
MSE2122 - Nonlinear Optimization Lecture Notes II

Fernando Dias (based on previous version by
Fabricio Oliveira)

September 3, 2023

Abstract

In this lecture, we discuss concepts related to sets, which will later in the course serve as a framework for dealing with constraints in optimisation problems. We will also see that the convexity of sets can be used to infer the convexity of functions, a topic we will discuss in detail in the next class. We describe how to identify convexity in sets and discuss set-related concepts practically when discussing optimality conditions later. We also discuss the concepts of separation and support, which are central to several results in optimisation theory, such as the Farkas theorem.

Contents

1	Convexity and optimisation	2
2	Identifying convexity of sets	2
2.1	Convexity-preserving set operations	3
2.2	Examples of convex sets	3
2.2.1	Hyperplanes and halfspaces	4
2.2.2	Norm balls and norm cones	5
3	Convex hulls	5
4	Closure and interior of sets	6
4.1	Closure, interior and boundary of a set	6
4.2	The Weierstrass theorem	7
5	Separation and support of sets	8
5.1	Hyperplanes and closest points	8
5.2	Halfspaces and separation	8
5.3	Farkas' theorem	10
5.4	Supporting hyperplanes	10

1 Convexity and optimisation

Convexity is perhaps the most important property that the elements forming an **optimization problem** can present. Paraphrasing Tyrrell Rockafellar:

... in fact, the great watershed in optimization is not between linearity and nonlinearity but convexity and nonconvexity.

In a nutshell, convexity allows us to infer the **global properties** of a solution (i.e., that holds for all of its domain) by considering exclusively local information (such as gradients, for example). Such property is critical in **optimization** since most methods we know to perform well in practice are designed to find solutions that satisfy local optimality conditions. Once convexity is attested, one can then guarantee that these local solutions are, in fact, **globally optimal** without exhaustively exploring the solution space.

For a problem of the form:

$$(P) : \min. f(x) \\ \text{subject to: } x \in X$$

to be convex, we need to verify whether f is a **convex function** and X is a **convex set**. If both statements hold, we can conclude that P is a **convex problem**. We start looking into how to identify convex sets since we can use the convexity of sets to infer the convexity of functions.

2 Identifying convexity of sets

Before formally defining convex sets, let us look at the idea of **combinations**. For that, let $S \subseteq \mathbb{R}^n$ be a set and $x_j \in S$ for $j = 1, \dots, k$ be a collection of vectors (i.e., n -dimensional “points”) belonging to S . Then, we have that:

- A **linear combination** of x_j for $j = 1, \dots, k$ is the set

$$\{x \in \mathbb{R}^n : \sum_{j=1}^k \lambda_j x_j, \lambda_j \in \mathbb{R} \text{ for } j = 1, \dots, k\}. \quad (1)$$

- An **affine combination** is a linear combination, with the additional constraint that $\sum_{j=1}^k \lambda_j = 1$. That is,

$$\{x \in \mathbb{R}^n : \sum_{j=1}^k \lambda_j x_j, \sum_{j=1}^k \lambda_j = 1, \lambda_j \in \mathbb{R} \text{ for } j = 1, \dots, k\}. \quad (2)$$

- A **conic combination** is a linear combination with the additional condition that $\lambda_j \geq 0$ for $j = 1, \dots, k$.

$$\{x \in \mathbb{R}^n : \sum_{j=1}^k \lambda_j x_j, \lambda_j \geq 0 \text{ for } j = 1, \dots, k\}. \quad (3)$$

- And finally, a **convex combination** is the intersection between an affine and a conic combination, implying that $\lambda_j \in [0, 1]$.

$$\{x \in \mathbb{R}^n : \sum_{j=1}^k \lambda_j x_j, \sum_{j=1}^k \lambda_j = 1, \lambda_j \geq 0 \text{ for } j = 1, \dots, k\}. \quad (4)$$

We say a set is convex if it contains all points formed by the convex combination of any pair of points in this set, which is equivalent to saying that the set contains the line segment between any two points belonging to the set.

Definition 2.1. Convex sets

A set $S \subseteq \mathbb{R}^n$ is said to be convex if $\bar{x} = \sum_{j=1}^k \lambda_j x_j$ belongs to S , where $\sum_{j=1}^k \lambda_j = 1$, $\lambda_j \geq 0$ and $x_j \in S$ for $j = 1, \dots, k$.

