot 2^{-1}$.

Step k: Assume that we have obtained sets $A_0 \supseteq A_1 \supseteq \dots \supseteq A_k$ such that, for each $j \in [1, k]_{\mathbb{Z}}$, the following holds true:

- **(1)** $A_j = C_{j,m_j} \cap A_{j-1}$ for a suitable cube C_{j,m_j} of side length $2^{-j}\gamma$
- **(2)** $\delta_j := \operatorname{diam}(A_j) \leq \frac{\gamma\sqrt{n}}{2^j}$, [4]
- **(3)** A_j contains an infinite subsequence $(x_{j,n})_n = x_{j,1}, x_{j,2}, \dots$ of the original sequence $(x_n)_n$.

Step $k+1$: We subdivide the cube C_{k,m_k} into 2^n cubes $C_{k+1,1}, \dots C_{k+1,2^n}$ of side length

[4] (this actually follows from **(1)** by means of a computation like the one done in (14.5)),

$\gamma/2 = 2^{-(k+1)}\gamma$.

It follows from $A_k \subseteq C_{k,m_k}$ that $A_k = (C_{k+1,1} \cap A_k) \cup \dots \cup (C_{k+1,2^n} \cap A_k)$.

Since $(x_{k,n})_n$ is an infinite sequence, there is at least one index $m_{k+1} \in [1, 2^n]_{\mathbb{Z}}$ such that $C_{k+1,m_{k+1}} \cap A_k$ possesses infinitely many members

$$x_{k+1,1} := x_{k,n_1},\ x_{k+1,2} := x_{k,n_2},\ x_{k+1,3} := x_{k,n_3},\ \dots$$

of that subsequence.

Let $A_{k+1} := C_{k+1,m_{k+1}} \cap A_k$ and $\delta_{k+1} :=$ diam(A_{k+1}). Then (14.6) yields

$$\delta_{k+1} \leq \operatorname{diam}(C_{k+1,m_{k+1}}) = \frac{1}{2} \cdot \operatorname{diam}(C_{k,m_k}) = \frac{\gamma\sqrt{n}}{2^{k+1}}.$$

Since $x_n \in A$ for all n, we also have $x_{k+1,n} \in A$ for all n, thus We have an infinite subsequence $(x_{k+1,n})_n$ of the original sequence x_n which lives in the subset A_{k+1} of A with diameter $\leq (\gamma\sqrt{n}) \cdot 2^{-1}$.

To summarize, we have achieved what we set out to do at the beginning of the proof: We have constructed a sequence of sets $A = A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_n \downarrow 0$ such that each A_j contains infinitely members of the sequence $(x_n)_n$.

- Since A_1 contains infinitely members of $(x_n)_n$, there exists $n_1 > n_0$ such that $x_{n_1} \in A_1$.
- Since A_2 contains infinitely members of $(x_n)_n$, there exists $n_2 > n_1$ such that $x_{n_2} \in A_2$.
  -
- Since A_k contains infinitely members of $(x_n)_n$, there exists $n_k > n_{k-1}$ such that $x_{n_k} \in A_k$.

Let $z_k := x_{n_k}$. Then $(z_k)_k$ is a subsequence of $(x_n)_n$ such that $z_k \in A_k$. Since $\lim\limits_{k\to\infty} \delta_k = 0$, it follows from Proposition **??** on p.**??** that $\left(x_{n_j}\right)_j$ is Cauchy. ■

The proof that allows us to extend Proposition 14.2 to totally bounded sets in arbitrary metric spaces is almost identical to the one given above. The main difference is that we no longer can subdivide the set A_k into 2^n subsets of smaller diameters from which to choose A_{k+1}. Rather, we must work directly with the definition of total boundedness to obtain A_k.

Theorem 14.2.

> *Let A be a totally bounded subset of a metric space (X, d). Let $(x_n)_n$ be a sequence such that $x_n \in A$ for all n. Then there exists a subsequence x_{n_j} which is Cauchy.*

PROOF:

Let $A_0 := A$. As in the proof of Proposition 14.2, we will construct $A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_n :=$ diam$(A_n) \downarrow 0$ such that each A_j contains infinitely members of $(x_n)_n$.

Step 1: Let $\delta_1 := 2^{-1} = \frac{1}{2}$. Since A_0 is totally bounded, there exists a finite grid of length δ_1, i.e., there is a finite set $G_1 = \{g_{1,1}, g_{1,2}, \dots, g_{1,M_1}\}$ such that $A_0 \subset \bigcup [N_{\delta_1}(g_{1,j}) : j = 1, \dots, M_1]$.

Since $A_0 = \bigcup [N_{\delta_1}(g_{1,j}) \cap A_0 : j = 1, \dots, M_1]$ and A_0 contains the entire (infinite) sequence $(x_n)_n$, there exists $m_1 \in [1, M_1]_{\mathbb{Z}}$ such that $N_{\delta_1}(g_{1,m_1}) \cap A_0$ contains an infinite subsequence

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

of that sequence.

Let $A_1 := A_0 \cap N_{\delta_1}(g_{1,m_1})$. Then $\mathrm{diam}(A_1) \leq \mathrm{diam}(N_{\delta_1}(g_{1,k_1}))$, i.e., $\mathrm{diam}(A_1) \leq \delta_1$.

Step 2: Let $\delta_2 := 2^{-2}$. Since A_0 is totally bounded, there exists a finite grid of length δ_2, i.e., there exists a finite set $G_2 = \{g_{2,1}, g_{2,2}, \dots, g_{2,M_2}\}$ such that $A_0 \subset \bigcup [N_{\delta_2}(g_{2,j}) : j = 1, \dots, M_2]$.

Since $A_1 \subseteq A_0$, it follows that $A_1 = \bigcup [N_{\delta_2}(g_{2,j}) \cap A_1 : j = 1, \dots, M_2]$. Since A_1 contains the infinite sequence $(x_{1,n})_n$, there exists $m_2 \in [1, M_2]_{\mathbb{Z}}$ such that $N_{\delta_2}(g_{2,m_2}) \cap A_1$ contains an infinite subsequence

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

of that subsequence.

Let $A_2 := A_1 \cap N_{\delta_2}(g_{2,m_2})$. Then $\mathrm{diam}(A_2) \leq \mathrm{diam}(N_{\delta_2}(g_{2,m_2}))$, i.e., $\mathrm{diam}(A_2) \leq \delta_2$.

Step k: Assume that we have obtained sets $A_0 \supseteq A_1 \supseteq \cdots \supseteq A_k$ such that, for each $j \in [1, k]_{\mathbb{Z}}$, the following holds true:

- **(1)** Let $\delta_j := 2^{-j}$. Then $A_j = N_{\delta_j}(g_{j,m_j}) \cap A_{j-1}$ for a suitable $g_{j,m_j} \in X$.
- **(2)** A_j contains an infinite subsequence $(x_{j,n})_n = x_{j,1}, x_{j,2}, \dots$ of the original sequence $(x_n)_n$.

Step $k+1$: Let $\delta_{k+1} := 2^{-(k+1)}$. Since A_0 is totally bounded, there exists a finite grid of length δ_{k+1}, i.e., there exists a finite set $G_{k+1} = \{g_{k+1,1}, g_{k+1,2}, \dots, g_{k+1,M_{k+1}}\}$ such that
$A_0 \subset \bigcup \left[N_{\delta_{k+1}}(g_{k+1,j}) : j = 1, \dots, M_{k+1}\right]$.
Since $A_k \subseteq A_0$, it follows that $A_k = \bigcup \left[N_{\delta_{k+1}}(g_{k+1,j}) \cap A_k : j = 1, \dots, M_{k+1}\right]$. Since A_k contains the infinite sequence $(x_{k,n})_n$, there exists $m_{k+1} \in [1, M_2]_{\mathbb{Z}}$ such that $N_{\delta_{k+1}}(g_{k+1,m_{k+1}}) \cap A_k$ contains an infinite subsequence

$$x_{k+1,1} := x_{k,n_1},\ x_{k+1,2} := x_{k,n_2},\ x_{k+1,3} := x_{k,n_3},\ \dots$$

of that subsequence.

Let $A_{k+1} := A_k \cap N_{\delta_{k+1}}(g_{k+1,m_{k+1}})$. Then $\mathrm{diam}(A_{k+1}) \leq \mathrm{diam}\big(N_{\delta_{k+1}}(g_{k+1,m_{k+1}})\big)$. In other words, $\mathrm{diam}(A_{k+1}) \leq \delta_{k+1}$.

Since $x_n \in A$ for all n, we also have $x_{k+1,n} \in A$ for all n. This infinite subsequence $(x_{k+1,n})_n$ of the original sequence x_n lives in the subset A_{k+1} of A with diameter $\leq \delta_{k+1} = 2^{-k+1}$.

To summarize, we have achieved what we set out to do at the beginning of the proof: We have constructed a sequence of sets $A = A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$ with diameters $\delta_j \downarrow 0$ such that each A_j contains infinitely members of the sequence $(x_n)_n$.

- Since A_1 contains infinitely members of $(x_n)_n$, there exists $n_1 > n_0$ such that $x_{n_1} \in A_1$.
- Since A_2 contains infinitely members of $(x_n)_n$, there exists $n_2 > n_1$ such that $x_{n_2} \in A_2$.

  -
- Since A_k contains infinitely members of $(x_n)_n$, there exists $n_k > n_{k-1}$ such that $x_{n_k} \in A_k$.

Let $z_k := x_{n_k}$. Then $(z_k)_k$ is a subsequence of $(x_n)_n$ such that $z_k \in A_k$. Since $\lim\limits_{k\to\infty} \delta_k = 0$, it follows from Proposition **??** on p.**??** that $\left(x_{n_j}\right)_j$ is Cauchy. ■

Proposition 14.3.

> *Totally bounded subsets of metric spaces are bounded.*

PROOF:
Let A be a subset of a metric space (X, d). We show the contrapositive: We assume that A is not bounded, i.e., $\operatorname{diam}(A) = \infty$, and we will show that A is not totally bounded.

We may assume that A is not empty because otherwise there is nothing to prove.

Step 1: We prove by induction that there exists a sequence $x_n \in A$ such that $d(x_i, x_j) \geq 1$ for any $i \neq j$.

Base case: Let $x_0 \in A$. Since A is not bounded, there exists $x_1 \in A$ such that $r_1 := d(x_0, x_1) \geq 1$.

Induction step: We assume that n elements $x_1, \dots x_n$ such that $d(x_i, x_j) \geq 1$ for any $1 \leq i < j \leq n$ have aready been chosen. Let

$$\beta \ := \ \max\{d(x_0, x_j) : j \leq n\}\,, \qquad r \ := \ \beta + 1\,.$$

Since A is not bounded, we can pick $x_{n+1} \in A \setminus N_r(x_0)$. We obtain for each $j \in [1, n]_{\mathbb{Z}}$,

$$\beta + 1 \ \leq d(x_{n+1}, x_0) \leq d(x_{n+1}, x_j) + d(x_j, x_0) \ \leq \ d(x_{n+1}, x_j) + \beta, \quad \text{i.e.,} \ \ 1 \ \leq \ d(x_{n+1}, x_j).$$

We have constructed a sequence (x_n) for which any two items have distance no less than 1.

Step 2: It follows that $(x_n)_n$ does not possess a Cauchy subsequence. According to Theorem 14.2, all sequences in totally bounded sets have Cauchy subsequences. It follows that A is not totally bounded. ■

Corollary 14.1.

> *If $A \subseteq \mathbb{R}^n$, then $\quad A$ is bounded $\ \Leftrightarrow\ $ A is totally bounded.*

PROOF: This is immediate from Proposition 14.3 above and Theorem 14.1 on p.656.

Next we prove the reverse of Theorem 14.2 on p.658.

Theorem 14.3.

> *Let A be a subset of a metric space (X, d) such that each sequence in A contains a Cauchy subsequence. Then A is totally bounded.*

PROOF: [5] We prove the contrapositive: If a set is not totally bounded, then there exists a sequence without any Cauchy subsequences.

