# Convex sets in Euclidean space

Alvin Šipraga*

June 28, 2013

**Abstract**

Convex analysis concerns itself with—among other things—the study of convex sets, which distinguish themselves by being particularly well-behaved mathematical objects. In this text we give an introduction to the notion of convex sets in Euclidean space, and develop the theory through the study of various important principles in convex analysis. We motivate the discussion by providing enough background to enjoy some of the classical theorems of convex analysis.

## Contents



## List of Figures



* ~~as586@bath.ac.uk~~ alvin@pqrs.dk, Department of Mathematical Sciences, University of Bath

# 1 Convex and affine sets

This first chapter constitutes the majority of the background and theory. We shall be studying only subsets of finite dimensional Euclidean space $\mathbb{R}^n$, where $n > 0$ is arbitrary unless otherwise stated. We begin by introducing convex sets, and the closely related concept of an affine set. The rest of the chapter discusses some fundamental concepts in convex analysis, some classical theorems, and their applications. This provides the necessary background for the second chapter.

## 1.1 Convex sets[1]

**Definition 1.** Let $x, y \in \mathbb{R}^n$. The *line segment* with endpoints $x$ and $y$ is the set

$$[x, y] := \{\lambda x + (1 - \lambda)y \mid 0 \leq \lambda \leq 1\}.$$

If $x \neq y$, then the interior of $[x, y]$ is the set

$$(x, y) := \{\lambda x + (1 - \lambda)y \mid 0 < \lambda < 1\}.$$

A point $z \in (x, y)$ is said to be an *interior point* of the line segment $[x, y]$. The sets $[x, y)$ and $(x, y]$ are defined similarly.

**Definition 2.** A set $S \subseteq \mathbb{R}^n$ is said to be *convex* if

$$x, y \in S \implies [x, y] \subseteq S.$$

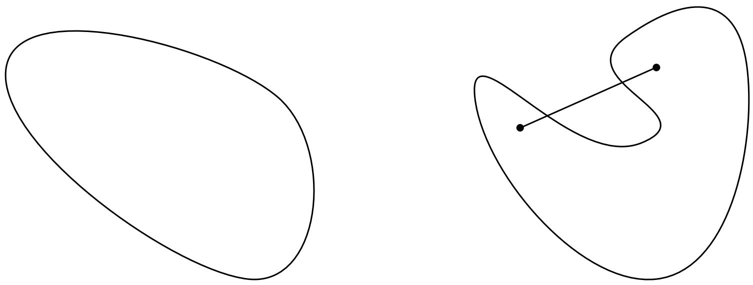



Figure 1: A simple example of two subsets of $\mathbb{R}^2$, one convex and one not convex.

**Theorem 1.** *Let $\{S_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of convex sets. Then their intersection $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ is convex.*

*Proof.* Let $x, y \in \bigcap_{\alpha \in \mathcal{A}} S_\alpha$ be arbitrary. Then $x, y \in S_\alpha$ for all $\alpha \in \mathcal{A}$, and by convexity, $[x, y] \subseteq S_\alpha$ for all $\alpha \in \mathcal{A}$. Hence $[x, y] \subseteq \bigcap_{\alpha \in \mathcal{A}} S_\alpha$ and the intersection is convex. $\square$

**Remark 1.** The same result does not hold for arbitrary unions of convex sets. Indeed, the union of two convex sets is not necessarily convex, e.g. the union $[-\infty, 0) \cup (0, \infty]$ of convex subsets (intervals) of $\mathbb{R}$.

[1] *For further treatment of this material, see also:* [1, §2], [3, §2], [4, §1], [6, §2], [8, §2,4].

**Definition 3.** A *convex combination* of the points $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$ is a point in $\mathbb{R}^n$ of the form

$$\sum_{i=1}^{m} \lambda_i x_i,$$

where $\lambda_i \geq 0$ for all $i \in \{1, \ldots, m\}$, and $\sum_{i=1}^{m} \lambda_i = 1$.

**Remark 2.** (a) We have that $0 \leq \lambda_i \leq 1$ for all $i \in \{1, \ldots, m\}$ (cf. Definition 1), since $\sum_{i=1}^{m} \lambda_i = 1$ is the sum of $m$ nonnegative terms.

(b) For $m = 2$, the above definition is analogous to the definition of a line segment (Definition 1) – the line segment $[x_1, x_2]$ (for $x_1, x_2 \in \mathbb{R}^n$) is the set of all convex combinations of $x_1, x_2$.

**Theorem 2.** *Let $S \subseteq \mathbb{R}^n$. Then $S$ is convex if, and only if, $S$ contains every convex combination of its elements.*

*Proof.* ($\Longleftarrow$). If $S$ contains every convex combination of its elements, then it contains every convex combination of pairs of elements in $S$. Then by Remark 2(b) it is clear that $S$ is convex.

($\Longrightarrow$). Suppose that $S$ is convex, and let

$$x = \sum_{i=1}^{m} \lambda_i x_i$$

be a convex combination of elements $x_1, \ldots, x_m \in S$, $m > 2$. As $S$ is convex, it contains all line segments between pairs of points in $S$, which is equivalent to containing all convex combinations of pairs of elements in $S$. We proceed by induction on $m$.

Assume that $S$ contains all convex combinations of fewer than $m$ elements in $S$. At least one $\lambda_i$ is not 1 (as otherwise $\sum_{i=1}^{m} \lambda_i = m \neq 1$). Without loss of generality (*wlog*), say $\lambda_1 \neq 1$. Set

$$y = \sum_{i=2}^{m} \lambda_i' x_i,$$

where $\lambda_i' := \lambda_i/(1 - \lambda_1)$. Then $\lambda_i' \geq 0$ for all $i \in \{2, \ldots, m\}$ (as $\lambda_i \geq 0$ for all $i \in \{1, \ldots, m\}$ (by Remark 2(a)) and $1 - \lambda_i > 0$), and

$$\sum_{i=2}^{m} \lambda_i' = \frac{\lambda_2 + \cdots + \lambda_m}{1 - \lambda_1} = \frac{\lambda_2 + \cdots + \lambda_m}{\lambda_2 + \cdots + \lambda_m} = 1.$$

Hence $y$ is a convex combination of $m - 1$ elements in $S$, so $y \in S$ by the inductive hypothesis. Then observe that

$$x = \lambda_1 x_1 + (1 - \lambda_1) y \in S,$$

since $x_1, y \in S$. As $x$ was an arbitrary convex combination of elements in $S$, we have shown that $S$ contains all convex combinations of its elements. $\square$

**Definition 4.** Let $S \subseteq \mathbb{R}^n$. The *convex hull* of $S$, denoted by $\mathrm{co}(S)$, is the smallest convex subset of $\mathbb{R}^n$ containing $S$.

**Remark 3.** The convex hull is well-defined. By Theorem 1 we have that the intersection of two convex sets is itself convex, and so there must be some notion of minimality. In particular, the convex hull of the $S$ is the intersection of every convex set containing $S$.

**Theorem 3.** *Let $S \subset \mathbb{R}^n$. $\text{co}(S)$ is the set of all convex combinations of elements in $S$.*

*Proof.* Let $C$ denote the set of all convex combinations of elements in $S$. $\text{co}(S)$ is convex and contains $S$ by definition, so by Theorem 2 $\text{co}(S)$ contains every convex combination of $S$, i.e. $C \subseteq \text{co}(S)$. Conversely, let $x, y \in C$ be convex combinations of the form

$$x = \sum_{i=1}^{m} \lambda_i x_i \qquad \text{and} \qquad y = \sum_{i=1}^{r} \mu_i y_i$$

for $x_1, \ldots, x_m, y_1, \ldots, y_r \in S$. Then for $0 \leq \lambda \leq 1$, the convex combination

$$(1-\lambda)x + \lambda y = \sum_{i=1}^{m} (1-\lambda)\lambda_i x_i + \sum_{i=1}^{r} \lambda \mu_i y_i$$

is a convex combination of elements in $S$ and therefore contained in $C$. Hence $C$ is convex (as $[x, y] \subseteq C$) and contains $S$, so $\text{co}(S) \subseteq C$ by the minimality of the convex hull. We have shown that $\text{co}(S) = C$. □

This theorem is very useful because it allows us to define the convex hull in terms of elements in the given set (whereas the original definition of the convex hull is intuitive, but perhaps less useful directly). We can think of the convex hull of $S \subseteq \mathbb{R}^n$ as the area enclosed by the points in $S$ (see Figure 2).

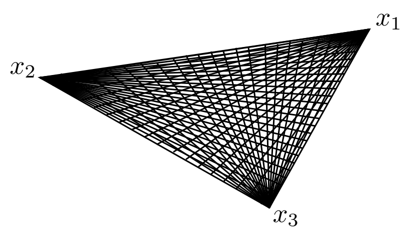



Figure 2: The convex hull of $S = \{x_1, x_2, x_3\} \subseteq \mathbb{R}^2$.

## 1.2 Affine sets[2]

**Definition 5.** Let $a \in \mathbb{R}^n$, and $L$ be a vector subspace of $\mathbb{R}^n$. A set of the form $a + L \subseteq \mathbb{R}^n$ is called an *affine set*.

**Proposition 4.** *All affine sets are convex.*

*Proof.* Consider an affine subset of the form $a + L$ ($a \in \mathbb{R}^n$, $L$ a vector subspace of $\mathbb{R}^n$). Choose $x, y \in a + L$ arbitrarily and define $x', y' \in L$ such that $x = a + x'$

[2] *See also:* [1, §1], [3, §2], [4, §1], [6, §1], [8, §2].

and $y = a + y'$. Noting that $L$ is closed under vector addition and scalar multiplication, we get:

$$\begin{aligned}[x, y] &= \{\lambda x + (1-\lambda)y \mid 0 \le \lambda \le 1\} \\ &= \{\lambda(a + x') + (1-\lambda)(a + y') \mid 0 \le \lambda \le 1\} \\ &= \{\lambda a + \lambda x' + a + y' - \lambda a - \lambda y' \mid 0 \le \lambda \le 1\} \\ &= \{a + \underbrace{\lambda x' + (1-\lambda)y'}_{\in L} \mid 0 \le \lambda \le 1\} \\ &\subseteq a + L.\end{aligned}$$

So $a + L$ is convex. □

**Remark 4.** In the above proof, we may replace $\mathbb{R}^n$ with any vector space $V$ and the result still holds.

**Definition 6.** An *affine combination* of the points $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$ is a point in $\mathbb{R}^n$ of the form

$$\sum_{i=1}^{m} \lambda_i x_i,$$

where $\lambda_i \in \mathbb{R}$ for all $i \in \{1, \ldots, m\}$, and $\sum_{i=1}^{m} \lambda_i = 1$ (cf. Definition 3).

**Remark 5.** Whereas a convex combination of two points $x, y \in \mathbb{R}^n$ lies in the line segment between $x$ and $y$, an affine combination of $x$ and $y$ has the weaker condition that it must lie anywhere on the line containing $x$ and $y$ (see Figure 3). The next theorem (much like Theorem 2) extends this idea beyond just lines.

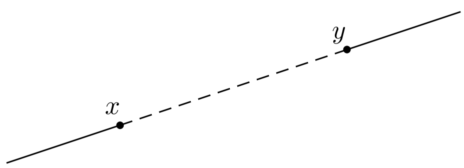



Figure 3: The line segment between $x$ and $y$ (denoted by a dashed line) must contain the convex combination, whereas the affine combination may lie anywhere on the line through $x$ and $y$.

**Theorem 5.** *Let $S \subseteq \mathbb{R}^n$. Then $S$ is affine if, and only if, $S$ contains every affine combination of its elements.*

*Proof.* The argument is analogous to the one seen in the proof to Theorem 2. □

**Theorem 6.** *Let $\{S_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of affine sets. Then their intersection $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ is affine.*

*Proof.* Let $x, y \in \bigcap_{\alpha \in \mathcal{A}} S_\alpha$ be arbitrary. Then $x, y \in S_\alpha$ for all $\alpha \in \mathcal{A}$, and by affineness,

$$\lambda x + (1-\lambda)y \in S_\alpha \qquad \forall\, \lambda \in \mathbb{R},$$

for all $\alpha \in \mathcal{A}$. Then

$$\lambda x + (1-\lambda)y \in \bigcap_{\alpha \in \mathcal{A}} S_\alpha \qquad \forall\, \lambda \in \mathbb{R},$$

so the intersection contains every affine combination of its elements, hence (by Theorem 5) it is affine. □

**Definition 7.** Let $S \subseteq \mathbb{R}^n$. The *affine hull* of $S$, denoted by $\mathrm{aff}(S)$, is the smallest affine subset of $\mathbb{R}^n$ containing $S$.

