Convex sets

We begin by defining the notion of a convex set.

**Definition (convex set).** The set $S \subseteq \mathbb{R}^n$ is convex if

$$x_1, x_2 \in S, \quad \lambda \in (0, 1) \implies \lambda x_1 + (1 - \lambda)x_2 \in S.$$ 

A set is thus convex if all convex combination of any two points in the set lies in the set, see Figure 1.

![Figure 1: A convex set](image)

Examples:

- The empty set is a convex set.
- The set $\{ x \in \mathbb{R}^n \mid \|x\| \leq a \}$ is convex for any $a \in \mathbb{R}$.
- The set $\{ x \in \mathbb{R}^n \mid \|x\| = a \}$ is non-convex for any $a > 0$.
- The set $\{ 0, 1, 2, 3 \}$ is non-convex.

**Proposition.** Let $S_k \subseteq \mathbb{R}^n, k \in K$ be a collection of convex sets. Then the intersection $\bigcap_{k \in K} S_k$ is also a convex set.

**Proof.** See Proposition 3.3. in the book.
We define the affine hull of a finite set $V = \{v^1, v^2, \ldots, v^k\}$ as

$$\text{aff } V := \left\{ \lambda_1 v^1 + \cdots + \lambda_k v^k \mid \lambda_1, \ldots, \lambda_k \in \mathbb{R}, \sum_{i=1}^{k} \lambda_i = 1 \right\}$$

We define the convex hull of a finite set $V = \{v^1, v^2, \ldots, v^k\}$ as

$$\text{conv } V := \left\{ \lambda_1 v^1 + \cdots + \lambda_k v^k \mid \lambda_1, \ldots, \lambda_k \geq 0, \sum_{i=1}^{k} \lambda_i = 1 \right\}$$

The sets are defined by all possible affine (convex) combinations of the $k$ points.

In general, we can define the convex hull of a set $S$ as

a) the unique minimal convex set containing $S$,

b) the intersection of all convex sets containing $S$, or

c) the set of all convex combinations of points in $S$.

Any point $x \in \text{conv } S$, where $S \subseteq \mathbb{R}^n$ can thus be expressed as a convex combination of points in $S$. But how many points are required? The answer is, according to the following theorem, $n + 1$ points.

**Theorem** (Caratheodory's theorem). Let $x \in \text{conv } S$, where $S \subseteq \mathbb{R}^n$. Then $x$ can be expressed as a convex combination of $n + 1$ or fewer points of $S$.

**Proof.** See Theorem 3.11. in the book.

We now define the notion of a polytope.

**Definition** (polytope). A set $P$ is a polytope if it is the convex hull of finitely many points in $\mathbb{R}^n$.

A cube and a tetrahedron are examples of polytopes in $\mathbb{R}^3$. In Figure 3, a polytope in $\mathbb{R}^2$ generated by seven points is illustrated.
Definition (extreme point). *A point* $v$ *of a convex set* $P$ *is an extreme point if whenever*

$$v = \lambda x^1 + (1 - \lambda)x^2, \quad x^1, x^2 \in P, \quad \lambda \in (0, 1) \implies v = x^1 = x^2.$$  

An extreme point of $P$ is thus a point in $P$ that can not be represented as a convex combination of two other points, see Figure 3 for an example. We can now formulate the following intuitive theorem.

**Theorem.** *Let* $P$ *be the polytope* $\text{conv } V$, *where* $V = \{v^1, \ldots, v^k\} \subset \mathbb{R}^n$. *Then* $P$ *is equal to the convex hull of its extreme points.*

The theorem basically says that the interesting points in a polytope are the extreme points. By combining the two last theorems, we can deduce that any point in a polytope $P$ can be described as a convex combination of at most $n + 1$ extreme points of $P$. This is often denoted as the *inner representation* of $P$. We now try to describe a polytope as linear inequalities instead of extreme points (*exterior representation*).

**Definition** (polyhedron). *A set* $P$ *is a polyhedron if there exists a matrix* $A \in \mathbb{R}^{m \times n}$ *and a vector* $b \in \mathbb{R}^m$ *such that*

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b\}.$$  

- $Ax \leq b \iff a_i x \leq b_i$, $i = 1, \ldots, m$. ($a_i$ row $i$ of $A$)
- $\{x \in \mathbb{R}^n \mid a_i x \leq b_i\}, i = 1, \ldots, m$ are half-spaces, so
- $P$ is the intersection of $m$ half-spaces.