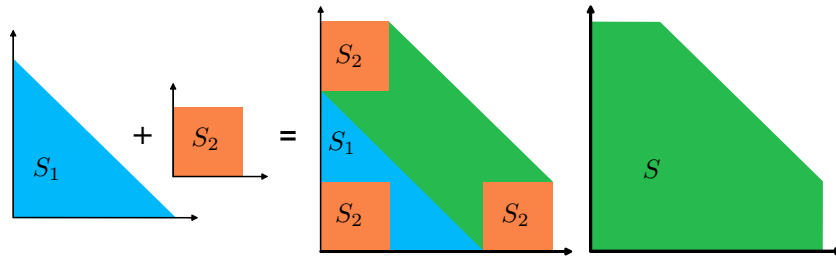


Figure 1: Minkowski sum of two convex sets.

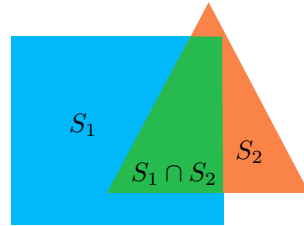


Figure 2: Intersection of two convex sets.

Definition 2.1 is useful as it shows that some set operations preserve convexity.

2.1 Convexity-preserving set operations

Lemma 2.2. Convexity-preserving operations Let S_1 and S_2 be convex sets in \mathbb{R}^n . Then, the sets resulting from the following operations are also convex.

1. Intersection: $S = S_1 \cap S_2$;
2. Minkowski addition: $S = S_1 + S_2 = \{x_1 + x_2 : x_1 \in S_1, x_2 \in S_2\}$;
3. Minkowski difference: $S = S_1 - S_2 = \{x_1 - x_2 : x_1 \in S_1, x_2 \in S_2\}$;
4. Affine transformation: $S = \{Ax + b : x \in S_1\}$.

Figures 1 and 2 illustrate the concept behind some of these set operations. Showing that the sets resulting from the operations in Lemma 2.2 are convex typically shows that convex combinations of elements in the resulting set S also belong to S_1 and S_2 .

2.2 Examples of convex sets

There are several familiar sets that are known to be convex. Knowing that these sets are convex is useful as a **building block** for determining the convexity of more complicated sets.

Some important examples of convex sets include:

- The empty set \emptyset , any singleton $\{\bar{x}\}$ and the whole space \mathbb{R}^n ;
- **halfspaces**: $S = \{x : p^\top x \leq \alpha\} \subset \mathbb{R}^n$;
- **hyperplanes**: $H = \{x : p^\top x = \alpha\} \subset \mathbb{R}^n$, where $p \neq 0^n$ is a normal vector and $\alpha \in \mathbb{R}$ is a scalar. Notice that H can be equivalently represented as $H = \{x \in \mathbb{R}^n : p^\top (x - \bar{x}) = 0\}$ for $\bar{x} \in H$;
- **polyhedral sets**: $P = \{x : Ax \leq b\} \subset \mathbb{R}^n$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$;
- **norm-induced sets (balls)**: $B = \{x : \|x - \bar{x}\| \leq \alpha\} \subseteq \mathbb{R}^n$, where $\|\cdot\|$ is any norm and α a scalar;
- **norm cones**: $C = \{(x, \alpha) \in \mathbb{R}^{n+1} : \|x\| \leq \alpha\}$;

For example, considering the polyhedral set $P = \{x \in \mathbb{R}^n : Ax \leq b\} \subset \mathbb{R}^n$ with A being a $m \times n$

matrix. Notice that S is the intersection of a collection of half-spaces $H_i = \{x \in \mathbb{R}^n : a_i^\top x \leq b_i\}$, where a_i are vectors from the rows of the matrix A and b_i are the components of the column vector b . We know that H_i are convex sets. Thus, $P = \cap_{i=1}^m H_i$ is also convex, as the intersection of sets is a convexity-preserving set operation.

2.2.1 Hyperplanes and halfspaces

Hyperplanes and **half-spaces** will play a central role in future topics. Therefore, let us discuss some important aspects related to these convex sets. First, notice that, geometrically, a hyperplane $H \subset \mathbb{R}^n$ can be interpreted as the set of points with a **constant** inner product to a given vector $p \in \mathbb{R}^n$, while \bar{x} determines the offset of the hyperplane from the origin. That is:

$$H = \{x : p^\top(x - \bar{x}) = 0\} \equiv \bar{x} + p^\perp,$$

where p^\perp is the orthogonal complement of p , i.e., the set of vectors orthogonal to p , which is given by $\{x \in \mathbb{R}^n : p^\top x = 0\}$.

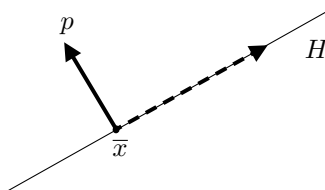


Figure 3: A hyperplane $H = \{x \in \mathbb{R}^n : p^\top(x - \bar{x}) = 0\}$ with normal vector p displaced to \bar{x} .