**Remark 6.** By the same argument as the one given in Remark 3 (but now with reference to Theorem 6), it follows that the affine hull is also well-defined. Note similarly that the affine hull of a set $S$ is the intersection of every affine set containing $S$.

**Definition 8.** Let $S \subseteq \mathbb{R}^n$ and suppose its affine hull is $\mathrm{aff}(S) = a + L$ for some $a \in \mathbb{R}^n$, $L$ a vector subspace of $\mathbb{R}^n$. The *dimension* of $S$, denoted by $\dim(S)$, is defined as follows:

$$\dim(S) := \dim(\mathrm{aff}(S)) := \dim(L).$$

**Remark 7.** (a) This notion of dimension is general for all sets in $\mathbb{R}^n$, and in particular it allows us to talk about the dimension of convex sets.

(b) It is worth justifying well-definedness of the above definition. Remark 6 tells us that the affine hull is well-defined. Moreover, an affine set in $\mathbb{R}^n$ is uniquely determined by a vector subspace $L$ of $\mathbb{R}^n$, and an element $a \in \mathbb{R}^n$ "modulo $L$", i.e.:

$$a + L = (a + x) + L \qquad \forall\, x \in L.$$

Hence it is clear that the dimension of an affine set is uniquely determined by its underlying vector subspace.

**Theorem 7.** *Let $S \subset \mathbb{R}^n$. $\mathrm{aff}(S)$ is the set of all affine combinations of elements in $S$.*

*Proof.* The argument is analogous to the one seen in the proof to Theorem 3. □

**Definition 9.** The convex hull of a finite number of points is called a *convex polytope*. By Theorem 3, this can be written:

$$\mathrm{co}(x_1, x_2, \ldots, x_m) = \left\{ \sum_{i=1}^m \lambda_i x_i \;\middle|\; \lambda_i \geq 0\ \forall\, i \in \{1, \ldots, m\},\ \sum_{i=1}^m \lambda_i = 1 \right\}.$$

Similarly, by Theorem 7 the affine hull of a finite number of points (which has no special name) can be written:

$$\mathrm{aff}(x_1, x_2, \ldots, x_m) = \left\{ \sum_{i=1}^m \lambda_i x_i \;\middle|\; \sum_{i=1}^m \lambda_i = 1 \right\}.$$

**Remark 8.** Figure 2 is an example of a convex polytope. The diagram suggests that there may be some level of compactness to convex polytopes and/or convex hulls in $\mathbb{R}^n$. We shall answer this question in §1.3.

**Proposition 8.** *Let* $S \subseteq \mathbb{R}^n$*. Then* $\text{aff}(\text{co}(S)) = \text{aff}(S)$.

*Proof.* $\text{co}(S)$ contains every convex combination of elements in $S$. A convex combination is also an affine combination, and $\text{aff}(S)$ contains every affine combination of elements in $S$, so $\text{co}(S) \subseteq \text{aff}(S)$ and hence

$$\text{aff}(\text{co}(S)) \subseteq \text{aff}(\text{aff}(S)) = \text{aff}(S).$$

Moreover, $S \subseteq \text{co}(S)$, so

$$\text{aff}(S) \subseteq \text{aff}(\text{co}(S)).$$

□

**Definition 10.** The points $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$ are said to be *affinely independent* if $\dim(\text{aff}(x_1, x_2, \ldots, x_m)) = m - 1$. If they are not affinely independent, then they are said to be *affinely dependent*.

**Proposition 9.** *The points* $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$ *are affinely independent if, and only if,*

$$\sum_{i=1}^{m} \alpha_i x_i = 0 \quad \textit{and} \quad \sum_{i=1}^{m} \alpha_i = 0 \qquad \Longrightarrow \qquad \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0,$$

*where* $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$.

*Proof.* Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$. Observe *wlog* that $\dim(\text{aff}(x_1, x_2, \ldots, x_m)) = m - 1 \iff x_2 - x_1, \ldots, x_m - x_1$ are linearly independent.

($\Longrightarrow$). Suppose that $x_1, x_2, \ldots, x_m$ are affinely independent. Let

$$\sum_{i=1}^{m} \alpha_i x_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{m} \alpha_i = 0,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$. Then

$$\alpha_1 = -(\alpha_2 + \cdots + \alpha_m),$$

and we may write

$$\begin{aligned} 0 = \sum_{i=1}^{m} \alpha_i x_i &= -(\alpha_2 + \cdots + \alpha_m)x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m \\ &= \alpha_2(x_2 - x_1) + \cdots + \alpha_m(x_m - x_1). \end{aligned}$$

By linear independence of $x_2 - x_1, \ldots, x_m - x_1$, it follows that

$$\alpha_2 = \cdots = \alpha_m = 0,$$

and hence also $\alpha_1 = 0$.

($\Longleftarrow$). Suppose that

$$\sum_{i=1}^{m} \alpha_i x_i = 0 \quad \text{and} \quad \sum_{i=1}^{m} \alpha_i = 0 \qquad \Longrightarrow \qquad \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0, \quad (1)$$

where $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$. Now consider the linear combination

$$\alpha_2(x_2 - x_1) + \cdots + \alpha_m(x_m - x_1) = 0,$$

where $\alpha_2, \dots, \alpha_m \in \mathbb{R}$, and set

$$\alpha_1 := -(\alpha_2 + \cdots + \alpha_m).$$

Then

$$\begin{aligned} 0 &= \alpha_2(x_2 - x_1) + \cdots + \alpha_m(x_m - x_1) \\ &= -(\alpha_2 + \cdots + \alpha_m)x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m \\ &= \sum_{i=1}^{m} \alpha_i x_i, \end{aligned}$$

and by construction,

$$\sum_{i=1}^{m} \alpha_i = 0.$$

So by (1),

$$\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0,$$

and $x_2 - x_1, \dots, x_m - x_1$ are linearly independent. Therefore, $x_1, x_2, \dots, x_m$ are affinely independent. □

**Remark 9.** Note that both the above proposition and Definition 10 refer to at least three points. We encounter problems in attempting to generalise for two distinct points, namely:

$$\alpha_1 + \alpha_2 = 0 \implies \alpha_1 = -\alpha_2,$$

and if

$$\alpha_1 x_1 + \alpha_2 x_2 = 0$$

then we reach the contradiction

$$\alpha_1 x_1 = \alpha_1 x_2 \implies x_1 = x_2.$$

**Theorem 10.** *Any subset of $\mathbb{R}^n$ consisting of at least $n+2$ distinct points is affinely dependent.*

*Proof.* Suppose that $x_1, \dots, x_m$ are distinct points in $\mathbb{R}^n$. Then *wlog* the $m-1$ points $x_2 - x_1, \dots, x_m - x_1$ are linearly dependent, and so there exist $\alpha_2, \dots, \alpha_m \in \mathbb{R}$ not all zero, such that

$$\sum_{i=2}^{m} \alpha_i (x_i - x_1) = 0.$$

Then in particular,

$$\begin{aligned} 0 &= \sum_{i=2}^{m} \alpha_i (x_i - x_1) \\ &= -(\alpha_2 + \cdots + \alpha_m)x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m, \end{aligned}$$

Set

$$\alpha_1 := -(\alpha_2 + \cdots + \alpha_m),$$

so that

$$\sum_{i=1}^{m} \alpha_i x_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{m} \alpha_i = 0.$$

As $\alpha_1, \dots, \alpha_m$ are not all zero, it follows that $x_1, \dots, x_m$ are affinely dependent. □

**Proposition 11.** *If $x_1, \dots, x_m \in \mathbb{R}^n$ are affinely independent, then every element of $\operatorname{aff}(x_1, \dots, x_m)$ can be written uniquely as an affine combination of $x_1, \dots, x_m$.*

*Proof.* Let $x \in \operatorname{aff}(x_1, \dots, x_m)$. By Theorem 7 we may write $x$ as an affine combination of $x_1, \dots, x_m$. Consider two such representations of $x$:

$$x = \sum_{i=1}^{m} \lambda_i x_i \qquad \text{and} \qquad x = \sum_{i=1}^{m} \lambda_i' x_i, \tag{2}$$

for some $\lambda_1, \dots, \lambda_m, \lambda_1', \dots, \lambda_m' \in \mathbb{R}$, where $\sum_{i=1}^{m} \lambda_i = 1$ and $\sum_{i=1}^{m} \lambda_i' = 1$.

Observe that

$$\sum_{i=1}^{m} (\lambda_i - \lambda_i') x_i = \sum_{i=1}^{m} \lambda_i x_i - \sum_{i=1}^{m} \lambda_i' x_i = x - x = 0,$$

and

$$\sum_{i=1}^{m} \lambda_i - \lambda_i' = \sum_{i=1}^{m} \lambda_i - \sum_{i=1}^{m} \lambda_i' = 1 - 1 = 0.$$

By Proposition 9 it follows that

$$\lambda_i = \lambda_i' \qquad \forall\, i \in \{1, \dots, m\}.$$

So the two representations in (2) are identical, whence uniqueness. □

## 1.3 Carathéodory's theorem and compactness[3]

We now introduce a theorem that has some powerful implications for convex sets, in particular with respect to compactness. It was discovered in 1907 by the Greek mathematician Constantin Carathéodory, who in 1904 completed his thesis at Göttingen under the supervision of Minkowski, another key player in the subject (as we shall see later).

Constantin Carathéodory (1873–1950)

[3] *See also:* [1, §2], [3, §2], [4, §1], [6, §17], [8, §4].

**Theorem 12** (Carathéodory)**.** *Let $S \subseteq \mathbb{R}^n$. Every $x \in \text{co}(S)$ can be expressed as a convex combination of $n+1$ or fewer elements of $S$.*

*Proof.* By Theorem 3, every element of $\text{co}(S)$ can be written as a convex combination of elements in $S$. Suppose that

$$x = \sum_{i=1}^{m} \lambda_i x_i \in \text{co}(S)$$

is a convex combination of elements $x_1, \ldots, x_m \in S$ for $m > n+1$. Observe that $x_1, \ldots, x_m$ must be affinely dependent, since $m > n+1$ and so their affine hull would have to have dimension $n+1$ or greater (depending on $m$) to ensure linear independence, and this is not possible for $S \subseteq \mathbb{R}^n$.

Note that $m \geq 3$. Now by affine dependence of $x_1, \ldots, x_m$ there exist (cf. Proposition 9) $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ not all zero, such that

$$\sum_{i=1}^{m} \alpha_i x_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{m} \alpha_i = 0.$$

It follows that

$$x = \sum_{i=1}^{m} (\lambda_i + t\alpha_i) x_i \qquad \forall\, t \in \mathbb{R}. \tag{3}$$

Set $t = \tilde{t} := -\lambda_k/\alpha_k$ for some $k \in \mathbb{N}$ such that $\alpha_k \neq 0$. Then by (3),

$$x = \sum_{i=1}^{m} (\lambda_i + \tilde{t}\alpha_i) x_i = \sum_{\substack{i=1 \\ i \neq k}}^{m} (\lambda_i + \tilde{t}\alpha_i) x_i,$$

which is a convex combination of $m-1$ elements $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_m \in S$.

If $m = n+2$ then the above argument shows that $x$ can be written as a convex combination of $m-1 = n+1$ elements. Hence any convex combination of elements of $S$ can be reduced to a convex combination of at least $n+1$ elements of $S$. This completes the proof. □

The following corollary is also known as Carathéodory's theorem, and is a direct result of the above statement.

**Corollary 13** (Carathéodory)**.** *Let $S \subseteq \mathbb{R}^n$, and set $m := \dim(S)$. Then $\text{co}(S)$ is the set of all convex combinations of precisely $m+1$ elements in $S$.*

*Proof.* Any convex combination has at most $m+1$ terms by Theorem 12. If the number of terms is $m+1$, then we are done. If not, terms of the form $\lambda x$ ($\lambda = 0$ and $x \in S$) can always be added to the convex combination to make up $m+1$ terms. □

**Remark 10.** Carathéodory's theorem can not be improved upon in general: consider for example a point in the centre of the convex hull described in Figure 2. It does not lie on any of the line segments between points in $S = \{x_1, x_2, x_3\}$ (which are themselves convex hulls of two points), and therefore must be a convex combination of all three points. Since we are working in $\mathbb{R}^2$, this shows that $n+1$ is the general upper bound for this result.