So assume that A is not totally bounded. Thus there is $\varepsilon > 0$ such that the following holds for any $n \in \mathbb{N}$: If $z_1, z_2, \dots z_n \in A$ then the union $\bigcup_{1 \leq j \leq n} N_\varepsilon(z_j)$ does not cover A: There exists $z \in A$ outside each one of those ε–neighborhoods, i.e., $z \in A \setminus \bigcup\ \big[N_\varepsilon(z_j) : 1 \leq j \leq n\ \big]$.

[5] The proof is similar to that of Proposition 14.3.

This allows us to create an infinite sequence $(x_j)_{j\in\mathbb{N}}$ such that $d(x_j, x_n) \geq \varepsilon$ for all $j, n \in \mathbb{N}$ such that $j \neq n$, say, $j < n$, as follows: We pick

$$x_1 \in A; \quad x_2 \in A \setminus N_\varepsilon(x_1); \quad x_3 \in A \setminus \big(N_\varepsilon(x_1) \cup N_\varepsilon(x_2)\big); \ \dots \quad x_n \in A \setminus \bigcup_{j<n} N_\varepsilon(x_j); \ \dots$$

Note that $x_n \in A \setminus \bigcup_{j<n} N_\varepsilon(x_j)$ implies $d(x_j, x_n) \geq \varepsilon$ for all indices $j < n$. Since this is true for arbitrary $n \in \mathbb{N}$, it is true that

$$d(x_i, x_j) \ \geq \ \varepsilon \quad \text{for all } i, j \in \mathbb{N} \text{ such that } i \neq j\,.$$

If $(x_{n_j})_j$ were a Cauchy subsequence of $(x_n)_n$ then there would be $n_0 \in \mathbb{N}$ such that

$$(\mathbf{A}) \qquad d(x_i, x_j) \ < \ \varepsilon \quad \text{for all } i, j \in \mathbb{N} \text{ such that } i, j \geq n_0\,.$$

But the x_n were constructed such that $d(x_m, x_k) \geq \varepsilon$ for **all** $m \neq k$, in particular for $m := n_i$ and $k := n_j$ if $i \neq j$. Since $n_i \neq n_j$ whenever $i \neq j$, it is not possible to construct a subsequence $(x_{n_j})_j$ which satisfies **(A)**. We have shown that sets which are not totally bounded possess sequences without any Cauchy subsequences. ■

Corollary 14.2.

Let A be a subset of a metric space (X, d). Then,

A is totally bounded $\Leftrightarrow$ every sequence in A possesses a Cauchy subsequence.

PROOF: This follows from Theorem 14.2 on p.658 and Theorem 14.3 above. ■

14.2 Sequence Compactness

We saw that totally bounded sets are those where every sequence possesses a Cauchy subsequence. Since all Cauchy sequences converge but the opposite usually is not true, total boundedness of a set A should be a weaker property than the following:

any sequence in A possesses a convergent subsequence.

In this subchapter we will examine sets with that property.

Definition 14.3 (Sequence compactness)**.**

Let (X, d) be a metric space and let $A \subseteq X$. We say that A is **sequence compact** or **sequentially compact** if it has the following property: Given any sequence (a_n) of elements of A, there exists $a \in A$ and a subset

$$n_1 < n_2 < \ldots < n_j < \ldots \quad \text{of indices such that} \quad a = \lim_{j\to\infty} a_{n_j},$$

In other words, there exists a subsequence [6](a_{n_j}) which converges to a. □

[6] See Definition ?? on p.??.

Remark 14.2.

It is important that you understand that is is not sufficient that $\lim\limits_{j\to\infty} a_{n_j} \in X$. Rather, we demand that this limit belongs to the smaller set A!

For example, the open unit interval $]0,1[$ is totally bounded as a bounded subset of $\mathbb{R} = \mathbb{R}^1$. (Can you prove this directly?) But this set is not sequence compact, since the sequence $x_n := \frac{1}{n}$ does not possess a convergent subsequence: Any such convergent subsequence would have limit zero, and zero is not an element of $]0,1[$. $\square$

Proposition 14.4 (Sequence compactness implies total boundedness)**.**

> *Let (X,d) be a metric space and let A be a sequentially compact subset of X.*
> *Then A is totally bounded.*

PROOF: Let $(x_n)_n$ be a sequence in A. Since convergent subsequences are Cauchy, there exists a Cauchy subsequence. It follows from Corollary 14.2 on p.661 that A is totally bounded. $\blacksquare$

Proposition 14.5 (Sequence compact implies completeness)**.**

> *Let (X,d) be a metric space and let A be a sequence compact subset of X.*
> *Then A is complete, i.e., any Cauchy sequence (x_{n_j}) in A converges to a limit $L \in A$.*

PROOF: Let (x_n) be a Cauchy sequence in A. Because A is sequence compact, we can extract a subsequence $z_j := x_{n_j}$ and find $L \in A$ such that $z_j \to L$ as $j \to \infty$. It follows from prop.**??** on p.**??** that the entire Cauchy sequence (x_n) converges to L. $\blacksquare$

The last two propositions have proved that any sequence compact set in a metric space is both totally bounded and complete. As the next theorem shows, the reverse is also true.

Theorem 14.4 (Sequence compact $\Leftrightarrow$ totally bounded and complete)**.**

> *Let A be a subset of a metric space (X,d).*
> *Then, A is sequence compact if and only if A is totally bounded and complete.*

PROOF: We have already seen in prop.14.4 on p.662 and prop.14.5 on p.662 that if A is sequentially compact then A is totally bounded and complete. We now show the other direction.

Let A be totally bounded and complete and let $(x_n)_n$ be a sequence in A. Since A is totally bounded, we can extract a Cauchy subsequence $(x_{n_j})_j$. Since A is complete, this subsequence has a limit $L \in A$. Thus $(x_n)_n$ is a convergent subsequence of $(x_n)_n$. $\blacksquare$

Theorem 14.5 (Sequence compact sets are closed and bounded)**.**

> *Let A be sequence compact subset of a metric space (X, d). Then A is a bounded and closed set.*

PROOF: Sequence compact spaces are totally bounded and complete by Theorem 14.4 on p.662.
Since they are totally bounded, they also are bounded by Proposition 14.3 on p.659.
Since they are complete, they also are closed by Theorem **??** on p.**??**. ■

Remark 14.3.
We obtain from the results of this and the previous subchapter the following:

> A subset of a metric space is sequentially compact
> $\Leftrightarrow$ it is totally bounded and complete
> $\Rightarrow$ it is bounded and closed. □

In subsets of $\mathbb{R}^n$ the last implication of Remark14.3 becomes an equivalence:

Theorem 14.6.

> *A subset of $\mathbb{R}^n$ is sequentially compact*
> $\Leftrightarrow$ *it is totally bounded and complete*
> $\Leftrightarrow$ *it is bounded and closed.*

PROOF:
Let $A \subseteq \mathbb{R}^n$. It suffices to prove that A is totally bounded and complete if A is bounded and closed since, as we mentioned in Remark14.3, everything else has already been established.
So assume that A is bounded and closed. Since A is closed and $\mathbb{R}^n$ is complete, A is complete by Theorem **??** on p.**??**. Since A is bounded, A also is totally bounded by Theorem 14.1. Thus A is both totally bounded and complete. ■

14.3 Open Coverings and the Heine–Borel Theorem

We now discuss families of open sets called "open coverings". You should review the concept of an indexed family and how it differs from that of a set (see (**??**) on p.**??**).

Definition 14.4 (Coverings and open coverings)**.**

> Let X be an arbitrary nonempty set and $A \subseteq X$. Let $U_i \in X$ $(i \in I)$ such that $A \subseteq \bigcup_{i \in I} U_i$. We call such a family a **covering** of A.

A **finite subcovering** of a covering $(U_i)_{i\in I}$ of the set A is a finite collection

(14.7) $$U_{i_1}, \dots, U_{i_n} \quad (i_j \in I \quad \text{for } 1 \leq j \leq n) \quad \text{such that} \quad A \subseteq U_{i_1} \cup U_{i_2} \cup \ldots \cup U_{i_n}.$$

Assume in addition that X is a topological space, e.g., a normed vector space or a metric space. If all members U_i are open then we call $(U_i)_{i\in I}$ an **open covering** of A.

We also write **cover**, **finite subcover**, **open cover** instead of covering, finite subcovering, open covering □

Remark 14.4.

(a) Partitions [7] are coverings.
(b) Formula 14.1 (ε–nets definition, p.654) tells us the following:
if G is an ε–grid for $A \subseteq (X,d)$, then $\{N_\varepsilon(g)\}_{g\in G}$ is an open covering of A.
(c) If $(U_i)_{i\in I}$ is a covering of A then $(U_i \cap A)_{i\in I}$ is a covering of A which satisfies

(14.8) $$\bigcup_{j\in I}(U_j \cap A) \;=\; A. \;\square$$

Definition 14.5 (Compact sets)**.**

Let $(X,\mathfrak{U})$ be a topological space and $K \subseteq X$.

• We call K **compact**, if K possesses the **"extract finite open subcovering" property**:

Given any **open** covering $(U_i)_{i\in I}$ of K, one can extract a finite subcovering. In other words, there is $n \in \mathbb{N}$ and indices

$$i_1, i_2, \dots, i_n \in I \qquad \text{such that} \quad A \subseteq \bigcup_{j=1}^{n} U_{i_j}. \;\square$$

Remark 14.5.

(a) An open covering for the entire space X is a collection of open sets $(U_i)_{i\in I}$ such that $X = \bigcup[U_i : i \in I]$.
(b) Any subcovering of an open covering necessarily consists exclusively of open sets, i.e., it is again an open covering of A.
(c) Let (X,d) be a metric space. Then

$$K \subseteq (X,d) \text{ is compact if and only if the metric subspace } (K, d\big|_{K\times K}) \text{ is compact,}$$

i.e., for any collection of subsets $(U_i)_{i\in I}$ of K which are open in K there exist finitely many indices $i_1,\dots,i_n \in I$ such that $K = U_{i_1} \cup \dots \cup U_{i_n}$. This is true because the open subsets of (K,d) are the traces in K of sets which are open in (X,d) (see Definition **??** on p.**??**). □

[7] see Definition **??** on p.**??**

Example 14.2. Here are some simple examples.

(a) Any finite topological space is compact.

(b) Any topological space that only contains finitely many open sets is compact. In particular a set with the indiscrete topology [8] is compact

(c) A space with the discrete metric [9] is compact if and only if it is finite.

And here is a counterexample.

> The open interval $]0,1[$ with the Euclidean metric is not compact because it is not possible to extract a finite covering from the open covering $\left(]\frac{1}{n},1[\right)_{n\in\mathbb{N}}$. ☐

Example 14.3. for sequence compactness in metric spaces we have the following results which correspond to the previous example.

(a) Any finite metric space is sequence compact.

(b) Any metric space that only contains finitely many open sets is sequence compact. [10]

(c) A space with the discrete metric is sequence compact if and only if it is finite.

The counterexample also fits in:

> The open interval $]0,1[$ with the Euclidean metric is not sequence compact because it is not possible to extract a convergent subsequence from the sequence $x_n := 1/n$ (the limit zero does not belong to $]0,1[$). ☐

We will now see that the correspondence in the above two examples is not a coincidence. The next two theorems show that (subspaces of) metric spaces are compact if and only if they are sequentially compact.

Theorem 14.7 (Compact metric spaces are sequence compact)**.**

> *Let (X,d) be a compact metric space. Then X is sequence compact.*

PROOF: We assume to the contrary that X is compact and that there is a sequence $(x_n)_n$ in X from which one cannot extract a convergent subsequence.

Let $F := \{x \in X : x = x_j \text{ for some } j \in \mathbb{N}\}$ [11] be the set of distinct(!) members of $(x_n)_n$. Let $z \in X$. There exists an open neighborhood U_z of z such that $U_z \cap F$ is finite, because otherwise one could construct a subsequence $(x_{j_m})_m$ of $(x_n)_n$ which converges to z. (See exercise **??** on p.**??**).

