TMA947 / MMG621 — Nonlinear optimization

Lecture 2 — Convexity

Emil Gustavsson, Zuzana Nedělková

October 20, 2017

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*

$$\begin{matrix} \boldsymbol{x}_1, \boldsymbol{x}_2 \in S \\ \lambda \in (0,1) \end{matrix} \Rightarrow \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in S.$$

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

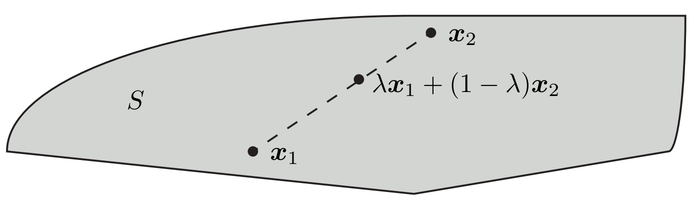


Figure 1: A convex set

Examples:

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

Two non-convex sets are shown in the below figure.

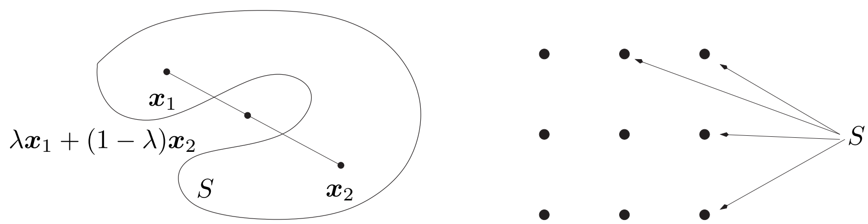


Figure 2: Two non-convex sets.

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. □

Convex and affine hulls

We define the *affine hull* of a finite set $V = \{\boldsymbol{v}^1, \boldsymbol{v}^2, \dots, \boldsymbol{v}^k\}$ as

$$\operatorname{aff} V := \left\{ \lambda_1 \boldsymbol{v}^1 + \cdots + \lambda_k \boldsymbol{v}^k \;\middle|\; \lambda_1, \dots, \lambda_k \in \mathbb{R},\ \sum_{i=1}^k \lambda_i = 1 \right\}$$

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

$$\operatorname{conv} V := \left\{ \lambda_1 \boldsymbol{v}^1 + \cdots + \lambda_k \boldsymbol{v}^k \;\middle|\; \lambda_1, \dots, \lambda_k \geq 0,\ \sum_{i=1}^k \lambda_i = 1 \right\}$$

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

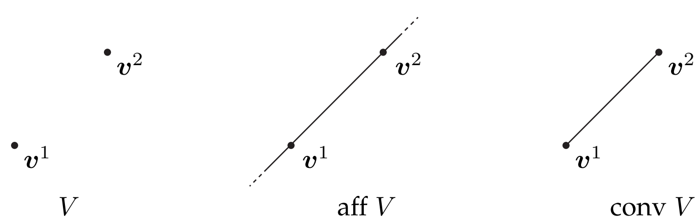


Figure 3: The set $V = \{\boldsymbol{v}^1, \boldsymbol{v}^2\}$, the affine hull of V, and the convex hull of V.

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.

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

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

Proof. See Theorem 3.11 in the book. $\square$

Polytope

We now define the notion of a *polytope*.

Definition (polytope)**.** *A set $P \subset \mathbb{R}^n$ 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 4, a polytope in $\mathbb{R}^2$ generated by seven points is illustrated.

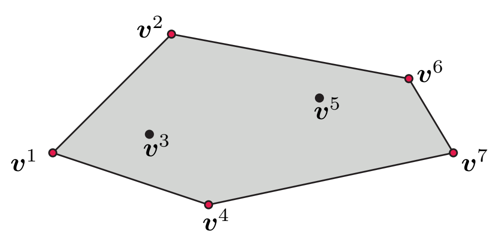


Figure 4: A polytope generated by seven point. The red dots are the extreme points

Definition (extreme point)**.** *A point $\boldsymbol{v}$ of a convex set P is an extreme point if whenever*

$$\left.\begin{aligned} \boldsymbol{v} = \lambda \boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2 \\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in P \\ \lambda \in (0,1) \end{aligned}\right\} \implies \boldsymbol{v} = \boldsymbol{x}^1 = \boldsymbol{x}^2.$$

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 4 for an example. We can now formulate the following intuitive theorem.

