CS2910 - Optimization	Summer 2023

Lecture 1 — Convex Sets and Functions

Lecturer: Yunwei Ren *Scribed by Zhidan Li*

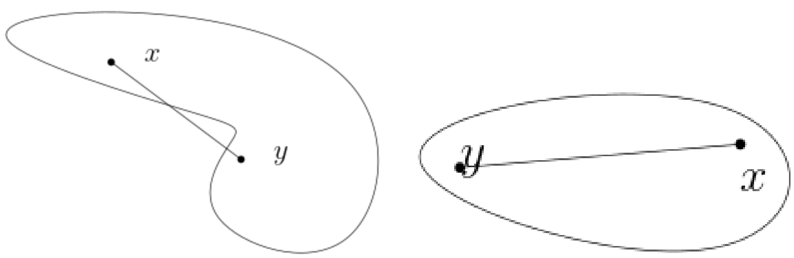
Contents

1 Convex Sets and Supporting Hyperplanes

Firstly, we introduce the definition of convex sets.

Definition 1. *A set $C \subseteq \mathbb{R}^d$ is said to a convex set if, for all $x, y \in C$, $\theta \in [0,1]$, the point $(1-\theta)x + \theta y \in C$. In words, C contains the segment connecting x and y.*

Examples: The following figures give an instance of a convex set and a non-convex set.



(a) a non-convex set (b) a convex set

Figure 1: a non-convex set and a convex one

Goal: We want to show that, every convex set $C \subseteq \mathbb{R}^d$ can be characterized by its 'supporting hyperplanes'.

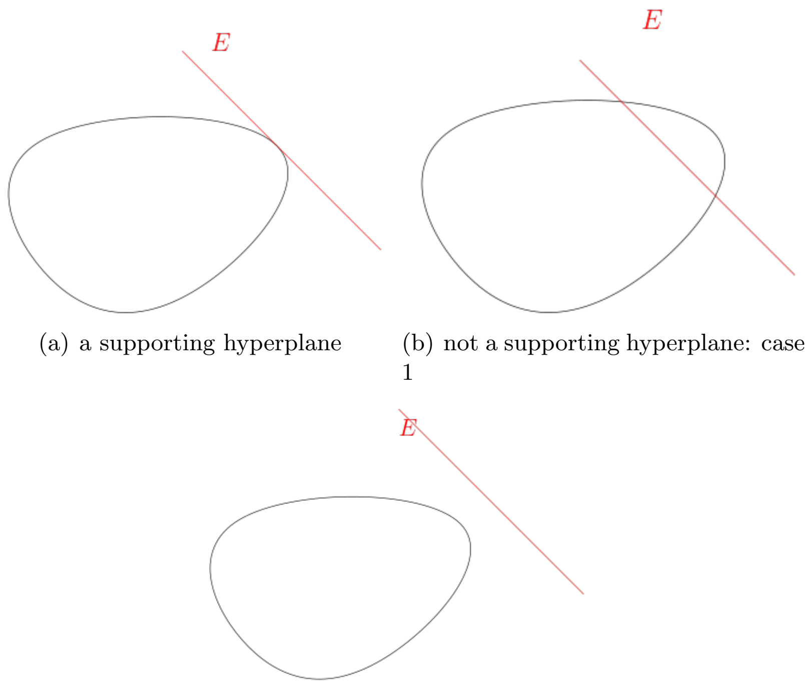
Rough Approach: Describe 'what is in C' $\iff$ describe the boundary of C $\iff$ describe the 'tangent planes' of C.

Remark 1. The reason we use the supporting hyperplanes is, hyperplanes are linear objects, and linear objects are easier to analyze than non-linear ones.

Definition 2. *Let $C \subseteq \mathbb{R}^d$ be a closed convex set. Let $E = \{f = \alpha\}$ be the hyperplane determined by $f : \mathbb{R}^d \to \mathbb{R}$ and the real number $\alpha \in \mathbb{R}$. We say E is a **supporting hyperplane** if $E \cap C \neq \varnothing$*

and C is contained either in $\{f \geq \alpha\}$ or in $\{f \leq \alpha\}$. We also call that C-containing half-space a supporting half-space.