**Definition 11.** A set $S \subseteq \mathbb{R}^n$ is said to be:

(a) *sequentially compact* if every sequence in $S$ has a convergent subsequence;

(b) *compact* is every open cover of $S$ has a finite subcover.

Sequential compactness and compactness are equivalent for all subsets of $\mathbb{R}^n$ (assuming the standard Euclidean metric when talking about convergence).

**Remark 11.** (a) The traditional definition of openness that is referred to above breaks down if $S$ has a lesser dimension to $\mathbb{R}^n$, and we should in fact speak of "relative openness": this is addressed in §1.5, but is not important for now.

(b) The Heine-Borel theorem states that a subset of $\mathbb{R}^n$ (the only kind we consider in this text) is compact if, and only if, it is closed and bounded. We do not use this result directly, but it aids in visualising compactness in this context. The Bolzano-Weierstrass theorem gives the same result in terms of sequential compactness.

**Theorem 14.** *Let $S \subseteq \mathbb{R}^n$ be compact. Then* $\mathrm{co}(S)$ *is also compact.*

*Proof.* Consider the sequence $(y_k)_{k\in\mathbb{N}}$ of points in $\mathrm{co}(S)$. We prove that $(y_k)_{k=1}$ has a convergent subsequence, thereby establishing compactness of $\mathrm{co}(S)$.

Let $m := \dim(S)$. By Corollary 13, every $y_k$ can be written as a convex combination

$$y_k = \sum_{i=1}^{m+1} \lambda_{k,i} x_{k,i},$$

where $x_{k,1}, \ldots, x_{k,(n+1)} \in S$.

Now consider the sequences

$$(x_{k,1})_{k\in\mathbb{N}},\ \ldots,\ (x_{k,(m+1)})_{k\in\mathbb{N}} \tag{4}$$

of points in $S$, and the sequences

$$(\lambda_{k,1})_{k\in\mathbb{N}},\ \ldots,\ (\lambda_{k,(m+1)})_{k\in\mathbb{N}} \tag{5}$$

of scalars in $[0,1]$. By compactness of $S$, there exists an infinite set $\Lambda_1 \subseteq \mathbb{N}$ such that $(x_{k,1})_{k\in\Lambda_1}$ is convergent. Now consider the sequence $(x_{k,2})_{k\in\Lambda_1}$. Again by compactness of $S$, there exists an infinite set $\Lambda_2 \subseteq \Lambda_1$ such that $(x_{k,2})_{k\in\Lambda_2}$.

Continuing this construction on the remainder of sequences in (4), and then on those in (5) (this time exploiting the compactness of $[0,1]$ rather than $S$), we get the index sets

$$\underbrace{\Lambda_1 \supseteq \Lambda_2 \supseteq \cdots \supseteq \Lambda_{m+1}}_{\text{by compactness of } S} \supseteq \underbrace{\Lambda_{m+2} \supseteq \cdots \supseteq \Lambda_{2m+1} \supseteq \Lambda_{2m+2}}_{\text{by compactness of } [0,1]},$$

such that $(x_{k,i})_{k\in\Lambda_i}$ and $(\lambda_{k,i})_{k\in\Lambda_{i+(m+1)}}$ are convergent for all $i \in \{1, \ldots, m+1\}$.

We want all of these sequences to converge simultaneously, and in fact it suffices to use the sequences defined by the index set $\Lambda_{2m+2}$. For clarity, we write the indices explicitly. Set $k_1 := \min \Lambda_{2m+2}$ and define inductively

$$k_{\ell+1} := \min\{k' \in \Lambda_{2m+2} \mid k' > k_\ell\} \qquad \forall\, \ell \in \mathbb{N} \setminus \{1\}.$$

Hence we have convergent sequences

$$(x_{k_\ell,1})_{\ell\in\mathbb{N}},\ \ldots,\ (x_{k_\ell,(m+1)})_{\ell\in\mathbb{N}},$$
$$(\lambda_{k_\ell,1})_{\ell\in\mathbb{N}},\ \ldots,\ (\lambda_{k_\ell,(m+1)})_{\ell\in\mathbb{N}}.$$

In particular, we may now take common limits of these sequences with respect to $\ell$, so for each $i \in \{1, \ldots, m+1\}$, set

$$x_{0,i} := \lim_{\ell\to\infty} x_{k_\ell,i},$$
$$\lambda_{0,i} := \lim_{\ell\to\infty} \lambda_{k_\ell,i}.$$

Observe that as

$$\sum_{i=1}^{m+1} \lambda_{k_\ell,i} = 1 \qquad \forall\, \ell \in \mathbb{N},$$

it follows that

$$\sum_{i=1}^{m+1} \lambda_{0,i} = 1.$$

Moreover, each $x_{0,i} \in S$ for all $i \in \{1, \ldots, m+1\}$ as $S$ is closed (since it is compact), and therefore,

$$y_0 := \sum_{i=1}^{m+1} \lambda_{0,i} x_{0,i}$$

is a convex combination, and $y_0 \in \mathrm{co}(S)$ by Theorem 3. By algebra of limits, clearly

$$\lim_{\ell\to\infty} y_{k_\ell} = y_0,$$

so the subsequence $(y_{k_\ell})_{\ell\in\mathbb{N}}$ of $(y_k)_{k\in\mathbb{N}}$ is convergent, whence compactness of $\mathrm{co}(S)$. □

We are now in a position to address the observation made in Remark 8 on the compactness of convex polytopes in $\mathbb{R}^n$.

**Corollary 15.** *All convex polytopes in $\mathbb{R}^n$ are compact.*

*Proof.* A convex polytope is the convex hull of a finite set. Any finite set is compact, so the result follows immediately from Theorem 14. □

## 1.4 Affine mappings[4]

Affine mappings are a useful tool in the study of convexity, as we shall see shortly. In words, an affine mapping "preserves lines".

**Definition 12.** A map $\varphi : S \subseteq \mathbb{R}^n \to \mathbb{R}^d$ is an *affine mapping* if it preserves affine combinations, i.e.

$$\varphi\left(\sum_{i=1}^{m} \lambda_i x_i\right) = \sum_{i=1}^{m} \lambda_i \varphi(x_i),$$

where $x_1, \ldots, x_m \in S$, $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, $\sum_{i=1}^{m} \lambda_i = 1$.

[4] *See also:* [1, §1], [4, §1], [6, §1].

**Theorem 16.** *Every affine mapping is continuous.*

*Proof.* Not given. See [1, p. 8]. ■

**Lemma 17.** *Let $S \subseteq \mathbb{R}^n$ be convex, and let $\varphi : S \to \mathbb{R}^d$ be an affine mapping. Then*

*(a) $\varphi(S) \subseteq \mathbb{R}^d$ is convex, and*

*(b) if $C \subseteq \mathbb{R}^d$ is convex, then $\varphi^{-1}(C) := \{x \in \mathbb{R}^n \mid \varphi(x) \in C\} \subseteq \mathbb{R}^n$ is convex.*

*Proof.* (a) Let $x, y \in S$. As $\varphi$ is an affine mapping,

$$\varphi([x, y]) = [\varphi(x), \varphi(y)]. \tag{6}$$

Convexity of $S$ implies that $[x, y] \subseteq S$, hence

$$\varphi(x), \varphi(y) \in \varphi(S). \tag{7}$$

Combining (6) and (7), it is clear that $\varphi([x, y]) \subseteq \varphi(S)$, and so $\varphi(S)$ is convex.

(b) If $\varphi^{-1}(C) = \emptyset$, then we are done, so suppose that $\varphi^{-1}(C)$ is non-empty. Let $x', y' \in C$ such that $x' = \varphi(x)$ and $y' = \varphi(y)$ for some $x, y \in \varphi^{-1}(C) \subseteq S$. Then

$$\underbrace{[x', y'] \subseteq C}_{\text{by convexity of } C} \implies [x, y] = \varphi^{-1}([x', y']) \subseteq \varphi^{-1}(C),$$

whence convexity of $\varphi^{-1}(S)$. □

**Remark 12.** The above Lemma lets us think about affine mappings as preserving convexity. In particular, straight lines are preserved, i.e. if $\varphi : \mathbb{R}^n \to \mathbb{R}^d$ is an affine mapping and $x, y \in \mathbb{R}^n$, $z \in [x, y]$, then $\varphi(z) \in \varphi([x, y]) = [\varphi(x), \varphi(y)]$.

**Theorem 18.** *Let $S_1, \ldots, S_k$ with $S_i \subseteq \mathbb{R}^{n_i}$. Then $S_1 \times \cdots \times S_k \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ is convex if, and only if, $S_1, \ldots, S_k$ are convex.*

*Proof.* ($\Longrightarrow$). Suppose that $S := S_1 \times \cdots \times S_k$ is convex. Consider the mappings $\varphi_i : S \to S_i$,

$$\varphi_i(x_1, \ldots, x_k) = x_i \qquad \forall\, i \in \{1, \ldots, k\}.$$

These projections are clearly affine mappings, and so by Lemma 17(a) it follows that each $S_i = \varphi_i(S)$ is convex.

($\Longleftarrow$). Suppose that $S_1, \ldots, S_k$ are convex and define $S := S_1 \times \cdots \times S_k$. Let $x, y \in S$ such that

$$x = (x_1, \ldots, x_k) \qquad \text{and} \qquad y = (y_1, \ldots, y_k),$$

where $x_i, y_i \in S_i$ for all $i \in \{1, \ldots, k\}$. Then

$$\lambda x + (1 - \lambda) y = (\lambda x_1, \ldots, \lambda x_k) + ((1 - \lambda) y_1, \ldots, (1 - \lambda) y_k) \in S,$$

whence convexity of $S$. □

**Corollary 19.** *Let $S_1, S_2 \subseteq \mathbb{R}^n$ be convex. Then*

$$S_1 + S_2 := \{x_1 + x_2 \mid x_1 \in S_1,\, x_2 \in S_2\}$$

*is convex.*

*Proof.* As $S_1$ and $S_2$ are convex, $S_1 \times S_2$ is convex by Theorem 18. Define the affine mapping $\varphi : S_1 \times S_2 \to \mathbb{R}^n$, $\varphi(x_1, x_2) = x_1 + x_2$. Then $C_1 + C_2 = \varphi(C_1 \times C_2)$ is convex by Lemma 17(a). □

**Remark 13.** The sum $S_1 + S_2$ as defined above is known as the *Minkowski sum* of $S_1$ and $S_2$. $S_1 - S_2$ is analogous:

$$S_1 - S_2 := \{x_1 - x_2 \mid x_1 \in S_1,\, x_2 \in S_2\}.$$

## 1.5 Relative interior of convex sets[5]

It is useful to characterise the interior of a convex set $S \subseteq \mathbb{R}^n$, but the usual definition of interior is insufficient if the dimension of $S$ is less than $n$ (as noted in Remark 11(a)). Recall that a point $x \in S \subseteq \mathbb{R}^n$ is in the interior of $S$ (denoted by $\mathring{S}$) if there exists a real number $\delta > 0$ such that $B_\delta(x) \subseteq S$, where $B_\delta(x)$ denotes a ball in $\mathbb{R}^n$ of radius $\delta$ and centre $x$. Clearly,

$$\dim(S) < n \implies \mathring{S} = \emptyset.$$

This motivates the notion of relative interior.

**Definition 13.** Let $S \subseteq \mathbb{R}^n$.

(a) The *relative interior* of $S$, denoted by $\mathrm{ri}(S)$, is the interior of $S$ regarded as a subset of $\mathrm{aff}(S)$. Specifically,

$$\mathrm{ri}(S) := \{x \in \mathrm{aff}(S) \mid \exists\, \delta > 0 : B_\delta(x) \cap \mathrm{aff}(S) \subseteq S\}.$$

(b) As well as relative interior, we have the notion of *relative openness* ($S$ is open in $\mathrm{aff}(S)$). However, there is no difference between "relative closure" and closure. The closure of $S$ is denoted by $\overline{S}$.

(c) The *relative boundary* of $S$, denoted by $\mathrm{rb}(S)$, is the boundary of $S$ regarded as a subset of $\mathrm{aff}(S)$. Specifically,

$$\mathrm{rb}(S) := \overline{S} \setminus \mathrm{ri}(A).$$

**Remark 14.** (a) For $x, y \in \mathbb{R}^n$, $\mathrm{ri}([x, y]) = (x, y)$ (cf. Definition 1).

(b) For $S \subseteq \mathbb{R}^n$, $\mathrm{ri}(S) \subseteq S \subseteq \overline{S}$.