It follows from $\{z\} \subseteq U_z$ that $(U_z)_{z\in X}$ is an open covering of X. X is compact, thus we can extract a finite subcovering $U_{z_1}, U_{z_2}, \dots, U_{z_k}$.

$$F = F \cap X = F \cap \bigcup_{j=1}^{k} U_{z_j} = \bigcup_{j=1}^{k} (U_{z_j} \cap F)$$

[8] See Definition **??** on p.**??**

[9] See Definition **??** on p.**??**

[10] We had to remove the example of the indiscrete topology because this topology does not come from a metric.

[11] We could have written more concisely $F := \{x_j : j \in \mathbb{N}\}$ but the above definition was chosen to remind you that F does not contain any duplicates. Note that F can be very small even if there are infinitely many indices j: If $x_j = (-1)^j$ then $F = \{-1, 1\}$ only contains two elements!

is a finite union of the finite sets $U_{z_j} \cap F \subseteq U_{z_j}$ and thus finite. We conclude that the entire sequence $(x_n)_n$ consists of only finitely many distinct members.

But then at least one of those members, say $x_{k^\star}$, will appear infinitely often in that sequence: there is $k_1 < k_2 < \dots$ such that $x_{k_1} = x_{k_2} = \dots = x_{k^\star}$. This constant subsequence converges (to $x_{k^\star}$). We have reached a contradiction. ■

The next proposition does most of the work in establishing the reverse direction: Sequence compactness implies compactness.

Proposition 14.6.

Let (X,d) be a sequence compact metric space. Let $(U_i)_{i \in I}$ be an open cover of X. Then, one can find for $(U_i)_{i \in I}$ a number $\rho > 0$ which possesses the following property:

- *For each $x \in X$ there exists $i \in I$ such that $N_\rho(x) \subseteq U_i$.*

PROOF: Assume to the contrary that no such $\rho > 0$ exists. We then can find for any $n \in \mathbb{N}$ some $x_n \in X$ such that $N_{1/n}(x_n)$ is not contained in any of the U_i. X is sequence compact, so there exists $x \in X$ and a subsequence $(x_{n_j})_j$ which converges to x. $(U_i)_{i \in I}$ covers X, so there exists $i_0 \in I$ such that $x \in U_{i_0}$.

$(\star)$ Because U_{i_0} is an open neighborhood of x there exists $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq U_{i_0}$.

$(\star\star)$ Because $(x_{n_j})_j$ converges to x there are infinitely many $j \in \mathbb{N}$ such that $d(x_{n_j}, x) < \varepsilon/2$, hence there is at least one j such that $j > 2/\varepsilon$. It follows from $n_j \geq j$ that $n_j > 2/\varepsilon$, i.e. $1/n_j < \varepsilon/2$.

$(\star\star\star)$ It follows from $d(x_{n_j}, x) < \varepsilon/2$ and lemma **??** on p.**??** that $N_{\varepsilon/2}(x_{n_j}) \subseteq N_{\varepsilon/2+\varepsilon/2}(x) = N_\varepsilon(x)$.

We apply first $(\star\star)$, then $(\star\star\star)$, then $(\star)$ and obtain

$$N_{1/n_j}(x_{n_j}) \subseteq N_{\varepsilon/2}(x_{n_j}) \subseteq N_\varepsilon(x) \subseteq U_{i_0}.$$

But this contradicts our assumption that each x_{n_j} was chosen in such a fashion that $N_{1/n}(x_n)$ is not contained in any of the U_i. ■

Remark 14.6.

Consider The number ρ of Proposition 14.6 above. Then $\lambda = 2\rho$ is called a **Lebesgue number** of $(U_i)_{i \in I}$. In other words, the Lebesgue number is the diameter of the ρ–neighborhoods. Note that if λ is a Lebesgue number of an open cover, then any λ' which satisfies $0 < \lambda' < \lambda$ also is a Lebesgue number. of that same cover. □

We now can prove the converse of thm.14.7.

Theorem 14.8.

Sequence compact metric spaces are compact.

PROOF: Let (X,d) be a sequence compact metric space and let $(U_i)_{i\in I}$ be an open covering of X. According to prop.14.6 there exists $\rho>0$ as follows: For each $x\in X$ there exists $i(x)\in I$ such that $N_\rho(x)\subseteq U_{i(x)}$.

Since X is totally bounded (see thm.14.4 on p.662) there exist finitely many $x_1,\dots,x_k\in X$ such that $\{N_\rho(x_j):j=1,..k\}$ forms an open covering of X. It then follows from $N_\rho(x_j)\subseteq U_{i(x_j)}$ that $U_{i(x_1)},U_{i(x_2)},\dots,U_{i(x_k)}$ also forms an open covering of X. We have extracted a finite subcovering from $(U_i)_{i\in I}$. ■

Theorem 14.9 (Sequence compactness coincides with compactness in metric spaces)**.**

Let (X,d) be a metric space and let A be a subset of X. Then,

$$A \text{ is sequence compact} \Leftrightarrow A \text{ is compact}, \quad \text{i.e.,}$$
$$A \text{ is sequence compact} \Leftrightarrow \text{every open cover of } A \text{ possesses a finite subcover.}$$

PROOF: Theorems 14.7 and 14.8. ■

An easy consequence is the Heine–Borel theorem.

Theorem 14.10 (Heine–Borel Theorem)**.**

A subset of Euclidean space $\mathbb{R}^n$ is compact $\Leftrightarrow$ this set is closed and bounded.

PROOF: We have seen in thm.14.6 on p.663 that closed and bounded subsets of $\mathbb{R}^n$ are sequence compact. Since sequence compact metric spaces are compact (thm.14.8) it follows that closed and bounded subsets of $\mathbb{R}^n$ are compact.

On the other hand, let K be a compact subset of $\mathbb{R}^n$. Then K is sequence compact by Theorem 14.9 on p.667. Thus K is closed and bounded according to Theorem 14.5 on p.663. ■

14.4 Continuous Functions and Compact Spaces

Theorem 14.11 (Closed subsets of compact topological spaces are compact)**.**

Let A be a closed subset of a compact topological space $(X,\mathfrak{U})$. Then A is a compact subspace.

In other words, the open sets

$$\mathfrak{U}_A = \{V\cap A : V\in\mathfrak{U}\}$$

of the subspace $(A,\mathfrak{U}_A)$ possess the "extract finite open subcovering" property of Definition 14.5 on p.664.

PROOF: Let $(U_j)_{j\in J}$ be a family of sets open in A whose union is A. According to Definition **??** on p.**??** there are open sets V_j in X such that $U_j = V_j \cap A$. It follows that $\bigcup\limits_{j\in J} V_j \supseteq A$, hence the family $(V_j)_{j\in J}$, augmented by the (open!) set $X \setminus A$ is an open cover of (X,d).
As X is compact, we can extract finitely many members from that extended family such that they still cover X. If one of them happens to be $X \setminus A$ then we remove it and we still obtain that the remaining ones, say, $V_{i_1}, V_{i_2}, \dots, V_{i_n}$, cover A. But then the traces in A (Definition **??** on p.**??**)

$$U_{i_1} = V_{i_1} \cap A,\ U_{i_2} = V_{i_2} \cap A,\ \dots,\ U_{i_n} = V_{i_n} \cap A$$

of those open sets in X are open in A and hence form an open covering of the subspace $(A, \mathfrak{U}_A)$. We have proved that the given open covering of A has contains a finite subcover of A. ∎

Corollary 14.3 (Closed subsets of compact metric spaces are compact)**.**

Let A be a closed subset of a compact metric space (X,d). Then $(A, d\big|_{A\times A})$ is a compact subspace.

PROOF: Immediate from thm.14.11. ∎

Let $(X,\mathfrak{U})$ and $(Y,\mathfrak{V})$ be topological spaces and $A \subseteq X$. We recall that continuity for functions $f : A \to (Y,\mathfrak{V})$ was defined in Definition **??** on p.**??**.

Theorem 14.12 (Continuous images of compact topological spaces are compact)**.**

Let $(X,\mathfrak{U})$ and $(Y,\mathfrak{V})$ be two topological spaces. and let $f : X \to Y$ be continuous on X.

- *If X is compact then the direct image $f(X)$ is compact.*

In other words, the topological subspace $(f(X), \mathfrak{V}_{f(X)})$ of Y is compact.

PROOF: Let $(V_j)_{j\in J}$ be a family of sets open in Y whose union contains $f(X)$. Let the sets $W_j := V_j \cap f(X)$ be the traces of V_j in $f(X)$. Then the W_j are open in the subspace $(f(X), \mathfrak{V}_{f(X)})$ of Y and they form an open cover of $f(X)$. We note that any open cover of $f(X)$ is obtained in this manner from open sets in Y.
Let $U_j := f^{-1}(V_j)$. Then

$$\bigcup_{j\in J} U_j \ = \ \bigcup_{j\in J} f^{-1}(V_j) \ = \ f^{-1}\Big(\bigcup_{j\in J} V_j\Big) \ \supseteq\ f^{-1}\big(f(X)\big) \ \supseteq\ X. \tag{14.9}$$

The second equation above follows from prop. **??** (f^{-1} is compatible with all basic set ops) on p.**??** and the last one follows from the fact that $f^{-1}\big(f((\Gamma)\big) \supseteq \Gamma$ for any subset Γ of the domain of f (see cor. **??** on p. **??**). The "$\supseteq$" relation follows from the assumption that $\bigcup \Big[V_j : j \in J\Big] \ \supseteq\ f(X)$
According to prop.**??** ("f^{-1}(open) $=$ open" continuity) on p.**??**, each U_j is open as the inverse image of the open set V_j under the continuous function f.

It follows from (14.9) that $(U_j)_{j \in J}$ is an open covering of the compact space X. We can extract a finite subcover $U_{i_1}, U_{i_2}, \dots, U_{i_n}$.

It follows from the interchangeability of unions with direct images (see (**??**) on p.**??**) that

$$\begin{aligned} f(X) &= f(U_{j_1} \cup \cdots \cup U_{j_n}) \;=\; f(U_{j_1}) \cup \cdots \cup f(U_{j_n}) \\ &= f\big(f^{-1}(V_{j_1})\big) \cup \cdots \cup f\big(f^{-1}(V_{j_n})\big) \;\subseteq\; V_{j_1} \cup \cdots \cup V_{j_n}. \end{aligned}$$

The inclusion relation above follows from the fact that $f(f^{-1}(B)) \;=\; B \cap f(X)$ for any subset B of the codomain of f (see prop.**??** on p. **??**).

We have proved that the arbitrary open cover $(V_j)_{j \in J}$ of $f(X)$ contains a finite subcover $V_{j_1}, \dots, V_{j_n}$ and it follows that $f(X)$ is indeed a compact metric subspace of Y. ■

Corollary 14.4 (Continuous images of compact metric spaces are compact)**.**

Let (X, d_1) and (Y, d_2) be two metric spaces. and let $f : X \to Y$ be continuous on X.
If X is compact, then its image $f(X)$ is compact, i.e., the metric subspace $(f(X), d_2)$ of Y is compact.

PROOF: Immediate from thm.14.12 ■

Read the following remark for an easier way to prove the above theorem.

Remark 14.7. We could have proved the last two theorems 14.11 14.12 in the special case of metric spaces more easily using sequence compactness instead of covering compactness, but the following proofs do not generalize to abstract topological spaces.

Alternate proof of cor.14.3 which uses sequence compactness.

Given is a sequence $x_n \in A$. X is compact, hence sequence compact and it follows that there is $x \in X$ and a subsequence $x_{n_j} \in A$ such that x_{n_j} converges to x. It follows from theorem **??** (Sequence criterion for contact points in metric spaces) on p.**??** that x is a contact point of A and hence $x \in \bar{A} = A$. This proves that A is (sequence) compact. ■

Alternate proof of cor.14.4 which uses sequence compactness (outline).

Given a sequence $y_n \in f(X)$ we construct a convergent subsequence y_{n_j} as follows: For each n there is some $x_n \in X$ such that $y_n = f(x_n)$ X is compact, hence sequence compact and it follows that there is $x \in X$ and a subsequence x_{n_j} such that x_{n_j} converges to x. We now use (sequence) continuity of f at x to conclude that $y_{n_j} = f(x_{n_j})$ converges to $f(x) \in f(X)$. ■

Corollary 14.5.

