Part III. Nonsmooth Convex Programming

Lecture 4.

General Convex Functions.

– Equivalent Definitions

– Closed Functions

– Continuity of Convex Functions

– Separation Theorems.

– Subgradients.

– Computation rules.

– Optimality Conditions.

General Convex Programming Problems

Problem:

$$\begin{aligned} &\min\ f_0(x), \\ \text{s.t.}\ &f_i(x) \leq 0,\ i = 1 \dots m, \\ &x \in Q \subseteq R^n, \end{aligned} \tag{4.1}$$

where

- Q is a convex set;
- $f_i(x)$ are *convex* functions; $i = 0 \dots m$.

In general, $f_i(x)$ are not differentiable.

Motivation:

1. Max-type functions:

$$f(x) = \max_{1 \leq j \leq p}\ \phi_j(x),$$

where $\phi_j(x)$ are convex and differentiable and p is *very large*.

2. Convex functions with *implicit* structure.

General Convex Functions

Denote by

$$\mathrm{dom}\, f = \{x \in R^n : \mid f(x) \mid < \infty\}$$

the *domain* of function f.

Definition A function $f(x)$ is called *convex* if its domain is convex and

$$\forall x, y \in \mathrm{dom}\, f,\ \forall \alpha \in [0, 1]$$

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y).$$

We call f *concave* if $-f$ is convex.

Lemma 4.1 (Jensen inequality)
For any $x_1, \ldots, x_m \in \mathrm{dom}\, f$ *and*

$$\alpha_1, \ldots, \alpha_m : \quad \alpha_i \ge 0, \quad \sum_{i=1}^{m} \alpha_i = 1, \qquad (4.2)$$

we have:

$$f\left(\sum_{i=1}^{m} \alpha_i x_i\right) \le \sum_{i=1}^{m} \alpha_i f(x_i).$$

For proof use induction and representation

$$\sum_{i=1}^{m} \alpha_i x_i = \alpha_1 x_1 + (1-\alpha_1) \sum_{i=2}^{m} \beta_i x_i,$$

where $\beta_i = \frac{\alpha_i}{1-\alpha_1}$. Note that $\sum_{i=2}^{m} \beta_i = 1$.

The point $x = \sum_{i=1}^{m} \alpha_i x_i$ with α_i satisfying (4.2) is called the *convex combination* of points x_i.

Corollary 4.1 *Let x be a convex combination of points $x_1, \ldots, x_m$. Then*

$$f(x) \le \max_{1 \le i \le m} f(x_i).$$

Proof. Indeed, in view of Jensen inequality and since

$$\alpha_i \ge 0, \quad \sum_{i=1}^{m} \alpha_i = 1,$$

we have:

$$\begin{aligned} f(x) &= f\left(\sum_{i=1}^{m} \alpha_i x_i\right) \\ &\le \sum_{i=1}^{m} \alpha_i f(x_i) \\ &\le \max_{1 \le i \le m} f(x_i). \end{aligned}$$

□

Corollary 4.2 *Let*

$$\begin{aligned} \Delta &= \mathrm{Conv}\,\{x_1, \ldots, x_m\} \\ &\equiv \{x = \sum_{i=1}^{m} \alpha_i x_i \mid \ \alpha_i \ge 0, \ \sum_{i=1}^{m} \alpha_i = 1\}. \end{aligned}$$

Then

$$\max_{x \in \Delta} f(x) = \max_{1 \le i \le n} f(x_i).$$

Theorem 4.1 *A function f is convex iff*

$$\forall x, y \in \operatorname{dom} f,\ \beta \geq 0: \quad y + \beta(y - x) \in \operatorname{dom} f$$

we have

$$f(y + \beta(y - x)) \geq f(y) + \beta(f(y) - f(x)). \qquad (4.3)$$

Proof. 1. Let f be convex. Denote

$$\alpha = \tfrac{\beta}{1+\beta}, \quad u = y + \beta(y - x).$$

Then

$$y = \tfrac{1}{1+\beta}(u + \beta x) = (1 - \alpha)u + \alpha x.$$

Therefore

$$\begin{aligned} f(y) &\leq (1 - \alpha) f(u) + \alpha f(x) \\ &= \tfrac{1}{1+\beta} f(u) + \tfrac{\beta}{1+\beta} f(x). \end{aligned}$$

