Chapter 1

Convex sets

This chapter is under construction; the material in it has not been proof-read, and might contain errors (hopefully, nothing too severe though).

We say a set C is *convex* if for any two points $x, y \in C$, the *line segment*

$$(1-\alpha)x + \alpha y, \qquad \lambda \in [0,1],$$

lies in C. The emptyset is also regarded as convex. Notice that while defining a convex set, we used addition and multiplication with a real scalar. This is so, because the above definition of convexity is set in a vector space (finite or infinite dimensional). We will focus on analysis and optimization problems in finite dimensional spaces, mostly in the Euclidean space $\mathbb{R}^n$ (notice that this does not lose too much generality, since any n dimensional vector space is isomorphic to $\mathbb{R}^n$).

Before we go further in our study of convex sets in $\mathbb{R}^n$, let us look at two alternative (but intimately related) views of convex sets.

Means and convexity.

Convexity is very closely related to the notion of *means*. For example, the point $(1-\alpha)x+\alpha y$ is just the (weighted) *arithmetic mean* of x and y. Over the reals $\mathbb{R}$ one may consider a variety of means. Let $x \leq y \in \mathbb{R}$, and $M : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a *mean*; typically, M is required to fulfill the following axiomatic properties:

(i) $x \leq M(x,y) \leq y$ with equality if and only if $x = y$; (interiority)

(ii) $M(x,x) = x$ for all $x \in \mathbb{R}$;

(iii) $M(\lambda x, \lambda y) = \lambda M(x,y)$ (homogeneity);

If instead of the arithmetic mean, we consider the (weighted) *geometric mean* $M(x,y) := x^{1-\alpha}y^{\alpha}$ for $\alpha \in (0,1)$, we can define "geometrically convex" sets (verify!).

Means have been very extensively studied and satisfy a large number of inequalities. Ultimately, most of these inequalities are in one way or the other, a reflection of convexity (we will revisit this idea in Chapter **??** when we discuss convex functions).

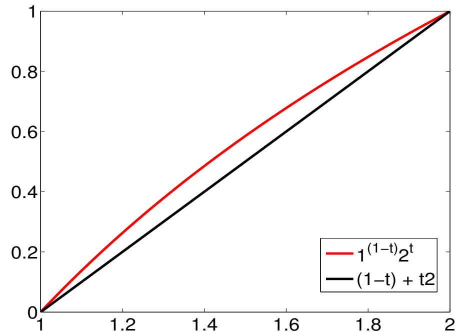

Figure 1.1: Lines based on arithmetic and geometric means

Lines in metric spaces

There is yet another, perhaps more natural, way to regard the concept of convexity. Convexity in a vector space is defined using lines between points; in a Euclidean space (or more generally in a Banach space), there is a line of shortest length that joins two points (the line need not be unique), and the length of this line is the distance between its two endpoints.

Karl Menger [**?**] realized that many properties of lines extend to more general metric spaces[1], where we have *geodesics*, i.e., paths whose length is equal to the distance between their endpoints. Menger developed geometric properties of such metric spaces, without appealing to local coordinates, or differentials. Instead, he relied on properties fulfilled by the distance function.

Let $\mathcal{X}$ be nonempty and let $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a function. We say that is a *distance* if

(i) $d(x, y) \geq 0$ for all $x, y \in \mathcal{X}$ and $d(x, y) = 0$ if and only if $x = y$ (*positive definiteness*);

(ii) $d(x, y) = d(y, x)$ for all $x, y \in \mathcal{X}$ (*symmetry*);

(iii) $d(x, y) \leq d(x, z) + d(y, z)$ for all $x, y, z \in \mathcal{X}$ (*triangle inequality*).

With a distance function in hand, we are ready to describe Menger's notion of convexity.

Definition 1.1 (Menger-convex)**.** A metric space $(\mathcal{X}, d)$ is called *Menger-convex* if for every pair of distinct points $x, y \in \mathcal{X}$, there exists a third point $z \in \mathcal{X}$ ($z \neq x, z \neq y$) such that

$$d(x, y) = d(x, z) + d(y, z). \tag{1.1}$$

Thus, z is not an "end point" like x and y, but makes the triangle inequality an equality.

Exercise 1.1. *Verify that the usual notion of convexity on* $\mathbb{R}^n$ *with* $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ *satisfies* (1.1).

Example 1.2. Consider the set of strictly positive reals $\mathbb{R}_{++} = (0, \infty)$. This set may be endowed with the *hyperbolic metric* (verify!)

$$d_h(x, y) := |\log x - \log y|, \qquad x, y > 0. \tag{1.2}$$

[1] The concept of a metric space had at that time recently been put forth in a brilliant PhD thesis by Maurice Fréchet [**?**].

With this metric (1.1) yields the so-called "multiplicative convexity" [NP06, pp. XX]. For distinct $x, y > 0$, let $\alpha \in (0,1)$. It is easy to verify that $z_\alpha := x^{1-\alpha}y^\alpha$ satisfies $d_h(x,y) = d_h(x,z_\alpha) + d_h(y,z_\alpha)$.

Definition 1.1 is remarkably rich, and we will return to it in Chapter **??** where we study geometric optimization. For the rest of this chapter, we will focus on convexity in $\mathbb{R}^n$ only. However, before we return to the familiar Euclidean domain, we would be remiss if we did not highlight the fundamental theorem on Menger-convexity.

Definition 1.3. A path $\gamma : [0,1] \mapsto \mathcal{X}$ between two points x, y in a metric space $(\mathcal{X}, d)$ is called a *geodesic* if it satisfies

$$d(\gamma(\alpha_1), \gamma(\alpha_2)) = |\alpha_1 - \alpha_2| d(x,y), \qquad \forall \alpha_1, \alpha_2 \in [0,1]. \tag{1.3}$$

In particular, $d(0) = x$ and $d(1) = y$ (we use the interval $[0,1]$ for clarity; one could also use $[0,\ell]$ for some $\ell > 0$).