Examples: The following figures illustrate the above definition. The first hyperplane is a supporting hyperplane and the second and third ones are not.



(a) a supporting hyperplane

(b) not a supporting hyperplane: case 1

(c) not a supporting hyperplane: case 2

Figure 2: a supporting hyperplane and not supporting cases

The following proposition defines the metric projection.

Proposition 3 (metric projection)**.** *Let $C \subseteq \mathbb{R}^d$ be a non-empty closed set. For any $x \in \mathbb{R}^d$, define its metric projection to C as the set*

$$\Pi_C(x) \triangleq \underset{y \in C}{\operatorname{argmin}} \|y - x\|_2 = \{\textit{the closest points in } C \textit{ to } x\} \subseteq \mathbb{R}^d.$$

If C is a convex, the above minimizer is unique, and we will abuse the above notation to denote this point.

Proof. For every $x \in C$, it is obvious that $\Pi_C(x) = x$. Now we assume that $x \notin C$. Since $C \neq \varnothing$, there exists some $r > 0$ such that $B_r(x) \cap C \neq \varnothing$. Since C is closed, the set $B_r(x) \cap C$ is compact. Note that

$$\inf_{y \in C} \|y - x\|_2^2 = \inf_{y \in B_r(x) \cap C} \|y - x\|_2^2$$

is continuous. Then the infimum can be achieved by some $y \in C$.

When C is convex, let $x'_* \neq x_*$ be two projection points and $\alpha \triangleq \|x_* - x\|_2$. Since C is convex, the point $x' = \frac{1}{2}(x_* + x'_*) \in C$. Then,

$$
\begin{aligned}
\|x' - x\|_2^2 &= \left\| \frac{1}{2}(x_* - x) + \frac{1}{2}(x'_* - x) \right\|_2^2 \\
&< 2\left(\frac{1}{4}\|x_* - x\|_2^2 + \frac{1}{4}\|x'_* - x\|_2^2 \right) \\
&= \alpha^2
\end{aligned}
$$

where the inequality comes from Cauchy-Schwarz, and a trivial observation that the equality cannot be achieved. This leads to a contradiction. □

With the definition of the metric projection, we can immediately specify a family of supporting hyperplanes.

Proposition 4. *Let $C \subseteq \mathbb{R}^d$ be a non-empty closed convex set. For all $x \notin C$, the hyperplane E through $\Pi_C(x)$ and orthogonal to $x - \Pi_C(x)$ supports C. Moreover, the half-space H bounded by E and not containing x is a supporting half-space.*

Proof. It is trivial to see $\Pi_C(x) \in E \cap C \neq \varnothing$, then we only need to show $C \subseteq H$. We claim that: for all $y \in C$, $\langle x - \Pi_C(x), y - \Pi_C(x) \rangle \leq 0$.

Consider the function $F : [0, 1] \to \mathbb{R}$ defined as

$$F(t) = \|(1 - t)\Pi_C(x) + ty - x\|_2^2.$$

Since C is convex, $(1 - t)\Pi_C(x) + ty \in C$. Then from the definition of the metric projection, it holds that

$$F(0) = \min_{t \in [0,1]} F(t).$$

By the continuity of F,

$$
\begin{aligned}
0 \leq F'(t)|_{t=0} &= 2\left\langle y - \Pi_C(x), (1 - t)\Pi_C(x) + ty - x \right\rangle|_{t=0} \\
&= 2\left\langle y - \Pi_C(x), \Pi_C(x) - x \right\rangle.
\end{aligned}
$$

Then we conclude $\langle y - \Pi_C(x), x - \Pi_C(x) \rangle \leq 0$. □

Additionally, it can be shown that we can use supporting hyperplanes to represent C.

Corollary 5. *Let $C \subsetneq \mathbb{R}^d$ be a non-empty closed convex set. Then*

$$C = \bigcap \{H : H \textit{ is a closed subspace containing } C\} = \bigcap \{H : H \textit{ is a supporting half-space}\}.$$

Proof. Define

$$
\begin{aligned}
C_1 &\triangleq \bigcap \{H : H \text{ is a closed subspace containing } C\}, \\
C_2 &\triangleq \bigcap \{H : H \text{ is a supporting half-space}\}.
\end{aligned}
$$

It is clear that $C \subseteq C_1$. Since every supporting half-space is a closed C-containing subspace, it holds that $C_1 \subseteq C_2$. Thus we prove $C \subseteq C_1 \subseteq C_2$.