Let $(X, \mathfrak{U})$ be a topological space and (Y, d) a metric space.

- *If X is compact and $f : X \to Y$ is continuous, then f is bounded.*
- *In particular, any continuous function on a closed interval of real numbers is bounded.*

The proof is left as exercise 14.3 (see p.671). ■

Corollary 14.6 (Continuous real–valued functions attain max and min on a compact domain)**.**

Let $(X,\mathfrak{U})$ be a topological space, $A \subseteq X$ a compact subspace and $f : A \to \mathbb{R}$ continuous on A. Then there exist $x_\star, x^\star \in A$ such that

$$f(x_\star) \; = \min_{x\in A} f(x) \quad \textit{and} \quad f(x^\star) \; = \max_{x\in A} f(x).$$

PROOF: It follows from thm.14.12 on p.668 and thm.14.5 on p.663 that $f(A)$ is closed and bounded in $\mathbb{R}$. It follows from exercise **??** on p.**??** that $\min(\big(f(A)\big)$ and $\max\big(f(A)\big)$ exist, i.e., according to the definition of preimages, there exist elements in the domain A of f which are mapped to those two values. ■

The following theorem relates compactness and uniform continuity. [12]

Theorem 14.13 (Uniform continuity on sequence compact spaces)**.**

Let (X,d_1), (Y,d_2) be metric spaces and let A be a compact subset of X. Then,

- *any continuous function $A \to Y$ is uniformly continuous on A.*

PROOF: Let us assume to the contrary that f is continuous but not uniformly continuous and find a contradiction. Because f is not uniformly continuous, there exists $\varepsilon > 0$ such that no $\delta > 0$, however small, will satisfy (**??**) on p.**??** for all pairs x,y such that $d_1(x,y) \; < \delta$. Looking specifically at $\delta := 1/j$ for all $j \in \mathbb{N}$, we can find $x_j, x'_j \in A$ such that

$$d_1(x_j,x'_j) \; < \; \frac{1}{j} \quad \text{but} \quad d_2(f(x_j),f(x'_j)) \; \geq \varepsilon. \tag{14.10}$$

Because A is compact, it is sequence compact. There is a subsequence (x_{j_k}) of the x_j which converges to an element $x \in A$ We have

$$d_1(x'_{j_k},x) \; \leq \; d_1(x'_{j_k},x_{j_k}) + d_1(x_{j_k},x) \; \leq \frac{1}{j_k} + d_1(x_{j_k},x). \tag{14.11}$$

Both right–hand terms converge to zero as $k \to \infty$. This is obvious for $1/j_k$ because $j_k \geq k$ for all k and it is true for $d_1(x_{j_k},x)$ because x_{j_k} converges to x.
It follows from (14.11) that (x'_{j_k}) also converges to x. The (ordinary) continuity of f gives us

$$f(x) \; = \; \lim_{k\to\infty} f(x'_{j_k}) \; = \; \lim_{k\to\infty} f(x_{j_k}).$$

Since $\lim\limits_{k\to\infty} f(x_{j_k}) = f(x)$ and $\lim\limits_{k\to\infty} f(x'_{j_k}) = f(x)$ there exist $N, N' \in \mathbb{N}$ such that

$$d_2(f(x),f(x_{j_k})) \; < \; \frac{\varepsilon}{2} \quad \text{for } k \geq N; \quad d_2(f(x),f(x'_{j_k})) \; < \; \frac{\varepsilon}{2} \quad \text{for } k \geq N'.$$

[12] See Definition **??** on p.**??**.

Both inequalities are true whenever $k \geq \max(N, N')$. It follows for all such k that

$$d_2(f(x_{j_k}), f(x'_{j_k})) \;<\; d_2(f(x_{j_k}), f(x)) + d_2(f(x), f(x'_{j_k})) \;<\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \;=\; \varepsilon$$

and we have a contradiction to (14.10). ■

Corollary 14.7 (Uniform continuity on closed intervals)**.**

Let $a, b \in \mathbb{R}$ such that $a \leq b$.
Any continuous real–valued function on the closed interval $[a, b]$ is uniformly continuous:
For any $\varepsilon > 0$, there exists $\delta > 0$ such that

(14.12) $$|f(x) - f(y)| \;<\; \varepsilon \quad \text{for all } x, y \in [a, b] \text{ such that } |x - y)| < \delta$$

PROOF: This follows from the previous theorem (14.13) because closed intervals $[a, b]$ are closed and bounded sets and, in $\mathbb{R}$, any closed and bounded set is sequence compact . ■

14.5 Exercises for Ch.14

Exercise 14.1. Let $N \in \mathbb{N}$. Let $X := \{x_1, x_2, \dots, x_N\}$ be a finite set with a metric $d(\cdot, \cdot)$ (so (X, d) is a metric space). Prove that X is compact three different ways:

(a) Show sequence compactness to prove that X is compact.
(b) Show that X has the "extract finite open subcovering" property to prove that it is compact.
(c) Show that X is complete and totally bounded to prove that it is compact. □

Hints:
(a) ANY sequence in X possesses a constant subsequence (WHY?)
(b) If $(U_i)_i$ covers X then for each x there exists (at least one) i such that $x \in U_i$ (WHY?) How many of those U_i do you need to cover X if X has only N elements?
(c) Prop.**??** on p.**??** should prove useful.

Exercise 14.2. Prove the following which was used in the proof of thm.14.7 (Compact metric spaces are sequence compact) on p.665: Let (X, d) be a metric space and let $(x_n)_n$ be a sequence in X. Let $F := \{x \in X : x = x_j \text{ for some } j \in \mathbb{N}\}$. Let $z \in X$ be such that any neighborhood U of z contains infinitely many points of F. Then one can extract a subsequence $(x_{n_j})_j$ of $(x_n)_n$ such that $d(x_{n_j}, z) < \frac{1}{j}$. □

Exercise 14.3.

Prove Corollary 14.5 on p.669 of this document:

Let $(X, \mathfrak{U})$ be a topological space, let (Y, d) be a metric space, and let $f : X \to Y$ be continuous. If X is compact then f is bounded.

In particular, if $h : [a, b] \to \mathbb{R}$ is a continuous function on the closed interval $[a, b]$ of real numbers (and both $[a, b]$ and $\mathbb{R}$ carry the Euclidean metric), then h is bounded. □

15 Applications of Zorn's Lemma

15.1 More on Partially Ordered Sets

Some of the following was copied almost literally from [2] Dudley.

Definition 15.1.

Let $(X, \preceq)$ be a POset (partially ordered set), $A \subseteq X$, and $m \in A$.

- m is called **maximal** for A iff there is no $a \in A$ such that $a \neq m$ and $m \preceq a$. m is called a **maximum** of A if $a \in A$ and $a \preceq m$ for all $a \in A$.
- m is called **minimal** for A iff there is no $a \in A$ such that $a \neq m$ and $m \succeq a$. m is called a **minimum** of A if $a \in A$ and $a \succeq m$ for all $a \in A$.

Proposition 15.1 below shows that such a maximum or minimum is unique. Thus, we may write $\max(A)$ for the maximum of A and $\min(A)$ for the minimum of A. □

It will be proved in prop.15.1 below that if A possesses a maximum and/or a minimum then it is unique. Thus we may write $\max(A)$ for the maximum of A and $\min(A)$ for the minimum of A.

Proposition 15.1.

Let $(X, \preceq)$ be a nonempty POset and $A \subseteq X$. If A has a maximum then it is unique.

PROOF: The proof is left as exercise 15.1. ■

Note 15.1 (Notes on maximal elements and maxima)**.**

(a) If $(X, \preceq)$ is not linearly ordered, then its subsets may have many maximal elements. For example, for the trivial partial ordering $x \preceq y$ if and only if $x = y$, every element is maximal. A maximum is a maximal element, but the converse is often not true.
(b) If an ordering is not specified, then we always mean set inclusion.
(c) Let $A \subseteq X$. If $m \in A$ is a maximum of A then this implies that m must be related to all other elements of A. □

For the following example we recall from Definition **??** (Linear orderings) on p.**??** that a chain C in a POset $(X, \preceq)$ is a subset $C \subseteq X$ which is totally ordered, i.e., for any $x, x' \in C$ at least one of $x \preceq x'$ or $x' \preceq x$ is true.

Example 15.1 (Maximal elements, maxima and chains)**.** **(A)** Let X be the collection of all intervals $[a, b]$ of length $b - a \leq 2$ such that $a, b \in \mathbb{R}$ and $a \leq b$. These intervals are partially ordered by inclusion. Any interval of length equal to 2 is a maximal element in $(X, \subseteq)$. There is no maximum.

(B) Let

$$\begin{aligned} A &:= \{\, [\,3+1/n,\, 5-1/n\,] \;:\; n \in \mathbb{N} \}\,, \\ B &:= \{\, [\,4+1/n,\, 5+1/n\,] \;:\; n \in \mathbb{N} \}\,, \\ C &:= \{\, [\,8-1/n,\, 8+1/n\,] \;:\; n \in \mathbb{N} \}\,. \end{aligned}$$

Then A and C are chains in $(X, \subseteq)$, but B is not a chain in $(X, \subseteq)$. $\square$

A hundred years ago the following was seen as extremely controversial by mathematicians who specialize in the foundations of mathematics.

Axiom 15.1 (Zorn's Lemma)**.**

> **Zorn's Lemma**: Let $(X, \preceq)$ be a partially ordered set with the **ZL property**:
>
> Every chain $C \subseteq X$, possesses an upper bound $u \in X$, i.e., $c \preceq u$ for all $c \in C$. **(ZL)**
>
> Then X has a maximal element. $\square$

Remark 15.1.

Zorn's Lemma is an axiom rather than a theorem or a proposition in the following sense: It is impossible to verify its truth or falsehood from the axioms of "a" (meaning there are more than one) "reasonable" axiomatic set theory. In that sense mathematicians are free to accept or reject Zorn's Lemma when building their mathematical theories. Two notes on that remark:

(a) Today the mathematicians who refuse to accept proofs which make use, directly or indirectly, of Zorn's Lemma, are a very small minority.

(b) It can be proven that if one accepts (rejects) Zorn's Lemma as a mathematical tool then this is equivalent to accepting (rejecting) the **Axiom of Choice** which states the following.

Let $\mathscr{A}$ be a collection of nonempty sets and let Ω be a set such that $\bigcup[A : A \in \mathscr{A}] \subseteq \Omega$. Then there exists a choice function on $\mathscr{A}$, i.e., a function $c : \mathscr{A} \to \Omega$ that satisfies $c(A) \in A$ for all $A \in \mathscr{A}$, $c(\cdot)$ picks or chooses an element $c(A)$ for each argument $A \in \mathscr{A}$: See Definition **??** on p.**??**).

(c) Moreover the Axiom of Choice, hence Zorn's Lemma, is equivalent to prop.**??(b)** on p.**??**: For **arbitrary**, not empty sets A, B the following is true. If $\varphi : A \to B$ is surjective then φ has a right inverse, i.e., a function $\psi : B \to A$ such that $\varphi \circ \psi = id_B$. For a proof see the optional Chapter **??** (Right Inverses and the Axiom of Choice). $\square$

We will see now how Zorn's Lemma allows a surprisingly simple proof to the effect that **any** vector space has a basis.

15.2 Existence of Bases in Vector Spaces

The following is thematically a continuation of the material in chapter **??** (Vectors and vector spaces).

We will prove that every vector space, **even if it does not possess a finite subset which spans the entire space**, possesses a basis (see Definition **??** (Basis of a vector space) on p.**??**).

For the remainder of this chapter we assume that V is a vector space and define

$$\mathfrak{B} := \{A \subseteq V : A \text{ is linearly independent } \}. \tag{15.1}$$

Obviously $\mathfrak{B}$ is a partially ordered set with respect to set inclusion. The next lemma allows us to apply Zorn's Lemma. For the definition of a chain, see Definition **??** on p.672.

Lemma 15.1.

Every chain $\mathfrak{C}$ in $(\mathfrak{B}, \subseteq)$ possesses an upper bound.