Theorem. *Let P be the polytope* $\text{conv}\,V$, *where $V = \{\boldsymbol{v}^1, \ldots, \boldsymbol{v}^k\} \subset \mathbb{R}^n$. 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*).

Polyhedron

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

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

– $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \iff \boldsymbol{a}_i\boldsymbol{x} \leq b_i,\ i = 1, \ldots, m.$ $\quad$ ($\boldsymbol{a}_i$ row i of $\boldsymbol{A}$)

– $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i\boldsymbol{x} \leq b_i\},\ i = 1, \ldots, m$ are half-spaces, so

– P is the intersection of m half-spaces.

– Figure 5 shows the bounded polyhedron defined by $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ -2 & 1 \\ 0 & -1 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 6 \\ -2 \\ -1 \end{pmatrix}$

– Figure 6 shows the undounded polyhedron defined by $\boldsymbol{A} = \begin{pmatrix} -2 & 1 \\ 0 & -1 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$.

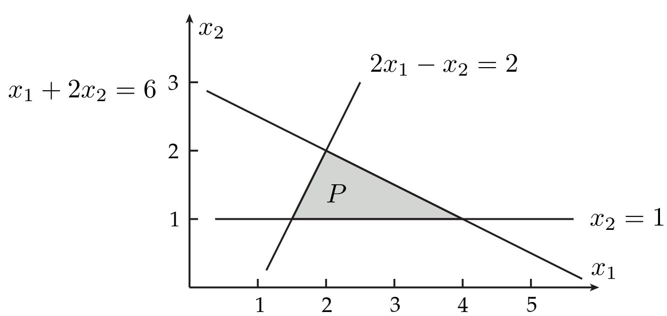


Figure 5: A bounded polyhedron described by three linear inequalities.

To make the distinction between a polytope and a polyhedron clearer, we have that

– A polytope = The convex hull of finitely many points.

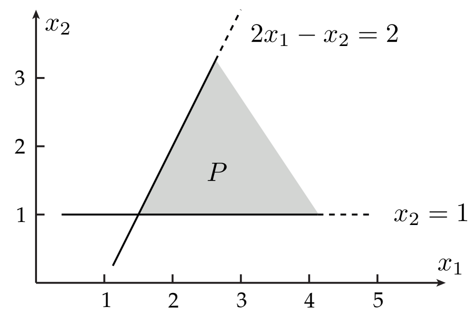


Figure 6: An unbounded polyhedron described by two linear inequalities.

– A polyhedron = The intersection of finitely many half-spaces.

We can now define the extreme points of a polyhedron algebraically

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

Proof. See Theorem 3.21 in the book. □

Note that the *equality subsystem* is obtained if one strikes out all rows i with $\boldsymbol{a}_i\bar{\boldsymbol{x}} < b_i$, and require equality for the rest of the rows.

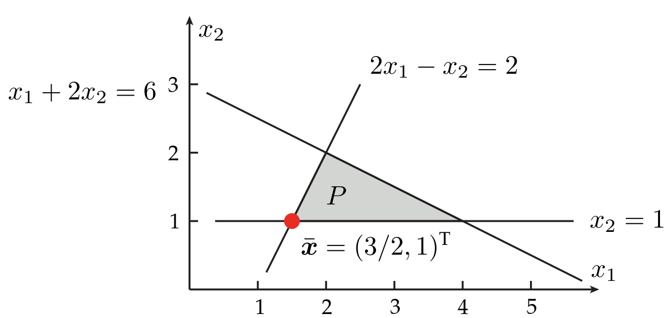


Figure 7: $\bar{\boldsymbol{x}} = (3/2, 1)^{\mathrm{T}}$ is an extreme point of P