Analogously, a half-spaces can be represented as $S = \{x \in \mathbb{R}^n : p^\top(x - \bar{x}) \leq 0\}$ where $p^\top \bar{x} = \alpha$ is the hyperplane that forms the boundary of the half-space. This definition suggests a simple geometrical interpretation: the half-space S consists of \bar{x} plus any vector with an obtuse or right angle (i.e., greater or equal to 90°) with the outward normal vector p .

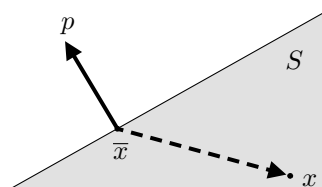


Figure 4: A halfspace $S = \{x \in \mathbb{R}^n : p^\top(x - \bar{x}) \leq 0\}$ defined by the same hyperplane H . Notice how the vectors p (or $p - \bar{x}$, which is fundamentally the same vector but translated to \bar{x}) and $x - \bar{x}$ form angles greater or equal than 90° .

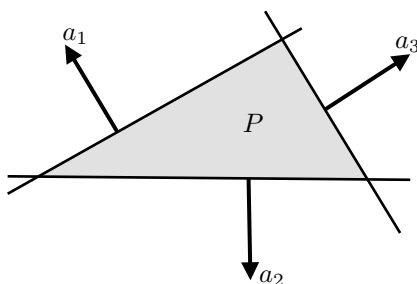


Figure 5: A polyhedron P formed by the intersection of three half-space. Each hyperplane $H_i = \{x \in \mathbb{R}^n : a_i^\top x \leq b_i\}$, for $i = 1, 2, 3$, has a normal vector a_i , and has an offset from the origin b_i (which cannot be seen once project on a 2-dimensional plane as in the picture).

2.2.2 Norm balls and norm cones

An **Euclidean ball** (or **simply ball**) of radius ϵ in \mathbb{R}^n has the form:

$$B(\bar{x}, \epsilon) = \{x \in \mathbb{R}^n : \|x - \bar{x}\|_2 \leq \epsilon\} \equiv \{x \in \mathbb{R}^n : (x - \bar{x})^\top (x - \bar{x}) \leq \epsilon^2\}$$

As one might suspect, balls are **convex**, which can be proved by noting that:

$$\begin{aligned} \|\lambda x_1 + (1 - \lambda)x_2 - \bar{x}\|_2 &= \|\lambda(x_1 - \bar{x}) + (1 - \lambda)(x_2 - \bar{x})\|_2 \\ &\leq \lambda\|x_1 - \bar{x}\|_2 + (1 - \lambda)\|x_2 - \bar{x}\|_2 \leq \epsilon. \end{aligned}$$

Notice that between the first and the second line, we use the triangle inequality, which states that $\|x + y\| \leq \|x\| + \|y\|$ for any two vectors x and y and any norm (including the Euclidean norm).

Euclidean balls are a special case of norm balls, which are defined as $B(\bar{x}, \epsilon) = \{x \in \mathbb{R}^n : \|x - \bar{x}\| \leq \epsilon\}$ where $\|\cdot\|$ is any norm on \mathbb{R}^n .

A related set is the norm cone, defined as $C(x, \alpha) = \{(x, \alpha) \in \mathbb{R}^{n+1} : \|x\| \leq \alpha\}$, where α is a scalar. For example, the second-order cone (also known as the ice cream cone or Lorentz cone) is the norm cone for the Euclidean norm.

Remark. Norm induced sets (balls or cones) are convex for any norm $\|x\|_p = (\sum_{i=1}^n x_i^p)^{\frac{1}{p}}$ for $x \in \mathbb{R}^n$ and $p \geq 1$.

3 Convex hulls

A **convex hull** of a set S , denoted $\mathbf{conv}(S)$, is the set formed by all convex combinations of all points in S . As the name suggests, $\mathbf{conv}(S)$ is a convex set, regardless of S being or not convex.

Another interpretation for $\mathbf{conv}(S)$ is to consider it the tightest enveloping (convex) set containing S . Notice that, if S is convex, then $S = \mathbf{conv}(S)$. Formally, convex hulls are defined as follows.

Definition 3.1. Convex hull of a set

Let $S \subseteq \mathbb{R}^n$ be an arbitrary set. The convex hull of S , denoted by $\mathbf{conv}(S)$, is the collection of all convex combinations of S . That is, for $x_j \in S$, with $j = 1, \dots, k$, $x \in \mathbf{conv}(S)$ if and only if

$$x = \sum_{j=1}^k \lambda_j x_j : \sum_{j=1}^k \lambda_j = 1, \lambda_j \geq 0, \text{ for } j = 1, \dots, k.$$

From Definition 3.1, one can show that the convex hull $\mathbf{conv}(S)$ can also be defined as the **intersection** of all convex sets containing S . Perhaps the easiest way to visualize this is to think of the infinitely many half-spaces containing S and their intersection, which can only be S . Figures 6 illustrates the convex hull $\mathbf{conv}(S)$ of a nonconvex set S .