PROOF: Let $U := \bigcup \left[C : C \in \mathfrak{C} \right]$. We will show that U is linearly independent, i.e., $U \in \mathfrak{B}$. As $U \supseteq C$ for all $C \in \mathfrak{C}$ it then follows that U is an upper bound of $\mathfrak{C}$ and the proof is finished.

Let $x_1, x_2, \dots x_k \in U$ and $\alpha_1, \alpha_2, \dots \alpha_k \in \mathbb{R}$ ($k \in \mathbb{N}$) such that

$$\sum_{j=1}^{k} \alpha_j x_j = 0. \tag{15.2}$$

We must show that each α_j is zero. For each $0 \leq j \leq k$ there is some $C_j \in \mathfrak{C}$ such that $x_j \in C_j$. $\mathfrak{C}$ is totally ordered, hence $C_i \subseteq C_j$ or $C_j \subseteq C_i$ for any two indices $0 \leq i, j \leq k$. But then there exists an index j_0 such that $C_{j_0} \supseteq C_j$ for all j, hence $x_1, x_2, \dots x_k \in C_{j_0}$. The set C_{j_0} is linearly independent because $C_{j_0} \in \mathfrak{C} \subseteq \mathfrak{B}$. It follows that $\alpha_1 = \dots = \alpha_k = 0$. ■

Theorem 15.1.

Every vector space V has a basis.

PROOF: It follows from lemma 15.1 and Zorns Lemma (axiom 15.1 on p. 673) that the set $\mathfrak{B}$ of all independent subsets of the vector space V contains a maximal element (subset of V) which we denote by B. As membership in $\mathfrak{B}$ guarantees its linear independence we only need to prove that $\mathit{span}(B) = V$.

Let us assume to the contrary that there exists $y \in \mathit{span}(B)$ It follows from lemma **??** on p.**??** that the set $B' := B \cup \{y\}$ is linearly independent, hence $B' \in \mathfrak{B}$. Clearly, $B \subsetneq B'$. This contradicts the maximality of B in the partially ordered set $(\mathfrak{B}, \subseteq)$. ■

15.3 The Cardinal Numbers are a totally ordered set

As another application of Zorn's Lemma we now prove thm.**??**, p.**??**, of ch.**??** (Cardinality as a Partial Ordering).

Theorem 15.2.

> *Let $X, Y \subseteq \Omega$. Then* $\mathbf{card}(X) \le \mathbf{card}(Y)$ *or* $\mathbf{card}(Y) \le \mathbf{card}(X)$

PROOF: [13] The result is immediate if $X = \emptyset$ or $Y = \emptyset$. Assume X and Y are not empty. Let

$$\mathscr{F} := \{D_f \xrightarrow{f} C_f : D_f \subseteq X,\ C_f \subseteq Y, \text{ and } f \text{ is bijective}\}$$

be the set of all bijective functions with domain contained in X and codomain contained in Y. We define a partial order $\preceq$ on $\mathscr{F}$ as follows: Let $D_f \xrightarrow{f} C_f$ and $D_g \xrightarrow{g} C_g$. Then

$$f \preceq g \quad \text{if and only if } D_f \subseteq D_g,\ C_f \subseteq C_g, \text{ and } g\Big|_{D_f} = f.$$

We will prove that $\mathscr{F}$ has the ZL property: Let $\mathfrak{C}$ be a chain in $\mathscr{F}$. Let $D_u \xrightarrow{u} C_u$ be defined as follows:

$$D_u := \bigcup[D_f : f \in \mathfrak{C}], \quad C_u := \bigcup[C_f : f \in \mathfrak{C}], \quad u(x) := g(x) \text{ for } x \in D_g.$$

Note that the assignment $u(x) = g(x)$ for $x \in D_g$ is unambiguous: If there also is $g' \in \mathfrak{C}$ such that $x \in D_{g'}$ then we obtain from the total ordering of $\mathfrak{C}$ that $g' \preceq g$, i.e., g is an extension of g', or $g \preceq g'$, i.e., g' is an extension of g. In either case it follows that $g'(x) = g(x)$.

Moreover u is injective: If $x_1, x_2 \in D_u$ and $x_1 \neq x_2$ then there exist $f, g \in \mathfrak{C}$ such that $x_1 \in D_f$ and $x_2 \in D_g$. Since f extends g or vice versa, say, f extends g, we may assume that both $x_1, x_2 \in D_f$. It follows from the injectivity of f that $f(x_1) \neq f(x_2)$, hence $u(x_1) \neq u(x_2)$.

Also note that u is surjective: If $y \in C_u$ then $y \in C_f$ for some $f \in \mathfrak{C}$ and surjectivity gives us $x \in D_f$ such that $f(x) = y$. But then $x \in D_u$ and $u(x) = f(x) = y$.

A bijective function u has been constructed in such a fashion that it extends any $g \in \mathfrak{C}$ and hence is an upper bound of $\mathfrak{C}$. It follows that the partially ordered set $(\mathscr{F}, \preceq)$ possesses the ZL property (see axiom 15.1 (Zorn's Lemma) on p.673), hence there exists a maximal element $D_h \xrightarrow{h} C_h$ in $\mathscr{F}$.

We claim that $D_h = X$ or $C_h = Y$ (or both). Otherwise there would be $x_0 \in D_h^{\complement}$ and $y_0 \in C_h^{\complement}$ and the function

$$\psi : D_h \cup \{x_0\} \longrightarrow C_h \cup \{y_0\}, \qquad \begin{cases} x \mapsto h(x) & \text{if } x \in D_h, \\ x \mapsto y_0 & \text{if } x = x_0 \end{cases}$$

is a bijective extension of h, thus, $\psi \in \mathscr{F}$, and $h \preceq \psi$. Since $\psi \neq h$, this contradicts the maximality of h.

Case 1: $D_h = X$. Then changing C_h to Y makes h an injective function which maps X into Y, i.e., $\mathbf{card}(X) \le \mathbf{card}(Y)$.

Case 2: $C_h = Y$. Then $h : D_h \xrightarrow{\sim} Y$ is a bijective function whose inverse $h^{-1} : Y \xrightarrow{\sim} D_h$ is an injection from Y into the subset D_h of X, i.e., $\mathbf{card}(Y) \le \mathbf{card}(X)$. ■

15.4 Extensions of Linear Functions in Arbitrary Vector Spaces

In this section we turn our attention to extending a linear real–valued function f from a subspace $F \subseteq V$ to the entire vector space V. Note that setting $f(x) = 0$ for all $x \in F^{\complement}$ does not yield a linear extension of f. See exercise 15.3 on p.687.

[13] See [4] Haaser/Sullivan: Real Analysis.

Lemma 15.2.

Let V be a vector space and let F be a (linear) subspace of V. Let $f: F \to \mathbb{R}$ be linear. Let

$$\begin{aligned} \mathscr{G} \ := \ \{(W, f_W) : & W \text{ is a subspace of } V, W \supseteq F, \\ & f_W : W \to \mathbb{R} \text{ is a linear extension of } f \text{ to } W\}. \end{aligned}$$

Then the following defines a partial ordering on $\mathscr{G}$:

$$(U, f_U) \preceq (W, f_W) \ \Leftrightarrow \ V \subseteq W \text{ and } f_W\big|_U = f_U .$$

Moreover this ordering satisfies the requirements of Zorn's Lemma:
Every chain in $(\mathscr{G}, \preceq)$ possesses an upper bound (in $\mathscr{G}$).

PROOF:

Reflexivity and transitivity of "$\preceq$" are trivial. The latter is true because the extension of an extension is again an extension.

Antisymmetry: If both $(U, f_U) \preceq (W, f_W)$ and $(W, f_W) \preceq (U, f_U)$ then both $U \subseteq W$ and $W \subseteq U$, hence $U = W$. But then f_W is an extension of f_U to itself, i.e., $f_U = f_W$. It follows that $\preceq$ is indeed a partial order on $\mathscr{G}$.

Now let $\mathscr{C}$ be a chain in $\mathscr{G}$. We must find an upper bound for $\mathscr{C}$. Let $W := \bigcup \big[U : (U, f_U) \in \mathscr{C}\big]$.

We show that W is a subspace of E: If $x, y \in W$ and $\lambda \in \mathbb{R}$ then there are $(C_1, f_1), (C_2, f_2) \in \mathscr{C}$ such that $x \in C_1$ and $y \in C_2$. Because $\mathscr{C}$ is a chain we have $C_1 \subseteq C_2$ or $C_2 \subseteq C_1$, say, $C_1 \subseteq C_2$. It follows that $x, y \in C_2$. But C_2 is a subspace of V and we conclude that $x + \lambda y \in C_2$, hence $x + \lambda y \in W$. It follows that W is a subspace of V.

Let $f_W : W \to \mathbb{R}$ be defined as follows: If $x \in W$ then there is some $(C, f_C) \in \mathscr{C}$ such that $x \in C$. We define $f_W(x) := f_C(x)$. This definition is unambiguous even if x belongs to (possibly infinitely) many elements of $\mathscr{C}$. To see this let $(C, f_C), (D, f_D) \in \mathscr{C}$ such that $x \in C$ and $x \in D$. Then $C \subseteq D$ or $D \subseteq C$. We may assume that $C \subseteq D$. But as $f_D\big|_C = f_C$ we conclude that $f_C(x) = f_D(x)$, i.e., the definition of $f_W(\cdot)$ is unambiguous. The above specifically holds for $x \in W$ and we note that f_W is an extension of f.

Next we show linearity of f_W. Let $x, y \in W$ and $\alpha \in \mathbb{R}$. Then there are $(C, f_C), (D, f_D) \in \mathscr{C}$ such that $x \in C$ and $y \in D$. Again we may assume that $C \subseteq D$. It follows from the linearity of f_D that

$$f_W((x + \alpha y) \ = f_D((x + \alpha y) \ = \ f_D((x) + \alpha f_D(y) \ = \ f_W((x) + \alpha f_W(y).$$

and we have proved that f_W is linear (on all of W).

To summarize, W is a subspace of V and f_W is a linear extension of f to W. But then $(W, f_W) \in \mathscr{G}$ and $(W, f_W) \succeq (C, f_C)$ for all $(C, f_C) \in \mathscr{C}$. It follows that (W, f_W) is an upper bound of $\mathscr{C}$. ■

Theorem 15.3 (Extension theorem for linear real–valued functions)**.**

Let V be a vector space and let F be a (linear) subspace of V. Let $f: F \to \mathbb{R}$ be a linear mapping. Then there is an extension of f to a linear mapping $\tilde{f}: V \to \mathbb{R}$.

PROOF:

Let $\mathscr{G} := \{(W, f_W) : W \text{ is a subspace of } V, W \supseteq F,\ f_W : W \to \mathbb{R} \text{ is a linear extension of } f \text{ to } W\}$ and let $(U, f_U) \preceq (W, f_W) \Leftrightarrow U \subseteq W$ and $f_W\big|_U = f_U$.

We have seen in lemma 15.2 that $\preceq$ is a partial ordering on $\mathscr{G}$ such that any chain in $(\mathscr{G}, \preceq)$ possesses an upper bound. We apply Zorn's Lemma (axiom 15.1 on p.673) and conclude that $\mathscr{G}$ possesses a maximal element (F', f') .

We show that $F' = V$.

If this was not true then we could find $a \in V \setminus F'$ and, according to prop.**??** on p.**??**, applied with $V' = \mathbb{R}$ and $y_0 = \alpha$, extend f' to a linear function $\tilde{f}$ on $\mathit{span}(F' \uplus \{a\})$. It follows that $\big(\mathit{span}(F)' \uplus \{a\}, \tilde{f}\big) \in \mathscr{G}$ and $(F', f') \precneqq \big(\mathit{span}(F)' \uplus \{a\}, \tilde{f}\big)$.

This contradicts the maximality of (F', f'). and we have reached a contradiction. ■

Next, we will make a digression and talk about dual vector spaces and dual linear functions. All the material is $\boxed{\star}$ optional reading.

Definition 15.2 (Dual vector space)**.**

$\boxed{\star}$

Let V and W be vector spaces, and let $L : V \to W$ be linear.

(a) We call $V^* := \{f : f \text{ is a linear function } V \to \mathbb{R}\}$ the **dual** or **algebraic dual** of V.

