Babeş-Bolyai University, Cluj-Napoca
Faculty of Mathematics and Computer Science

Optimization Techniques

Prof. Nicolae Popovici

2021–2022, 2nd Semester

Lecture 1

Optimization problems in general setting

Let $f : D \to \mathbb{R}$ be a function defined on a nonempty set D and let $S \subseteq D$, $S \neq \emptyset$.

Definition 1.1 An element $x^0 \in S$ is called:

- *minimum point* of f w.r.t. S, if

$$f(x^0) \leq f(x), \ \forall\, x \in S.$$

- *maximum point* of f w.r.t. S, if

$$f(x^0) \geq f(x), \ \forall\, x \in S.$$

The sets of all minimum points and maximum points of f w.r.t. S will be denoted by

$$\begin{aligned}\operatorname*{argmin}_{x\in S} f(x) &:= \{x^0 \in S \mid f(x^0) \leq f(x), \ \forall\, x \in S\}\\ \operatorname*{argmax}_{x\in S} f(x) &:= \{x^0 \in S \mid f(x^0) \geq f(x), \ \forall\, x \in S\}.\end{aligned}$$

Remark 1.2 Since S is nonempty, its image by f, i.e., the set

$$f(S) := \{f(x) \mid x \in S\}$$

is a nonempty subset of $\mathbb{R}$, hence the following (extended-)real numbers

$$\inf f(S) \in \mathbb{R} \cup \{-\infty\} \quad \text{and} \quad \sup f(S) \in \mathbb{R} \cup \{+\infty\}$$

are well-defined. It is easily seen that

$$\begin{aligned}\operatorname*{argmin}_{x\in S} f(x) &:= \{x^0 \in S \mid f(x^0) = \inf f(S)\}\\ &= \{x^0 \in S \mid f(x^0) = \min f(S)\};\\ \operatorname*{argmax}_{x\in S} f(x) &:= \{x^0 \in S \mid f(x^0) = \sup f(S)\}\\ &= \{x^0 \in S \mid f(x^0) = \max f(S)\}.\end{aligned}$$

Remark 1.3 $\inf f(S) \in \mathbb{R}$ if and only if $f(S)$ is bounded from below. However, the lower boundedness of $f(S)$ does not guarantee the existence of the least element $\min f(S)$ of $f(S)$. For instance, if $S = D = \mathbb{R}$ and $f : D \to \mathbb{R}$ is the exponential function

$$f(x) = e^x, \ \forall\, x \in D,$$

we have $f(S) =]0, +\infty[$, hence $\inf f(S) = 0 \in \mathbb{R}$, but $f(S)$ does not possess a least element. In this case we have $\operatorname*{argmin}_{x \in S} f(x) = \emptyset$.

Remark 1.4 $\sup f(S) \in \mathbb{R}$ if and only if $f(S)$ is bounded from above. However, the upper boundedness of $f(S)$ does not guarantee the existence of the largest element $\max f(S)$ of $f(S)$. For instance, if $S = [0, +\infty[\subseteq D = \mathbb{R}$ and $f : D \to \mathbb{R}$ is defined by

$$f(x) = \arctan x, \ \forall\, x \in D,$$

we have $f(S) = [0, \pi/2[$, hence $\sup f(S) = \pi/2 \in \mathbb{R}$, but $f(S)$ does not possess a largest element. In this case we have $\operatorname*{argmax}_{x \in S} f(x) = \emptyset$.

Definition 1.5 The problem of finding the value $\inf\ f(S) \in \mathbb{R} \cup \{-\infty\}$ and the set $\operatorname*{argmin}_{x \in S} f(x)$ (or, in practice, at least one element of this set, if any) is called *minimization problem* with *objective function* f and *feasible set* S. We denote this problem by

$$\begin{cases} \text{Minimize } f(x) \\ \\ x \in S. \end{cases} \tag{1.1}$$

The elements of S are called *feasible points* (or *feasible solutions*) of problem (1.1) while the elements of the set $\operatorname*{argmin}_{x \in S} f(x)$ are called *optimal solutions* of problem (1.1).