**Theorem 20.** *Let $S \subseteq \mathbb{R}^n$ be convex and non-empty. Then* $\mathrm{ri}(S) \neq \emptyset$.

[5] *See also:* [1, §3], [3, §2], [4, §2], [6, §6], [8, §4].

*Proof.* Set $m := \dim(S)$. Then there exists an affinely independent collection of elements $x_1, \dots, x_{m+1} \in S$ (cf. Definition 10). Set

$$C := \operatorname{co}(x_1, \dots, x_{m+1}).$$

By Proposition 8,

$$\operatorname{aff}(C) = \operatorname{aff}(x_1, \dots, x_{m+1}) = \left\{ \sum_{i=1}^{m+1} \lambda_i x_i \;\middle|\; \sum_{i=1}^{m+1} \lambda_i = 1 \right\} \subseteq \mathbb{R}^n. \tag{8}$$

Proposition 11 tells us that every element in $\operatorname{aff}(C)$ can be written as a unique affine combination in $x_1, \dots, x_{m+1}$, and so we may define an injective mapping

$$\varphi : \operatorname{aff}(C) \to \mathbb{R}^{m+1},$$

where

$$\varphi\left( \sum_{i=1}^{m+1} \lambda_i x_i \right) := (\lambda_1, \dots, \lambda_{m+1}).$$

This is an affine mapping, and so continuous by Theorem 16. Now set

$$K_i := \left\{ (\lambda_1, \dots, \lambda_{m+1}) \in \mathbb{R}^{m+1} \;\middle|\; \lambda_i > 0, \sum_{j=1}^{m+1} \lambda_j = 1 \right\} \qquad \forall\, i \in \{1, \dots, m+1\}.$$

The sets $K_1, \dots, K_{m+1}$ are clearly open in $\mathbb{R}^{m+1}$, so by continuity of $\varphi$, the sets

$$\varphi^{-1}(K_1),\ \dots,\ \varphi^{-1}(K_{m+1})$$

are open in $\operatorname{aff}(C)$. Then their intersection

$$\bigcap_{i=1}^{m+1} \varphi^{-1}(K_i) \tag{9}$$

is also open in $\operatorname{aff}(C)$. Observe that

$$\bigcap_{i=1}^{m+1} \varphi^{-1}(K_i) = \left\{ \sum_{i=1}^{m+1} \lambda_i x_i \;\middle|\; \lambda_i > 0 \;\forall\, i \in \{1, \dots, m+1\}, \sum_{i=1}^{m+1} \lambda_i = 1 \right\} \neq \emptyset.$$

Moreover, every element in the set (9) is a convex combination and therefore

$$\emptyset \neq \bigcap_{i=1}^{m+1} \varphi^{-1}(K_i) \subseteq \operatorname{co}(x_1, \dots, x_{m+1}) = C.$$

Hence $C$ contains a non-empty, relatively open (that is, open in $\operatorname{aff}(C)$) set, i.e. $\operatorname{ri}(C) \neq \emptyset$.

Now, $x_1, \dots, x_{m+1} \in S$ are affinely independent and $\dim(S) = m + 1$, so

$$\operatorname{aff}(x_1, \dots, x_{m+1}) = \operatorname{aff}(S),$$

and combining this with (8) we get in particular that

$$\operatorname{aff}(C) = \operatorname{aff}(S).$$

Therefore, $\operatorname{ri}(C) \neq \emptyset$ implies that $C$ has a non-empty interior relative to $\operatorname{aff}(S)$. But $C \subseteq S$, so $S$ has a non-empty interior relative to $\operatorname{aff}(S)$, i.e. $\operatorname{ri}(S) \neq \emptyset$. □

**Theorem 21.** *Let $S \subseteq \mathbb{R}^n$ be convex. Then $\overline{S}$ is convex.*

*Proof.* If $S = \emptyset$, then we are done, so suppose that S is non-empty. Let $x, y \in \overline{S}$ and set $u := \lambda x + (1-\lambda)y$, where $0 \le \lambda \le 1$. Consider the (relatively) open ball $B_\delta(u)$. As $x, y \in \overline{S}$, there exist points $x_0 \in B_\delta(x) \cap S$ and $y_0 \in B_\delta(y) \cap S$. Set $u_0 := \lambda x_0 + (1-\lambda)y_0$. Then,

$$\begin{aligned}
|u - u_0| &= |(\lambda x + (1-\lambda)y) - (\lambda x_0 + (1-\lambda)y_0)| \\
&\le |(\lambda x + (1-\lambda)y) - (\lambda x_0 + (1-\lambda)y)| \\
&\qquad + |(\lambda x_0 + (1-\lambda)y) - (\lambda x_0 + (1-\lambda)y_0)| \\
&= \lambda|x - x_0| + (1-\lambda)|y - y_0| \\
&< \lambda\delta + (1-\lambda)\delta \\
&= \delta,
\end{aligned}$$

and hence $u_0 \in B_\delta(u)$. Moreover, $u_0 \in [x_0, y_0] \subseteq S$ (by convexity of $S$) and so $u \in \overline{S}$, whence convexity of $\overline{S}$. □

**Lemma 22.** *Let $S \subseteq \mathbb{R}^n$ be convex, $x \in \mathrm{ri}(S)$, and $y \in \overline{S}$ such that $x \neq y$. Then $[x, y) \subseteq \mathrm{ri}(S)$.*

*Proof.* Set

$$z := \lambda x + (1-\lambda)y$$

for some $0 < \lambda < 1$. As $x \in \mathrm{ri}(S)$, there exists a relatively open set $U \subseteq S$ containing $x$. Now set

$$V := (1-\lambda)^{-1}(z - \lambda U).$$

Observe that

$$(1-\lambda)^{-1} - (1-\lambda)^{-1}\lambda = 1,$$

hence $V \subseteq \mathrm{aff}(S)$ (the sum of coefficients of the linear combination is 1), and $V$ is relatively open (as $U$ is relatively open). Since $x \in U$ and $y = (1-\lambda)^{-1}(z - \lambda x)$, it follows that $y \in V$. As $y \in \overline{S}$, there exists a point $z_1 \in V \cap S$. Set

$$W := \lambda U + (1-\lambda)z_1.$$

Then $W \subset \mathrm{aff}(S)$ is also relatively open, and since $U \subseteq S$ and $z_1 \in S$, convexity of $S$ implies that $W \subseteq S$. From the definition of $V$, there exists a point $z_0 \in U$ such that

$$z_1 = (1-\lambda)^{-1}(z - \lambda z_0).$$

Then

$$z = \lambda z_0 + (1-\lambda)z_1 \in \lambda U + (1-\lambda)z_1 = W \subseteq \mathrm{ri}(S),$$

and as $z$ is an arbitrary point in $(x, y)$ and $x \in \mathrm{ri}(S)$, we conclude that $[x, y) \subseteq \mathrm{ri}(S)$. □

**Theorem 23.** *Let $S \subseteq \mathbb{R}^n$ be convex. Then $\mathrm{ri}(S)$ is convex.*

*Proof.* If $S = \emptyset$, then we are done, so assume that $S$ is non-empty. Let $x, y \in \mathrm{ri}(S)$. If $x = y$, then $[x, y] = \{x\} \subseteq \mathrm{ri}(S)$. Otherwise note that $y \in \mathrm{ri}(S) \subseteq \overline{S}$, and so by Lemma 22, $[x, y) \subseteq \mathrm{ri}(S)$ and hence $[x, y] = [x, y) \cup \{y\} \subseteq \mathrm{ri}(S)$. □

**Theorem 24.** *Let $S \subseteq \mathbb{R}^n$ be convex. Then*

*(a)* $\overline{S} = \overline{\mathrm{ri}(S)}$*, and*

*(b)* $\mathrm{ri}(S) = \mathrm{ri}(\overline{S})$*.*

*Proof.* As before, the results are trivial if $S = \emptyset$, so assume that $S$ is non-empty.

(a) Clearly $\overline{\mathrm{ri}(S)} \subseteq \overline{S}$ (cf. Remark 14). Conversely, let $y \in \overline{S}$ and $x \in \mathrm{ri}(S)$ (cf. Theorem 20). If $x = y$, then

$$y \in \mathrm{ri}(S) \subseteq \overline{\mathrm{ri}(S)},$$

and we are done. If $x \neq y$, then by Lemma 22,

$$[x, y) \subseteq \mathrm{ri}(S).$$

As every neighbourhood of $y$ contains points from $[x, y)$, it follows that $y \in \overline{\mathrm{ri}(S)}$, hence $\overline{S} \subseteq \overline{\mathrm{ri}(S)}$.

(b) Note that

$$\mathrm{aff}(S) = \mathrm{aff}(\overline{S}), \tag{10}$$

as $\mathrm{aff}(S)$ is closed. Then clearly $\mathrm{ri}(S) \subseteq \mathrm{ri}(\overline{S})$. On the other hand, let $x \in \mathrm{ri}(\overline{S})$ and choose a point $x_0 \in \mathrm{ri}(S)$ (cf. Theorem 20). If $x_0 = x$, then $x \in \mathrm{ri}(S)$ and we are done. If $x \neq x_0$, then $\mathrm{aff}(x_0, x)$ is a line in $\mathbb{R}^n$, and, more specifically,

$$\mathrm{aff}(x_0, x) \subseteq \mathrm{aff}(\overline{S}) \overset{(10)}{=} \mathrm{aff}(S).$$

Since $x \in \mathrm{ri}(\overline{S})$, there exists a point $x_1 \in \mathrm{aff}(x_0, x)$ such that $x_1 \in \overline{S}$, and $x \in (x_0, x_1)$. Then Theorem 20 yields $x \in \mathrm{ri}(S)$, hence $\mathrm{ri}(\overline{S}) \subseteq \mathrm{ri}(S)$. □

**Theorem 25.** *Let $S \subseteq \mathbb{R}^n$ be convex, and let $\varphi : \mathbb{R}^n \to \mathbb{R}^d$ be an affine mapping. Then*

*(a)* $\varphi(\overline{S}) \subseteq \overline{\varphi(S)}$*, and*

*(b)* $\mathrm{ri}(\varphi(S)) = \varphi(\mathrm{ri}(S))$*.*

*Proof.* The results hold trivially for $S = \emptyset$, so assume $S \neq \emptyset$.

(a) Let $y \in \varphi(\overline{S})$ and set $x := \varphi^{-1}(y) \in \overline{S}$ ($\varphi$ need not necessarily be injective – in this context $\varphi^{-1}(y)$ denotes any element of the set $\{x \in \overline{S} \mid \varphi(x) = y\}$). If $x \in S$, then we are done. If $x \in \overline{S} \setminus S$, then there exists a sequence $(x_k)_{k \in \mathbb{N}}$ in $S$ such that

$$x = \lim_{k \to \infty} x_k.$$

Define the sequence $(y_k)_{k \in \mathbb{N}}$ in $\varphi(S)$ by $y_k := \varphi(x_k)$. By Theorem 16, $\varphi$ is continuous, and so

$$\lim_{k \to \infty} y_k = \lim_{k \to \infty} \varphi(x_k) = \varphi\left(\lim_{k \to \infty} x_n\right) = \varphi(x) = y \in \overline{\varphi(S)}.$$

Hence $\varphi(\overline{S}) \subseteq \overline{\varphi(S)}$.