To prove the corollary, it suffices to show $C_2 \subseteq C$, in other words, $C^c \subseteq C_2^c$. For every $x \notin C$, from Proposition 4, there exists a supporting hyperplane H such that $x \notin H$, $C \in H$. Then we know $x \notin C_2$, i.e., $x \in C_2^c$. Then we conclude $C_2 \subseteq C$. □

2 Convex Functions and Equivalent Definitions

In this section, we introduce another fundamental definition — convex functions.

Definition 6. *A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is said to be convex if $\forall x, y \in \mathbb{R}^d$, $\theta \in [0, 1]$,*

$$f((1-\theta)x + \theta y) \leq (1-\theta)f(x) + \theta f(y).$$

Exercise 1. *Prove that, $f : \mathbb{R}^d \to \mathbb{R}$ is a convex function if and only if its epigraph*

$$\mathrm{epi}(f) \triangleq \left\{(x, y) \in \mathbb{R}^{d+1} : f(x) \leq y\right\}$$

is a convex set.

Now we show some equivalent definitions for convex functions

Proposition 7 (first-order condition)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ and $f \in C^1(\mathbb{R}^d)$. Then, f is a convex function if and only if f has lower linear bound, i.e., for all $x, y \in \mathbb{R}^d$,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle .$$

Proof. Since $f \in C^1(\mathbb{R}^d)$,

$$\begin{aligned}
\langle \nabla f(x), y - x \rangle &= \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon(y - x)) - f(x)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{f((1-\varepsilon)x + \varepsilon y) - f(x)}{\varepsilon} \\
&\leq \lim_{\varepsilon \to 0} \frac{(1-\varepsilon)f(x) + \varepsilon f(y) - f(x)}{\varepsilon} \\
&= f(y) - f(x)
\end{aligned}$$

where the inequality holds since f is convex. This implies the lower linear bound.

Suppose that f has lower linear bound. Set $z = (1-\theta)x + \theta y$ where $\theta \in (0, 1)$. Then,

$$\begin{aligned}
f(x) &\geq f(z) + \langle \nabla f(z), x - z \rangle , \\
f(y) &\geq f(z) + \langle \nabla f(z), y - z \rangle .
\end{aligned}$$

Then we conclude

$$\begin{aligned}
(1-\theta)f(x) + \theta f(y) &\geq (1-\theta)\left(f(z) + \langle \nabla f(z), x - z \rangle\right) + \theta \left(f(z) + \langle \nabla f(z), y - z \rangle\right) \\
&= f(z) + \langle \nabla f(z), (1-\theta)(x - z) + \theta(y - z) \rangle \\
&= f(z) + \langle \nabla f(z), (1-\theta)x + \theta y - z \rangle \\
&= f(z).
\end{aligned}$$

This implies the convexity of f. □

Remark 2. The following three concepts are equivalent: *first-order Taylor expansion*, *local linear approximation* and *tangent space*.

Lemma 8. *$f : \mathbb{R}^d \to \mathbb{R}$ is a convex if and only if it is convex along all segments, i.e., for all $x, y \in \mathbb{R}^d$, the 1-dimensional function $t \mapsto f((1-t)x + ty)$ on $[0, 1]$ is convex.*

Proof. The lemma holds directly from the definition. □

Proposition 9 (monotone gradient)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a $C^1(\mathbb{R}^d)$. Then, f is a convex function if and only if ∇f is monotone, i.e., $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$.*

Proof. When f is convex, for all $x, y \in \mathbb{R}^d$, by Proposition 7,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle ,$$
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle .$$

Then,

$$0 \geq \langle \nabla f(x) - \nabla f(y), y - x \rangle .$$

Thus we obtain the monotone gradient.