2. Let (4.3) holds. Let us fix

$$x, y \in \operatorname{dom} f, \quad \alpha \in (0, 1].$$

Denote $\quad \beta = \frac{1-\alpha}{\alpha}, \quad u = \alpha x + (1 - \alpha) y.$

Then $x = \frac{1}{\alpha}(u - (1 - \alpha)y) = u + \beta(u - y)$. Therefore

$$\begin{aligned} f(x) &\geq f(u) + \beta(f(u) - f(y)) \\ &= \tfrac{1}{\alpha} f(u) - \tfrac{1-\alpha}{\alpha} f(y). \end{aligned}$$

□

Theorem 4.2 *Function f is convex iff its* epigraph

$$\text{epi}\,(f) = \{(x,t) \in \text{dom}\, f \times R \mid\ t \geq f(x)\}$$

is a convex set.

Proof. 1. Indeed, if

$$(x_1, t_1) \in \text{epi}\,(f), \quad (x_2, t_2) \in \text{epi}\,(f),$$

then for any $\alpha \in [0,1]$ we have:

$$\begin{aligned} \alpha t_1 + (1-\alpha) t_2 &\geq \alpha f(x_1) + (1-\alpha) f(x_2) \\ &\geq f(\alpha x_1 + (1-\alpha) x_2). \end{aligned}$$

Thus, $(\alpha x_1 + (1-\alpha)x_2, \alpha t_1 + (1-\alpha) t_2) \in \text{epi}\,(f)$.

2. Let epi (f) be convex. Note that for $x_1, x_2 \in \text{dom}\, f$

$$(x_1, f(x_1)) \in \text{epi}\,(f), \quad (x_1, f(x_2) \in \text{epi}\,(f).$$

Therefore

$$(\alpha x_1 + (1-\alpha)x_2, \alpha f(x_1) + (1-\alpha) f(x_2)) \in \text{epi}\,(f).$$

That is

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2). \quad \square$$

Theorem 4.3 *If function f is convex then all its sublevel sets*

$$\mathcal{L}_f(\beta) = \{x \in \operatorname{dom} f \mid \ f(x) \leq \beta\}$$

are convex.

Proof. Indeed, if

$$x_1 \in \mathcal{L}_f(\beta), \quad x_2 \in \mathcal{L}_f(\beta),$$

then for any $\alpha \in [0, 1]$ we have:

$$\begin{aligned} f(\alpha x_1 + (1-\alpha)x_2) &\leq \alpha f(x_1) + (1-\alpha)f(x_2) \\ &\leq \alpha\beta + (1-\alpha)\beta = \beta. \end{aligned}$$

□

Definition. A convex function f is called *closed* if its epigraph is a closed set.

Theorem 4.4 *If convex function f is closed then all its sublevel sets are closed.*

Proof. Note that

$$(\mathcal{L}_f(\beta), \beta) = \operatorname{epi}(f) \bigcap \{(x, t) \mid \ t = \beta\}.$$

Therefore, $\mathcal{L}_f(\beta)$ is closed as an intersection of two closed sets. □

Note: If f is convex and continuous and $\operatorname{dom} f$ is closed then f is a closed function.

Operations with convex functions

Theorem 4.5 *Let* f_1, f_2 *are convex and* $\beta \geq 0$. *Then*

1. $f(x) = \beta f_1(x)$ *is convex and* $\operatorname{dom} f = \operatorname{dom} f_1$.
2. $f(x) = f_1(x) + f_2(x)$ *is convex.*
3. $f(x) = \max\{f_1(x), f_2(x)\}$ *is convex.*

For 2) and 3) $\operatorname{dom} f = (\operatorname{dom} f_1) \cap (\operatorname{dom} f_2)$.

Proof. 1.