Definition 1.6 The problem of finding the value $\sup\ f(S) \in \mathbb{R} \cup \{+\infty\}$ and the set $\operatorname*{argmax}_{x \in S} f(x)$ (or, in practice, at least one element of this set, if any) is called *maximization problem* with *objective function* f and *feasible set* S. We denote this problem by

$$\begin{cases} \text{Maximize } f(x) \\ \\ x \in S. \end{cases} \tag{1.2}$$

The elements of S are called *feasible points* (or *feasible solutions*) of problem (1.2) while the elements of the set $\operatorname*{argmax}_{x \in S} f(x)$ are called *optimal solutions* of problem (1.2).

Remark 1.7 It is easy to check that

$$\begin{aligned}\operatorname*{argmax}_{x\in S} f(x) &= \operatorname*{argmin}_{x\in S}(-f)(x);\\ \operatorname*{argmin}_{x\in S} f(x) &= \operatorname*{argmax}_{x\in S}(-f)(x).\end{aligned}$$

These realtions show that any maximization problem of type (1.2) can be transformed into a minimization problem of type (1.1) and vice-versa.

Remark 1.8 A minimization problem (1.1) has no optimal solutions, i.e.,

$$\operatorname*{argmin}_{x\in S} f(x) = \emptyset,$$

in one and only one of the following situations:

- f is not bounded from below on S, i.e.,

$$\inf f(S) = -\infty;$$

- f is bounded from below on S, but f does not attain its minimal value, i.e.,

$$\inf f(S) \in \mathbb{R} \setminus f(S).$$

Remark 1.9 A maximization problem (1.2) has no optimal solutions, i.e.,

$$\operatorname*{argmax}_{x\in S} f(x) = \emptyset,$$

in one and only one of the following situations:

- f is not bounded from above on S, i.e.,

$$\sup f(S) = +\infty;$$

- f is bounded from above on S, but f does not attain its maximal value, i.e.,

$$\sup f(S) \in \mathbb{R} \setminus f(S).$$

Lecture 2

Level sets; characterizations of optimal solutions

Let $f : D \to \mathbb{R}$ be a function defined on a nonempty set D and let $S \subseteq D$, $S \neq \emptyset$.

Definition 2.1 For any $\lambda \in \mathbb{R}$, the following sets are called *level sets* of f (w.r.t. S and λ):

$$\begin{aligned}
S_f(\lambda) &:= \{x \in S \mid f(x) = \lambda\}, \\
S_f^{\leqslant}(\lambda) &:= \{x \in S \mid f(x) \leq \lambda\}, \\
S_f^{<}(\lambda) &:= \{x \in S \mid f(x) < \lambda\} = S_f^{\leqslant}(\lambda) \setminus S_f(\lambda), \\
S_f^{>}(\lambda) &:= \{x \in S \mid f(x) > \lambda\} = S \setminus S_f^{\leqslant}(\lambda), \\
S_f^{\geqslant}(\lambda) &:= \{x \in S \mid f(x) \geq \lambda\} = S \setminus S_f^{<}(\lambda).
\end{aligned}$$

Proposition 2.2 *The following characterizations of optimal solutions hold:*

$$\operatorname*{argmin}_{x \in S} f(x) = \{x^0 \in S \mid S \subseteq D_f^{\geqslant}(f(x^0))\}, \tag{2.1}$$

$$\operatorname*{argmax}_{x \in S} f(x) = \{x^0 \in S \mid S \subseteq D_f^{\leqslant}(f(x^0))\}. \tag{2.2}$$

Proposition 2.4 *Let $\lambda \in \mathbb{R}$. If $A \in \{S_f^{\leqslant}(\lambda), S_f^{<}(\lambda)\}$ and $B \in \{S_f^{\geqslant}(\lambda), S_f^{>}(\lambda)\}$ are nonempty sets, then the following relations hold:*

$$\operatorname*{argmin}_{x \in S} f(x) = \operatorname*{argmin}_{x \in A} f(x), \tag{2.3}$$

$$\operatorname*{argmax}_{x \in S} f(x) = \operatorname*{argmax}_{x \in B} f(x). \tag{2.4}$$

Lecture 3

Existence and unicity of optimal solutions

Let $f : D \to \mathbb{R}$ be a function defined on a nonempty set D and let $S \subseteq D$, $S \neq \emptyset$.