The notion of convex hulls is a powerful tool in optimization. One important application is using

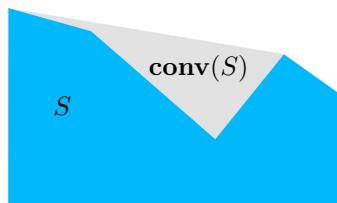


Figure 6: Example of an arbitrary set S (in solid blue) and its convex hull $\mathbf{conv}(S)$ (combined blue and grey areas).

$\mathbf{conv}(S)$ to obtain approximations for a nonconvex S that can be exploited to solve an optimization problem with a constraint set defined by S . This is the underpinning technique in many important optimization methods, such as branch-and-bound-based methods for nonconvex problems and decomposition methods (i.e., methods that solve large problems by breaking them into smaller parts that are presumably easier to solve).

Let us consider the convex hull of a finite collection of discrete points. Some of these sets are so

important in optimization that they have their own names.

Definition 3.2

Let $S = \{x_1, \dots, x_{n+1}\} \subset \mathbb{R}^n$. Then $\text{conv}(S)$ is called a **polytope**. If x_1, \dots, x_{n+1} are affinely independent (i.e., $x_2 - x_1, \dots, x_{n+1} - x_1$ are linearly independent) then $\text{conv}(S)$ is called a **simplex** with vertices x_1, \dots, x_{n+1} .

4 Closure and interior of sets

Many of the set-related results we will see in this course depend on the characteristics of the set itself. Often, assuming properties such as closedness or compactness considerably eases technical derivations.

4.1 Closure, interior and boundary of a set

Let us define some properties that will be useful in this course. For that, we will use an ϵ -neighbourhood of $x \in \mathbb{R}^n$ (which is a norm ball of radius ϵ centred in x) defined as:

$$N_\epsilon(x) = \{y : \|y - x\| < \epsilon\}.$$

Let $S \subseteq \mathbb{R}^n$ be an arbitrary set. We can use N_ϵ to define the following elements related to S .

1. **Closure of S** : the set defined by the closure of S , denoted $\text{clo}(S)$, is defined as

$$\text{clo}(S) = \{x \in \mathbb{R}^n : S \cap N_\epsilon(x) \neq \emptyset \text{ for every } \epsilon > 0\}.$$

Notice that the closure might contain points not belonging to S . We say that a set is **closed** if $S = \text{clo}(S)$; the set itself is its own closure.

2. **Interior of S** : the interior of S , denoted $\text{int}(S)$, is the set

$$\text{int } S = \{x \in S : N_\epsilon(x) \subset S \text{ for some } \epsilon > 0\}.$$

If S is the same as its own interior, then we say that S is **open**. Some authors say that S is solid if it has a nonempty interior (that is, $\text{int}(S) \neq \emptyset$). Notice that the interior of S is a subset of S , that is, $\text{int}(S) \subseteq S$.

3. **Boundary of S** : the boundary of S , denoted $\text{bou}(S)$ is the collection of points defined by:

$$\text{bou}(S) = \{x \in \mathbb{R}^n : N_\epsilon(x) \text{ contains some } x_i \in S \text{ and some } x_j \notin S \text{ for every } \epsilon > 0\}.$$

We say that S is bounded if exists $N_\epsilon(x)$, $x \in \mathbb{R}^n$, for some $\epsilon > 0$ such that $S \subset N_\epsilon(x)$.

We say that a set is **compact** if it is both **closed** and **bounded**. Compact sets appear very frequently in real-world optimization applications since, typically, one can assume the existence of bounds for decision variables (such as nonnegativity or maximum physical bounds or, in an extreme case, smallest/largest computational constants). Another frequent example of a bounded set is the convex hull of a collection of discrete points, which is called by some authors **polytopes** (effectively bounded polyhedral sets).

Let us consider the following example. Let $S = \{(x_1, x_2) \in \mathbb{R}^n : x_1^2 + x_2^2 \leq 1\}$. Then, we have that:

1. $\text{clo}(S) = \{(x_1, x_2) \in \mathbb{R}^n : x_1^2 + x_2^2 \leq 1\}$. Since $S = \text{clo}(S)$, S is closed.
2. $\text{int}(S) = \{(x_1, x_2) \in \mathbb{R}^n : x_1^2 + x_2^2 < 1\}$.
3. $\text{bou}(S) = \{(x_1, x_2) \in \mathbb{R}^n : x_1^2 + x_2^2 = 1\}$. Notice that it makes S bounded.
4. S is compact since it is closed and bounded.

Notice that, if S is closed, then $\text{bou}(S) \subset S$. That is, its boundary is part of the set itself. Moreover, it can be shown that $\text{clo}(S) = \text{bou}(S) \cup S$ is the smallest closed set containing S .