Theorem 1.4 ([Pap05, Thm. 2.6.2])**.** *Let $(\mathcal{X}, d)$ be a complete, locally compact metric space. Then the following are equivalent:*

(i) $(\mathcal{X}, d)$ is Menger-convex.

(ii) For all $x, y \in \mathcal{X}$, there exists a point $m \in \mathcal{X}$, called a midpoint*, such that*

$$d(x,m) = d(y,m) = \tfrac{1}{2} d(x,y).$$

(iii) Any two points $x, y \in \mathcal{X}$ are joined by a geodesic*.*

Notice that as suggested by this theorem, we can consider *geodesically convex* sets. We say a set $C \subset \mathcal{X}$, where $(\mathcal{X}, d)$ is a Menger-convex geodesic metric space, is called geodesically convex if for any pair $x, y \in C$, the entire geodesic from x to y lies inside C. Henceforth, we will denote this geodesic using the suggestive notation

$$\gamma(\alpha) := (1-\alpha)x \oplus \alpha y, \quad \text{for } \alpha \in [0,1].$$

Example 1.5. Consider again $\mathbb{R}_{++}$ with d_h. It is easy to verify that for any two points $x, y \in \mathbb{R}_{++}$, the geometric mean $\sqrt{xy}$, which is the midpoint of the geodesic $(1-\alpha)x \oplus \alpha y \equiv x^{1-\alpha}y^\alpha$ furnishes such a midpoint. $\Diamond$

Example 1.6. A fancier example is obtained by considering the space $\mathbf{S}_+^n$ of Hermitian strictly positive definite matrices. We will later see (Chap. **??**) that on $\mathbf{S}_+^n$, one has the *Riemannian distance*

$$d_R(X,Y) := \|\log(Y^{-1/2}XY^{-1/2})\|_{\mathrm{F}}, \qquad X, Y \in \mathbf{S}_+^n;$$

Here, $\log(\cdot)$ denotes the matrix logarithm, and $\|M\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(M^*M)}$ denotes the *Frobenius norm*. Under this distance, we have the (unique) geodesic (*cf.* Example 1.5)

$$\gamma(\alpha) := X^{1/2}(X^{-1/2}YX^{-1/2})^\alpha X^{1/2}, \quad t \in [0,1].$$

The midpoint of this geodesic is $\gamma(\frac{1}{2})$, and the reader is invited to verify that

$$d_R(X, \gamma(\tfrac{1}{2})) = d_R(Y, \gamma(\tfrac{1}{2})) = \tfrac{1}{2} d(X,Y). \qquad \Diamond$$

After this excursion into convexity in metric spaces, we will largely shift our focus to convex sets in $\mathbb{R}^n$, and return to metric space convexity in the next chapter.

1.1 Constructing convex sets

We mention below a few standard operations that yield convex sets.

- *Intersection.* Let $\{C_j\}_{j\in J}$ be an arbitrary collection of convex sets. Then, their intersection $C := \cap_{j\in J} C_j$ is also convex (verify!).
- *Cartesian Product.* Let $\{C_j\}_{j\in J}$ be an arbitrary collection of convex sets. Then, their Cartesian product $C := \prod_{j\in J} C_j$ is also convex (verify!).

Exercise 1.2. *If C_1 and C_2 are geodesically convex sets in a geodesic metric space, is it true that their intersection $C_1 \cap C_2$ is also geodesically convex? What about the Cartesian product $C_1 \times C_2$?*

Since we are studying convexity in a vector space, the following result is not surprising, though it is of fundamental importance.

Proposition 1.7. *Let $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^m$ be an affine mapping and $C \subset \mathbb{R}^n$ be convex. Then the image $\boldsymbol{A}(C) := \{\boldsymbol{A}(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n\} \subset \mathbb{R}^m$ is also convex. Conversely, if $D \subset \mathbb{R}^m$ is convex, then the inverse image $\boldsymbol{A}^{-1}(D) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}(\boldsymbol{x}) \in D\} \subset \mathbb{R}^n$ is also convex.*

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in C$; then $A(x), A(y) \in A(C)$. Since $(1-\alpha)\boldsymbol{x} + \alpha\boldsymbol{y} \in C$ for $\alpha \in [0,1]$, and A is affine we see that $A((1-\alpha)\boldsymbol{x} + \alpha\boldsymbol{y}) = (1-\alpha)\boldsymbol{A}(\boldsymbol{x}) + \alpha\boldsymbol{A}(\boldsymbol{y}) \in A(C)$. Similarly, if $A(x), A(y) \in D$, then $x, y \in A^{-1}(D)$. Since D is convex, $(1-\alpha)A(x) + \alpha A(y) \in D$, which shows that $(1-\alpha)x + \alpha y \in A^{-1}(D)$. □

Prop. 1.7 has three immediate consequences.

Corollary 1.8. *(i) If C is convex, then αC is also convex for all $\alpha \in \mathbb{R}$*

(ii) Let C_1, C_2 be convex. Then, their Minkowski sum $C_1 + C_2 := \{x + y \mid x \in C_1, y \in C_2\}$ is also convex (proof: *apply the affine map $(x,y) \mapsto x + y$).*

(iii) The projection of a convex set onto some of its coordinates is convex. That is, if $C \subset \mathbb{R}^n \times \mathbb{R}^m$ is convex. Then, $C_1 := \{\boldsymbol{x}_1 \in \mathbb{R}^n \mid (\boldsymbol{x}_1, \boldsymbol{x}_2) \in C$ for some $\boldsymbol{x}_2 \in \mathbb{R}^m\}$ is also convex.