Theorem 3.1 *The minimization problem* (1.1) *has at least one optimal solution, i.e.,*

$$\operatorname*{argmin}_{x \in S} f(x) \neq \emptyset,$$

if one of the following conditions is fulfilled:

(C1) *There exists $\mu \in \mathbb{R}$ such that $S_f^{\leqslant}(\mu)$ is nonempty and bounded, and $S_f^{\leqslant}(\lambda)$ is closed for every $\lambda \in]-\infty, \mu]$.*

(C2) *S is closed, f is continuous, and there exists $\mu \in \mathbb{R}$ such that $S_f^{\leqslant}(\mu)$ is nonempty and bounded.*

(C3) *S is compact and $S_f^{\leqslant}(\lambda)$ is closed for each $\lambda \in \mathbb{R}$.*

Corollary 3.2 *The maximization problem* (1.2) *has at least one optimal solution, i.e.,*

$$\operatorname*{argmax}_{x \in S} f(x) \neq \emptyset,$$

if one of the following conditions is fulfilled:

(C1) *There exists* $\mu \in \mathbb{R}$ *such that* $S_f^{\geqslant}(\mu)$ *is nonempty and bounded, and* $S_f^{\geqslant}(\lambda)$ *is closed for every* $\lambda \in [\mu, +\infty[$.

(C2) S *is closed,* f *is continuous, and there exists* $\mu \in \mathbb{R}$ *such that* $S_f^{\geqslant}(\mu)$ *is nonempty and bounded.*

(C3) S *is compact and* $S_f^{\geqslant}(\lambda)$ *is closed for each* $\lambda \in \mathbb{R}$.

Remark 3.3 In the particular case when f is continuous on the nonempty compact set S, then both level sets $S_f^{\leqslant}(\lambda)$ and $S_f^{\geqslant}(\lambda)$ are closed for every $\lambda \in \mathbb{R}$. In this case we recover from $(C3)$ in Theorem 3.1 and Corollary 3.2 the conclusion of the classical Weierstrass Theorem.

Theorem 3.4 *Let $f\colon S \to \mathbb{R}$ be a function defined on a nonempty set $S \subseteq \mathbb{R}^n$. The following assertions are equivalent*:

1° *The minimization problem* (1.1) *has at most one optimal solution.*

2° *For all $x^1,\ x^2 \in S,\ x^1 \neq x^2,$ there exists $\ x^* \in S\ $ such that $f(x^*) < \max\{f(x^1), f(x^2)\}$.*

Proof. 1° $\Rightarrow$ 2°. Assume that $\mathrm{card}(\mathrm{argmin}_{x\in S} f(x)) \leq 1$ and suppose to the contrary that there exist two distinct points $x^1, x^2 \in S$ satisfying the inequality $f(x) \geq \max\{f(x^1), f(x^2)\}$ for every $x \in S$. We infer that $f(x^1) \leq f(x)$ and $f(x^2) \leq f(x)$ for any point $x \in S$, i.e., $x^1, x^2 \in \mathrm{argmin}_{x\in S} f(x)$, contradicting the hypothesis.

2° $\Rightarrow$ 1°. Assume that for every distinct points $x^1, x^2 \in S$ there exists $x^* \in S$ such that $f(x^*) < \max\{f(x^1), f(x^2)\}$, and suppose to the contrary that $\mathrm{card}(\mathrm{argmin}_{x\in S} f(x)) > 1$. Then we can choose $x^1, x^2 \in \mathrm{argmin}_{x\in S} f(x)$, $x^1 \neq x^2$. By hypothesis, we can find $x^* \in S$ such that $f(x^*) < \max\{f(x^1), f(x^2)\}$. We infer that $f(x^*) < f(x^1) = f(x^2) = \inf f(S) \leq f(x^*)$, a contradiction. □

Corollary 3.5 *Let $f\colon S \to \mathbb{R}$ be a function defined on a nonempty set $S \subseteq \mathbb{R}^n$. The following assertions are equivalent*:

1° *The maximization problem* (1.2) *has at most one optimal solution.*

2° *For all $x^1,\ x^2 \in S,\ x^1 \neq x^2,$ there exists $\ x^* \in S\ $ such that $f(x^*) > \min\{f(x^1), f(x^2)\}$.*

Proof. Follows by Theorem 3.4 applied to $-f$ in the role of f. □

Lecture 4

Convex sets and their extreme subsets/points

For any points $x, y \in \mathbb{R}^n$ we denote

$$
\begin{aligned}
[x, y] &:= \{(1-t)x + ty \mid t \in [0,1]\}; \\
]x, y[&:= \{(1-t)x + ty \mid t \in]0,1[\}.
\end{aligned}
$$

Notice that if $x \neq y$, then $]x, y[= [x, y] \setminus \{x, y\}$; otherwise,if $x = y$, then $[x, y] =]x, y[= \{x\}$.