In case S is convex, one can infer the convexity of the interior $\text{int}(S)$ and its closure $\text{clo}(S)$. The following theorem summarises this result.

Theorem 4.1

Let $S \subseteq \mathbb{R}^n$ be a convex set with $\text{int}(S) \neq \emptyset$. Let $x_1 \in \text{clo}(S)$ and $x_2 \in \text{int}(S)$. Then $x = \lambda x_1 + (1 - \lambda)x_2 \in \text{int}(S)$ for all $\lambda \in (0, 1)$.

Theorem 4.1 helps infer the convexity of the elements related to S . We summarise the key results in the following corollary.

Corollary 4.2. Let S be a convex set with $\text{int}(S) \neq \emptyset$. Then

1. $\text{int}(S)$ is convex;
2. $\text{clo}(S)$ is convex;
3. $\text{clo}(\text{int}(S)) = \text{clo}(S)$;
4. $\text{int}(\text{clo}(S)) = \text{int}(S)$.

4.2 The Weierstrass theorem

The **Weierstrass theorem** is a result that guarantees the existence of optimal solutions for optimization problems. To make it more precise, let:

$$(P) : z = \min. \{f(x) : x \in S\}$$

be our optimization problem. If an optimal solution x^* exists, then $f(x^*) \leq f(x)$ for all $x \in S$ and $z = f(x^*) = \min\{f(x) : x \in S\}$.

Notice the difference between $\min.$ (an abbreviation for **minimize**) and the operator \min . The first is meant to represent the problem of **minimizing** the function f in the domain S , while \min is shorthand for **minimum**, in this case z , assuming it is attainable.

It might be that an optimal solution is not attainable, but a bound can be obtained for the optimal solution value. The greatest lower bound for z is its **infimum** (or **supremum** for maximization problems), denoted by \inf . That is, if $z = \inf\{f(x) : x \in S\}$, then $z \leq f(x)$ for all $x \in S$ and there is no $\bar{z} > z$ such that $\bar{z} \leq f(x)$ for all $x \in S$. We might sometimes use the notation:

$$(P) : z = \inf\{f(x) : x \in S\}$$

to represent optimization problems for which one cannot be sure whether an optimal solution is attainable. The Weierstrass theorem describes the situations in which those minimums (or maximums) are guaranteed to be attained, which is the case whenever S is compact.

Theorem 4.3. Weierstrass theorem

Let $S \neq \emptyset$ be a compact set, and let $f : S \rightarrow \mathbb{R}$ be continuous on S . Then there exists a solution $\bar{x} \in S$ to $\min. \{f(x) : x \in S\}$.

Figure 7 illustrates three examples. In the first (on the left), the domain $[a, b]$ is compact; thus, the minimum of f is attained at b . In the other two, $[a, b]$ is open; therefore, the Weierstrass theorem does not hold. In the middle example, one can obtain $\inf f$, which is not the case for the last example on the right.

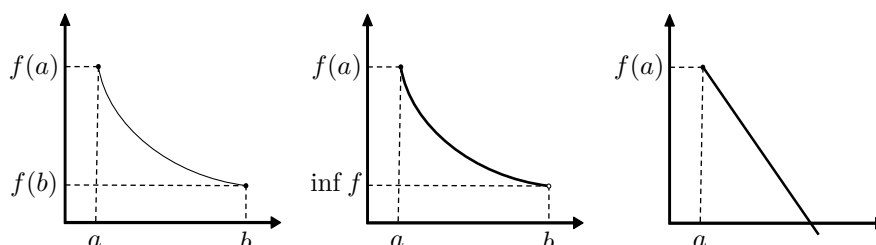


Figure 7: Examples of attainable minimum (left) and infimum (centre) and an example where neither are attainable (right).

5 Separation and support of sets

The concepts of **separation** and **support** of sets are vital for establishing optimality conditions later in this course. We are interested in mechanisms that allow one to infer whether there exists hyperplanes separating points from sets (or sets from sets). We will also be interested in means to, given a point $x \notin S$, find the closest point not belonging to S .

5.1 Hyperplanes and closest points

We start with how to identify the closest points to sets.

Theorem 5.1. Closest-point theorem

Let $S \neq \emptyset$ be a closed convex set in \mathbb{R}^n and $y \notin S$. Then, there exists a unique point $\bar{x} \in S$ with minimum distance from y . In addition, \bar{x} is the minimising point if and only if

$$(y - \bar{x})^\top (x - \bar{x}) \leq 0, \text{ for all } x \in S$$

Simply put, if S is a closed convex set, then $\bar{x} \in S$ will be the closest point to $y \notin S$ if the vector $y - \bar{x}$ is such that it forms an angle that is greater or equal than 90° with all other vectors $x - \bar{x}$ for $x \in S$. Figure 8 illustrates this logic.