Exercise 1.3. *Let $C_1 \subset \mathbb{R}^{n_1}$ and $C_2 \subset \mathbb{R}^{n_2}$ be convex. Prove that if $C_1 \times C_2$ is convex, then C_1 and C_2 must be convex.*

> **Minkowski sums** show up in number of areas, both pure and applied. One example is the entire body of questions dealing with *sumsets*, which are nothing but Minkowski sums of sets comprised of elements from a group—e.g., say $A \subset \mathbb{Z}$. A basic question on sumsets is estimating their cardinality, e.g., we easily have the trivial bounds $|A| \le |A + A| \le |A|^2$, but given some more information about the structure of A, more refined bounds can be obtained. See the recent book by Tao [**?**] on *Arithmetic Combinatorics* for a rich coverage.

An important operation that goes beyond affine transformations but still preserves convexity is the so-called perspective transform.

Proposition 1.9. *The perspective transform on $\mathbb{R}^n \times \mathbb{R}_{++}$ is given by the nonlinear map $(\boldsymbol{x}, t) \mapsto \boldsymbol{x}/t$. If $C \subset \mathbb{R}^n \times \mathbb{R}_{++}$ is a convex set, then its image $P(C)$ is also convex.*

Proof. See [BV04, §2.3.3]. Alternatively, observe that the intersection of C with the hyperplane $\{(\boldsymbol{x}, t) \mid t = 1\}$ is a convex set. Now project this down to $\mathbb{R}^n$ by dropping the last coordinate, which results in a convex set (See Corollary **??**-(iii)). □

1.1.1 Convex Hulls

An important method of constructing a convex set from an arbitrary set of points is that of taking their convex hull (see Fig. **TODO**). Formally, if $X := \{x_i \in \mathbb{R}^n \mid 1 \le i \le m\}$ is an arbitrary set of points, then its *convex hull* is the set obtained by taking all possible convex combinations of the points in X. That is,

$$\mathrm{co}X := \Big\{\sum\nolimits_{i=1}^{m} \alpha_i x_i \mid \alpha_i \ge 0, \sum\nolimits_i \alpha_i = 1\Big\}. \tag{1.4}$$

More generally, we can also define convex hulls of sets containing an infinite number of points. In this case the following three equivalent definitions of $\mathrm{co}\, X$ may be used:

(a) the (unique) minimal convex set containing X;

(b) the intersection of all convex sets containing X;

(c) the set of all convex combinations of points in X.

The last definition is a generalization of (1.4).

Remark 1.10. The term $\sum_i \alpha x_i$ in (1.4) is called a convex combination. The vector $\boldsymbol{\alpha}$ of "convex coefficients" may also be interchangeably called a probability vector.

We warn the computationally oriented reader that it is computing the convex hull of a given set of points is in general very difficult.

1. lower bound via sorting reduction
2. list of methods used for convex hulls on plane
3. potential troubles in higher-D; link to methods, papers
4. remarks about vertices to faces—polymake etc—difficult
5. also notice that convex combinations involve real numbers, standard computational complexity is set in integers (or rational) arithmetic

There is a geometric property worth noting. Even if $X \subset \mathbb{R}^n$ contains an infinite number of points, each point in co$X$ can be represented as a convex combination of at most $n+1$ points from X. This connection is formally described by a famous result of Carathéodory.

Theorem 1.11 (Carathéodory)**.**

It is important to note that a *different* set of $n+1$ points may be needed to encode each point, that is, Carathéodory's theorem does not provide us a "basis."

1.1.2 Basic convex sets

We briefly mention some of the most important convex sets. We do not dwell on the details, and refer the reader to [Roc70, BV04, HUL01] for a more detailed treatment.

(i) *Subspace.* A set $S \subset \mathbb{R}^n$ is a subspace if for any $x, y \in S$, any linear combination $\alpha x + \beta y \in S$. We know from linear algebra that any subspace in $\mathbb{R}^n$ may be identified with the set of solutions to a *homogenous system* of linear equations, i.e., the set $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = 0\}$. Observe that subspaces contain the origin.

(ii) *Affine Manifold.* A translated subspace, i.e., a set of the form $M = x + S$ where S is a subspace. Note that $\dim(M) = \dim(S)$.

(iii) *Affine hull.* For a nonempty set S, we define its affine hull aff S as the intersection of all affine manifolds containing S. The dimension of a convex set C is the dimension of aff C (since an affine manifold is just a translated subspace).

(iv) *Hyperplane.* A special affine manifold, parameterized by a linear functional $\boldsymbol{a}^T$ and a scalar $\gamma \in \mathbb{R}$, given by the set $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^T\boldsymbol{x} = \gamma\}$. Observe that the "row-vector" $\boldsymbol{a}^T$ is better viewed as an element of $(\mathbb{R}^n)^*$ (called the *dual space* of $\mathbb{R}^n$), which is the set of all linear functionals on $\mathbb{R}^n$. Henceforth, we write $\langle \boldsymbol{a},\, \boldsymbol{x}\rangle \equiv \boldsymbol{a}^T\boldsymbol{x}$.

(v) *Halfspace.* The set $\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a},\, \boldsymbol{x}\rangle \leq \gamma\}$. Observe that any hyperplane divides $\mathbb{R}^n$ into two parts, those lying to the *left* $\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a},\, \boldsymbol{x}\rangle \leq \gamma\}$ and those to the *right* $\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a},\, \boldsymbol{x}\rangle \geq \gamma\}$.

(vi) *Polyhedra.* Solution sets to *finite* system of linear equations and inequalities, e.g., $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\}$ (the finiteness is crucial in the definition). One of the most important polyhedra is the *unit simplex* $\Delta_n := \{\boldsymbol{x} \in \mathbb{R}^n \mid x_i \geq 0, \sum_i x_i = 1\}$.