Definition 4.1 A set $S \subseteq \mathbb{R}^n$ is said to be *convex* if

$$[x, y] \subseteq S \quad \text{for all} \quad x, y \in S.$$

Remark 4.2 For any set $S \subseteq \mathbb{R}^n$ the following assertions are equivalent:

1° S is convex.

2° $(1-t)x + ty \in S, \ \forall\, x, y \in S, \ \forall\, t \in [0,1]$.

Example 4.3 By a *hyperplane* in $\mathbb{R}^n$ we mean any set of type

$$H(c, \lambda) := \{x \in \mathbb{R}^n \mid \langle c, x \rangle = \lambda\},$$

where $c \in \mathbb{R}^n \setminus \{0_n\}$ and $\lambda \in \mathbb{R}$. The sets

$$\begin{aligned} H^{\leqslant}(c,\lambda) &:= \{x \in \mathbb{R}^n \mid \langle c, x\rangle \leq \lambda\} \\ H^{\geqslant}(c,\lambda) &:= \{x \in \mathbb{R}^n \mid \langle c, x\rangle \geq \lambda\} \end{aligned}$$

are called *closed half-spaces* while the sets

$$\begin{aligned} H^{<}(c,\lambda) &:= \{x \in \mathbb{R}^n \mid \langle c, x\rangle < \lambda\} \\ H^{>}(c,\lambda) &:= \{x \in \mathbb{R}^n \mid \langle c, x\rangle > \lambda\} \end{aligned}$$

are called *open half-spaces*. It is a simple exercise to check that all hyperplanes, as well as closed and open half-spaces, are convex sets.

Proposition 4.4 *If $\mathscr{F}$ is a family of convex sets in $\mathbb{R}^n$, then the set $\bigcap_{S\in\mathscr{F}} S$ is also convex.*

Proof. Let $t \in [0,1]$. For any $S' \in \mathscr{F}$ we have

$$(1-t)\bigcap_{S\in\mathscr{F}} S + t\bigcap_{S\in\mathscr{F}} S \subseteq (1-t)S' + tS' \subseteq S',$$

hence $(1-t)\bigcap_{S\in\mathscr{F}} S + t\bigcap_{S\in\mathscr{F}} S \subseteq \bigcap_{S'\in\mathscr{F}} S' = \bigcap_{S\in\mathscr{F}} S$. Thus $\bigcap_{S\in\mathscr{F}} S$ is a convex set. □

Definition 4.5 The *convex hull* of an arbitrary set $M \subseteq \mathbb{R}^n$ is defined by

$$\operatorname{conv} M := \bigcap\{S \subseteq \mathbb{R}^n \mid S \text{ is convex and } M \subseteq S\}.$$

Remark 4.6 $\operatorname{conv} M$ is a convex set as an intersection of a family of convex sets. Therefore, M is convex if and only if $M = \operatorname{conv} M$.

Definition 4.7 For any $k \in \mathbb{N}$, the set

$$\Delta_k := \{(t_1,\dots,t_k) \in (\mathbb{R}_+)^k \mid t_1 + \cdots + t_k = 1\}$$

is called the *standard simplex* of $\mathbb{R}^k$. It is easily seen that Δ_k is convex.

Definition 4.8 Given an arbitrary nonempty set $M \subseteq \mathbb{R}^n$, a point $x \in \mathbb{R}^n$ is said to be a *convex combination* of elements of $M \subseteq \mathbb{R}^n$, if there exist $k \in \mathbb{N}^*$, $x^1,\dots,x^k \in M$, and $(t_1,\dots,t_k) \in \Delta_k$, such that $x = t_1x^1 + \cdots + t_kx^k$.