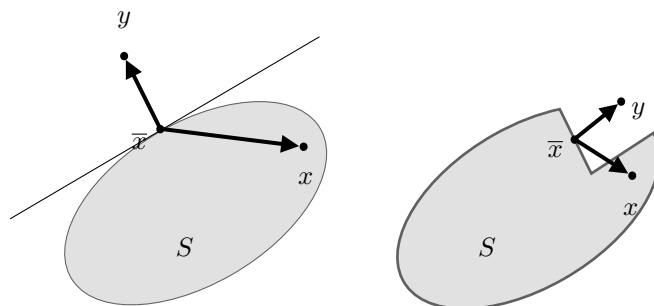


Figure 8: Closest-point theorem for a closed convex set (on the left). On the right is an illustration of how the absence of convexity invalidates the result.

Notice that S lies in the half-space $(y - \bar{x})^\top (x - \bar{x}) \leq 0$ defined by the hyperplane $p^\top (x - \bar{x}) = 0$ with normal vector $p = (y - \bar{x})$. We will next revise the concepts of half-spaces and hyperplanes since they will play a central role in the derivations in this course.

5.2 Halfspaces and separation

We can use half-spaces to build the concept of separation. Let us start recalling that a hyperplane $H = \{x : p^\top x = \alpha\}$ with normal vector $p \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ defines two half-spaces $H^+ = \{x : p^\top x \geq \alpha\}$ and $H^- = \{x : p^\top x \leq \alpha\}$. Figure 9 illustrates the concept. Notice how the vector p lies in the half-space H^+ .

Any hyperplane H can be defined in reference to a point $\bar{x} \in H$ by noticing that:

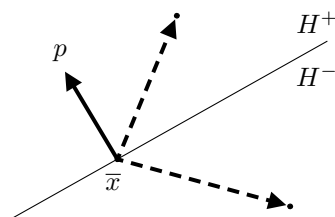


Figure 9: Normal vectors, hyperplane and halfspaces

$$p^\top(x - \bar{x}) = p^\top x - p^\top \bar{x} = \alpha - \alpha = 0.$$

From that, the half-spaces defined by H can be equivalently stated as $H^+ = \{x : p^\top(x - \bar{x}) \geq 0\}$ and $H^- = \{x : p^\top(x - \bar{x}) \leq 0\}$.

We can now define the separation of convex sets.

Definition 5.2

Let S_1 and S_2 be nonempty sets in \mathbb{R}^n . The hyperplane $H = \{x : p^\top x = \alpha\}$ is said to **separate** S_1 and S_2 if $p^\top x \geq \alpha$ for each $x \in S_1$ and $p^\top x \leq \alpha$ for each $x \in S_2$. In addition, the following apply:

1. **Proper separation:** $S_1 \cup S_2 \not\subset H$;
2. **Strict separation:** $p^\top x < \alpha$ for each $x \in S_1$ and $p^\top x > \alpha$ for each $x \in S_2$;
3. **Strong separation:** $p^\top x \geq \alpha + \epsilon$ for some $\epsilon > 0$ and $x \in S_1$, and $p^\top x \leq \alpha$ for each $x \in S_2$.

Figure 10 illustrates the three types of separation in Definition 5.2. On the left, proper separation is illustrated, obtained by any hyperplane that does not contain both S_1 and S_2 but might contain points from either or both. In the middle, sets S_1 and S_2 belong to two distinct half-spaces in a strict sense. On the right, strict separation holds with an additional margin $\epsilon > 0$, which is defined as strong separation.

A powerful yet simple result that we will use later is that, for a closed convex set S , there always exists

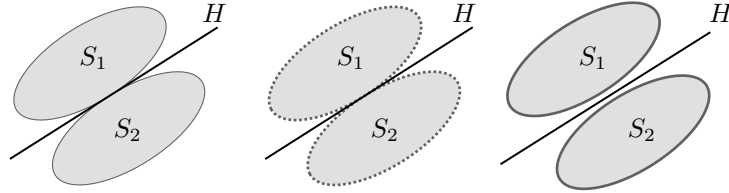


Figure 10: Three types of separation between S_1 and S_2 .

a hyperplane separating S and a point y that does not belong to S .

Theorem 5.3. Separation theorem

Let $S \neq \emptyset$ be a closed convex set in \mathbb{R}^n and $y \notin S$. Then, there exists a nonzero vector $p \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that $p^\top x \leq \alpha$ for each $x \in S$ and $p^\top y > \alpha$.