(b) By Theorem 23, ri($S$) is convex and so we may apply part (a):

$$\varphi(\overline{\mathrm{ri}(S)}) \subseteq \overline{\varphi(\mathrm{ri}(S))}. \tag{11}$$

Combining this with Theorem 24(a) and Remark 14(b) yields

$$\overline{\varphi(\mathrm{ri}(S))} \overset{(11)}{\supseteq} \varphi(\overline{\mathrm{ri}(S)}) \overset{\mathrm{T24}(a)}{=} \varphi(\bar{S}) \overset{\mathrm{R14}(b)}{\supseteq} \varphi(S) \overset{\mathrm{R14}(b)}{\supseteq} \varphi(\mathrm{ri}(S)). \tag{12}$$

Then,

$$\overline{\varphi(S)} \overset{(a)}{\supseteq} \varphi(\bar{S}) \overset{(12)}{\supseteq} \varphi(\mathrm{ri}(S)),$$

and taking closures (note this has no effect on the left hand side),

$$\overline{\varphi(S)} \supseteq \overline{\varphi(\mathrm{ri}(S))}. \tag{13}$$

Also,

$$\overline{\varphi(\mathrm{ri}(S))} \overset{(12)}{\supseteq} \varphi(S),$$

and taking closures again,

$$\overline{\varphi(\mathrm{ri}(S))} \supseteq \overline{\varphi(S)}. \tag{14}$$

Combining (13) and (14), we get

$$\overline{\varphi(\mathrm{ri}(S))} = \overline{\varphi(S)}. \tag{15}$$

By Lemma 17(a), $\varphi(S)$ and $\varphi(\mathrm{ri}(S))$ are convex, so we may apply Theorem 24(b) to get

$$\mathrm{ri}(\varphi(S)) \overset{\mathrm{T24}(b)}{=} \mathrm{ri}(\overline{\varphi(S)}) \overset{(15)}{=} \mathrm{ri}(\overline{\varphi(\mathrm{ri}(S))}) \overset{\mathrm{T24}(b)}{=} \mathrm{ri}(\varphi(\mathrm{ri}(S))) \overset{\mathrm{R14}(b)}{\subseteq} \varphi(\mathrm{ri}(S)).$$

So, in particular,

$$\mathrm{ri}(\varphi(S)) \subseteq \varphi(\mathrm{ri}(S)). \tag{16}$$

Conversely, recall that $\mathrm{ri}(S) \neq \emptyset$ and $\mathrm{ri}(\varphi(S)) \neq \emptyset$ by Theorem 20. Let $z \in \varphi(\mathrm{ri}(S))$ and set $z' := \varphi^{-1}(z) \in \mathrm{ri}(S)$. Let $y \in \mathrm{ri}(\varphi(S))$ and set $y' := \varphi^{-1}(y) \in S$ (this is well-defined because $\mathrm{ri}(\varphi(S)) \subseteq \varphi(S)$). As $z' \in \mathrm{ri}(S)$, there exists a point $x' \in S$ such that $z' \in (x', y')$. Set $x := \varphi(x') \in \varphi(S)$. We then have $z \in (x, y)$ (cf. Remark 12). Noting that $x \in \overline{\varphi(S)}$ and $y \in \varphi(S)$, Lemma 22 gives

$$(x, y) \subseteq \mathrm{ri}(\varphi(S)),$$

and hence

$$\varphi(\mathrm{ri}(S)) \subseteq \mathrm{ri}(\varphi(S)). \tag{17}$$

Combining (16) and (17), we conclude that

$$\varphi(\mathrm{ri}(S)) = \mathrm{ri}(\varphi(S)). \qquad \square$$

**Lemma 26.** *Let $S_1, S_2 \subseteq \mathbb{R}^n$. Then*

*(a)* $\operatorname{aff}(S_1 \times S_2) = \operatorname{aff}(S_1) \times \operatorname{aff}(S_2)$, *and*

*(b)* $\operatorname{ri}(S_1 \times S_2) = \operatorname{ri}(S_1) \times \operatorname{ri}(S_2)$.

*Proof.* (a) This follows immediately from Theorem 7 and some simple manipulation of sets. Here we permit the slightly ambiguous notation $(1 \leq i \leq m)$, $i \in \mathbb{N}$ being omitted, to denote $\forall\, i \in \{1, \dots, m\}$.

$$\begin{aligned}
\operatorname{aff}(S_1 \times S_2) &= \left\{ \sum_{i=1}^{m} \lambda_i x_i \;\middle|\; m \in \mathbb{N},\, x_i \in S_1 \times S_2 \ (1 \leq i \leq m), \sum_{i=1}^{m} \lambda_i = 1 \right\} \\
&= \Bigg\{ \sum_{i=1}^{m} \lambda_i (x_i^{(1)}, x_i^{(2)}) \;\Bigg|\; m \in \mathbb{N},\, x_i^{(1)} \in S_1, \\
&\qquad\qquad x_i^{(2)} \in S_2 \ (1 \leq i \leq m), \sum_{i=1}^{m} \lambda_i = 1 \Bigg\} \\
&= \Bigg\{ \left( \sum_{i=1}^{m} \lambda_i x_i^{(1)}, \sum_{i=1}^{m} \lambda_i x_i^{(2)} \right) \;\Bigg|\; m \in \mathbb{N},\, x_i^{(1)} \in S_1, \\
&\qquad\qquad x_i^{(2)} \in S_2 \ (1 \leq i \leq m), \sum_{i=1}^{m} \lambda_i = 1 \Bigg\} \\
&= \left\{ \sum_{i=1}^{m} \lambda_i x_i \;\middle|\; m \in \mathbb{N},\, x_i \in S_1 \ (1 \leq i \leq m), \sum_{i=1}^{m} \lambda_i = 1 \right\} \\
&\qquad \times \left\{ \sum_{i=1}^{m} \lambda_i x_i \;\middle|\; m \in \mathbb{N},\, x_i \in S_2 \ (1 \leq i \leq m), \sum_{i=1}^{m} \lambda_i = 1 \right\} \\
&= \operatorname{aff}(S_1) \times \operatorname{aff}(S_2).
\end{aligned}$$

(b) In lieu of part (a), this result follows immediately from the definition of the relative interior (Definition 13(a)).

□

**Theorem 27.** *Let $S_1, S_2 \subseteq \mathbb{R}^n$ be convex. Then* $\operatorname{ri}(S_1 + S_2) = \operatorname{ri}(S_1) + \operatorname{ri}(S_2)$.

*Proof.* Define the affine mapping $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$,

$$\varphi(x, y) = x + y. \tag{18}$$

$S_1 + S_2$ is convex by Corollary 19, so

$$\begin{aligned}
\operatorname{ri}(S_1 + S_2) &= \operatorname{ri}(\varphi(S_1 \times S_2)) && \text{(by (18))} \\
&= \varphi(\operatorname{ri}(S_1 \times S_2)) && \text{(by Theorem 25(b))} \\
&= \varphi(\operatorname{ri}(S_1) \times \operatorname{ri}(S_2)) && \text{(by Lemma 26(b))} \\
&= \operatorname{ri}(S_1) + \operatorname{ri}(S_2), && \text{(by (18))}
\end{aligned}$$

and we are done. □

## 1.6 Hyperplanes and separation[6]

Hyperplanes are a generalisation of planes in arbitrary dimension Euclidean space. This leads to the notion of separation, which we shall see demonstrates the pleasant behaviour of convex sets particularly well.

A hyperplane in $\mathbb{R}^3$ is a plane, in $\mathbb{R}^2$ it is a line, and in $\mathbb{R}$ it is a point. The following definition formalises the concept.

**Definition 14.** A set $H \subseteq \mathbb{R}^n$ is a *hyperplane* in $\mathbb{R}^n$ if it is a dimension $n-1$ affine subset of $\mathbb{R}^n$.

**Remark 15.** Every hyperplane $H$ in $\mathbb{R}^n$ can be written in the form

$$H = a + V,$$

where $a \in \mathbb{R}^n$, and $V \subseteq \mathbb{R}^n$ is a dimension $n-1$ vector subspace of $\mathbb{R}^n$.

**Theorem 28.** *For every hyperplane $H$ in $\mathbb{R}^n$, there exists an affine mapping $\vartheta : \mathbb{R}^n \to \mathbb{R}$ such that $H$ is a level set of $\vartheta$, i.e. $H = \vartheta^{-1}(\tau)$ for some $\tau \in \mathbb{R}$. Moreover, the mapping is linear.*

*Proof.* Let $H = a + V$ be a hyperplane in $\mathbb{R}^n$ (assume the conditions on $a$ and $V$ as in Remark 15). As $V$ has dimension $n-1$, there exists a point $p \in \mathbb{R}^n \setminus V$ such that

$$\mathbb{R}^n = V + \mathbb{R}p = \{v + \lambda p \mid v \in V,\ \lambda \in \mathbb{R}\}.$$

Define the mapping $\vartheta : \mathbb{R}^n \to \mathbb{R}$ where

$$\vartheta(x) = \vartheta(v + \lambda p) := \lambda \qquad \forall\, x \in \mathbb{R}^n = V + \mathbb{R}p. \tag{19}$$

Note that $\vartheta$ is defined for a fixed $p \in \mathbb{R}^n$. Then

$$\begin{aligned} x \in H &\iff x = a + v \qquad \text{for } v \in V \\ &\iff \vartheta(x) = \vartheta(a), \end{aligned}$$

and so we may write

$$H = \vartheta^{-1}(\vartheta(a)) = \{x \in \mathbb{R}^n \mid \vartheta(x) = \vartheta(a)\}.$$

Typically we set $\tau := \vartheta(a) \in \mathbb{R}$ and write

$$H = \vartheta^{-1}(\tau).$$

We now proceed to show that $\vartheta$ is both a linear and an affine mapping.

($\vartheta$ is linear). Let $x, y \in \mathbb{R}^n$ such that

$$x := x' + \lambda p \qquad \text{and} \qquad y := y' + \mu p,$$

where $x', y' \in V$, $\lambda, \mu \in \mathbb{R}$, and $p \in \mathbb{R}^n \setminus V$ as defined above. Let $\alpha \in \mathbb{R}$. Then

$$\vartheta(\beta x) = \vartheta(\alpha x' + \alpha\lambda p) = \alpha\lambda = \alpha\vartheta(x),$$

and also,

$$\vartheta(x + y) = \vartheta(x' + y' + (\lambda + \mu)p) = \lambda + \mu = \vartheta(x) + \vartheta(y),$$

[6] *See also:* [1, §4], [3, §3,4], [4, §1,2], [6, §1,11], [8, §3,4].

whence linearity.

($\vartheta$ is affine). Let $x_1, \ldots, x_m \in \mathbb{R}^n$ such that

$$x_i := x_i' + \lambda_i p \qquad \forall\, i \in \{1, \ldots, m\},$$

where $x_1', \ldots, x_m' \in V$, $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, and $\sum_{i=1}^m \lambda_i = 1$. Then

$$\begin{aligned}\vartheta\left(\sum_{i=1}^m \mu_i x_i\right) &= \vartheta\left(\sum_{i=1}^m \mu_i (x_i' + \lambda_i p)\right)\\ &= \vartheta\left(\sum_{i=1}^m \mu_i x_i' + \sum_{i=1}^m \mu_i \lambda_i p\right)\\ &= \sum_{i=1}^m \mu_i \lambda_i = \sum_{i=1}^m \mu_i \vartheta(x_i' + \lambda_i p) = \sum_{i=1}^m \mu_i \vartheta(x_i),\end{aligned}$$

whence affineness. □

**Remark 16.** Whenever we talk about a hyperplane described by a mapping (by convention always denoted $\vartheta$), we assume (safely, by the above theorem) that it is of the form seen in (19).

**Definition 15.** The *orthogonal projection* from $\mathbb{R}^n$ onto $S \subseteq \mathbb{R}^n$ is a mapping that sends each $x \in \mathbb{R}^n$ to a unique point $y \in S$ such that $(x - y) \perp S \iff \forall\, z \in S : \langle (x - y), z \rangle = 0$, where $\langle\,\cdot\,,\,\cdot\,\rangle : \mathbb{R}^n \to \mathbb{R}$ is the inner product.

**Lemma 29.** *Let $S \subseteq \mathbb{R}^n$ be convex, relatively open, and non-empty, and let $x \in \mathbb{R}^n \setminus S$. Then there exists a hyperplane $H$ in $\mathbb{R}^n$ such that $x \in H$ and $H \cap S = \emptyset$.*

*Proof.* Clearly if $S = \mathbb{R}^n$, then no such $x$ exists and there is nothing to prove, so assume that $S \neq \mathbb{R}^n$. We proceed by induction on $n$, but first deal with the trivial case $n = 1$, and the case $n = 2$ (which then forms the inductive base case).

($n = 1$). As $S \neq \mathbb{R}$, it must be an open interval in $\mathbb{R}$. If $S = (x, y) \subseteq \mathbb{R}$ for $x \in \mathbb{R}$, $y \in \mathbb{R} \cup \{\infty\}$ then $H = \{x\}$ is a suitable hyperplane (hyperplanes in $\mathbb{R}$ are just single points), and if $S = (x, y) \subseteq \mathbb{R}$ for $x \in \mathbb{R} \cup \{-\infty\}$, $y \in \mathbb{R}$ then choose $H = \{y\}$.