Theorem 4.9 (Characterization of the convex hull via convex combinations) *The convex hull of a nonempty set $M \subseteq \mathbb{R}^n$ admits the following representation:*

$$\operatorname{conv} M = \left\{\sum_{i=1}^k t_ix^i \mid k \in \mathbb{N}^*,\ x^1,\dots,x^k \in M,\ (t_1,\dots,t_k) \in \Delta_k\right\}.$$

Proof. Denote by

$$C(M) := \left\{ \sum_{i=1}^{k} t_i x^i \mid k \in \mathbb{N}^*,\ x^1, \dots, x^k \in M,\ (t_1, \dots, t_k) \in \Delta_k \right\}. \tag{4.1}$$

For the equality $\operatorname{conv} M = C(M)$ it suffices to show that the following conditions are fulfilled:

(i) $M \subseteq C(M)$;

(ii) $C(M)$ is convex;

(iii) $C(M) \subseteq S$ for every convex set $S \subseteq \mathbb{R}^n$ with $M \subseteq S$.

Condition (i) holds, since one obtains in $C(M)$ the elements of M considering $k = 1$.

In order to prove (ii) pick $x, y \in C(M)$ and $\alpha \in [0,1]$. Then there exist $k, \ell \in \mathbb{N}^*$, $x^1, \dots, x^k, y^1, \dots, y^\ell \in M$, $(t_1, \dots, t_k) \in \Delta_k$ and $(s_1, \dots, s_\ell) \in \Delta_\ell$ such that $x = \sum_{i=1}^{k} t_i x^i$ and $y = \sum_{i=1}^{\ell} s_i y^i$. Thus

$$(1-\alpha)x + \alpha y = \sum_{i=1}^{k} (1-\alpha) t_i x^i + \sum_{i=1}^{\ell} \alpha s_i y^i.$$

Since $\sum_{i=1}^{k}(1-\alpha)t_i + \sum_{i=1}^{\ell} \alpha s_i = (1-\alpha)\sum_{i=1}^{k} t_i + \alpha \sum_{i=1}^{\ell} s_i = 1 - \alpha + \alpha = 1$, it follows that $(1-\alpha)x + \alpha y$ is also a convex combination of elements of M, that is, it belongs to $C(M)$. Thus (ii) holds.

For proving (iii) consider a convex subset $S \subseteq \mathbb{R}^n$ such that $M \subseteq S$. We get the inclusion $C(M) \subseteq S$ by performing an induction argument. We prove that proposition

$$\mathscr{P}(k) :\ \text{“}\sum_{i=1}^{k} t_i x^i \in S,\ \forall\, x^1, \dots, x^k \in M,\ \forall\, (t_1, \dots, t_k) \in \Delta_k\text{”}$$

is true for every $k \in \mathbb{N}^*$. Obviously $\mathscr{P}(1)$ is true (since $M \subseteq S$). Assume now that $\mathscr{P}(h)$ is true for a natural number $h \in \mathbb{N}^*$. We are going to prove that $\mathscr{P}(h+1)$ is also true. Let $x^1, \dots, x^h, x^{h+1} \in M$ and $(t_1, \dots, t_h, t_{h+1}) \in \Delta_{h+1}$. Without loss of generality we may assume that $t := \sum_{i=1}^{h} t_i > 0$ (otherwise $\mathscr{P}(h+1)$ obviously would be true). Then $t + t_{h+1} = 1$ and $\frac{1}{t}(t_1 + \cdots + t_h) = 1$. By the induction hypothesis and using the convexity of S, we get

$$\sum_{i=1}^{h+1} t_i x^i = t\left(\sum_{i=1}^{h} \frac{t_i}{t} x^i\right) + (1-t)x^{h+1} \in S.$$

Hence $\mathscr{P}(h+1)$ is true. It follows that $\mathscr{P}(k)$ is true for every $k \in \mathbb{N}^*$. Thus $C(M) \subseteq S$. $\square$

Theorem 4.12 (Carathéodory) *If S is a nonempty subset of $\mathbb{R}^n$, then every point $x \in \operatorname{conv} S$ can be expressed as a convex combination of at most $n+1$ elements of S.*

Definition 4.13 Let $S \subseteq \mathbb{R}^n$ be a convex set. We say that $E \subseteq S$ is an *extremal subset* (or *face*) of S if E is convex and

$$\forall\ x, y \in S\ :\ ]x, y[\cap E \neq \emptyset \quad \Rightarrow \quad [x, y] \subseteq E.$$

An element $v \in S$ is called *extreme point* (or *vertex*) of S if the singleton $\{v\}$ is an extremal subset of S. The set of all extreme points of of S is denoted by $\mathrm{ext}S$, i.e.,

$$\mathrm{ext}S := \{v \in S \mid \forall\ x, y \in S\ :\ v \in]x, y[\Rightarrow\ x = y = v\}.$$

Theorem 4.14 (Characterization of extreme points) *Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set . For any point $v \in S$ the following assertions are equivalent:*

1° $v \in \mathrm{ext}S$.