(b) We call $L^* : W^* \to V^*$, defined by $L^*(f) := f \circ L$,
i.e., $L^*(f)(x) = f(Lx)\ \forall\, x \in V$, the **dual function** or **dual mapping** of L. □

This is not a course about linear algebra, so we give here some elementary properties of algebraic duals of vector spaces [14] and dual functions without proof.

Proposition 15.2.

$\boxed{\star}$ *For the following see Definition **??** (Transposed matrix) on p.**??**.*

(a) *V^* is a vector space, i.e., $f, g, \in V^*$ and $\alpha, \beta \in \mathbb{R} \Rightarrow \alpha f + \beta g \in V^*$.*

(b) *Since V^* is a vector space, its dual $V^{**} := \big(V^*\big)^*$ exists.*

(c) *Assume that $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$. For every linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ there exists a matrix $A = \big((a_{ij})\big)$ such that for every column vector $\vec{x}$, $L(\vec{x}) = A\vec{x}$, i.e., the function value $\vec{y} = L(\vec{x})$ has coordinates $y_i = \sum_{j=1^n} a_{ij} x_j$.*

(d) *If V is a finite dimensional vector space, then there is a bijection $V \to V^*$ which is linear in both directions.* [15] *This allows us to "identify" $\big(\mathbb{R}^n\big)^*$ with $\mathbb{R}^n$, thus, the dual function of L from Definition 15.2 is a linear function $L^* : \mathbb{R}^m \to \mathbb{R}^n$ (Careful: Switched dimensions!).*

[14] If you work with vector spaces that carry a topology, e.g., normed vector spaces $(V, \|\cdot\|)$, then there also is the concept of the topological dual $V' := \{ f \in V^* : f \text{ is continuous} \}$. If V is a finite dimensional normed space then one can show that $V' = V^*$, but this will not necessarily be true for normed spacea of infinite dimension,

> *According to part **(c)** of this remark, there exists a matrix $A^* = \big((a_{k\ell})\big)$ such that the following is true. If $\vec{y}^* \in \mathbb{R}^m$ and $\vec{x}^* \in \mathbb{R}^n$ are column vectors such that $\vec{x}^* = L^*\big(\vec{y}^*\big)$, then $\vec{x}^* = A^*\vec{y}$, the product of the matrix A^* and the column vector $\vec{y}$.*
> *This matrix A^* is the transpose $A^\top$ of A: $a^*_{k\ell} = a_{\ell k}$ for $k = 1, \dots n$ and $\ell = 1, \dots, m$.*
> □

PROOF: To be found in text books on linear algebra. ■

The following can be found as an exercise in [3] Friedberg, Insel, Spence: Linear Algebra.

Theorem 15.4.

> ⋆ *Let $L : V \to W$ be a linear function between two vector spaces V and W.*
> *Let $L^* : W^* \to V^*$ be the associated dual function of L. Then,*
> $$L \text{ is injective } \Leftrightarrow \ L^* \text{ is surjective}; \qquad L^* \text{ is injective } \Leftrightarrow \ L \text{ is surjective}.$$

PROOF:“ We will show that L^* is injective $\Leftrightarrow$ L is surjective.

Be aware that 0 can be the zero vector in any of V, W, V^*, W^*. In the latter two cases 0 denotes the (linear) zero function which maps any vector to zero.

PROOF of “$\Rightarrow$):”

We prove the contrapositive. We assume that L is not surjective and will show that L^* is not injective. Let $R := L(V)$ be the range of L. Since L is not surjective, $R \subsetneq W$. Let $w_0 \in W \setminus R$. Let $\psi \in W^*$ be defined as follows.

$$\psi(w) \ = \ 0 \ \text{ for } w \in R, \ \ \psi(\alpha w_0) \ = \ \alpha\,.$$

Then ψ is defined, so far, on $span(W \biguplus \{w_0\})$. According to Theorem 15.3 (Extension theorem for linear real–valued functions), ψ can be extended to a real–valued, linear function on all of W, which we call again ψ.

Obviously, $\psi \neq 0$, since $\psi(\alpha) = 1$. On the other hand. for any $v \in V$,

$$L^*(\psi)(x) \ = \ (\psi \circ L)(x) \ = \ \psi\big(L(x)\big) \ = \ 0\,.$$

since $L(x) \in L(V) = R$, and we assumed that $\psi = 0$ on R. Thus $L^*(\psi) = 0$. Since also $L^*(0) = 0$ and $\psi \neq 0$, we conclude that L^* is not injective.

PROOF of “$\Leftarrow$):”

We show directly that L is surjective $\Rightarrow$ L^* is injective. Let $\varphi, \psi \in W^*$ such that $L^*(\varphi) = l^*(\psi)$. We must show that $\varphi = \psi$. It follows from $L^*(\varphi) = l^*(\psi)$ and the definition of the dual function that, for all $v \in V$,

$$\textbf{(A)} \qquad \varphi\big(L(v)\big) \ = L^*(\varphi)(v) \ = \ L^*(\psi)(v) \ = \ \psi\big(L(v)\big)\,.$$

[15] One calls such bijection which is structure compatible a **linear isomorphism** or a **vector space isomorphism**.

Let $w \in W$. Since L is surjective, there exists $v \in V$ such that $w = L(v)$. It follows from **(A)** that

$$\varphi(w) \;=\; \varphi\big(L(v)\big) \;=\; \psi\big(L(v)\big) \;=\; \psi(w)\,.$$

Since $w \in W$ was arbitrary, this shows that $\varphi = \psi$. This concludes the proof. ■

Corollary 15.1.

*Let $A = \big((a_{ij})\big)$ be a matrix with m rows and n columns. Then **(a)** $\Leftrightarrow$ **(b)**, where*

(a) *The set of m linear equations in n unknowns $\vec{x} = (x_1, \dots, x_n)^\top$,*

$$A\,\vec{x} \;=\; \vec{y}\,,$$

has a solution $\vec{x}$ for any choice of right hand side $\vec{y} = (y_1, \dots, y_m)^\top$.

(b) *the set of n linear equations in m unknowns $\vec{\xi} = (\xi_1, \dots, \xi_m)^\top$,*

$$A^\top\,\vec{\xi} \;=\; \vec{\eta}\,,$$

has at most one solution $\vec{\xi}$ for any $\vec{\eta} = (\eta_1, \dots, \eta_n)^\top$.

PROOF: This follows from Theorem 15.4 and the fact that the linear function $L : \vec{x} \mapsto A\,\vec{x}$ has as its dual the function $L^* : \vec{\xi} \mapsto \vec{\xi}\,A^\top$. ■

15.5 The Hahn-Banach Extension Theorem ⋆

Note that this chapter is starred, hence optional. The proof given here is a more detailed version of the one found in [1] Choquet.

Let V be a vector space and let F be a (linear) subspace of V. Let $f : F \to \mathbb{R}$ be linear function. The Hahn-Banach Extension Theorem shows how to extend f from its domain F to a linear function on entire space V, subject to some majorization condition.

If V is a normed space (hence a metric space) and if f is continuous on F then this majorization condition can be chosen in such a fashion that the linear extension will be continuous on all of V.

In preparation for this subject matter we must study sublinear functions, which generalize both linear functions and norms.

15.5.1 Sublinear Functionals

Definition 15.3 (Sublinear functionals)**.**

Let V be a vector space and $p : V \to \mathbb{R}$ such that

(a) if $\lambda \in \mathbb{R}_{\geq 0}$ and $x \in V$ then $p(\lambda x) \;=\; \lambda p(x)$ (positive homogeneity)

(b) if $x, y \in V$ then $p(x + y) \;\leq\; p(x) + p(y)$ (subadditivity)

Then we call p a **sublinear functional** on V. □

The term "functional" has several meanings in mathematics. We use it here to denote a real–valued function whose domain is a vector space.

Proposition 15.3.

Let V be a vector space and $p : V \to \mathbb{R}$ sublinear. Let $x \in V$. Then
(a) $p(0) = 0$,
(b) $-p(x) \le p(-x)$,

PROOF of **(a)**: $p(0) = p(0 \cdot 0) = 0 \cdot p(0) = 0$.
PROOF of **(b)**: This follows from $0 = p(0) = p(x + (-x)) \le p(x) + p(-x)$. ■

Example 15.2 (Norms are sublinear)**.**
Let $(V, \|, \cdot\|)$ be a normed vector space.
Then the function $p(x) := \|x\|$ is sublinear.
Indeed, norms are absolutely homogenous: We have $\|\lambda x\| = |\lambda| \cdot \|x\|$ not only for $\lambda \ge 0$ but for all $\lambda \in \mathbb{R}$. Further, subadditivity is just the validity of the triangle inequality. □

Example 15.3 (Linear functions are sublinear)**.**

Let V be a vector space and let $f := V \to \mathbb{R}$ be a linear function. Then f is sublinear.
Indeed, linear functions f satisfy $f(\lambda x) = \lambda \cdot f(x)$ not only for $\lambda \ge 0$ but for all $\lambda \in \mathbb{R}$.
Further linear functions satisfy additivity: $f(x + y) = f(x) + f(y)$,
hence also subadditivity $f(x + y) \le f(x) + f(y)$. □

More about sublinearity can be found in chapter 15.6 on p.684

15.5.2 The Hahn-Banach extension theorem and its Proof

This chapter follows closely [1] Choquet.

As mentioned previously, the subject of this chapter is the extension of a linear, real–valued function from a subspace to the entire vector space in such a fashion that some majorization condition will be preserved. The Hahn-Banach extension theorem (theorem 15.5 below) states that if p is a sublinear functional defined on all of a vector space V and if a linear, real–valued function f is defined on a subspace F of V such that $f \le p$ on F, then f can be linearly extended to V in such a way that p dominates this extension everywhere on V. Once we have that, it is not very difficult to prove what we truly want, thm.15.6 (Continuous extensions of continuous linear functions).

The following remark is about first extending f to "one more dimension".

Remark 15.2.
Let V be a vector space, let F be a linear subspace of V and let $f := F \to \mathbb{R}$ be a linear function. Let $a \in V \setminus F$. We saw in prop.**??** on p.**??** that any linear extension $\tilde{f}$ of f to $span(F \uplus \{a\})$ is uniquely determined by its value $k := \tilde{f}(a)$.

Indeed, any $x \in \mathit{span}(F \uplus \{a\})$ can be written as $u + \lambda a$ for some $u \in F$ and $\lambda \in \mathbb{R}$. It follows from the linearity of $\tilde{f}$ that

$$\tilde{f}(x+\lambda a) \;=\tilde{f}(x)+\lambda\tilde{f}(a) \;=\; f(x)+\lambda k. \;\Box \tag{15.3}$$

Theorem 15.5 (Hahn–Banach extension theorem)**.**

Let V be a vector space and $p : V \to \mathbb{R}$ a sublinear function.
Suppose F is a (linear) subspace of V and $f : F \to \mathbb{R}$ is a linear mapping such that $f \leq p$ on F.
Then there is an extension of f to a linear map $\tilde{f} : V \to \mathbb{R}$ such that $\tilde{f} \leq p$ on V.

Before proving this theorem, first we prove two lemmata.