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \beta(\alpha f_1(x_1) + (1-\alpha) f_1(x_2)).$$

2. $\forall x_1, x_2 \in (\operatorname{dom} f_1) \cap (\operatorname{dom} f_2)$ and $\alpha \in [0,1]$ we have:

$$\begin{aligned} & f_1(\alpha x_1 + (1-\alpha)x_2) + f_2(\alpha x_1 + (1-\alpha)x_2) \\ \leq\ & \alpha f_1(x_1) + (1-\alpha) f_1(x_2) \\ & + \alpha f_2(x_1) + (1-\alpha) f_2(x_2) \\ =\ & \alpha(f_1(x_1) + f_2(x_1)) + (1-\alpha)(f_1(x_2) + f_2(x_2)). \end{aligned}$$

3. $\forall x_1, x_2 \in (\operatorname{dom} f_1) \cap (\operatorname{dom} f_2)$ and $\alpha \in [0,1]$ we have:

$$\begin{aligned} \max\ & \{f_1(\alpha x_1 + (1-\alpha)x_2), f_2(\alpha x_1 + (1-\alpha)x_2)\} \\ \leq\ & \max\{\alpha f_1(x_1) + (1-\alpha) f_1(x_2), \\ & \alpha f_2(x_1) + (1-\alpha) f_2(x_2)\} \\ \leq\ & \alpha \max\{f_1(x_1), f_2(x_1)\} \\ & + (1-\alpha)\max\{f_1(x_2), f_2(x_2)\}. \qquad \square \end{aligned}$$

Theorem 4.6 *Let $\phi(y)$, $y \in R^m$, be convex and*

$$\mathcal{A}(x) = Ax + b : \quad R^n \to R^m.$$

Then $f(x) = \phi(\mathcal{A}(x)$ is convex and

$$\operatorname{dom} f = \{x \in R^n \mid \ \mathcal{A}(x) \in \operatorname{dom} \phi\}.$$

Proof. For $x_1, x_2 \in \operatorname{dom} f$ denote

$$y_1 = \mathcal{A}(x_1), \quad y_2 = \mathcal{A}(y_2).$$

Then for $\alpha \in [0, 1]$ we have:

$$\begin{aligned} f(\alpha x_1 \ &+ \ (1-\alpha)x_2) \\ &= \ \phi(\mathcal{A}(\alpha x_1 + (1-\alpha)x_2)) \\ &= \ \phi(\alpha y_1 + (1-\alpha)y_2) \\ &\le \ \alpha\phi(y_1) + (1-\alpha)\phi(y_2) \\ &= \ \alpha f(x_1) + (1-\alpha)f(x_2). \end{aligned}$$

□

Examples

1. Linear function is convex.
2. $f(x) = \mid x \mid$, $x \in R$, is convex.
3. $f(x) = \max\limits_{1 \leq i \leq n} \{x^{(i)}\}$ is convex.
4. All examples of *differentiable* convex functions.

All these functions are closed with $\mathrm{dom}\, f = R^n$.

5. $f(x) = \frac{1}{x}$, $x \in R^0_+$, is convex and closed.

However, $\mathrm{dom}\, f = R^0_+$ is open.

6. $f(x) = \| x \|$, where $\| \cdot \|$ is a *norm*, is convex:

$$\begin{aligned} f(\alpha x_1 + (1-\alpha)x_2) &= \| \alpha x_1 + (1-\alpha)x_2 \| \\ &\le \| \alpha x_1 \| + \| (1-\alpha)x_2 \| \\ &= \alpha \| x_1 \| + (1-\alpha) \| x_2 \| . \end{aligned}$$

Important norms (l_p norms):

$$\begin{aligned} \| x \|_p &= [\sum_{i=1}^{n} | x^{(i)} |^p]^{1/p}, & p \ge 1, \\ \| x \| &= [\sum_{i=1}^{n} (x^{(i)})^2]^{1/2}, & p = 2, \\ \| x \|_1 &= \sum_{i=1}^{n} | x^{(i)} |, & p = 1, \\ \| x \|_\infty &= \max_{1 \le i \le n} | x^{(i)} |, & p = \infty. \end{aligned}$$

l_p-balls of the radius r:

$$B_p(x_0, r) = \{x \in R^n \mid \| x - x_0 \|_p \le r\}.$$

Note that

$$B_1(x_0, r) \subseteq B_2(x_0, r) \subseteq B_1(x_0, r\sqrt{n})$$

since

$$\sum_{i=1}^{n} (x^{(i)})^2 \le \left(\sum_{i=1}^{n} | x^{(i)} | \right)^2,$$

$$\frac{1}{n} \sum_{i=1}^{n} | x^{(i)} |^2 \ge \left(\frac{1}{n} \sum_{i=1}^{n} | x^{(i)} | \right)^2.$$

Theorem 4.7 *Let Δ be a set and*

$$f(x) = \sup\{\phi(y, x) \mid \ y \in \Delta\}.$$

Suppose that for any fixed $y \in \Delta$ the function $\phi(y, x)$ is convex in x.