($n = 2$). Let $C \subseteq \mathbb{R}^2$ be a circle with centre $x$. Set

$$C' := C \cap \bigcup_{u \in S} \underbrace{\{\lambda u + (1 - \lambda)x \mid \lambda > 0\}}_{\text{halfline from } x \text{ through } u}.$$

The set $C'$ is an open (due to the relative openness of $S$) arc in $C$ (see Figure 4). As $x \notin S$, and $S$ is convex, no two opposite points of $C$ can be in $C'$. Hence, the angle between the two halflines from $x$ through the endpoints of $C'$ ($L_1$ and $L_2$ in Figure 4) is at most $\pi$. Any of the two lines determined by $L_1$ or $L_2$ is then a suitable candidate for $H$. If the angle is $\pi$, then these two lines coincide.

($n > 2$). Assume that the statement holds for all dimensions less than $n$. Let $A \subseteq \mathbb{R}^n$ be a dimension 2 affine subspace such that $x \in A$ and $A \cap S \neq \emptyset$. Then $A \cap S$ is convex (Theorem 1 – intersection of convex sets is convex), relatively open (open in $A$), non-empty, and $x \neq A \cap S$. Identifying $A$ with $\mathbb{R}^2$, this reduces to the above case, and so there exists a line (the hyperplane in $A$) $L \subseteq A$ such that $x \in L$ and $L \cap (A \cap S) = \emptyset$. As $L \subseteq A$, it follows that

$$L \cap S = \emptyset. \tag{20}$$

Now, let $B$ be a hyperplane in $\mathbb{R}^n$ that is orthogonal to $L$, and let $\pi : \mathbb{R}^n \to B$ be the orthogonal projection onto $B$. Then $\pi(S)$ is convex, relatively open (open in $B$), and non-empty. Moreover, as $\pi^{-1}(\pi(x)) = L$ and $x \notin S$, we have $\pi(x) \notin \pi(S)$. The dimension of $B$ is $n-1$, so by the inductive hypothesis there exists a hyperplane $H' \subset B$ such that $\pi(x) \in H'$ and

$$H' \cap \pi(S) = \emptyset. \tag{21}$$

Then we have found a suitable hyperplane, namely:

$$H := \pi^{-1}(H') = \operatorname{aff}(H' \cup L) \subseteq \mathbb{R}^n,$$

where $x \in H$ (since $x \in L$), and $H \cap S = \emptyset$ (by (20) and (21)).

This completes the proof. □

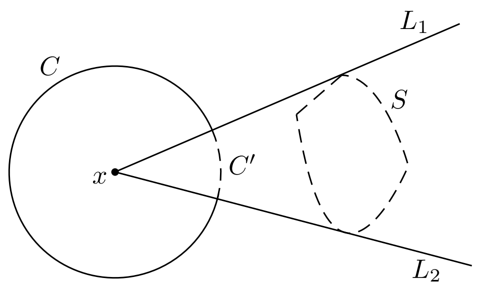



Figure 4: Illustration of the proof to Lemma 29 for the case $n = 2$. The sets $C'$ and $S$ are dashed to emphasise openness (in $C$ and $\operatorname{aff}(S)$ respectively).

**Notation.** We write $\vartheta(X) \leq \tau$ as shorthand for $\forall\, x \in X : \vartheta(x) \leq \tau$, and similarly for $\vartheta(X) \geq \tau$.

**Definition 16.** Let $A, B \subseteq \mathbb{R}^n$ and let $H = \vartheta^{-1}(a)$ be a hyperplane in $\mathbb{R}^n$. $H$ is said to *separate* $A$ and $B$ if

(a) $\vartheta(A) \leq \tau$ and $\vartheta(B) \geq \tau$, or

(b) $\vartheta(A) \geq \tau$ and $\vartheta(B) \leq \tau$.

$H$ is said to *separate* $A$ and $B$ *properly* if $H$ separates $A$ and $B$, and there exists a point $x \in A \cup B$ such that $\vartheta(x) \neq \tau$ (i.e. $H$ does not contain both $A$ and $B$ in their entirety).

**Theorem 30** (Separation). *Let $S_1, S_2 \subseteq \mathbb{R}^n$ be convex and non-empty. Then there exists a hyperplane in $\mathbb{R}^n$ that properly separates $S_1$ and $S_2$ if, and only if,* $\operatorname{ri}(S_1) \cap \operatorname{ri}(S_2) = \emptyset$.

*Proof.* Let $A = S_1 - S_2$. Then $A$ is convex by Corollary 19, and $\mathrm{ri}(A) = \mathrm{ri}(S_1) - \mathrm{ri}(S_2)$ by Theorem 27. Observe that

$$\mathrm{ri}(S_1) \cap \mathrm{ri}(S_2) = \emptyset \iff 0 \notin \mathrm{ri}(A). \tag{22}$$

($\Longleftarrow$). Suppose that $\mathrm{ri}(S_1) \cap \mathrm{ri}(S_2) = \emptyset$. Let $B = \mathrm{ri}(A)$. Then $B$ is relatively open (open in $\mathrm{aff}(B)$), convex (by Theorem 23), and non-empty (by Theorem 20), and also $0 \notin B$ (by (22)). By Lemma 29, there exists a hyperplane

$$H = \vartheta^{-1}(0)$$

such that $0 \in H$ and $H \cap B = \emptyset$. As $B$ is convex and $H \cap B = \emptyset$, either $\vartheta(B) > 0$, or $\vartheta(B) < 0$. We treat the former case and assume that $\vartheta(B) > 0$. The latter case follows similarly. As $\vartheta(\mathrm{ri}(B)) = \vartheta(B) > 0$, it follows that $\vartheta(A) \geq 0$, so

$$\forall\, x_1 \in S_1,\, x_2 \in S_2 : \vartheta(x_1) \geq \vartheta(x_2),$$

and

$$\exists\, x_1 \in S_1,\, x_2 \in S_2 : \vartheta(x_1) > \vartheta(x_2),$$

as equality would imply $\vartheta(A) = 0$, but we have $\vartheta(B) > 0$ and $\emptyset \neq B \subseteq A$, a contradiction ※. Now set

$$z := \inf\{\vartheta(x_1) \mid x_1 \in S_1\}.$$

$S_1$ and $S_2$ are then separated properly by the hyperplane $\vartheta^{-1}(z)$.

($\Longrightarrow$). Suppose that $H = \vartheta^{-1}(z)$ is a hyperplane that separates $S_1$ and $S_2$ properly. Then assume *wlog* that $\vartheta(S_1) \geq z$, $\vartheta(S_2) \leq z$, and $\vartheta(x_1) > z$ for some $x_1 \in S_1$ (other cases follow similarly, cf. Definition 16). We have $\vartheta(A) \geq 0$ and $\vartheta(a) > 0$ for some $a \in A$. Let $x \in \mathrm{ri}(A)$. Then there exists a $\delta > 0$ such that

$$[x, x + \delta(x - a)] \subseteq A,$$

which implies that

$$\vartheta(x + \delta(x - a)) \geq 0.$$

By the linearity of $\vartheta$, we have

$$(1 + \delta)\vartheta(x) \geq \delta\vartheta(a).$$

As $\vartheta(a) > 0$ and $\delta > 0$, this means $\vartheta(x) > 0$, and so we conclude that

$$\vartheta(\mathrm{ri}(A)) > 0,$$

hence $0 \notin \mathrm{ri}(A) \overset{(22)}{\Longrightarrow} \mathrm{ri}(S_1) \cap \mathrm{ri}(S_2)$. □

**Definition 17.** Let $S \subseteq \mathbb{R}^n$ and let $x \in \mathrm{rb}(S)$. A hyperplane $H = \vartheta^{-1}(\tau)$ in $\mathbb{R}^n$ is said to be a *supporting hyperplane* for $S$ at $x$ if

(a) $x \in H$, and

(b) $\vartheta(S) \leq \tau$, or $\vartheta(S) \geq \tau$.

Moreover, it is said to be *non-trivial* if $A \not\subseteq H$.

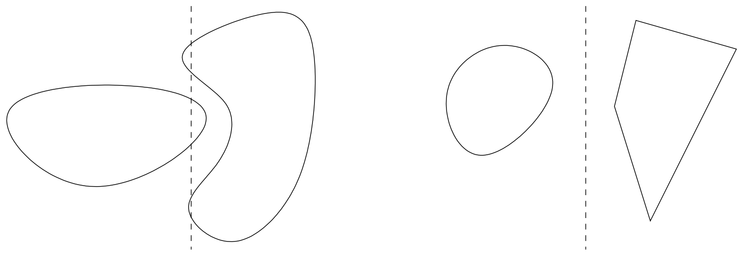


Figure 5: The well-behaved nature of convex sets is particularly evident in this example in $\mathbb{R}^2$. It is a necessary condition for the separation theorem above for obvious reasons, and the properties of convexity are used extensively in the proof.

**Theorem 31.** *Let $S \subseteq \mathbb{R}^n$ be convex and non-empty, and let $x \in \mathrm{rb}(S)$. Then there exists a non-trivial supporting hyperplane for $S$ at $x$.*

*Proof.* By Theorem 30, there exists a hyperplane $H = \vartheta^{-1}(\tau)$ in $\mathbb{R}^n$ that separates $\{x\}$ and $S$ properly. Assume *wlog* that $\vartheta(S) \leq \tau$ and $\vartheta(x) \geq \tau$. Since $x \in \mathrm{rb}(S)$, we have that $\vartheta(x) \leq \tau$, hence $\vartheta(x) = \tau$ and $x \in H$. Now, as the separation is proper, we have $S \cup \{x\} \not\subseteq H$, hence $S \not\subseteq H$. Thus $H$ is a non-trivial supporting hyperplane for $S$ at $x$. □

## 1.7 Extreme points and Minkowski's theorem[7]

**Definition 18.** Let $S \subseteq \mathbb{R}^n$. A point $x \in S$ is called an *extreme point* if it is not an interior point of any line segment in $A$.

This section relates the theorem of Hermann Minkowski, one of the fathers of modern convex analysis. In words, it can be described as follows: compact convex sets are "spanned" by their extreme points. Here, span means roughly the same as in linear algebra, but is restricted to convex combinations rather than all linear combinations.

Hermann Minkowski (1864–1909)

**Lemma 32.** *Let $S \subseteq \mathbb{R}^n$ be convex and bounded, containing at least two points. Then $\mathrm{ri}(S) \subseteq \mathrm{co}(\mathrm{rb}(S))$.*

[7] *See also:* [4, §2], [8, §4].

*Proof.* Let $x \in \mathrm{ri}(S)$ (cf. Theorem 20). As there are at least two points in $S$, there exists a point $y \in S$, $y \neq x$. Let $L \subseteq \mathbb{R}^n$ be the line through $x$ and $y$. The set $L \cap S$ is bounded, and contains (by convexity of $S$) a line segment with $x$ as an interior point. Then there exist points $p, q \in L \cap \mathrm{rb}(S)$ such that $x \in [p, q] \subseteq \mathrm{co}(\mathrm{rb}(S))$. We have shown that $\mathrm{ri}(S) \subseteq \mathrm{co}(\mathrm{rb}(S))$. □

**Theorem 33** (Minkowski)**.** *Let $S \subseteq \mathbb{R}^n$ be compact, convex, and non-empty. Let $E$ be the set of extreme points of $S$. Then $E \neq \emptyset$, and $S = \mathrm{co}(E)$.*

*Proof.* Let $d = \dim(S)$. We induct on $d$. If $d = 0$, there $S$ consists of only one point and there is nothing to prove, so assume that $d > 0$ and that the statement holds for all dimensions less than $d$. As $S$ is bounded, $\mathrm{rb}(S) \neq \emptyset$ and so we may choose a point $x_0 \in \mathrm{rb}(S)$. By Theorem 31, there exists a non-trivial supporting hyperplane $H = \vartheta^{-1}(\tau)$ for $S$ at $x_0$. Assume *wlog* that $\vartheta(S) \leq \tau$ (cf. Definition 17). The intersection $H \cap S$ is compact, convex, non-empty, and $\dim(H \cap S) \leq d$. If $\dim(H \cap S) = d$, then it would be that $S \subseteq H$ (as $H$ is an affine subset of $\mathbb{R}^n$), but $H$ is non-trivial and so this is not the case, so

$$\dim(H \cap S) < d.$$

Now by the induction hypothesis, the set $E_1$ of extreme points of $H \cap S$ is non-empty, and $H \cap S = \mathrm{co}(E_1)$. Let $e \in E_1$. Suppose that there exist points $x, y \in S$ and $\lambda \in \mathbb{R}$, $0 < \lambda < 1$, such that

$$e = \lambda x + (1 - \lambda) y.$$

Since $e \in H$, it follows that $\vartheta(e) = \tau$ and so

$$\vartheta(e) = \vartheta(\lambda x + (1 - \lambda) y) = \lambda \vartheta(x) + (1 - \lambda) \vartheta(y) = \tau.$$

As $x, y \in S$ and $\vartheta(S) \leq \tau$, we have $\vartheta(x) \leq \tau$ and $\vartheta(y) \leq \tau$. Moreover, as $\lambda < 1$ we have $\vartheta(x) = \vartheta(y) = e$, hence $x, y \in H$ also, so

$$x, y \in H \cap S.$$

Since $e \in E_1$ is an extreme point, it follows that $x = y = e$, as otherwise $e \in (x, y) \subseteq H \cap S$, which would be contradictory. Thus $e \in E$, which implies that $E_1 \subseteq E$. $E_1$ is non-empty, so $E \neq \emptyset$.