- Figure 4 shows the bounded polyhedron defined by $A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \\ 0 & -1 \end{pmatrix}$ and $b = \begin{pmatrix} 6 \\ -2 \\ -1 \end{pmatrix}$

- Figure 5 shows the undounded polyhedron defined by $A = \begin{pmatrix} -2 & 1 \\ 0 & -1 \end{pmatrix}$ and $b = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$.

To make the distinction between a polytope and a polyhedron clearer, we have that
A polytope = The convex hull of finitely many points.
A polyhedron = The intersection of finitely many half-spaces.

We can now define the extreme points of a polyhedron algebraically

**Theorem.** Let \( \bar{x} \in P = \{ x \in \mathbb{R}^n \mid Ax \leq b \} \), where \( \text{rank } A = n \) and \( b \in \mathbb{R}^m \). Further, let \( \bar{A} \bar{x} = \bar{b} \) be the equality subsystem of \( Ax \leq b \). Then \( \bar{x} \) is an extreme point of \( P \) if and only if \( \text{rank } \bar{A} = n \).

**Proof.** See Theorem 3.21 in the book. \( \square \)

Note that the equality subsystem is obtained if one strikes out all rows \( i \) with \( a_i \bar{x} < b_i \), and require equality for the rest of the rows.

That the equality subsystem has rank \( n \) basically means that there should be at least \( n \) linearly independent half-spaces going through the point in order for it to be an extreme point. In Figure 6, two half-spaces goes through the point \( \bar{x} \). Hence, it is an extreme point. In Figure 7, however, only one half-space goes through \( \bar{x} \), meaning that it is not an extreme point.

How many extreme points can there exist in a polyhedron \( P = \{ x \in \mathbb{R}^n \mid Ax \leq b \} \)? How many ways can we choose \( n \) linearly independent rows from the \( m \) existing rows? At most \( \binom{m}{n} \).

**Definition (cone).** A set \( C \subseteq \mathbb{R}^n \) is a cone if \( \lambda x \in C \) whenever \( x \in C \) and \( \lambda > 0 \).
Example: The set \( \{ x \in \mathbb{R}^n \mid Ax \leq 0 \} \) is a polyhedral cone. In Figure 11 one convex and one non-convex cone are illustrated.

We can now formulate a very important theorem which basically says that

\[
\text{a polyhedron} = \text{a polytope} + \text{a polyhedral cone}
\]

**Theorem** (The representation theorem). Let the polyhedron \( Q = \{ x \in \mathbb{R}^n \mid Ax \leq b \} \) and let \( \{ v^1, \ldots, v^k \} \) be its extreme points. Define \( P := \text{conv} \ (\{ v^1, \ldots, v^k \}) \) and \( C := \{ x \in \mathbb{R}^n \mid Ax \leq 0 \} \). Then \( Q = P + C = \{ x \in \mathbb{R}^n \mid x = u + v \ \text{for some} \ u \in P \ \text{and} \ v \in C \} \).

**Proof.** See Theorem 3.26 in the book.
Now we will present a very useful theorem for convex sets which says that: "If a point \( y \) does not lie in a closed convex set \( S \), then there exist a hyperplane separating \( y \) from \( S \)." Mathematically, this amounts to the following.

**Theorem (The separation theorem).** Suppose that the set \( S \subseteq \mathbb{R}^n \) is closed and convex, and that the point \( y \) does not lie in \( S \). Then there exist \( \alpha \in \mathbb{R} \) and \( \pi \neq 0 \) such that \( \pi^T x \leq \alpha \) for all \( x \in S \).

**Proof.** Later in the course. \( \square \)

One last important theorem related to convex set is Farkas’ Lemma.

**Theorem (Farkas’ Lemma).** Let \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \). Then exactly one of the systems

\[
\begin{align*}
Ax &= b, \\
x &\geq 0,
\end{align*}
\]

and

\[
\begin{align*}
A^T \pi &\leq 0, \\
b^T \pi &> 0,
\end{align*}
\]

has a feasible solution, and the other is inconsistent.

**Proof.** See Theorem 3.37 in the book. \( \square \)

Farkas’ Lemma says that:

(A) Either \( b \) lies in the cone spanned by the columns of \( A \), i.e.,
\[
b = Ax, \quad \text{for some} \quad x \geq 0,
\]

(B) or, \( b \) does not lie in the cone, meaning that there exists a hyperplane \( \pi \) separating \( b \) from the cone, i.e.,
\[
A^T \pi \leq 0, \quad b^T \pi > 0.
\]
Convex functions

**Definition (convex function).** Suppose $S \subseteq \mathbb{R}^n$ is convex. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex on $S$ if

$$x^1, x^2 \in S \quad \lambda \in (0, 1) \implies f(\lambda x^1 + (1 - \lambda) x^2) \leq \lambda f(x^1) + (1 - \lambda) f(x^2).$$

- A function is **strictly convex** on $S$ if $<$ holds in place of $\leq$ for all $x^1 \neq x^2$.
- A function $f$ is **concave** if $-f$ is convex.