Then $f(x)$ is convex and

$$\operatorname{dom} f = \{x \in R^n \mid \ \exists \gamma : \phi(y, x) \leq \gamma \ \forall y \in \Delta\}. \quad (4.4)$$

Proof. If $x \notin$ RHS(4.4) then

$$\exists \{y_k\} : \quad \phi(y_k, x) \to \infty.$$

Therefore $x \notin \operatorname{dom} f$.

If $x \in$ RHS(4.4) then $f(x) < \infty \Rightarrow x \in \operatorname{dom} f$.

Let us fix $x_1, x_2 \in \operatorname{dom} f$ and $\alpha \in [0, 1]$. Consider

$$\{y_k\} : \phi(y_k, \alpha x_1 + (1 - \alpha)x_2) \to f(\alpha x_1 + (1 - \alpha)x_2).$$

Note that

$$\begin{aligned} \phi(y_k, \alpha x_1 \ &+ \ (1 - \alpha)x_2) \\ &\leq \ \alpha\phi(y_k, x_1) + (1 - \alpha)\phi(y_k, x_2) \\ &\leq \ \alpha f(x_1) + (1 - \alpha)f(x_2). \end{aligned}$$

Therefore

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2).$$

$\square$

Examples

7. Let Q be a closed convex set. Consider the function

$$\psi_Q(x) = \max\{\langle g, x\rangle \mid \ g \in Q\}.$$

Function $\psi_Q(x)$ is called the *support* function of the set Q.

Note:

- $\psi_Q(x)$ is convex (Theorem 4.7).
- It is homogeneous of degree one:

$$\psi_Q(tx) = t\psi_Q(x), \quad x \in \operatorname{dom} Q, \ t \geq 0.$$

- If Q is bounded then $\operatorname{dom} \psi_Q = R^n$.

8. Consider the function

$$\psi(g,\gamma) = \sup_{y \in Q} \phi(y,g,\gamma),$$

$$\phi(y,g,\gamma) = \langle g, y\rangle - \tfrac{\gamma}{2} \parallel y \parallel^2 .$$

This function is convex in (g,γ) in view of T. 4.7.

If Q is bounded then $\mathrm{dom}\,\psi = R^{n+1}$.

Let $Q = R^n$. Then

- If $\gamma < 0$ then for any $g \neq 0$ we can take

$$y_\alpha = \alpha g \quad \Rightarrow \quad \phi(y_\alpha, g, \gamma) \to \infty, \; \alpha \to \infty.$$

- If $\gamma = 0$ then we can take only $g = 0$.
- If $\gamma > 0$ then

$$y^*(g,\gamma) = \tfrac{1}{\gamma} g \quad \Rightarrow \quad \psi(g,\gamma) = \tfrac{\|g\|^2}{2\gamma}.$$

Thus, $$\psi(g,\gamma) = \begin{cases} 0, & \text{if } g = 0, \gamma = 0, \\ \frac{\|g\|^2}{2\gamma}, & \text{if } \gamma > 0. \end{cases}$$

Note: 1. $\mathrm{dom}\,\psi = (R^n \times \{\gamma > 0\}) \cup (0,0)$ is convex.

2. $\psi(g,\gamma)$ is not continuous at $(0,0)$:

$$\lim_{\gamma \downarrow 0} \psi(\sqrt{\gamma} g, \gamma) = \tfrac{1}{2} \parallel g \parallel^2 .$$

9. Consider the following function of two variables:

$$f(x,y) = \begin{cases} 0, & \text{if } x^2 + y^2 < 1, \\ \phi(x,y), & \text{if } x^2 + y^2 = 1, \end{cases}$$

where $\phi(x,y)$ is an *arbitrary* nonnegative function defined on the unit sphere.