When ∇f is monotone, for all $x, y \in \mathbb{R}^d$, define $g(t) = f((1-t)x + ty)$. Then

$$g'(t) = \langle \nabla f((1-t)x + ty), y - x \rangle .$$

For $t_1, t_2 \in [0, 1]$, by elementary calculation

$$(g'(t_1) - g'(t_2))(t_1 - t_2) = \langle \nabla f((1-t_1)x + t_1 y) - \nabla f((1-t_2)x + t_2 y), (t_1 - t_2)(y - x) \rangle .$$

Note that $((1-t_1)x + t_1 y) - ((1-t_2)x + t_2 y) = (t_1 - t_2)(y - x)$, then

$$(g'(t_1) - g'(t_2))(t_1 - t_2) \geq 0.$$

This means, we can assume $d = 1$. For all $x, y \in \mathbb{R}$, $\theta \in (0, 1)$, define $z = (1-\theta)x + \theta y$. Without loss of generality, let $x < y$. By the mean value theorem, there exists $0 \leq \theta_1 \leq \theta \leq \theta_2 \leq 1$ such that

$$\frac{f(x) - f(z)}{x - z} = f'(\theta_1) \leq f'(\theta_2) = \frac{f(z) - f(y)}{z - y}.$$

Organizing each terms we obtain

$$f(x)(z - y) + f(y)(x - z) \leq f(z)(x - y).$$

Plugging $z = (1-\theta)x + \theta y$ into above, we get

$$(1-\theta)f(x)(x - y) + \theta f(y)(x - y) \leq f(z)(x - y).$$

This leads to the desired inequality $(1-\theta)f(x) + \theta f(y) \leq f(z)$. □

Proposition 10 (second-order condition)**.** *A $C^2(\mathbb{R}^d)$ function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^d$.*

Proof. Firstly we prove the always positive semidefinite Hessian implies convexity. For all $x, y \in \mathbb{R}^d$, by Taylor's theorem, there exists $z \in [x, y]$ (z lies in the segment through xy) such that

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \left\langle y - x, \nabla^2 f(z)(y - x) \right\rangle \\ &= f(x) + \langle \nabla f(x), y - x \rangle . \end{aligned}$$

Then by Proposition 7 f is convex.

Suppose that f is a convex function. Assume that there exists a point $x_0 \in \mathbb{R}^d$ such that $\nabla^2 f(x_0) \prec 0$. Suppose that $(\lambda, \boldsymbol{v})$ is a pair of eigenvalue and eigenvector of $\nabla^2 f(x_0)$, with $\lambda < 0$ and $\|v\|_2 = 1$. By Taylor's theorem,

$$\begin{aligned} f(x_0 + t\boldsymbol{v}) &= f(x_0) + t \left\langle \nabla f(x_0), \boldsymbol{v} \right\rangle + \frac{t^2}{2} \left\langle \boldsymbol{v}, \nabla^2 f(x_0) v \right\rangle + o(t^2) \\ &= f(x_0) + t \left\langle \nabla f(x_0), \boldsymbol{v} \right\rangle + \frac{t^2}{2} \lambda + o(t^2). \end{aligned}$$

Since $\lambda < 0$, when t is sufficiently small, we know

$$f(x_0 + t\boldsymbol{v}) < f(x_0) + t \left\langle \nabla f(x_0), \boldsymbol{v} \right\rangle$$

which leads to a contradiction to the convexity of f. □

Remark 3. The following four concepts are equivalent: *positive semidefinite Hessian*, *local second-order approximation*, *positive curvature* and *the pace it deviates from the tangent*.

Sometimes we need the function f much more 'convex'. Now we introduce the definition of strongly convex functions.

Proposition 11 (strong convexity)**.** *Given a $C^2(\mathbb{R}^d)$ function $f : \mathbb{R}^d \to \mathbb{R}$ and a real $\mu > 0$, the followings are equivalent:*

(a) f is μ-strongly convex, i.e., $x \mapsto f(x) - \frac{\mu}{2}\|x\|_2^2$ is convex.

(b) $f((1-\theta)x + \theta y) \le (1-\theta)f(x) + \theta f(y) - \frac{\theta(1-\theta)\mu}{2}\|x - y\|_2^2$ for all $x, y \in \mathbb{R}^d$ and $\theta \in (0, 1)$.