2° *There are no distinct points $x, y \in S$ such that $v = \frac{1}{2}(x + y)$.*

Lecture 5

Convex functions

Definition 5.1 Let $f : S \to \mathbb{R}$ be a function defined o a nonempty set $S \subseteq \mathbb{R}^n$. We say that f is a *convex function* if its domain S is a convex set and

$$f((1-t)x^1 + tx^2) \le (1-t)f(x^1) + t\, f(x^2),\ \forall\, x^1, x^2 \in S,\ \forall\, t \in [0,1].$$

Example 5.2 (Distance function) Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set. Consider the distance function w.r.t. C, $d_C : \mathbb{R}^n \to \mathbb{R}$, defined by

$$d_C(x) \coloneqq \inf\{||x - c||\ :\ c \in C\},\ \forall\, x \in \mathbb{R}^n.$$

We will prove that this function is convex. In particular, for any $a \in \mathbb{R}^n$, the distance function $d(\cdot, a) : \mathbb{R}^n \to \mathbb{R}$, defined by

$$d(x,a) \coloneqq \|x - a\|,\ \forall\, x \in \mathbb{R}^n,$$

is convex (in this case the set $C ::= \{a\}$ is a singleton).

Indeed, let $x^1, x^2 \in \mathbb{R}^n$ and $t \in [0,1]$. Consider two sequences $(c^k)_{k\in\mathbb{N}}$ and $(\tilde{c}^k)_{k\in\mathbb{N}}$ of points of C such that

$$\lim_{k\to\infty} \|x^1 - c^k\| = d_C(x^1), \quad \lim_{k\to\infty} \|x^2 - \tilde{c}^k\| = d_C(x^2). \tag{5.1}$$

Since C is a convex set, it follows that for any $k \in \mathbb{N}$ we have $(1-t)c^k + t\tilde{c}^k \in C$, hence

$$\begin{aligned} d_C((1-t)x^1 + tx^2) &\le \|(1-t)x^1 + tx^2 - (1-t)c^k - t\tilde{c}^k\| \\ &\le (1-t)\|x^1 - c^k\| + t\|x^2 - \tilde{c}^k\|. \end{aligned}$$

Letting $k \to \infty$ and recalling (5.1), we infer

$$d_C((1-t)x^1 + tx^2) \le (1-t)d_C(x^1) + td_C(x^2).$$

Thus d_C is a convex function.

Definition 5.4 Let $f\colon M \to \mathbb{R}$ be a function defined on a nonempty set $M \subset \mathbb{R}^n$. The set

$$\mathrm{epi} f \ := \ \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} \mid x \in M,\ f(x) \leq \lambda\}$$

is called the *epigraph* of f.

Theorem 5.5 (Characterization of convex functions by means of their epigraph) *Let $f : S \to \mathbb{R}$ be a function defined on a nonempty convex set $S \subseteq \mathbb{R}^n$. Then the following assertions are equivalent:*

1° *Function f is convex.*

2° *The epigraph of f (i.e., $\mathrm{epi} f$) is a convex set (in the space $\mathbb{R}^n \times \mathbb{R}$).*

Proof. $1^\circ \Rightarrow 2^\circ$. Assume that 1° holds and consider any points $(x^1, \lambda_1), (x^2, \lambda_2) \in \mathrm{epi} f$ and any number $t \in [0, 1]$. Then we have $f(x^1) \leq \lambda_1$ and $f(x^2) \leq \lambda_2$. By 1° it follows that

$$f((1-t)x^1 + tx^2) \leq (1-t)f(x^1) + tf(x^2) \leq (1-t)\lambda_1 + t\lambda_2$$

which shows that

$$((1-t)x^1 + tx^2, (1-t)\lambda_1 + t\lambda_2) = (1-t)(x^1, \lambda_1) + t(x^2, \lambda_2) \in \mathrm{epi}f.$$

Thus epi$f$ is a convex set, i.e., 3° holds.