- $\operatorname{dom} f$ is the unit Euclidean ball;
- f is convex;
- It is not closed unless $\phi(x,y) = 0$;
- It has no reasonable properties on the boundary.

Continuity of Convex Functions

Lemma 4.2 *Let f be convex and $x_0 \in \operatorname{int}(\operatorname{dom} f)$. Then f is locally upper bounded at x_0.*

Proof. Let us choose $\epsilon > 0$:

$$x_0 \pm \epsilon e_i \in \operatorname{int}(\operatorname{dom} f), \quad i = 1 \dots n,$$

where e_i are the coordinate orts.

Denote $\Delta = \operatorname{Conv}\{x_0 \pm \epsilon e_i,\ i = 1 \dots n\}$.

Let us show that $\Delta \supset B_2(x_0, \bar{\epsilon})$, $\bar{\epsilon} = \frac{\epsilon}{\sqrt{n}}$. Indeed, let

$$x = x_0 + \sum_{i=1}^{n} h_i e_i, \quad \sum_{i=1}^{n} (h_i)^2 \leq \bar{\epsilon}.$$

We can assume that $h_i \geq 0$. Then

$$\beta \equiv \sum_{i=1}^{n} h_i \leq \sqrt{n} \sum_{i=1}^{n} (h_i)^2 \leq \epsilon.$$

Therefore for $\bar{h}_i = \frac{1}{\beta} h_i$ we have:

$$\begin{aligned} x &= x_0 + \beta \sum_{i=1}^{n} \bar{h}_i e_i = x_0 + \frac{\beta}{\epsilon} \sum_{i=1}^{n} \bar{h}_i \epsilon e_i \\ &= \left(1 - \frac{\beta}{\epsilon}\right) x_0 + \frac{\beta}{\epsilon} \sum_{i=1}^{n} \bar{h}_i (x_0 + \epsilon e_i) \in \Delta. \end{aligned}$$

Thus, using Corollary 4.2, we obtain:

$$\begin{aligned} M &\equiv \max_{x \in B_2(x_0, \bar{\epsilon})} f(x) \leq \max_{x \in \Delta} f(x) \\ &\leq \max_{1 \leq i \leq n} f(x_0 \pm \epsilon e_i). \end{aligned}$$

$\square$

Theorem 4.8 *Let f be convex and $x_0 \in \operatorname{int}(\operatorname{dom} f)$. Then f is locally Lipshitz continuous at x_0.*

Proof. Let $B_2(x_0, \epsilon) \subseteq \operatorname{dom} f$ and

$$\sup_{x \in B_2(x_0,\epsilon)} f(x) \le M.$$

Let $y \in B_2(x_0, \epsilon)$, $y \neq x_0$. Denote

$$\alpha = \tfrac{1}{\epsilon} \parallel y - x_0 \parallel, \quad z = x_0 + \tfrac{1}{\alpha}(y - x_0).$$

Note that $\parallel z - x_0 \parallel = \frac{1}{\alpha} \parallel y - x_0 \parallel = \epsilon$,

$$y = \alpha z + (1-\alpha)x_0, \quad \alpha \le 1.$$

Therefore

$$\begin{aligned} f(y) &\le \alpha f(z) + (1-\alpha) f(x_0) \\ &\le f(x_0) + \alpha(M - f(x_0)) \\ &= f(x_0) + \tfrac{M - f(x_0)}{\epsilon} \parallel y - x_0 \parallel . \end{aligned}$$

Denote $u = x_0 + \frac{1}{\alpha}(x_0 - y)$.

Then $\parallel u - x_0 \parallel = \epsilon$ and $y = x_0 + \alpha(x_0 - u)$.

Therefore, in view of T. 4.1 we have:

$$\begin{aligned} f(y) &\ge f(x_0) + \alpha(f(x_0) - f(u)) \\ &\ge f(x_0) - \alpha(M - f(x_0)) \\ &= f(x_0) - \tfrac{M - f(x_0)}{\epsilon} \parallel y - x_0 \parallel . \end{aligned}$$

Thus, $\mid f(y) - f(x_0) \mid \le \frac{M - f(x_0)}{\epsilon} \parallel y - x_0 \parallel$. □

Directional Differentiability

Definition. Let $x_0 \in \operatorname{int}(\operatorname{dom} f)$. We call f *differentiable in the direction* p if the limit

$$f'(x_0; p) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[f(x_0 + \alpha p) - f(x_0)] \qquad (4.5)$$

exists.