(c) $\langle \nabla f(x) - \nabla f(y), x - y \rangle \ge \mu\|x - y\|_2^2$ for all $x, y \in \mathbb{R}^d$.

(d) $f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2$ for all $x, y \in \mathbb{R}^d$.

(e) $\nabla f(x) \succeq \mu I_d$ for all $x \in \mathbb{R}^d$.

Example: All linear functions are *not* strongly convex. The function $x \mapsto \frac{1}{2}x^\top Ax$ for positive definite matrix A is $\lambda_{\min}(A)$-strongly convex.

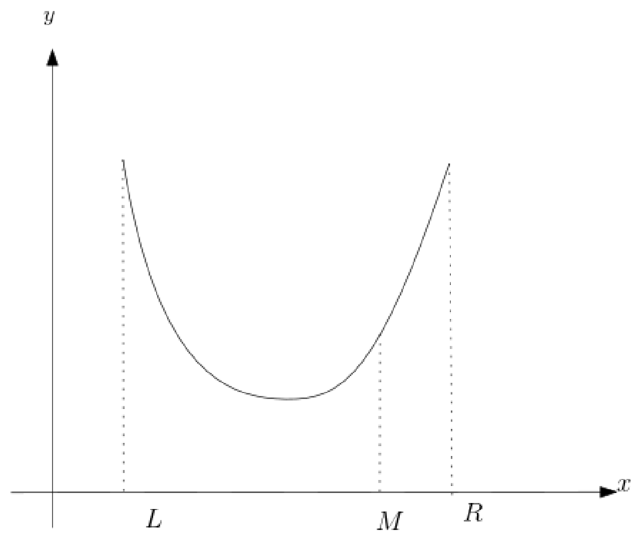


Figure 3: an example of bisection.

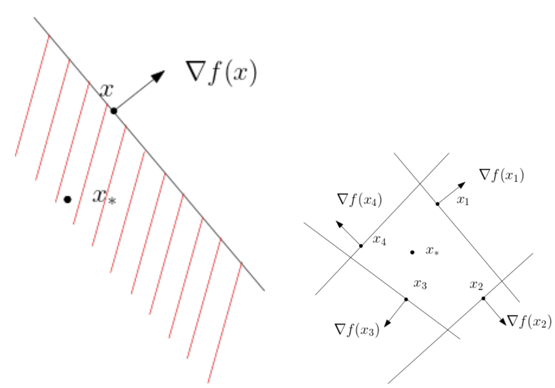
2.1 Ellipsoid method

A typical question is, how to minimize an one-dimensional function f? The answer is, bisection. Figure 3 shows an example of the bisection. We want to find the minimum in $[L, R]$. Firstly we choose $M \in [L, R]$ and calculate $f'(M)$. If $f'(M) \le 0$ we let $R \leftarrow M$ otherwise $L \leftarrow M$. When $|R - L| \le \varepsilon$ we find the (approximate) minimum.

For the high-dimensional case, a natural-thinking idea is to extend the bisection method. Let $f : \mathbb{R}^d \to \mathbb{R}$ be the objective function we want to minimize. Similarly to the bisection, at each step, we choose a hyperplane, and select the subspace bounded by this hyperplane containing the optimal point x_*. To choose the proper subspace, recall the monotone gradient:

$$0 \le \langle \nabla f(x) - \nabla f(x_*), x - x_* \rangle = \langle \nabla f(x), x - x_* \rangle .$$

Then we can properly pick the subspace containing x_*.



(a) select the proper subspace

(b) several bisections

Figure 4: bisection on $\mathbb{R}^d$

Ideally, at each step, we want to choose a 'good' x_{k+1} such that the volume of the polytope containing x_* decreases by half (or a constant ratio) — However, it is hard.

To solve this problem, we replace polytopes with a family of ellipsoids $\{E_k\}_{k\in\mathbb{N}}$, and set x_{k+1} as the center of E_k.

Claim 12. *We can construct* $\{E_k\}$ *such that* $x_* \in E_k$ *and*

$$\mathrm{vol}\left(E_{k+1}\right) \leq \left(1 - \Omega\left(\frac{1}{d^2}\right)\right) \mathrm{vol}\left(E_k\right).$$