2° $\Rightarrow$ 1°. Assume that 2° holds and consider any $x^1, x^2 \in S$ and $t \in [0,1]$. Since $(x^1, f(x^1)), (x^2, f(x^2)) \in \mathrm{epi}f$, we have

$$((1-t)x^1 + tx^2, (1-t)f(x^1) + tf(x^2)) = (1-t)(x^1, f(x^1)) + t(x^2, f(x^2)) \in \mathrm{epi}f,$$

hence $f((1-t)x^1 + tx^2) \leq (1-t)f(x^1) + tf(x^2)$. Thus function f is convex, i.e., 1° holds. □

Theorem 5.6 (Characterization of differentiable convex functions) *Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and let $f : S \to \mathbb{R}$ be a differentiable function. The following assertions are equivalent*:

1° *Function f is convex.*

2° $\langle x - x^0, \nabla f(x^0)\rangle \leq f(x) - f(x^0)$ *for any* $x, x^0 \in S$.

3° $\langle x^1 - x^2, \nabla f(x^1) - \nabla f(x^2)\rangle \geq 0$ *for any* $x^1, x^2 \in S$.

Theorem 5.7 (Characterization of twice differentiable convex functions) *Assume that $S \subseteq \mathbb{R}^n$ is a nonempty open convex set and let $f : S \to \mathbb{R}$ be a twice differentiable function. Then the following assertions are equivalent:*

1° *Function f is convex.*

2° *For any point $x^0 \in S$, the Hessian matrix $\nabla^2 f(x^0) = \left(\dfrac{\partial^2 f}{\partial x_i \partial x_j}(x^0)\right)_{1 \le i,j \le n}$ is positive semidefinite, i.e.,*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0) x_i x_j \geq 0, \ \forall\, x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

22

Lecture 6

Local/global extrema of convex functions

Definition 6.1 Let $f : S \to \mathbb{R}$ be a function defined on a nonempty set $S \subseteq \mathbb{R}^n$. An element $x^0 \in S$ is said to be a:

- *local minimum point* of f, if there exists a neighborhood $V \in \mathcal{V}(x^0)$ of the point x^0 such that
$$f(x^0) \leq f(x),\ \forall\, x \in V \cap S.$$
- *local maximum point* of f, if there exists a neighborhood $V \in \mathcal{V}(x^0)$ of the point x^0 such that
$$f(x^0) \geq f(x),\ \forall\, x \in V \cap S.$$
- *global minimum point* of f (or, simply, *minimum point* of f), if x^0 is a minimum point of f w.r.t. S in the sense of Definition 1.1, i.e., $x^0 \in \operatorname*{argmin}_{x\in S} f(x)$, which actually means that
$$f(x^0) \leq f(x),\ \forall\, x \in S\, (= V \cap S \text{ with } V := \mathbb{R}^n \in \mathcal{V}(x^0)).$$
- *global maximum point* of f (or, simply, *maximum point* of f), if x^0 is a maximum point of f w.r.t. S in the sense of Definition 1.1, i.e., $x^0 \in \operatorname*{argmax}_{x\in S} f(x)$, which actually means that
$$f(x^0) \geq f(x),\ \forall\, x \in S\, (= V \cap S \text{ with } V := \mathbb{R}^n \in \mathcal{V}(x^0)).$$

The local/global minimum points and maximum points are generically called *local/global extremum points* (or *local/global extrema*) of f.

Remark 6.2 For any point $a \in \mathbb{R}$ and any real number $r > 0$, we denote by $B(a, r)$ the open Euclidean ball centered at a with radius r, i.e.,
$$B(a, r) := \{x \in \mathbb{R}^n \mid \|x - a\| < r\}.$$

It is easy to see that:

- x^0 is a local minimum point of f if and only if there exists $\varepsilon > 0$ such that

$$f(x^0) \leq f(x),\ \forall\, x \in B(x^0, \varepsilon) \cap S.$$

- x^0 is a local maximum point of f if and only if there exists $\varepsilon > 0$ such that

$$f(x^0) \geq f(x),\ \forall\, x \in B(x^0, \varepsilon) \cap S.$$

Remark 6.3 Every global minimum (resp. maximum) point of f is a local minimum (resp. maximum) point of f. The converse is not true, as the following example shows.

Example 6.4 Consider the function $f : S = [0, 3] \to \mathbb{R}$, defined by

$$f(x) := \lfloor x \rfloor,\ \forall\, x \in [0, 3],$$

where $\lfloor x \rfloor$ denotes the integer part (floor) of x. It is a simple exercise to check that:

- The set of all global minimum points of f is $\underset{x \in S}{\operatorname{argmin}} f(x) = [0, 1[$.
- The set of all global maximum points of f is $\underset{x \in S}{\operatorname{argmax}} f(x) = \{3\}$.
- The set of all local minimum points of f is $[0, 1[\cup]1, 2[\cup]2, 3[$.
- The set of all local maximum points of f is $[0, 3]$.