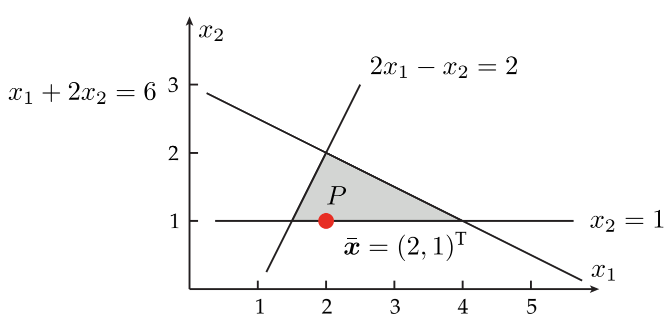


Figure 8: $\bar{\boldsymbol{x}} = (2, 1)^{\mathrm{T}}$ is not an extreme point of P

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 7, two half-spaces goes through the point $\bar{\boldsymbol{x}}$. Hence, it is an extreme point. In Figure 8, however, only one half-space goes through $\bar{\boldsymbol{x}}$, meaning that it is not an extreme point.

How many extreme points can there exist in a polyhedron $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$?

The number of extreme points of P is finite (very important for linear programming).

Definition (cone)**.** *A set $C \subseteq \mathbb{R}^n$ is a cone if $\lambda \boldsymbol{x} \in C$ whenever $\boldsymbol{x} \in C$ and $\lambda > 0$.*

Example: The set $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}\}$ is a polyhedral cone. In Figure 9 one convex and one non-convex cone are illustrated.

(a) A convex cone

(b) A non-convex cone

Figure 9: Two cones.

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 $Q = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ and let $\{\boldsymbol{v}^1, \ldots, \boldsymbol{v}^k\}$ be its extreme points. Define $P := \text{conv}\left(\{\boldsymbol{v}^1, \ldots, \boldsymbol{v}^k\}\right)$ and $C := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}\}$. Then $Q = P + C = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$ for some $\boldsymbol{u} \in P$ and $\boldsymbol{v} \in C\}$*

Proof. See Theorem 3.26 in the book. □

Farkas' Lemma

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

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}, \\ \boldsymbol{x} &\geq \boldsymbol{0}^n, \end{aligned} \tag{I}$$

and

$$\begin{aligned} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\pi} &\leq \boldsymbol{0}^n, \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\pi} &> 0, \end{aligned} \tag{II}$$

has a feasible solution, and the other system is inconsistent.

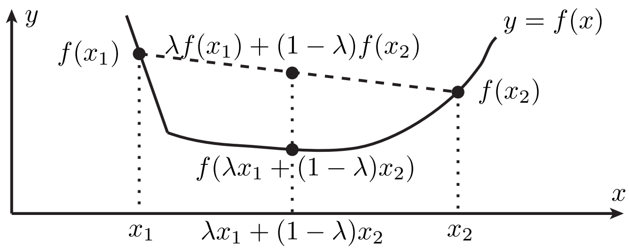


Figure 10: A convex function

Proof. See Theorem 3.32 in the book. □

Farkas' Lemma says that:

(A) Either $\boldsymbol{b}$ lies in the cone spanned by the columns of $\boldsymbol{A}$, i.e.,

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}, \quad \text{for some} \quad \boldsymbol{x} \geq \boldsymbol{0},$$

(B) or, $\boldsymbol{b}$ does not lie in the cone (and can then be separated from the cone).

Farkas' Lemma is crucial for LP theory and optimality conditions.

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*

$$\left.\begin{array}{c} \boldsymbol{x}^1, \boldsymbol{x}^2 \in S \\ \lambda \in (0,1) \end{array}\right\} \implies f\left(\lambda \boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2\right) \leq \lambda f(\boldsymbol{x}^1) + (1-\lambda) f(\boldsymbol{x}^2).$$

– The linear interpolation between two points on the function never is lower than the function.
– A function is *strictly convex* on S if $<$ holds in place of $\leq$ for all $\boldsymbol{x}^1 \neq \boldsymbol{x}^2$.
– A function f is *concave* if $-f$ is convex.