Proof. Theorem 5.1 guarantees the existence of a unique minimising $\bar{x} \in S$ such that $(y - \bar{x})^\top(x - \bar{x}) \leq 0$ for each $x \in S$. Let $p = (y - \bar{x}) \neq 0$ and $\alpha = \bar{x}^\top(y - \bar{x}) = p^\top \bar{x}$. Then we get $p^\top x \leq \alpha$ for each $x \in S$, while $p^\top y - \alpha = (y - \bar{x})^\top(y - \bar{x}) = \|y - \bar{x}\|^2 > 0$. 🤔

This is the first proof we look at in these notes, and the reason for that is its importance in many of the results we will discuss further. The proof first looks at the problem of finding a minimum distance point as an optimization problem and uses the Weierstrass theorem (our Theorem 5.1 is a consequence of the Weierstrass theorem stated in Theorem 4.3) to guarantee that such a \bar{x} exists. Being a minimum distance point, we know from Theorem 5.1 that $(y - \bar{x})^\top(x - \bar{x}) \leq 0$ holds. Now, by defining p and α as in the proof, one might notice that:

$$(y - \bar{x})^\top(x - \bar{x}) \leq 0 \Leftrightarrow (y - \bar{x})^\top x \leq (y - \bar{x})^\top \bar{x} \Leftrightarrow p^\top x \leq p^\top \bar{x} = \alpha.$$

The inequality $p^\top y > \alpha$ is demonstrated to hold in the final part by noticing that:

$$\begin{aligned} p^\top y - \alpha &= (y - \bar{x})^\top y - \bar{x}^\top(y - \bar{x}) \\ &= y^\top(y - \bar{x}) - \bar{x}^\top(y - \bar{x}) \\ &= (y - \bar{x})^\top(y - \bar{x}) = \|y - \bar{x}\|^2 > 0. \end{aligned}$$

Theorem 5.3 has interesting consequences. For example, one can apply it to every point in the boundary $\text{bou}(S)$ to show that S is formed by the intersection of all half-spaces containing S .

Another interesting result is the existence of strong separation. If $y \notin \text{clo}(\text{conv}(S))$, then one can show that a strong separation between y and S exists since there will surely be a distance $\epsilon > 0$ between y and S .

5.3 Farkas' theorem

Farkas' theorem plays a central role in deriving optimality conditions. It can assume several alternative forms, typically referred to as Farkas' lemmas. In essence, Farkas' theorem demonstrates that a given system of linear equations has a solution if and only if a related system can be shown to have no solutions and vice-versa.

Theorem 5.4


Let A be an $m \times n$ matrix and c be an n vector. Then exactly one of the following two systems has a solution:

$$(1) : Ax \leq 0, \quad c^\top x > 0, \quad x \in \mathbb{R}^n$$

$$(2) : A^\top y = c, \quad y \geq 0, \quad y \in \mathbb{R}^m.$$

Proof. Suppose (2) has a solution. Let x be such that $Ax \leq 0$. Then $c^\top x = (A^\top y)^\top x = y^\top Ax \leq 0$. Hence, (1) has no solution.

Next, suppose (2) has no solution. Let $S = \{x \in \mathbb{R}^n : x = A^\top y, y \geq 0\}$. Notice that S is closed and convex and that $c \notin S$. By Theorem 5.3, there exists $p \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that $p^\top c > \alpha$ and $p^\top x \leq \alpha$ for $x \in S$.

As $0 \in S$, $\alpha \geq 0$ and $p^\top c > 0$. Also, $\alpha \geq p^\top A^\top y = y^\top Ap$ for $y \geq 0$. This implies that $Ap \leq 0$, and thus p satisfies (1). 

The first part of the proof shows that if we assume that system (2) has a solution, then $c^\top x > 0$ cannot hold for $y \geq 0$. The second part uses the separation theorem (Theorem 5.3) to show that c can be seen as a point not belonging to the closed convex set S for which there is a separation hyperplane and that the existence of such plane implies that system (1) must hold. The set S is closed and convex since it is a conic combination of rows a_i , for $i = 1, \dots, m$. Using the $0 \in S$, one can show that $\alpha \geq 0$. The last part uses the identity $p^\top A^\top = (Ap)^\top$ and the fact that $(Ap)^\top y = y^\top Ap$. Notice that, since y can be arbitrarily large and α is a constant, $y^\top Ap \leq \alpha$ can only hold if $y^\top Ap \leq 0$, requiring that $p \leq 0$ since $y \geq 0$ from the definition of S .