Lemma 6.5 (Structure of lower level sets of convex functions) *Let $f : S \to \mathbb{R}$ be a convex function, defined on a nonempty convex set $S \subseteq \mathbb{R}^n$. Then, for any $\lambda \in \mathbb{R}$, the level set $S_f^{\leqslant}(\lambda)$ is convex.*

Proof. Let $\lambda \in \mathbb{R}$. Recall that

$$S_f^{\leqslant}(\lambda) := \{x \in S \mid f(x) \leq \lambda\},$$

according to Definition 2.1 of Lecture 2. Let $x^1, x^2 \in S_f^{\leqslant}(\lambda)$ and $t \in [0, 1]$. It follows that $f(x^1) \leq \lambda$ and $f(x^2) \leq \lambda$, hence

$$(1 - t)f(x^1) + tf(x^2) \leq (1 - t)\lambda + t\lambda = \lambda.$$

Since function f is convex, the following inequality also holds:

$$f((1 - t)x^1 + tx^2) \leq (1 - t)f(x^1) + tf(x^2).$$

Therefore, we have $f((1 - t)x^1 + tx^2) \leq \lambda$. Taking into account that S is a convex set, we deduce that $(1 - t)x^1 + tx^2 \in S_f^{\leqslant}(\lambda)$. Thus the level set $S_f^{\leqslant}(\lambda)$ is convex. ☐

Theorem 6.6 (Structure of the set of minimum points of convex functions) *Let $f : S \to \mathbb{R}$ be a convex function, defined on a nonempty convex set $S \subseteq \mathbb{R}^n$. Then, the set of all global minimum points of f, i.e., $\underset{x \in S}{\operatorname{argmin}} f(x)$, is convex.*

Proof. In view of Remark 1.8, we distinguish two cases.

Case 1: $\inf f(S) = -\infty$.

In this case, the set $\operatorname*{argmin}_{x\in S} f(x) = \emptyset$ is obviously convex.

Case 2: $\inf f(S) =: \lambda \in \mathbb{R}$.

In this case, the set $\operatorname*{argmin}_{x\in S} f(x) = S_f^{\leqslant}(\lambda)$ is convex, by Lemma 6.5. $\square$

Theorem 6.7 (Coincidence of local and global minimum points of convex functions) *Let $f : S \to \mathbb{R}$ be a convex function, defined on a nonempty convex set $S \subseteq \mathbb{R}^n$. For any point $x^0 \in S$, the following assertions are equivalent:*

1° *x^0 is a global minimum point of f.*

2° *x^0 is a local minimum point of f.*

Lemma 6.8 (Fermat's necessary optimality condition) *Let $f : S \to \mathbb{R}$ be a function, defined on a nonempty set $S \subseteq \mathbb{R}^n$ and let $x^0 \in S$ be a local extremum point of f. If $x^0 \in \operatorname{int} S$ and f is partially derivable at x^0, then x^0 is a stationary point of f, i.e., $\nabla f(x^0) = 0_n$.*

Theorem 6.9 (Characterization of minimum points of differentiable convex functions) *Let $f : S \to \mathbb{R}$ be a differentiable convex function, defined on a nonempty open convex set $S \subseteq \mathbb{R}^n$. For any point $x^0 \in S$ the following assertions are equivalent:*

1° *x^0 is a global minimum point of f.*

2° *x^0 is a stationary point of f.*

Proof. The implication $1^\circ \Rightarrow 2^\circ$ holds by Fermat's optimality condition (Lemma 6.8) even when f is not convex.

For proving the implication $2^\circ \Rightarrow 1^\circ$ we will use Theorem 5.6. More precisely, since function f is differentiable and convex, we have $\langle x - x^0, \nabla f(x^0)\rangle \le f(x) - f(x^0)$ for all $x \in S$. On the other hand, by hypothesis 2° we have $\nabla f(x^0) = 0_n$. Thus we have $0 \le f(x) - f(x^0)$, i.e., $f(x^0) \le f(x)$, for all $x \in S$, which actually means 1°. $\square$