(vii) *Norm ball.* Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. The norm-ball of radius $r \geq 0$ centered at $\boldsymbol{x}_0$ is the set $B_r(\boldsymbol{x}_0) := \{\boldsymbol{x} \in \mathbb{R} \mid \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq r\}$. From the triangle inequality for norms, convexity of $B_r(\boldsymbol{x}_0)$ follows. More generally, let $d(\cdot,\cdot)$ be any metric on $\mathbb{R}^n$. Then the *metric-ball* $B_r^d(\boldsymbol{x}_0) := \{\boldsymbol{x} \in \mathbb{R}^n \mid d(\boldsymbol{x}_0, \boldsymbol{x}) \leq r\}$ is convex.

(viii) *Convex cones.* A set $K \subset \mathbb{R}^n$ is called a *cone* if for $\boldsymbol{x} \in K$, the ray $\alpha\boldsymbol{x}$ is in K for all $\alpha > 0$. The origin may or *may not* be included. A few distinguished cones are $\mathbb{R}^n_+$ (nonnegative orthant), the *Lorentz cone* $\{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \|\boldsymbol{x}\|_2 \leq t\}$, the *positive semidefinite cone* $\mathbf{S}^n_+ := \{\boldsymbol{X} \in \mathbb{R}^{n\times n} \mid \boldsymbol{X} = \boldsymbol{X}^T,\ \boldsymbol{X} \succeq 0\}$. We note that the $\mathbb{R}^n_+$ is a polyhedral cone, while $\mathbf{S}^n_+$ is nonpolyhedral.

Exercise 1.4. *Verify the following claims:*

(i) The intersection of an arbitrary collection of convex cones is a convex cone

(ii) Let $\{\boldsymbol{b}_j\}_{j\in J}$ be vectors in $\mathbb{R}^n$. Then,

$$P := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x},\, \boldsymbol{b}_j\rangle \leq 0,\ j \in J\}$$

is a convex cone (if J is finite, then this cone is polyhedral).

(iii) A cone K is convex if and only if $K + K \subset K$.

(iv) Verify that $\{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \|\boldsymbol{x}\| \leq t\}$ is a cone for any norm $\|\cdot\|$ on $\mathbb{R}^n$.

(v) A real symmetric matrix $\boldsymbol{A}$ is called copositive if for every nonegative vector $\boldsymbol{x}$ we have $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0$. Verify that the set CP_n of $n \times n$ copositive matrices forms a convex cone.

(vi) Spectrahedron: the set $S := \{\boldsymbol{x} \in \mathbb{R}^n \mid x_1A_1 + \ldots + x_nA_n \succeq 0\}$ is convex for symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m\times m}$. Additionally, observe that the spectrahedron is the inverse image of $\mathbf{S}^m_+$ under the affine map $A(\boldsymbol{x}) = \sum_i x_iA_i$.

(vii) The convex hull of $S = \{\boldsymbol{x}\boldsymbol{x}^T \mid \boldsymbol{x} \in \mathbb{R}^n\}$ is $\mathbf{S}^n_+$.

The following exercise shows that in the nonlinear convex world, convex cones are analogous to subspaces in the linear world.

Exercise 1.5. *A set $K \subset \mathbb{R}^n$ is a convex cone if and only it is closed under addition and positive scalar multiplication, i.e., if $x, y \in K$ then $x + y \in K$ and $\alpha x \in K$ for all $\alpha > 0$.*

Exercise 1.5 implies that $K \subset \mathbb{R}^n$ is a convex cone if and only if it contains all positive linear combinations of its elements. Thus, if $S \subset \mathbb{R}^n$ is arbitrary, then the set of all positive linear combinations of S is the smallest convex cone that includes S.

Hence, akin to convex hulls we can also define *conic hulls*. If we adjoin the origin to the smallest convex cone containing a set S, we obtain the conic hull of S; we denote this by con S. The following result shows that convex sets have particularly simple conic hulls.

Proposition 1.12. *Let C be a convex set. Then,*

$$\operatorname{con} C = \{\alpha x \mid \alpha \geq 0,\ x \in C\}.$$

Exercise 1.6. *Determine the conic hulls of the following sets:*

1. *The unit simplex.*
2. *The set $\{\boldsymbol{e}_i \in \mathbb{R}^n \mid 1 \leq i \leq n\}$, where $\boldsymbol{e}_i$ denotes the ith standard basis vector.*
3. *TODO*
4. *...*

Convex cones are extremely important in both convex geometry and optimization. They are usually simpler to handle than general convex sets, so it can be useful to turn a question about convex sets into a question about convex cones. The following proposition suggests why—it is essentially a generalization of the fact that a circle is just a slice of a 3D cone (Fig. **TODO**).

Proposition 1.13. *Every convex set $C \subset \mathbb{R}^n$ can be regarded as the cross-section of some convex cone K in $\mathbb{R}^{n+1}$.*

Proof. Let $S := \{(\boldsymbol{x}, 1) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in C\}$, and let K be the conic hull of S. Consider the hyperplane $H = \{(\boldsymbol{x}, \lambda) \mid \lambda = 1\}$. Since K consists of points $(\lambda\boldsymbol{x}, \lambda)$, intersecting K with H may be then regarded as C (by dropping the extra dimension). □

TODO: Make a picture here of a polyhedral 2D-set, etc.

Remark 1.14. Alternatively, we may obtain C by the perspective transform of the smallest cone containing C, i.e., $\{(\lambda\boldsymbol{x}, \lambda) \mid \lambda > 0, x \in C\}$.

1.1.3 Polars, dual cones

We briefly introduce the notions of polars and dual cones here. We will revisit these later. Let K be a cone. It *polar* K° is defined as

$$K^\circ := \{y \mid \forall x \in K, \langle x,\, y\rangle \leq 0\}. \tag{1.5}$$

If K is nonempty, closed and convex then $K^{\circ\circ} = K$. Polars are to the world of convexity what orthogonal complements are to the linear world. Indeed, if K is a subspace (which is a cone, albeit not a pointed one), then K° is nothing but the orthogonal complement $K^\perp$.