Lemma 15.3. *Suppose F is a subspace of V, $f : F \to \mathbb{R}$ is a linear mapping, $a \in V \setminus F$, and $k \in \mathbb{R}$. Let $\tilde{f}$ be the linear extension of f to $\mathit{span}(F \uplus \{a\})$ given in prop.**??** on p.**??**, choosing $V' = \mathbb{R}$ and $y_0 = k$:*

$$\tilde{f}(x+\lambda a) \;:=\; f(x)+\lambda k, \quad \textit{i.e.,} \;\; \tilde{f}(a) \;=\; k. \tag{15.4}$$

Then,

$$k \leq \inf_{u\in F}\{p(u+a)-f(u)\} \;\;\Leftrightarrow\;\; \tilde{f}(x+\lambda a) \;\leq\; p(x+\lambda a) \;\; \textit{for all } \lambda > 0 \textit{ and } x \in F, \tag{15.5}$$

$$k \geq \sup_{v\in F}\{f(v)-p(v-a)\} \;\;\Leftrightarrow\;\; \tilde{f}(x+\lambda a) \;\leq\; p(x+\lambda a) \;\; \textit{for all } \lambda < 0 \textit{ and } x \in F. \tag{15.6}$$

Further,

$$\begin{aligned} &\sup_{v\in F}\{f(v)-p(v-a)\} \;\leq k \leq\; \inf_{u\in F}\{p(u+a)-f(u)\} \\ &\qquad\qquad \Leftrightarrow\;\; \tilde{f}(x+\lambda a) \;\leq\; p(x+\lambda a) \;\; \textit{for all } \lambda \in \mathbb{R} \textit{ and } x \in F, \end{aligned} \tag{15.7}$$

Proof of (15.5)**, $\Rightarrow$):** Let us assume that $\lambda > 0$ and $x \in F$. Then $u := \frac{x}{\lambda} \in F$ because F is a subspace. On account of the left–hand side of (15.5),

$$\begin{aligned} \tilde{f}(x+\lambda a) \;&= f(x)+\lambda k \;=\; \lambda\big(f(x/\lambda)+k\big) \;=\; \lambda\big(f(u)+k\big) \;\leq\; \lambda\Big(f(u)+\big(\,p(u+a)-f(u)\big)\Big) \\ &= \lambda p(u+a) \;=\; \lambda p(x/\lambda+a) \;=\; p(x+\lambda a) \end{aligned}$$

The inequality follows from the left–hand side of (15.5), and we used the positive homogeneity of p for the last equation.

Proof of (15.5)**, $\Leftarrow$):** We assume $\tilde{f}(x+\lambda a) \;\leq\; p(x+\lambda a)$ for all $\lambda > 0$ and $x \in F$. We will show that $k = \tilde{f}(a) \;\leq\; p(u+a)-f(u)$ for all $u \in F$.

$$p(u+a)-f(u) \;\geq \tilde{f}(u+a)-f(u) \;=\; \tilde{f}(u)+\tilde{f}(a)-f(u) \;=\; f(u)+\tilde{f}(a)-f(u) \;=\; \tilde{f}(a) \;=\; k.$$

Proof of (15.6), $\Rightarrow$): Let us assume that $\lambda < 0$ and $x \in F$. Then $v := \frac{x}{\lambda} \in F$ because F is a subspace.

Because of the left–hand side of (15.6) and $\lambda < 0$ and positive homogeneity of p,

$$\begin{aligned} k \;\geq\; f(v) - p(v-a) \;&\Rightarrow \lambda k \;\leq\; f(\lambda v) - \lambda p(v-a) \\ &\Rightarrow\; -f(\lambda v) + \lambda k \;\leq (-\lambda)p(v-a) \;=\; p\big((-\lambda)(v-a)\big) \;=\; p\big((-\lambda)v + \lambda a\big). \end{aligned}$$

Since $x = \lambda v$, and since linearity of f implies $-f(x) = f(-x)$, that last inequality yields

$$f(-x) + \lambda k \;=\; -f(x) + \lambda k \;\leq p(-x + \lambda a), \quad \text{hence} \quad \tilde{f}(-x + \lambda a) \;=\; f(-x) + \lambda k \;\leq\; p(-x + \lambda a)$$

We can switch from $-x$ to x as the above holds for all x in the subspace F and because $-x \in F$ if and only if $x \in F$. It follows that p indeed dominates $\tilde{f}$ for all $x \in F$ and $\lambda < 0$.

Proof of (15.6), $\Leftarrow$): We assume $\tilde{f}(x + \lambda a) \;\leq\; p(x + \lambda a)$ for all $\lambda < 0$ and $x \in F$. We now show that $k = \tilde{f}(a) \;\geq\; f(v) - p(v-a)$ for all $v \in F$.

We apply $\tilde{f}(x + \lambda a) \;\leq\; p(x + \lambda a)$ with $x := v$ and $\lambda := -1$ and obtain

$$-p(v-a) + f(v) \;\leq\; -\tilde{f}(v-a) + f(v) \;=\; \tilde{f}(a-v) + f(v) \;=\; \tilde{f}(a) - \tilde{f}(v) + f(v) \;=\; \tilde{f}(a) \;=\; k.$$

Proof of (15.7), $\Rightarrow$): Let us assume that $\lambda \in \mathbb{R}$ and $x \in F$. For $\lambda \neq 0$, validity of the right–hand side of (15.7) follows from (15.5) and (15.6). If $\lambda = 0$ then we must show that $\tilde{f}(x) \leq p(x)$. This is true because $x \in F$ implies $\tilde{f}(x) = f(x)$ and we assumed that $f \leq p$ on F.

Proof of (15.7), $\Leftarrow$): This is immediate from (15.5) and (15.6). ■

Lemma 15.4.

Let V be a vector space and $p : V \to \mathbb{R}$ a sublinear function. Let $F \subsetneq V$ be a genuine subspace of V and $a \in V \setminus F$. Let $f : F \to \mathbb{R}$ be a linear mapping with $f \leq p$ on F. Let $G := span(F \uplus \{a\})$ be the subspace of all linear combinations that can be created by a and/or vectors in F. Then

- ***(a)*** *there exists a linear extension $\tilde{f}$ of f to G such that $\tilde{f} \leq p$ on G,*
- ***(b)*** *This extension is unique if and only if $\sup_{v \in F}\{f(v) - p(v-a)\} \;=\; \inf_{u \in F}\{p(u+a) - f(u)\}$.*

Proof of a. For $u, v \in F$ we have

$$f(u) + f(v) \;= f(u+v) \leq p(u+v) \;=\; p\big((u+a) + (v-a)\big) \;\leq\; p(u+a) + p(v-a),$$

and hence $f(v) - p(v-a) \;\leq\; p(u+a) - f(u)$. Therefore

$$\sup_{v \in F}\{f(v) - p(v-a)\} \;\leq\; \inf_{u \in F}\{p(u+a) - f(u)\}. \tag{15.8}$$

For a fixed $k \in \mathbb{R}$, we define $\tilde{f}(x + \lambda a) \;=\; f(x) + \lambda k$. We claim that $\tilde{f} \leq p$ if and only if we have

$$\sup_{v \in F}\{f(v) - p(v-a)\} \;\leq k \;\leq\; \inf_{u \in F}\{p(u+a) - f(u)\} \tag{15.9}$$

which will conclude the proof of **(a)** since such a k exists by 15.8.

Our claim holds because $f(x) + \lambda k \;=\; \tilde{f}(x + \lambda a) \;\leq\; p(x + \lambda a)$ for all λ if and only if

$$\begin{aligned} k \;&\leq p(u+a) - f(u) \quad \text{for all } u \in F \\ \text{and } \; k \;&\geq f(v) - p(v-a) \quad \text{for all } v \in F \end{aligned}$$

(the cases $\lambda > 0$ and $\lambda < 0$ respectively). This was proved in lemma 15.3.

Proof of b. From (15.9) we deduce that k is unique if and only if

$$\sup_{v \in F}\{f(v) - p(v-a)\} \;=\; \inf_{u \in F}\{p(u+a) - f(u)\}.$$

Because the extension $\tilde{f}(x+\lambda a) \;=\; f(x) + \lambda k$ of f to G is uniquely determined by k and vice versa, we have proven b. ■

PROOF of thm.15.5 (Hahn–Banach Extension Theorem):

Let $\mathscr{G} \;=\; \{(V', g):\ V'$ is a subspace of $V, V' \supseteq F,\ g : V' \to \mathbb{R}$ is linear, $g\big|_F = f,$ and $g\ \le p$ on $V'\}$. We define a partial order "$\preceq$" on $\mathscr{G}$ as follows:

$$(V_1, g_1) \preceq (V_2, g_2) \;\Leftrightarrow\; V_1 \subseteq V_2 \text{ and } g_2 \text{ is an extension of } g_1. \tag{15.10}$$

Note that $\mathscr{G}$ is not empty because $(F, f) \in \mathscr{G}$.

(a) We first prove that any chain $\mathscr{C} \subseteq (\mathscr{G}, \preceq)$ has an upper bound: Let $W := \bigcup \big[V' : (V', g) \in \mathscr{C}\big]$. Then W is a subspace of V because if $x, y \in W$ and $\lambda \in \mathbb{R}$ then there are $(V_1, g_1), (V_2, g_2) \in \mathscr{C}$ such that $x \in V_1$ and $y \in V_2$. Because $\mathscr{C}$ is a chain we have $V_1 \subseteq V_2$ or $V_2 \subseteq V_1$, say, $V_1 \subseteq V_2$. It follows that $x, y \in V_2$. But V_2 is a subspace, and we conclude that $x + \lambda y \in V_2$, hence $x + \lambda y \in W$. It follows that W is a subspace.

Next we construct a linear $h : W \to \mathbb{R}$ such that $h \le p$ on W and $h\big|_{V'} = g$ for all $(V', g) \in \mathscr{C}$, i.e., h is a linear extension of g for all $(V', g) \in \mathscr{C}$. If we find such h then it follows that $(W, h) \in \mathscr{G}$ and (W, h) is an upper bound of $\mathscr{C}$.

Let $x \in W$. Then $x \in V_1$ for some $(V_1, g_1) \in \mathscr{C}$. We define $h(x) := g_1(x)$. This assignment is unambiguous because if $x \in V_2$ for some other $(V_2, g_2) \in \mathscr{C}$ then one of them, say V_1, is contained in the other and g_2 is an extension of g_1, i.e., $h(x) = g_1(x) = g_2(x)$. As $(V_1, g_1) \in \mathscr{G}$ we conclude that $h(x) = g_1(x) \le p(x)$, i.e., $h \le p$ on W.

Next we show that h is linear. Let $x, y \in W$ and $\lambda \in \mathbb{R}$. We repeat the argument from the proof that W is a subspace to conclude that both x, y belong to some subspace V' such that $(V', g) \in \mathscr{C}$. We obtain

$$h(x + \lambda y) \;= g(x + \lambda y) \;=\; g(x) + \lambda g(y) \;=\; h(x) + \lambda h(y)$$

This completes the proof that $(W, h) \in \mathscr{G}$. Let $(V', g) \in \mathscr{C}$. Clearly, $V' \subseteq W = \bigcup \big[U : U \in \mathscr{C}\big]$. We have seen that h is linear, dominated by p, and constructed in such a manner that $h(x) = g(x)$ for all $x \in V'$. It follows that $(W, h) \succeq (V', g)$ for all $(V', g) \in \mathscr{C}$, and we have proved that $\mathscr{C}$ has an upper bound in $(\mathscr{G}, \preceq)$.

(b) It follows from **(a)** that we can apply Zorn's Lemma (axiom 15.1 on p.673), and hence conclude that $(\mathscr{G}, \preceq)$ possesses a maximal element (M, m).

We show that $M = V$. Assume to the contrary that there exists $a \in V \setminus M$. According to lemma 15.4 on p.682, we can extend m to a linear function $\tilde{m}$ on $M \uplus \{a\}$ in such a fashion that $\tilde{m} \le p$ on $M \uplus \{a\}$. It follows that $(M \uplus \{a\}, \tilde{m}) \in \mathscr{G}$. Further, $(M, m) \precneqq (M \uplus \{a\}, \tilde{m})$ because $(M \subsetneqq M \uplus \{a\}$. This contradicts the maximality of (M, m). We conclude that $M = V$.