Farkas' theorem has an interesting geometrical interpretation from this proof, as illustrated in Figures 11. Consider the cone C formed by the rows of A :

$$C = \{c \in \mathbb{R}^n : c_j = \sum_{i=1}^m a_{ij} y_i, \quad j = 1, \dots, n, \quad y_i \geq 0, \quad i = 1, \dots, m\}$$

The **polar cone** of C , denoted C^0 , is formed by all vectors having angles of 90° or more with vectors in C . That is:

$$C^0 = \{x : Ax \leq 0\}.$$

Notice that (1) has a solution if the intersection between the polar cone C^0 and the positive (H^+ as defined earlier) half-space $H^+ = \{x \in \mathbb{R}^n : c^\top x > 0\}$ is not empty. If (2) has a solution, as at the beginning of the proof, then $c \in C$ and the intersection $C^0 \cap H^+ = \emptyset$. Now, if (2) does not have a solution, that is, $c \notin C$, then one can see that $C^0 \cap H^+$ cannot be empty, meaning that (1) has a solution.

5.4 Supporting hyperplanes

There is an important connection between the existence of hyperplanes that support a whole set and the optimality conditions of points. Let us first define supporting hyperplanes.

Definition 5.5. Supporting hyperplane

Let $S \neq \emptyset$ be a set in \mathbb{R}^n , and let $\bar{x} \in \text{bou}(S)$. $H = \{x \in \mathbb{R}^n : p^\top (x - \bar{x}) = 0\}$ is a supporting hyperplane of S at \bar{x} if either $S \subseteq H^+$ (i.e., $p^\top (x - \bar{x}) \geq 0$ for $x \in S$) or $S \subseteq H^-$.

Figures 12 illustrates the concept of supporting hyperplanes. Notice that supporting hyperplanes

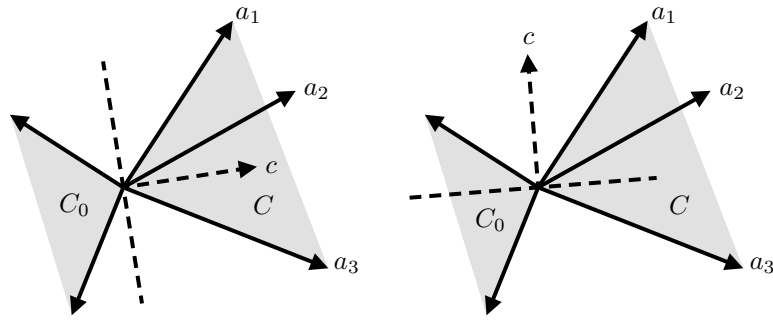


Figure 11: Geometrical illustration of the Farkas' theorem. On the left, system (2) has a solution, while on the right, system (1) has a solution

might not be unique, with the geometry of the set S playing an important role.

Let us define the function $f(x) = p^\top x$ with $x \in S$. One can see that the optimal solution \bar{x} given by:

$$\bar{x} = \operatorname{argmax}_{x \in S} f(x)$$

is a point $x \in S$ for which p is a supporting hyperplane. A simple geometric analogy is to think that the f increases value as one moves toward p . The constraint $x \in S$ will eventually prevent the movement further from S , and this last contact point is precisely \bar{x} . This is a useful concept for optimizing problems using gradients of functions, as we will discuss later in the course.

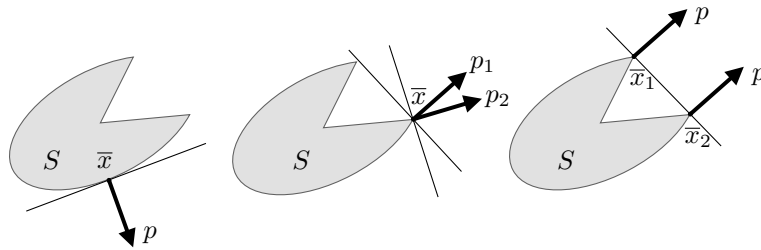


Figure 12: Supporting hyperplanes for an arbitrary set. Notice how a single point might have multiple supporting planes (middle) or different points might have the same supporting hyperplane (right)

One characteristic that convex sets present that will be of great importance when establishing optimality conditions is supporting hyperplanes at every boundary point.

Theorem 5.6

Let $S \neq \emptyset$ be a convex set in \mathbb{R}^n , and let $\bar{x} \in \mathbf{bou}(S)$. Then there exists $p \neq 0$ such that $p^\top (x - \bar{x}) \leq 0$ for each $x \in \mathbf{clo}(S)$.

The proof follows immediately from Theorem 5.3, without explicitly considering a point $y \notin S$ and by noticing that $\mathbf{bou}(S) \subset \mathbf{clo}(S)$. Figures 13 illustrates the theorem.

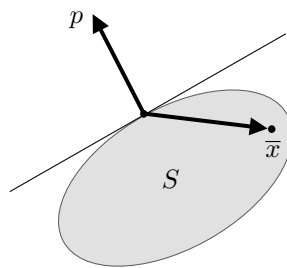


Figure 13: Supporting hyperplanes for convex sets. Notice how every boundary point has at least one supporting hyperplane