Observe that the polar (1.5) depends on the choice of inner product. If we impose a different inner product on the Euclidean space, then the polar also changes. From its definition, one sees that K° is always a closed convex cone (closed, since inner product is continuous). Polarity is an important *involutory order reversing transform*: if K_1 and K_2 are closed convex cones such that $K_1 \subset K_2$, then $K_1^{\circ\circ} = K_1$ ($K_2^{\circ\circ} = K_2$) and $K_1^\circ \supset K_2^\circ$.

Example 1.15. Let $K = \mathbb{R}^n_+$; then, $K^\circ = -K$, the nonpositive orthant. $\Diamond$

Exercise 1.7. *Prove the order reversing property of polars for closed convex cones.*

Exercise 1.8. *What is $K \cap K^\circ$?*

Example 1.16. Let $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ be m points in $\mathbb{R}^n$. Then,

$$(\operatorname{con} X)^\circ = \{y \in \mathbb{R}^n \mid \langle y,\, x_i\rangle \le 0,\ 1 \le i \le m\}.$$

TODO $\Diamond$

Exercise 1.9. *Consider the* ordered orthant*. Prove that it is a convex cone. Determine its polar.* **TODO**

We may also define the *dual cone* is defined as $-K^\circ$ (this sign-flip in the definition stems from various reasons, but those are not important for now). Thus, for a cone K, its dual cone is

$$K^* := \{y \mid \forall x \in K, \langle x,\, y\rangle \ge 0\}. \tag{1.6}$$

Observe that K need not be convex in either (1.5) or (1.6) but that K° and K^* are always convex cones.

Exercise 1.10. *Prove that the semidefinite cone $\mathbf{S}^n_+$ is self-dual under the inner-product $\langle \boldsymbol{X},\, \boldsymbol{Y}\rangle = \operatorname{tr}(\boldsymbol{X}\boldsymbol{Y})$.*

The nonnegative orthant is self-dual, and Ex. 1.10 shows that the semidefinite cone is also self-dual. The reader may wonder which other examples of self-dual cones are known? A deep theorem of **TODO** shows that the only self-dual convex cones are:

- The nonnegative orthant $\mathbb{R}^n_+$
- The Lorentz cone (second-order cone)
- The semidefinite cone $\mathbf{S}^n_+$
- **quaternions** etc.
- Cartesian products of the above.

1.1.3.1 Polars of convex sets

The idea of polars also extends to general convex sets (one way to extend it is to use Prop. 1.13 that identifies convex sets with cones, and then compute polars of cones). If C is a convex set, then its polar is

$$C^\circ := \{y \mid \forall x \in C,\ \langle x,\, y\rangle \le 1\}. \tag{1.7}$$

If C is closed, convex, and contains the origin, then its polar C° is also a closed convex set containing the origin, and in fact $C^{\circ\circ} = C$.

Exercise 1.11. *Determine the polars of the following closed convex sets.*

1. $C = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_1 \le 1\}$
2. $C = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 \le \sqrt{1 + x_2^2}\}$

1.11 *(i)* $C^\circ = \{\boldsymbol{x} \in \mathbb{R}^n \mid |x_i| \le 1,\ 1 \le i \le n\}$
(ii) $C^\circ = \mathrm{co}(D \cup \{0\})$, *where* $D = \{\boldsymbol{x} \in \mathbb{R}^2 \mid 2x_1 \ge (1 + x_2^2)\}$

Exercise 1.12. *Describe the dual cones to the following:*

1. *The copositive cone* CP_n

2. *A matrix* $\boldsymbol{A}$ *is called* doubly nonnegative *(DNN) if it can be written as* $\boldsymbol{S} + \boldsymbol{N}$, *where* $\boldsymbol{S} \in \mathbf{S}_+^n$ *and* $\boldsymbol{N} \in \mathbb{R}_+^{n \times n}$. *What is the dual cone?*

Convex bodies, Polars, Mahler conjecture

Let $C \subset \mathbb{R}^n$ be a symmetric convex body, thus C is closed, convex, bounded, and symmetric about the origin. The *volume* of C is given by the (Lebesgue) integral

$$V_n(C) := \int_{C \subset \mathbb{R}^n} d\boldsymbol{x}.$$

The polar C° is also closed, convex, contains the origin, and is also a convex body with volume $V_n(C^\circ)$. The *Mahler volume* is the product $M(C) := V_n(C)V_n(C^\circ)$. One may easily verify that the Mahler volume is affine invariant, i.e., if A is any invertible linear transformation, then $M(AC) = M(C)$.

Exercise 1.13. *If* $C_1 \subset \mathbb{R}^{n_1}$ *and* $C_2 \subset \mathbb{R}^{n_2}$ *are convex bodies, then* $M(C_1 \times C_2) = M(C_1)M(C_2)/\binom{n_1+n_2}{n_1}$

Exercise 1.14. *What is the Mahler volume of (i) the unit Euclidean ball; (ii) the unit cube?*

The upper-bound on $M(C)$ was established by Santaló, who showed that the maximum is achieved at the Euclidean ball B_2^n. The corresponding inequality is known as the Blasckhe-Santaló inequality [? ? ?]

$$M(C) \le M(B_2^n),$$

with equality if and only if C is an ellipsoid. Mahler conjectured in 1939 [?] that $M(C)$ is minimized at the cube B_∞^n (unit ball of the ℓ_∞-norm), i.e.,

$$M(C) \ge \frac{4^n}{n!}.$$

There has been a lot of work on this problem ever since it was posed, and we refer the reader to [Kim12]. There are reasons [?] to believe that the minimum has to be achieved at a polytope. The lower-bound has eluded proof, because unlike the upper bound, there seems to be no "essentially unique" class of convex bodies that minimizes the Mahler volume.

1.1.4 Relative interiors

For convex sets, standard topological operations such as interior, closure, etc. again yield convex sets. Frequently, convex sets have empty interior—e.g., a 2D-rectangle in $\mathbb{R}^3$ has empty interior, which is nonempty when the convex set is regarded as a subset of $\mathbb{R}^2$. This suggests that when the dimension of the affine hull of a convex set S is not full, then it has an empty interior, so it is more useful to consider interiors relative to the affine manifold within which the set actually lies.