Theorem 4.9 *A convex function f is differentiable in any direction at any $x \in \operatorname{int}(\operatorname{dom} f)$.*

Proof. Let $x_0 \in \operatorname{int}(\operatorname{dom} f)$. Consider the function

$$\phi(\alpha) = \frac{1}{\alpha}[f(x_0 + \alpha p) - f(x_0)], \quad \alpha > 0.$$

Let $\gamma \in (0, 1]$ and $\alpha \in (0, \epsilon] : x_0 + \epsilon p \in \operatorname{dom} f$. Then

$$\begin{aligned} f(x_0 + \alpha\beta p) &= f((1-\beta)x_0 + \beta(x_0 + \alpha p)) \\ &\leq (1-\beta) f(x_0) + \beta f(x_0 + \alpha p). \end{aligned}$$

Therefore

$$\begin{aligned} \phi(\alpha\beta) &= \tfrac{1}{\alpha\beta}[f(x_0 + \alpha\beta p) - f(x_0)] \\ &\leq \tfrac{1}{\alpha}[f(x_0 + \alpha p) - f(x_0)] = \phi(\alpha). \end{aligned}$$

Thus, $\phi(\alpha)$ decreases as $\alpha \downarrow 0$. Hence, the limit in (4.5) exists. $\square$

Lemma 4.3 *Let* $x_0 \in \mathrm{int}\,(\mathrm{dom}\, f)$*. Then* $f'(x_0; p)$ *is a closed convex homogeneous (of degree one) function of* p*. For any* $y \in \mathrm{dom}\, f$ *we have:*

$$f(y) \geq f(x_0) + f'(x_0; y - x_0). \tag{4.6}$$

Proof. 1. $\forall p \in R^n$ and $\tau > 0$ we have:

$$f'(x_0; \tau p) = \lim_{\alpha \downarrow 0} \tfrac{1}{\alpha}[f(x_0 + \tau\alpha p) - f(x_0)]$$

$$= \tau \lim_{\beta \downarrow 0} \tfrac{1}{\beta}[f(x_0 + \beta p) - f(x_0)] = \tau f'(x_0; p).$$

2. For any p_1, $p_2 \in R^n$ and $\beta \in [0, 1]$

$$f'(x_0; \beta p_1 + (1 - \beta)p_2)$$

$$= \lim_{\alpha \downarrow 0} \tfrac{1}{\alpha}[f(x_0 + \alpha(\beta p_1 + (1 - \beta)p_2)) - f(x_0)]$$

$$\leq \lim_{\alpha \downarrow 0} \tfrac{1}{\alpha}\{\beta[f(x_0 + \alpha p_1) - f(x_0)]$$

$$+(1 - \beta)[f(x_0 + \alpha p_2) - f(x_0)]\}$$

$$= \beta f'(x_0; p_1) + (1 - \beta)f'(x_0; p_2).$$

3. Let $\alpha \in (0, 1]$, $y \in \mathrm{dom}\, f$ and $y_\alpha = x_0 + \alpha(y - x_0)$. Then (see T. 4.1),

$$f(y) = f(y_\alpha + \tfrac{1}{\alpha}(1 - \alpha)(y_\alpha - x_0))$$

$$\geq f(y_\alpha) + \tfrac{1}{\alpha}(1 - \alpha)[f(y_\alpha) - f(x_0)],$$

and we get (4.6) taking the limit in $\alpha \downarrow 0$. $\square$

Separation Theorems

Definition. Let Q be a convex set.

- We say that the hyperplane

$$\mathcal{H}(g,\gamma) = \{x \in R^n \mid \langle g, x\rangle = \gamma\}, \quad g \neq 0,$$

is *supporting* to Q if any $x \in Q$ satisfies inequality

$$\langle g, x\rangle \leq \gamma.$$

- We say that the hyperplane $\mathcal{H}(g,\gamma)$ *separates* a point x_0 from Q if

$$\langle g, x\rangle \leq \gamma \leq \langle g, x_0\rangle \tag{4.7}$$

for all $x \in Q$.