The linear interpolation between two points on the function never is lower than the function. Two important examples:

- $f(x) = c^T x + d$, where $c \in \mathbb{R}^n$, $d \in \mathbb{R}$ is both convex and concave.
- $f(x) = ||x||$ is convex, $f(x) = ||x||^2$ is strictly convex.

**Proposition (sum of convex functions).** The non-negative linear combination of convex functions is convex.

**Proposition (composite functions).** Suppose $S \subseteq \mathbb{R}^n$ and $P \subseteq \mathbb{R}$. Let $g : S \to \mathbb{R}$ be convex on $S$ and $f : P \to \mathbb{R}$ be convex and non-decreasing on $P$. Then the composite function $f(g)$ is convex on $\{x \in S \mid g(x) \in P\}$.

To show the connection between convex sets and convex functions, we introduce the following notion.

**Definition (epigraph).** The **epigraph** of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\text{epi } f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}.$$

The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ restricted to the set $S \subseteq \mathbb{R}^n$ is defined as

$$\text{epi}_S f := \{(x, \alpha) \in S \times \mathbb{R} \mid f(x) \leq \alpha\}.$$

**Theorem.** Suppose $S \subseteq \mathbb{R}^n$ is a convex set. Then, the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex on $S$ if and only if its epigraph restricted to $S$ is a convex set in $\mathbb{R}^{n+1}$.
To check whether a function is convex or not, one would have to verify that the linear interpolation of any two points does not lie below the function itself. This is, in general, not possible to do, so we often rely on the fact that we know more about the function we consider. We start by assuming that the function is once differentiable with continuous partial derivatives, i.e., that \( f \in C^1 \).

**Theorem \( f \in C^1 \).** A function \( f \in C^1 \) is convex on the open, convex set \( S \) if and only if
\[
f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}),
\]
for all \( x, \bar{x} \in S \).

**Proof.** See Theorem 3.61 in the book.

This theorem basically says that every tangent plane to the graph of \( f \) lies on, or below, the epigraph of \( f \), or that each first-order approximation of \( f \) lies below \( f \). See Figure 12.

Another equivalent way of writing this is the following: \( f \in C^1 \) is convex on the open, convex set \( S \) if and only if
\[
[\nabla f(x) - \nabla f(y)]^T(x - y) \geq 0
\]
- the gradient of \( f \) is monotone on \( S \), or
- the angle between \( \nabla f(x) - \nabla f(y) \) and \( x - y \) should be between \( -\pi/2 \) and \( \pi/2 \).
Theorem \((f \in C^2)\). Let \(S\) be a convex set.

\begin{align*}
  a) & \quad f \text{ is convex on } S \iff \nabla^2 f(x) \succeq 0 \text{ for all } x \in S \\
  b) & \quad \nabla^2 f(x) \succ 0 \text{ for all } x \in S \implies f \text{ is strictly convex on } S.
\end{align*}

**Proof.** See Theorem 3.62 in the book. \(\square\)

Note that in b), \(\iff\) does not hold. Take for example \(f(x) = x^4\).

An important example of a function in \(C^2\) is the quadratic function

\[f(x) = \frac{1}{2} x^T Q x - q^T x,
\]

which is convex on \(\mathbb{R}^n\) if and only if \(Q \succeq 0\). This because \(\nabla^2 f(x) = Q\) is independent of \(x\).

**Convex problems**

A *convex problem* is an optimization problem where one wants to minimize a convex function over a convex set. The general optimization problem

\[
\begin{align*}
\text{minimize} & \quad f(x), \\
\text{subject to} & \quad g_i(x) \leq 0, \quad i \in I, \\
& \quad g_i(x) = 0, \quad i \in E, \\
& \quad x \in X,
\end{align*}
\]

is convex if

- \(f\) is a convex function,
- \(g_i, i \in I\) are convex functions,
- \(g_i, i \in E\) are affine functions, and
- \(X\) is a convex set.

(Then the sets \(\{x \mid g_i(x) \leq 0, \ i \in I\}\) and \(\{x \mid g_i(x) = 0, \ i \in E\}\) are convex).