$S$ is compact and therefore closed (by the Heine-Borel theorem, cf. Remark 11(b)), so we have $x_0 \in H \cap S$, which implies $x_0 \in \mathrm{co}(E_1) \subseteq \mathrm{co}(E)$. As $x_0 \in \mathrm{rb}(S)$ was chosen arbitrarily, we see that

$$\mathrm{rb}(S) \subseteq \mathrm{co}(E). \tag{23}$$

By Lemma 32 we have

$$\mathrm{ri}(S) \overset{\text{L32}}{\subseteq} \mathrm{co}(\mathrm{rb}(S)) \overset{(23)}{\subseteq} \mathrm{co}(E),$$

and it follows that $S \subseteq \mathrm{co}(E)$ (since clearly $S \subseteq \mathrm{co}(\mathrm{rb}(S))$). Conversely, it is obvious that $\mathrm{co}(E) \subseteq S$ (since $E \subseteq S$, and $S$ is convex). We conclude that $S = \mathrm{co}(E)$, and we are done. □

**Corollary 34.** *Let $S \subseteq \mathbb{R}^n$ be convex, compact, and non-empty. Then every $x \in S$ can be expressed as a convex combination of $n+1$ or fewer extreme points of $S$.*

*Proof.* Let $E$ be the set of extreme points of $S$. By Minkowski's theorem (Theorem 33), $S = \mathrm{co}(E)$. Then by Carathéodory's theorem (Theorem 12), every element in $S$ can be expressed as a convex combination of $n+1$ or fewer elements of $E$. □

# 2 Helly's theorem and its applications

We have done the heavy lifting and demonstrated the well-behaved nature of convex sets. Now we conclude by offering some results that further demonstrate their beauty.

Eduard Helly (1884–1943)

Johann Radon (1887–1956)

The main theorem discussed in this chapter was discovered by the Austrian mathematician Eduard Helly in 1913, but was not published until 1921 by fellow Austrian mathematician Johann Radon. During World War I, Helly served as a soldier in the Austrian army, and was taken prisoner in Russia in 1914. It was during this time, or shortly before—sources differ in their accounts—that he communicated his result to Radon. In his 1921 publication, Radon used his own result (now known as Radon's theorem) to prove Helly's theorem. Helly later published his own proof in 1923.

## 2.1 Radon's theorem and Helly's theorem[8]

There are numerous ways to prove Helly's theorem, and here we follow the method of Radon by first proving Radon's theorem, from which Helly's theorem neatly follows.

**Theorem 35** (Radon)**.** *Let $S \subseteq \mathbb{R}^n$ contain at least $n+2$ points. Then $S$ can be partitioned into two disjoint subsets $S_1$ and $S_2$ such that* $\mathrm{co}(S_1) \cap \mathrm{co}(S_2) \neq \emptyset$.

*Proof.* Write $S =: \{x_1, x_2, \ldots, x_m\}$. As $m \geq n+2$, it follows from Theorem 10 that $S$ is affinely dependent. Thus, there exist $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ not all zero, such that

$$\sum_{i=1}^{m} \alpha_i x_i = 0 \qquad \text{and} \qquad \sum_{i=1}^{m} \alpha_i = 0.$$

[8] *See also:* [1, §2], [2, §1,2], [3, §6], [4, §1], [7, §18], [8, §4].

At least two of $\alpha_1, \ldots, \alpha_m$ must have opposite signs, so assume *wlog* that

$$\alpha_1 \geq 0, \ldots, \alpha_k \geq 0, \alpha_{k+1} < 0, \ldots, \alpha_m < 0,$$

for some $k \in \{1, \ldots, m-1\}$. Note that

$$\alpha := \alpha_1 + \cdots + \alpha_k = -(\alpha_{k+1} + \cdots + \alpha_m),$$

and then $\alpha > 0$, so set

$$x := \underbrace{\sum_{i=1}^{k} \frac{\alpha_i}{\alpha} x_i = \sum_{i=k+1}^{m} \frac{-\alpha_i}{\alpha} x_i}_{\text{equal since } \sum_{i=1}^{m} \alpha_i x_i = 0} .$$

Then $x$ is a convex combination of $x_1, \ldots, x_k$ (notice the coefficients sum to 1), so by Theorem 3 we have

$$x \in \operatorname{co}(x_1, \ldots, x_k),$$

and similarly,

$$x \in \operatorname{co}(x_{k+1}, \ldots, x_m).$$

Setting

$$S_1 := \{x_1, \ldots, x_k\} \qquad \text{and} \qquad S_2 := \{x_{k+1}, \ldots, x_m\},$$

we have $S_1 \cap S_2 = \emptyset$, $S_1 \cup S_2 = S$, and $\operatorname{co}(S_1) \cap \operatorname{co}(S_2) \neq \emptyset$ (as at least $x$ is in this intersection). This completes the proof. □

**Theorem 36** (Helly)**.** *Let $\mathscr{F} := \{S_1, \ldots, S_m\}$ be a (finite) family of $m \geq n+1$ convex sets in $\mathbb{R}^n$. If every subfamily of $n+1$ sets in $\mathscr{F}$ has a non-empty intersection, then $\bigcap_{i=1}^{m} S_i \neq \emptyset$.*

*Proof.* We proceed by induction on $m$. If $m = n+1$ then the result holds trivially, so assume $m > n+1$ and that the result holds for families of fewer than $m$ sets. Then it follows immediately that

$$\forall\, i \in \{1, \ldots, m\} : \exists\, x_i \in S_1 \cap \cdots \cap S_{i-1} \cap S_{i+1} \cap \cdots \cap S_m. \tag{24}$$

Since $m \geq n+2$, we may apply Radon's theorem to the set $S := \{x_1, \ldots, x_m\}$, yielding a parition of $S$ into two disjoint subsets $C_1$ and $C_2$ with $\operatorname{co}(C_1) \cap \operatorname{co}(C_2) \neq \emptyset$. We set *wlog*

$$C_1 := \{x_1, \ldots, x_k\} \qquad \text{and} \qquad C_2 := \{x_{k+1}, \ldots, x_m\},$$

for some $k \in \{1, \ldots, m-1\}$, and choose a point

$$x \in \operatorname{co}(x_1, \ldots, x_k) \cap \operatorname{co}(x_{k+1}, \ldots, x_m) = \operatorname{co}(C_1) \cap \operatorname{co}(C_2).$$

We wish to show that $x \in S_i$ for all $i \in \{1, \ldots, m\}$. By (24), we have that

$$\begin{aligned} x_i &\in S_{k+1} \cap \cdots \cap S_m && \forall\, i \in \{1, \ldots, k\}, \\ x_i &\in S_1 \cap \cdots \cap S_k && \forall\, i \in \{k+1, \ldots, m\}, \end{aligned}$$

and since each $S_i$ is convex,

$$\begin{aligned} x &\in \operatorname{co}(x_1, \ldots, x_k) \subseteq S_{k+1} \cap \cdots \cap S_m \\ x &\in \operatorname{co}(x_{k+1}, \ldots, x_m) \subseteq S_1 \cap \cdots \cap S_k, \end{aligned}$$

so we are done. (See Figure 6 for an illustration of this proof applied to four sets in the plane.) □

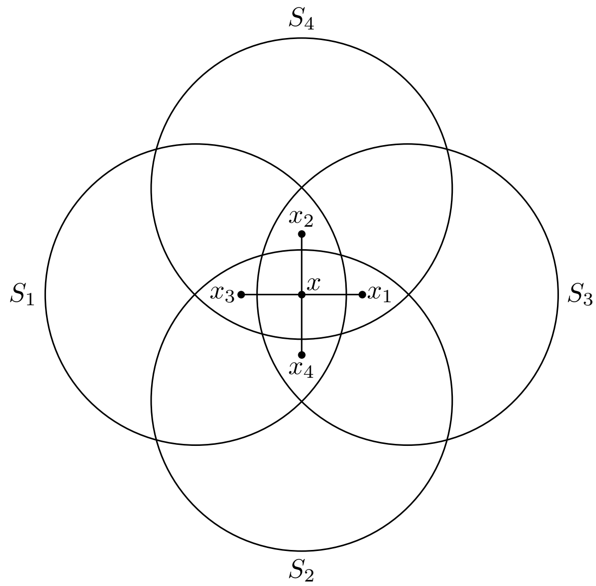



Figure 6: Illustation of the proof to Helly's theorem for four sets in the plane. In this case, $C_1 = \{x_1, x_3\}$ and $C_2 = \{x_2, x_4\}$.

**Remark 17.** Helly's theorem is a classical result in convex analysis, and as such, the condition of convexity is as usual one of the most important necessary conditions. An example of a family of sets that are not all convex, but satisfy all other conditions of Helly's theorem, is illustrated in Figure 7.

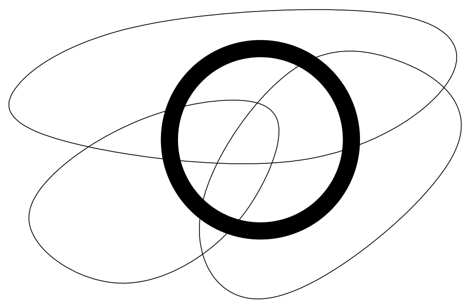


Figure 7: Here, all the sets in the plane satisfy the conditions of Helly's theorem *except* convexity, so the result does not hold in general for all sets in $\mathbb{R}^n$.

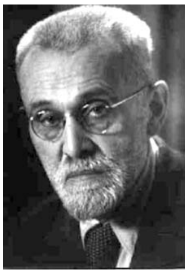
Helly's theorem has an analogue for infinite families of sets, with the extra condition that they are all compact. This is not proven here, but the result is given to complement the discussion in later sections.

**Theorem 37** (Helly)**.** *Let $\mathscr{F} := \{S_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of convex, compact sets in $\mathbb{R}^n$ containing at least $n+1$ members. If every subfamily of $n+1$ sets in $\mathscr{F}$ has a non-empty intersection, then $\bigcap_{\alpha \in \mathcal{A}} S_\alpha \neq \emptyset$.*

The rest of this chapter showcases some interesting applications of Helly's theorem.

## 2.2 Jung's theorem[9]

Take any finite collection of points in the plane with pairwise distance no greater than 1. Then they are contained entirely in a disc of radius $1/\sqrt{3}$. This fact is a particular case of a theorem of the German mathematician Heinrich Jung.



Heinrich Jung (1876–1953)

We shall prove the general case of this theorem for arbitrary dimensions, but first we require one extra definition for the diameter of a given set in $\mathbb{R}^n$, which is how we refer to the "pairwise distance".

**Definition 19.** Let $S \subseteq \mathbb{R}^n$ be non-empty. Then the *diameter* of $S$ is the number

$$\operatorname{diam}(S) := \sup_{x,y \in S} |x - y|.$$

If $S$ is unbounded, then $\operatorname{diam}(S) = \infty$.

**Theorem 38** (Jung)**.** *Let $S \subseteq \mathbb{R}^n$ be finite, and set $m := \operatorname{diam}(S)$. Then $S$ is contained in a closed ball of radius $r \leq d\sqrt{\frac{n}{2(n+1)}}$.*

*Proof.* We split the proof into two parts, depending on the cardinality of the set $S$.