The proof is finished: m is the linear extension of f to V we were looking for. ■

Theorem 15.6 (Continuous extensions of continuous linear functions)**.**

Let $(V, \|\cdot\|)$ be a normed vector space. Let F be a (linear) subspace of V. Then,

- *any continuous, linear $f : F \to \mathbb{R}$ possesses a continuous, linear extension $\tilde{f} : V \to \mathbb{R}$.*

PROOF: Let $p(x) := \|f\| \cdot \|x\|$ [16] Because p is a positive multiple of the norm $\|\cdot\|$ on all of V, it also is a norm on V (see Remark **??**) on p.**??**), hence sublinear by example 15.2 on p.680. According to the Hahn-Banach extension theorem there exists a linear extension $\tilde{f}$ of f to all of V such that

$$\tilde{f}(x) \leq p(x) \text{ for all } x \in V. \tag{15.11}$$

We replace x with $-x$ and obtain from the linearity of $\tilde{f}$ that $-\tilde{f}(x) = \tilde{f}(-x) \leq p(-x)$.
We note that $p(x) = p(-x)$ because p is a norm. Hence

$$-\tilde{f}(x) \leq p(-x) = p(x), \text{ i.e., } \tilde{f}(x) \geq -p(x).$$

Together with (15.11) this shows that

$$-p(x) \leq \tilde{f}(x) \leq p(x) \text{ for all } x \in V,$$

and thus

$$|\tilde{f}(x)| \ \leq \ p(x) \ = \ \|f\| \cdot \|x\| \text{ for all } x \in V \tag{15.12}$$

It follows from (15.12) that f has been extended in such a way that $\|\tilde{f}\| \leq \|f\|$. We aply the continuity criterion for linear functions (thm.**??** on p.**??**) twice in a row to finish the proof as follows: It follows from the continuity of f that $\|f\| < \infty$. But then $\|\tilde{f}\| < \infty$ and this proves the continuity of $\tilde{f}$. ■

15.6 Convexity ⋆

Note that this chapter is starred, hence optional.

Definition 15.4 (Concave-up and convex functions)**.**

Let $-\infty \leq \alpha < \beta \leq \infty$ and let $I := \,]\alpha, \beta[$ be the open interval of real numbers with endpoints α and β. Let $f : I \to \mathbb{R}$.

(a) The **epigraph** of f is the set $epi(f) := \{(x_1, x_2) \in I \times \mathbb{R} : x_2 \geq f(x_1)\}$ of all points in the plane that lie above the graph of f.

(b) f is **convex** if for any two vectors $\vec{a}, \vec{b} \in$ $epi(f)$the entire line segment $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \leq \lambda \leq 1$ is contained in $epi(f)$. See Figure 15.1.

(c) Let f be differentiable at all points $x \in I$. Then f is **concave-up**, if the function $f' : x \mapsto f'(x) = \frac{df}{dx}(x)$ is increasing. □

Proposition 15.4 (Convexity criterion)**.**

[16] See Definition **??** on p.**??** for the definition of $\|f\|$

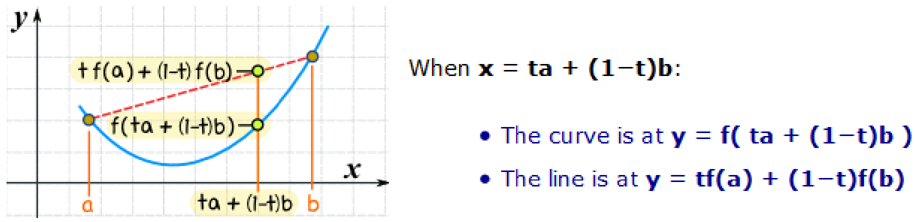

Figure 15.1: Convex function
17

f is convex if and only if the following is true: For any

$$\alpha \;<\; a \;\leq\; x_0 \;\leq\; b \;<\; \beta$$

let $S(x_0)$ be the unique number such that the point $(x_0, S(x_0))$ is on the line segment that connects the points $(a, f(a))$ and $(b, f(b))$. Then

$$f(x_0) \;\leq\; S(x_0). \tag{15.13}$$

Note that any x_0 between a and b can be written as $x_0 = \lambda a + (1-\lambda)b$ for some $0 \leq \lambda \leq 1$ and that the corresponding y-coordinate $S(x_0) = S(\lambda a + (1-\lambda)b)$ on the line segment that connects $(a, f(a))$ and $(b, f(b))$ then is $S(\lambda a + (1-\lambda)b) = \lambda f(a) + (1-\lambda)f(b)$. Hence we can rephrase the above as follows:

f is convex if and only if for any $a < b$ such that $a, b \in I$ and $0 \leq \lambda \leq 1$ it is true that

$$f(\lambda a + (1-\lambda)b) \;\leq \lambda f(a) + (1-\lambda)f(b). \tag{15.14}$$

PROOF of "$\Rightarrow$": Any line segment S that connects the points $(a, f(a))$ and $(b, f(b))$ in such a way that S is entirely contained in the epigraph of f will satisfy $(x_0, S(x_0)) \in$ $epi(f)$and hence $f(x_0) \leq S(x_0)$ for all $a \leq x_0 \leq b$. It follows that convexity of f implies (15.13).

PROOF of "$\Leftarrow$": Let (15.13) be valid for all $a, b \in I$. Let $\vec{a} = (a_1, a_2), \vec{b} = (b_1, b_2) \in$ $epi(f)$Then

$$a_2 \geq f(a_1) \;\text{ and }\; b_2 \geq f(b_1). \tag{15.15}$$

We must show that the entire line segment $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \leq \lambda \leq 1$ is contained in $epi(f)$.

Let $\vec{a}' := (a_1, f(a_1))$. Let $S' := \{\lambda\vec{a}' + (1-\lambda)\vec{b} : 0 \leq \lambda \leq 1\}$ be the line segment obtained by leaving the right endpoint $\vec{b}$ unchanged and pushing the left one downward until a_2 matches $f(a_1)$. Clearly, S' nowhere exceeds S.

Let $\vec{b}'' := (b_1, f(b_1))$. Let $S'' := \{\lambda\vec{a}' + (1-\lambda)\vec{b}' : 0 \leq \lambda \leq 1\}$ be the line segment obtained by leaving the left endpoint $\vec{a}'$ unchanged and pushing the right one downward until the b_2 matches $f(b_1)$. Clearly, S'' nowhere exceeds S'.

We view any line segment T between two points with abscissas a_1 and b_1 as a function $T(\cdot) : [a_1, b_1] \to \mathbb{R}$ which assigns to $x \in [a_1, b_1]$ that unique value $T(x)$ for which the point $\big(x, T(x)\big)$ lies on T.

The segment S'' connects the points $(a, f(a))$ and $(b, f(b))$ and it follows from assumption **(b)** that for any $a \le x_0 \le b$ we have $f(x_0) \le S''(x_0)$. We conclude from $S(\cdot) \ge S'(\cdot) \ge S''(\cdot)$ that $S(x_0) \ge f(x_0)$, i.e. $(x_0, S(x_0)) \in \; \mathit{epi}(f)$ As this is true for any $a \le x_0 \le b$ it follows that the line segment S is entirely contained in the epigraph of f. ■

Proposition 15.5 (Convex vs concave-up)**.**

Let $f : \mathbb{R} \to \mathbb{R}$ be concave-up. Then f is convex.

PROOF: Assume to the contrary that f is (differentiable and) concave-up and that there are $a, b, x_0 \in I$ such that $a < x_0 < b$ and $f(x_0) > S(x_0)$. Here $S(x_0)$ denotes the unique number such that the point $(x_0, S(x_0))$ is on the line segment that connects the points $(a, f(a))$ and $(b, f(b))$. Let m be the slope of the linear function $S(\cdot) : x \mapsto S(x)$, i.e.,

$$m \;=\; \frac{S(b) - S(a)}{b - a}.$$

It follows that

$$m \;=\; \frac{S(b) - S(x_0)}{b - x_0} \;>\; \frac{S(b) - f(x_0)}{b - x_0} \;=\; \frac{f(b) - f(x_0)}{b - x_0} \;=\; f'(\xi) \tag{15.16}$$

for some $x_0 < \xi < b$ (according to the mean value theorem for derivatives). Further

$$m \;=\; \frac{S(x_0) - S(a)}{x_0 - a} \;<\; \frac{f(x_0) - S(a)}{x_0 - a} \;=\; \frac{f(x_0) - f(a)}{x_0 - a} \;=\; f'(\eta) \tag{15.17}$$

for some $a < \eta < x_0$ (according to the mean value theorem for derivatives).
Because f is concave up we have

$$f'(a) \;\le\; f'(\eta) \;\le\; f'(x_0) \;\le\; f'(\xi) \;\le\; f'(b).$$

From (15.16) and (15.17) we obtain

$$m \;<\; f'(\eta) \;\le\; f'(x_0) \;\le\; f'(\xi) \;<\; m,$$

and we have reached a contradiction. ■

Proposition 15.6 (Sublinear functions are convex)**.**

Let $f : \mathbb{R} \to \mathbb{R}$ be sublinear. Then f is convex.

PROOF: Let $0 \le \lambda \le 1$ and $x, y \in \mathbb{R}$. Then

$$p(\lambda x + (1 - \lambda)y) \;\le p(\lambda x) \;+\; p((1 - \lambda)y) \;=\; \lambda p(x) \;+\; (1 - \lambda)p(y). \tag{15.18}$$

It follows from prop.15.4 that f is concave-up. ■

15.7 Exercises for Ch.15

Exercise 15.1.

Prove prop.15.1 on p.672: Let $(X, \preceq)$ be a nonempty POset and $A \subseteq X$. If A has a maximum then it is unique. $\square$

Exercise 15.2.

Let $A \subseteq \mathbb{R}^2$ and $\vec{x}_1 = (x_1, y_1), \vec{x}_2 = (x_2, y_2) \in A$. Let

$$\vec{x}_1 \ \preceq \ \vec{x}_2 \ \Leftrightarrow \ x_1 \leq x_2 \text{ and } y_1 \leq y_2.$$

(a) Prove that "$\preceq$" defines a partial order on A.
(b) Prove that no maximal elements exist if $A = \mathbb{R}^2$.
(c) What are the maximal elements of $A = \{\vec{x} \in \mathbb{R}^2 : \|\vec{x}\|_2 = 1\}$? $\square$

Exercise 15.3.

The following was stated at the beginning of ch.15.4 (Extensions of Linear Functions in Arbitrary Vector Spaces) on p.675.

Let $F \subseteq V$ be a subspace of a vector space V and let $f : F \to \mathbb{R}$ be linear. Let

$$g : F \to \mathbb{R}; \qquad x \ \mapsto \begin{cases} f(x) & \text{if } x \in F, \\ 0 & \text{if } x \in F^{\complement}. \end{cases}$$

Then g is linear only if $f(x) = 0$ for all $x \in F$. Prove it. $\square$

References

[1] Gustave Choquet. Lectures on Analysis, Vol. 2: Representation Theory. Benjamin, New York, 1st edition, 1969.

[2] Richard M. Dudley. Real Analysis and Probability. Cambridge University Press, Cambridge, New York, 2nd edition, 2002.

[3] S.H. Friedberg, A.J. Insel, and L.E. Spence. Linear Algebra. Pearson, 5th edition, 2018.

[4] Norman B. Haaser and Joseph A. Sullivan. Real Analysis. publisher = Van Nostrand, year = 1971, 1st edition.

List of Symbols

$\mathit{epi}(f)$ – epigraph , 684
$\max(A), \max A$ – maximum of A , 672
$\min(A), \min A$ – minimum of A , 672

Index

ε–grid, 654
ε–net, 654

algebraic dual, 677
Axiom of Choice, 673

compact, 664
- covering compact, 664
- sequentially, 661

concave-up, 684
convex, 684
cover, 664
covering, 663
- extract finite open subcovering property, 664

dual, 677
- algebraic, 677

dual function, 677
dual mapping, 677

epigraph, 684
extract finite open subcovering property, 664

finite subcover, 664
finite subcovering, 664
function
- dual function, 677

grid point, 654

isomorphism
- linear, 678
- vector space, 678

Lebesgue number, 666
linear isomorphism, 678

mapping
- dual mapping, 677

maximal element, 672
maximum, 672
minimal element, 672
minimum, 672

open cover, 664
open covering, 664
- Lebesgue number, 666

POset
- maximal element, 672
- maximum, 672
- minimal element, 672
- minimum, 672

sequence compact, 661
sequentially compact, 661
set
- compact, 664
- cover, 664
- covering, 663
- finite subcover, 664
- finite subcovering, 664
- open cover, 664
- open covering, 664

sublinear functional, 679

totally bounded, 654

vector space
- algebraic dual, 677
- dual, 677

vector space isomorphism, 678

ZL property (Zorn's Lemma), 673
Zorn's Lemma, 673