Definition 1.17. The *relative interior* $\operatorname{ri} C$ of a convex set $C \subset \mathbb{R}^n$ is the interior of C relative to the affine hull of C. That is,

$$x \in \operatorname{ri} C \quad \text{iff} \quad x \in \text{aff}\, C, \text{and } \exists \delta > 0 \text{ s.t. } (\text{aff}\, C) \cap B_\delta(x) \subset C.$$

Example 1.18. Say $C = \{x\}$, then $\operatorname{ri} C = x$. If $C = (1-\alpha)x + \alpha y$ for $x, y \in \mathbb{R}^n$ and $\alpha \in [0, 1]$ (a line-segment), then $\operatorname{ri} C = (1-\alpha)x + \alpha y$ with $\alpha \in (0, 1)$. ◊

Exercise 1.15. *Determine the relative interiors of the following sets:*

1. *The unit simplex*

Exercise 1.16. *Is it true that if $C_1 \subset C_2$ then $\operatorname{ri} C_1 \subset \operatorname{ri} C_2$?*

We have the following theorem showing the usefulness of relative interiors.

Theorem 1.19. *Let $C \neq \emptyset$ be a convex set. Then, $\operatorname{ri} C \neq \emptyset$. Moreover, $\dim(\operatorname{ri} C) = \dim C$.*

Proof. $\dim(\operatorname{ri} C) = \dim(\text{aff}\, C)$, because the relative interior is formed by intersecting full-dimensional balls $B_\delta(x)$ with the affine hull of C. □

For optimization, a very important result on relative interiors is the following.

Proposition 1.20. *Let C_1, C_2 be convex sets for which $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 \neq \emptyset$. Then,*

$$\operatorname{ri}(C_1 \cap C_2) = \operatorname{ri} C_1 \cap \operatorname{ri} C_2.$$

Proof. See [HUL01, Prop. 2.1.10]. □

Two other basic properties of relative interiors are presented below.

Proposition 1.21. *Relative interiors are preserved under Cartesian products, and also under affine and inverse affine maps. That is,*

$$\begin{aligned} \operatorname{ri}(C_1 \times \cdots C_k) &= (\operatorname{ri} C_1) \times \cdots \times (\operatorname{ri} C_k) \\ \operatorname{ri}(A(C)) &= A(\operatorname{ri} C) \\ \operatorname{ri}(A^{-1}(D)) &= A^{-1}(\operatorname{ri} D). \end{aligned}$$

Proof. TODO □

1.2 Projection onto a convex set

We have now arrived at one of the most important optimization problems: projection onto a convex set. To begin, recall from linear-algebra that projection onto a subspace S was defined (say $u_1, \ldots, u_k$ is an orthonormal basis for S; let $\boldsymbol{S} = [u_1, \ldots, u_k]$ be the $n \times k$ matrix with $u_1, \ldots, u_k$ as its columns, then $P_S \equiv \boldsymbol{S}\boldsymbol{S}^T$). The key properties of the projection operator $x \mapsto P_S(x)$ are: linearity, symmetry, semidefiniteness, idempotency ($P_S \circ P_S = P_S$), and nonexpansivitity $\|P_S(x)\|_2 \leq \|x\|_2$. Moreover, any $x \in \mathbb{R}^n$ can be decomposed as $x = P_S(x) + P_{S^\perp}(x)$.

We will see below that an operator with similar properties may be associated with projection onto convex sets, not just subspaces.

Definition 1.22. Consider the metric space $(\mathbb{R}^n, \|\cdot\|)$. Let $C \subset \mathcal{X}$ be a closed convex set. Then, the (norm) *projection* of any point $y \in \mathbb{R}^n$ onto the set C is the solution to the optimization problem

$$\inf_{x \in C} \{\|x - y\|\}. \tag{1.8}$$

If the norm is Euclidean, we call this *orthogonal projection* (or simply projection for short).

Before proceeding further, let us recall a basic result from real analysis.

Theorem 1.23 (Weierstraß)**.** *A continuous function on a compact set in $\mathbb{R}^n$ attains its minimum and maximum.*

Suppose C is a compact set in $\mathbb{R}^n$. Then, from Theorem 1.23 it follows that the projection problem (1.8) has a solution (since the function $x \mapsto \|x - y\|$ is continuous). That is, there exists a point $x^{‘} \in C$ such that $\|x^* - y\| = \min_{x \in C} \|x - y\|$. Notice that this existence result does not depend on convexity of C. However, for uniqueness of x^* convexity plays a crucial role. Let us investigate this a little further.

Definition 1.24. A subset $S \subset \mathbb{R}^n$ is said to be a *Chebyshev set* if for each point $y \in \mathbb{R}^n$, there is a unique point $P_S(y) \in S$ such that $\|P_S(y) - y\|_2 = \inf_{x \in S} \|x - y\|_2$ (thus, the 'inf' is attained and is actually a 'min').

Theorem 1.25. *A set $S \subset \mathbb{R}^n$ is Chebyshev if and only if it is convex.*

Proof. Let us proof the sufficiency part. The necessity is more involved, and we refer the reader to [Bor07] for an illuminating proof.