Two important examples:

– $f(\boldsymbol{x}) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + d$, where $\boldsymbol{c} \in \mathbb{R}^n$, $d \in \mathbb{R}$ is both convex and concave.
– $f(\boldsymbol{x}) = ||\boldsymbol{x}||$ is convex, $f(\boldsymbol{x}) = ||\boldsymbol{x}||^2$ 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 $\{\boldsymbol{x} \in S \mid g(\boldsymbol{x}) \in P\}$.*

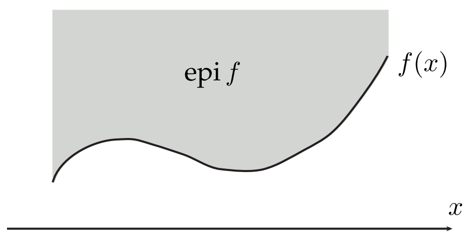
To show the connection between convex sets and convex functions, we introduce the following notion.

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

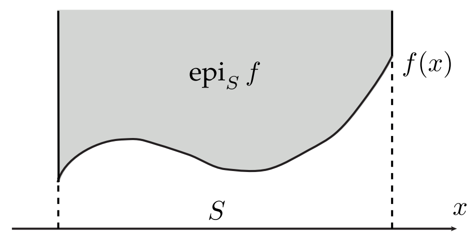
$$\operatorname{epi} f := \{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(\boldsymbol{x}) \leq \alpha\}.$$

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*

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



(a) The epigraph of f.



(b) The epigraph of f restricted to S.

Figure 11: Epigraphs.

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* $\mathbb{R}^{n+1}$.

Proof. See Theorem 3.47 in the book. □

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 (convexity characterization in C^1)**.** *Let* $f \in C^1$ *on an open convex set* S

$$f \text{ is convex on } S \iff f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}), \text{ for all } \boldsymbol{x}, \bar{\boldsymbol{x}} \in S.$$

Proof. See Theorem 3.48 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

$$[\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})]^T (\boldsymbol{x} - \boldsymbol{y}) \geq 0$$

– the gradient of f is monotone on S, or

– the angle between $\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})$ and $\boldsymbol{x} - \boldsymbol{y}$ should be between $-\pi/2$ and $\pi/2$.

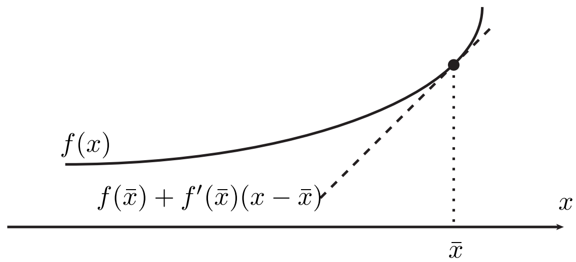


Figure 12: A convex function. The first-order approximation lies below the function.

Now we add even more information about the function f.

Theorem (convexity characterization in C^2)**.** *Let S be a convex set.*

a) *f is convex on $S \iff \nabla^2 f(\boldsymbol{x}) \succeq 0$ for all $\boldsymbol{x} \in S$*

b) *$\nabla^2 f(\boldsymbol{x}) \succ 0$ for all $\boldsymbol{x} \in S \implies f$ is strictly convex on S.*

Proof. See Theorem 3.49 in the book. □

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

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

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

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

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

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}), \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leq 0, \quad i \in \mathcal{I}, \\ & g_i(\boldsymbol{x}) = 0, \quad i \in \mathcal{E}, \\ & x \in X, \end{aligned}$$

is convex if

- f is a convex function,
- g_i, $i \in \mathcal{I}$ are convex functions,
- g_i, $i \in \mathcal{E}$ are affine functions, and
- X is a convex set.

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