($|S| \leq n+1$). Let $z \in \mathbb{R}^n$ denote the centre of the smallest closed ball in $\mathbb{R}^n$ that contains $S$, and let $r$ be its radius. Such a ball is well-defined, since $S$ is finite and therefore has finite diameter, for which a radius of half the diameter is the upper bound. Set

$$\{x_1, \ldots, x_m\} := \{x \in S \mid |z - x| = r\},$$

where obviously $m \leq n+1$ (this is important later). Notice then that $z \in \operatorname{co}(z_1, \ldots, z_m)$, as otherwise this would contradict the minimality of $r$. Assume *wlog* that $z = 0$. Then we have the convex combination

$$\sum_{i=1}^{m} \lambda_i x_i = 0, \tag{25}$$

where $\lambda_i \geq 0$ for all $i \in \{1, \ldots, m\}$, and $\sum_{i=1}^{m} \lambda_i = 1$. Now, for each $i, j \in \{1, \ldots, m\}$, set

$$d_{i,j} := |x_i - x_j| \overset{\text{D19}}{\leq} d. \tag{26}$$

[9] *See also:* [2, §2], [7, §8].

Then

$$\begin{aligned} d_{i,j}^2 &= |x_i - x_j|^2 \\ &= \langle x_i - x_j, x_i - x_j \rangle \\ &= \langle x_i, x_i \rangle + \langle x_j, x_j \rangle - 2\langle x_i, x_j \rangle \\ &= 2r^2 - 2\langle x_i, x_j, \rangle, \end{aligned} \tag{27}$$

and for each $j \in \{1, \dots, m\}$,

$$\begin{aligned} 1 - \lambda_j = \sum_{\substack{i=1 \\ i \neq j}}^{m} \lambda_i \geq \sum_{i=1}^{m} \lambda_i \frac{d_{i,j}^2}{d^2} & \qquad \text{(by (26), noting that } d_{j,j} = 0) \\ = \frac{2r^2}{d^2} - 2 \left\langle \sum_{i=1}^{m} \lambda_i x_i, x_j \right\rangle \bigg/ d^2 & \qquad \text{(by (27))} \\ = \frac{2r^2}{d^2}. & \qquad \text{(by (25))} \end{aligned}$$

We have

$$\sum_{i=1}^{m} 1 - \lambda_i = m - 1,$$

and also (from the inequality above),

$$m - 1 = \sum_{i=1}^{m} 1 - \lambda_i \geq \sum_{i=1}^{m} \frac{2r^2}{d^2} = \frac{2mr^2}{d^2},$$

hence (recalling that $m \leq n + 1$),

$$r^2 \leq \frac{d^2(m-1)}{2m} \leq \frac{d^2 n}{2(n+1)} \implies r \leq d\sqrt{\frac{n}{2(n+1)}}.$$

($|S| > n + 1$). This case can now be reduced to the previous case by application of Helly's theorem. Set $\rho := d\sqrt{\frac{n}{2(n+1)}}$ and consider for each $x \in S$ the closed ball

$$\overline{B}_\rho(x) := \left\{ y \in \mathbb{R}^n \;\middle|\; |x - y| \leq \rho = d\sqrt{\frac{n}{2(n+1)}} \right\}. \tag{28}$$

The result holds for all sets of $n+1$ or fewer elements, so any set $S' \subseteq S$ of $n+1$ elements of $S$ is contained in a closed ball of radius $\rho$ with centre $z \in \mathbb{R}^n$. Then observe that

$$S' \subseteq \overline{B}_\rho(z) \iff z \in \bigcap_{x \in S'} \overline{B}_\rho(x), \tag{29}$$

and hence the intersection of any $n+1$ of the balls defined in (28) have nonempty intersection. Applying Helly's theorem (Theorem 36) to the family $\mathscr{F} := \{\overline{B}_\rho(x) \mid x \in S\}$ then yields $\bigcap_{x \in S} \overline{B}_\rho(x) \neq \emptyset$. This is equivalent (by the same reasoning employed in (29)) to

$$\exists\, z \in \bigcap_{x \in S} \overline{B}_\rho(x) : S \subseteq \overline{B}_\rho(z),$$

i.e. $S$ is contained in a closed ball of radius $r \leq d\sqrt{\frac{n}{2(n+1)}}$. $\square$

**Remark 18.** (a) Setting $n = 2$ and $d = 1$, the claim at the beginning of this section is immediately clear.

(b) Jung's theorem has an analogue for infinite subsets of $\mathbb{R}^n$, just like the infinite version of Helly's theorem. Again, the extra condition is compactness. The proof for this is identical to the finite case, applying Theorem 37 instead.

## 2.3 Kirchberger's theorem[10]

Steven R. Lay offers in [3, §7] the following useful analogy to understand this theorem. Suppose we have a collection of sheep and wolves standing in a valley, and we wish to build a straight fence through the valley to separate the sheep from the wolves so as to prevent them being eaten. What simple condition ensures that such a fence can be built? In fact, it suffices that if every collection of 4 animals can be separated in this way by a straight fence, then we can build the single straight fence to separate them all like that.

This fact is a consequence of the theorem of the German mathematician Paul Kirchberger (1878–1945). Little seems to be known of the life of Kirchberger, but his result on the separation of points is impressive in that it predates many of the many classical results in the area of convex analysis, including those of Helly and Carathéodory.

Kirchberger's proof, from his 1903 publication (which was also his doctoral thesis, completed at Göttingen under the supervision of Hilbert in 1902), goes on for more than 20 pages. The proof we offer here of Kirchberger's theorem follows the method of Rademacher and Schoenberg [5], which is considerably shorter, and employs Helly's theorem to reach the desired result.

Before we begin, we give an alternate representation of hyperplanes in $\mathbb{R}^n$, and a more restrictive definition of separation.

**Definition 20.** A *hyperplane* in $\mathbb{R}^n$ is a set of the form

$$\{x \in \mathbb{R}^n \mid \langle x, y\rangle = \kappa\} \subseteq \mathbb{R}^n,$$

where $y \in \mathbb{R}^n \setminus \{0\}$, and $\kappa \in \mathbb{R}$.

**Definition 21.** Let $A, B \subseteq \mathbb{R}^n$ and let $H = \{x \in \mathbb{R}^n \mid \langle x, y\rangle = \kappa\}$ be a hyperplane in $\mathbb{R}^n$. $H$ is said to *separate* $A$ and $B$ *strictly* if

(a) $\langle a, y\rangle < \kappa$ and $\langle b, y\rangle > \kappa$ for all $a \in A$, $b \in B$, or

(b) $\langle a, y\rangle > \kappa$ and $\langle b, y\rangle < \kappa$ for all $a \in A$, $b \in B$.

**Theorem 39** (Kirchberger)**.** *Let $A, B \subseteq \mathbb{R}^n$ be finite, and suppose that for every subset $C \subseteq A \cup B$, containing at most $n + 2$ elements, there exists a hyperplane that separates $C \cap A$ and $C \cap B$ strictly. Then there exists a hyperplane that separates $A$ and $B$ strictly.*

*Proof.* Assume that the supposition in the statement of the theorem holds. Then define for each $x \in \mathbb{R}^n$ the sets

$$A(x) := \{(y, \kappa) \in \mathbb{R}^n \times \mathbb{R} \mid \langle x, y\rangle > \kappa\},$$
$$B(x) := \{(y, \kappa) \in \mathbb{R}^n \times \mathbb{R} \mid \langle x, y\rangle < \kappa\}.$$

[10] *See also:* [2, §2], [3, §7], [5], [8, §4].

Notice that $A(x)$ and $B(x)$ are convex sets, each corresponding to the set of all hyperplanes lying to one side or another of a given point $x \in \mathbb{R}^n$. Define the family of subsets of $\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$,

$$\mathscr{F} := \{A(x) \mid x \in A\} \cup \{B(x) \mid x \in B\}.$$

Now, every collection of $n+2$ or fewer points in $A \cup B$ can be separated strictly. Consider a corresponding collection of $n+2$ members of $\mathscr{F}$,

$$A(x_1), \ldots, A(x_k), B(x_{k+1}), \ldots, B(x_n + 2). \tag{30}$$

Then it follows that there exists a pair $(y, \kappa) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\begin{aligned} \langle y, x_i \rangle > \kappa &\implies (y, \kappa) \in A(x_i) \qquad \forall\, i \in \{1, \ldots, k\}, \\ \langle y, x_i \rangle < \kappa &\implies (y, \kappa) \in B(x_i) \qquad \forall\, i \in \{k+1, \ldots, n+2\}, \end{aligned}$$

i.e. $(y, \kappa)$ is contained in the intersection of the $n+2$ sets in (30). We have shown that an arbitrary intersection of $n+2$ members of $\mathscr{F}$ is nonempty, and so we may apply Helly's theorem (Theorem 36) and conclude that the members of $\mathscr{F}$ have a common element. This element then describes a hyperplane in $\mathbb{R}^n$ that separates $A$ and $B$ strictly. □

**Remark 19.** (a) Setting $n = 2$ yields the claim at the beginning of this section.

(b) Kirchberger's theorem also has an analogue for infinite subets of $\mathbb{R}^n$ (cf. Remark 18(b)), with the additional condition of compactness. Theorem 37 is used instead, and the proof is otherwise identical.

# A Appendix: Use of sources

The following list details the primary sources used for proofs of given theorems, lemmas, and so on. Typically the "idea" of the proof is found in the referenced article(s). Proofs have been reformulated and modified to suit the discussion; for example for Theorem 38, the proof in [2, Theorem 2.6] is for only the specific case $d = 2$, but is generalised in this text. Other (mostly early, short) proofs remain largely unchanged; for example for Theorem 1.

Unlisted results are, for the most part, own work in their entirety. This includes in particular all results labelled *Proposition*.

**Theorem 1** [3, Theorem 2.19]
**Theorem 2** [6, Theorem 2.2]
**Theorem 3** [6, Theorem 2.3], [8, Theorem 2.3]
**Theorem 10** [3, Theorem 2.18]
**Theorem 12** [4, Theorem 1.5], [9, Example 1.1]
**Theorem 14** [1, Theorem 2.8]
**Corollary 15** [1, Corollary 2.9]
**Theorem 16** [1, p. 8]
**Lemma 17** [4, Proposition 1.1(c)]
**Theorem 18** [4, Proposition 1.1(b)]
**Corollary 19** [4, p. 8]

**Theorem 20** [1, Theorem 3.1, Lemma 3.2]
**Theorem 21** [3, Theorem 2.11]
**Lemma 22** [1, Theorem 3.3]
**Theorem 23** [1, Theorem 3.4(b)]
**Theorem 24** [1, Theorem 3.4(c)]
**Theorem 25(b)** [8, Theorem 4.9]
**Lemma 26** [8, Theorem 4.10(b)]
**Theorem 27** [8, Theorem 4.10(b)]
**Theorem 28** [8, pp. 32–33]
**Lemma 29** [1, Lemma 4.4]
**Theorem 30** [8, Theorem 4.11]
**Theorem 31** [8, Theorem 4.12]
**Lemma 32** [8, Lemma 4.13]
**Theorem 33** [8, Theorem 4.13]
**Corollary 34** [8, p. 50]
**Theorem 35** [3, Theorem 6.1]
**Theorem 36** [3, Theorem 6.2]
**Theorem 38** [2, Theorem 2.6]
**Theorem 39** [8, Theorem 4.5], [5, p. 246]

In addition, the following chapters or sections were used in the general writing of this text: [1, §1,2,3,4], [3, §2,4,5,6,7], [4, §1,2], [6, §1,2,3,6,11,17], [7, §18], [8, §2,3,4], [9, §1].

Portraits of the key players were taken from *The MacTutor History of Mathematics* archive, found at `http://www-history.mcs.st-and.ac.uk/`.

# References


[1] Arne Brøndsted, *An introduction to convex polytopes*, Graduate texts in mathematics, no. 90, Springer-Verlag, 1982.

[2] Ludwig Danzer, Branko Grünbaum, and Victor Klee, *Helly's theorem and its relatives*, Proc. Sympos. Pure Math, vol. 7, 1963, pp. 101–180.

[3] Steven R. Lay, *Convex sets and their applications*, John Wiley & Sons, 1982.

[4] Stefan Nickel, *Convex analysis*, University of Kaiserslautern, 1998.

[5] Hans Rademacher and Isaac J. Schoenberg, *Helly's theorem on convex domains and Tchebycheff's approximation problem*, Canad. J. Math **2** (1950), 245–256.

[6] R. Tyrrell Rockafellar, *Convex analysis*, Princeton University Press, 1972.

[7] Alexander Soifer, *Geometric etudes in combinatorial mathematics*, Springer, 2010.

[8] Jan van Tiel, *Convex analysis*, John Wiley & Sons, 1984.

[9] Constantin Zălinescu, *Convex analysis in general vector spaces*, World Scientific Publishing Co., 2002.

# Index