Let $x_1 \neq x_2$ be two solutions to (1.8). Then, consider $x = \frac{1}{2}(x_1 + x_2)$. We have

$$\|x - y\|_2^2 = \|\left(\tfrac{x_1 - y}{2}\right) + \left(\tfrac{x_2 - y}{2}\right)\|_2^2 = \tfrac{1}{4}\|x_1 - y + x_2 - y\|_2^2$$
$$= \tfrac{1}{2}\|x_1 - y\|_2^2 + \tfrac{1}{2}\|x_2 - y\|_2^2 - \tfrac{1}{4}\|x_1 - x_2\|_2^2.$$

This implies the desired uniqueness. □

> The sufficiency part of Theorem 1.25 can be extended to certain infinite-dimensional Banach spaces. We say a Banach space $\mathcal{X}$ is *uniformly convex* if for every $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that for any two unit norm vectors $x, y \in \mathcal{X}$ that satisfy $\|x - y\| \geq \epsilon$, we have $\|\left(\frac{x+y}{2}\right)\| \leq 1 - \delta(\epsilon)$. Roughly speaking, this means that if two unit norm points in a uniformly convex space $\mathcal{X}$ are far apart, then their midpoint must be deep inside the space. It can be shown that every closed convex subset of a uniformly convex Banach space is Chebyshev. See Exercise **??** outlines a proof. **TODO**.
>
> Theorem 1.25 shows that a subset of a finite-dimensional Hilbert space is Chebyshev if and only if it is convex. For infinite-dimensional Hilbert spaces it is a famous open problem to determine whether every Chebysev set in a Hilbert space is convex.

Theorem 1.25 establishes the *projection operator*

$$P_S(y) \equiv y \mapsto \operatorname*{argmin}_{x \in S} \|x - y\|_2, \tag{1.9}$$

which associates to each vector $y \in \mathbb{R}^n$ the unique point in $P_S(y) \in S$ that is called the (orthogonal) projection of y onto S.

Some basic characterizations of the projection operator are summarized below.

Theorem 1.26. *Let C be a closed convex set in $\mathbb{R}^n$, $y \in \mathbb{R}^n$ arbitrary, $z \in C$ an arbitrary but fixed point, and P_C the projection operator. Then, the following are equivalent:*

(i) $x^* = P_C(y)$

(ii) For each $x \in C$, the map $t \mapsto \|(1-t)x + tx^* - y\|_2$ is decreasing on $[0,1]$.

(iii) $\langle x - x^*,\, y - x^* \rangle \leq 0$ *for all* $x \in C$

(iv) For each $x \in C$, $\|x^* - y\|_2 = \min\{\|(1-t)x + tx^* - y\|_2 \mid 0 \leq t \leq 1\}$.

(v) For each $x \in C$, $\|x - (2x^* - y)\|_2 \leq \|x - y\|_2$.

Proof. In (v), the operator $R_C := 2P_C - \mathrm{Id}$ is called the *reflection* operator. Thus, (v) states that $\|R_C y - x\| \leq \|y - x\|$ for all $x \in C$. □

Corollary 1.27. *The projection operator is nonexpansive and monotone.*

Proof. □

1.2.1 Projection onto a cone

- Projection characterization
- Moreau's decomposition

1.2.2 Separation

Theorem 1.28. *Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set. Any point $x \notin C$ can be* separated *from C. That is, there exists an $\boldsymbol{a} \in \mathbb{R}^n$ such that*

$$\langle \boldsymbol{a},\, \boldsymbol{x} \rangle > \sup\{\langle \boldsymbol{a},\, \boldsymbol{y} \rangle \mid \boldsymbol{y} \in C\}. \tag{1.10}$$

Proof. Use choice $\boldsymbol{a} = \boldsymbol{x} - P_C(\boldsymbol{x})$. Since $\boldsymbol{x} \notin C$, $\boldsymbol{a} \neq 0$. Theorem 1.26-(iii) tells us that

$$\begin{aligned} \forall y \in C, \quad & \langle \boldsymbol{y} - P_C(\boldsymbol{x}),\, \boldsymbol{a} \rangle \leq 0 \leftrightarrow \langle \boldsymbol{a},\, \boldsymbol{x} - \boldsymbol{a} - \boldsymbol{y} \rangle \geq 0 \\ & \langle \boldsymbol{a},\, \boldsymbol{x} \rangle - \langle \boldsymbol{a},\, \boldsymbol{y} \rangle \geq \|\boldsymbol{a}\|^2 > 0 \quad \implies \langle \boldsymbol{a},\, \boldsymbol{x} \rangle - \|\boldsymbol{a}\|^2 \geq \langle \boldsymbol{a},\, \boldsymbol{y} \rangle. \end{aligned}$$

This shows that $\boldsymbol{s}$ separates $\boldsymbol{x}$ from $\boldsymbol{y}$. Notice, we may choose $\|\boldsymbol{a}\| = 1$ if we wish. □

Corollary 1.29. *Strict separation of convex sets.*

Mention: proper separation.

Note: The complexity of optimization over a given convex set depends crucially on the cost to detect membership in the convex set and / or to find a separating hyperplane. There exist convex sets, where testing membership can be NP-Hard.

There are several consequences of this important separation property. We refer the reader to [**?** HUL01] for more details. We will appeal to the separation property in one of our examples below.

1.3 Some important convex sets

1.3.1 Löwner-John ellipsoid, John ellipsoid

TODO

Remark 1.30. Inner approximation and outer approximation.

1.3.2 Doubly-stochastic matrices

We now come to a very important convex set, the set of doubly stochastic matrices.

Definition 1.31. An $n \times n$ matrix $\boldsymbol{A} = [a_{ij}]$ is called *doubly stochastic* if

$$a_{ij} \geq 0 \quad \text{for all } i, j, \tag{1.11}$$

$$\sum\nolimits_{j=1}^{n} a_{ij} = 1 \quad \text{for all } i \tag{1.12}$$

$$\sum\nolimits_{i=1}^{n} a_{ij} = 1 \quad \text{for all } j. \tag{1.13}$$

That is, A is nonnegative, row-stochastic and column-stochastic.

Exercise 1.17. *Verify the following:*

(i) The set $\mathcal{D}_n$ of all $n \times n$ doubly stochastic matrices is a convex set.

(ii) $\mathcal{D}_n$ is closed under multiplication (and the adjoint operation). However, $\mathcal{D}_n$ is not a group (or semigroup).

(iii) Every permutation matrix is doubly stochastic, and is an extreme point of $\mathcal{D}_n$.

We prove below now the famous theorem of Birkhoff that says that every extreme point of $\mathcal{D}_n$ is a permutation matrix, which implies in particular that $\mathcal{D}_n$ is a polytope.

Theorem 1.32 (Birkhoff)**.** *The extreme points of $\mathcal{D}_n$ are the permutation matrices.*

Proof. From Minkowski's theorem (Theorem **??**) we know that **TODO** □

Volume of $\mathcal{D}_n$

Open problem. However, state of the art is **TODO**.
Application of knowing the volume of this polytope: *can sample uniformly from $\mathcal{D}_n$.*

Proposition 1.33. *Let $P := \{x \mid Ax = b, x \geq 0\}$ be a closed convex polyhedron. Prove that a nonzero $x \in P$ is extremal in P if and only if the columns of A corresponding to $x_i > 0$ are linearly independent.*

1.3.3 Numerical range

Given a complex matrix $A \in \mathbb{C}^{n \times n}$ the quadratic form $\frac{z^*Az}{z^*z}$ yields a complex number (Rayleigh quotient) that is reminiscent of an eigenvalue. **?**] introduced the notion of *field of values*

$$W(A) := \{z^*Az \mid \|z\| = 1\}. \tag{1.14}$$

Clearly, $W(A)$ is compact and connected. **?**] showed that $W(A)$ has a convex outer boundary and shortly thereafter **(author?)** [Hau19] showed that $W(A)$ itself is convex. This result is called the *Toeplitz-Hausdorff* theorem in their honor. We present a few related results below.

We begin with a simple exercise.

Exercise 1.18. *Let $A \in \mathbb{H}^n$ be an $n \times n$ Hermitian matrix. Prove that the set*

$$S_A := \{\langle Az,\, z\rangle \mid z \in \mathbb{C}^n\},$$

is a convex cone.

1.18 Clearly, S_A is a cone, for if a complex number $w \in S_A$, then αw for $\alpha > 0$ also lies in S_A. Now notice that $S_A \subset \mathbb{R}$. Depending on A, S_A is either a ray or the entire real line, in any case, a convex cone.

Theorem 1.34 (Dines)**.** *Let $A, B \in \mathbf{S}^n$ ($n \times n$ symmetric matrices). Then,*

$$K := \{(x^TAx, x^TBx) \mid x \in \mathbb{R}^n\} \subset \mathbb{R}^2,$$

is a convex cone.

Theorem 1.35 (Brickman)**.** *Let $A, B \in \mathbf{S}^n$ ($n \times n$ symmetric matrices). Then, the set*

$$R(A,B) := \{(x^TAx, x^TBx) \mid \|x\|_2 = 1\} \subset \mathbb{R}^2,$$

is a compact convex set for $n \geq 3$.

Exercise 1.19. *Show that Theorem 1.35 implies Theorem 1.34.*

Exercise 1.20. *Let $A_1, \ldots, A_m \in \mathbb{C}^{n\times n}$. Is the following set (which is a subset of $\mathbb{C}^m$) convex?*

$$W(A_1,\ldots,A_m) := \{(\langle A_1z,\, z\rangle, \ldots, \langle A_mz,\, z\rangle) \mid z \in \mathbb{C}^n,\ \ \|z\|_2 = 1\}.$$

?] actually shows equivalence of his result with the Toeplitz-Hausdorff theorem.

Numerical ranges are fascinating objects. There are several open problems concerning their geometry; we mention one of them below.

Open problem

Develop necessary and sufficient conditions for the origin to be a point of $W(A)$

1.4 Miscellaneous topics⋆

1.4.1 Volumes and areas of convex bodies⋆

Volumes of simplices: More generally, integrals over simplices can be nicely computed. Summarize here the contents of the two papers of Jesus Loera.

Notes

1. Brief history (Arichmedes times, then Minkowski)
2. See the software `polymake` for polyhedra; going between vertex and facet representation
3. Software for generating convex hulls
4. Computational complexity of Euclidean convex hulls; higher-dimensional convex hulls
5. Recovering a convex body from its support measurements
6. Material in convex geometry; basic results like Caratheodory, Helly's theorem, Radon's theorem, and other key material from Rockafellar, Gruenbaum, etc.; give a brief summary of several of these books.
7. Somewhere have to weave in Minkowski's theorem, Krein-Milman, Examples of extreme points of certain norms balls
8. See [Fan57] for the doubly stochastic matrix based proof of the HLP result
9. See [Bor07] for updated information about Chebyshev sets.
10. Concept of uniform convexity in Banach spaces was introduced by J. A. Clarkson in 1936.
11. Klee sets (then in next chapter discuss Wangs paper on Chebyshev and Klee functions)
12. Summarize Helton's paper on possible shapes of numerical ranges; mention both the results of that paper.

Bibliography

[Bor07] JonathanM. Borwein. Proximality and chebyshev sets. *Optimization Letters*, 1(1):21–32, 2007.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004.

[Fan57] Ky Fan. Existence theorems and extreme solutions for inequalities concerning convex functions or linear transformations. *Mathematische Zeitschrift*, 68(1):205–216, 1957.

[Hau19] Felix Hausdorff. Der Wertvorrat einer Bilinearform. *Math. Zeit.*, 3:314–316, 1919.

[HUL01] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer, 2001.

[Kim12] Jaegil Kim. Minimal volume product near Hanner polytopes. *arXiv:1212.2544*, 2012.

[NP06] C. Niculesu and L. E. Persson. *Convex functions and their applications: a contemporary approach*, volume 13 of *Science & Business*. Springer, 2006.

[Pap05] A. Papadopoulos. *Metric spaces, convexity and nonpositive curvature*. Europ. Math. Soc., 2005.

[Roc70] R. T. Rockafellar. *Convex Analysis*. Princeton Univ. Press, 1970.