- If the second inequality in (4.7) is strict, we call the separation *strict*.

Motivation:

Minimization methods need some *directions* replacing the gradient.

Projection:

Let Q be a set and $x_0 \in R^n$.
Denote

$$\pi_Q(x_0) = \arg\min\{\| x - x_0 \|: \; x \in Q\}.$$

We call $\pi_Q(x_0)$ the *projection* of x_0 on Q.

Theorem 4.10 *If Q is a closed convex set, then the point $\pi_Q(x_0)$ is unique and well-defined.*

Proof. Indeed,

$$\pi_Q(x_0) = \arg\min\{\phi(x) \mid \; x \in Q\},$$

where

$$\phi(x) = \frac{1}{2} \| x - x_0 \|^2$$

is a function from $\mathcal{S}^{1,1}_{1,1}(R^n)$.

Therefore $\pi_Q(x_0)$ is unique and well-defined in view of Theorem 3.6. □

Note:

$\pi_Q(x_0) = x_0$ iff $x_0 \in Q$ and Q is closed.

Lemma 4.4 *Let Q be a closed convex set and $x_0 \notin Q$.*
Then for any $x \in Q$ we have:

$$\langle \pi_Q(x_0) - x_0, x - \pi_Q(x_0) \rangle \geq 0. \tag{4.8}$$

Proof. Note that $\pi_Q(x_0)$ is the solution to

$$\min_{x \in Q} \phi(x)$$

with $\phi(x) = \frac{1}{2} \| x - x_0 \|^2$.
Therefore, in view of Theorem 3.5

$$\langle \phi'(\pi_Q(x_0)), x - \pi_Q(x_0) \rangle \geq 0$$

for all $x \in Q$.
It remains to note that $\phi'(x) = x - x_0$. □

Lemma 4.5 *For any $x \in Q$ we have*

$$\| x - \pi_Q(x_0) \|^2 + \| \pi_Q(x_0) - x_0 \|^2 \leq \| x - x_0 \|^2 .$$

Proof. Indeed, in view of (4.8), we have:

$$\begin{aligned} &\| x - \pi_Q(x_0) \|^2 - \| x - x_0 \|^2 \\ &= \langle x_0 - \pi_Q(x_0), 2x - \pi_Q(x_0) - x_0 \rangle \\ &\leq - \| x_0 - \pi_Q(x_0) \|^2 . \end{aligned}$$

□

Theorem 4.11 *Let Q be a closed convex set and $x_0 \notin Q$.*

Then there exists a hyperplane $\mathcal{H}(g, \gamma)$ strictly separating x_0 from Q. Namely, we can take

$$g = x_0 - \pi_Q(x_0) \neq 0, \quad \gamma = \langle x_0 - \pi_Q(x_0), \pi_Q(x_0) \rangle.$$

Proof. Indeed, in view of (4.8), for any $x \in Q$ we have:

$$\begin{aligned} \langle x_0 - \pi_Q(x_0), x \rangle &\leq \langle x_0 - \pi_Q(x_0), \pi_Q(x_0) \rangle \\ &= \langle x_0 - \pi_Q(x_0), x_0 \rangle - \| x_0 - \pi_Q(x_0) \|^2 . \end{aligned}$$

□

Corollary 4.3 *Let Q_1, Q_2 are some closed convex sets.*

1. If for any $g \in \operatorname{dom} \psi_{Q_2}$ we have

$$\psi_{Q_1}(g) \leq \psi_{Q_2}(g)$$

then $Q_1 \subseteq Q_2$.

2. If $\operatorname{dom} \psi_{Q_1} = \operatorname{dom} \psi_{Q_2}$ and for any $g \in \operatorname{dom} \psi_{Q_1}$ we have $\psi_{Q_1}(g) = \psi_{Q_2}(p)$, then $Q_1 \equiv Q_2$.

Proof. 1. Assume that $\exists x_0 \in Q_1$, $x_0 \notin Q_2$.

Then $\exists g$:

$$\langle g, x_0 \rangle > \gamma \geq \langle g, x \rangle, \quad \forall x \in Q_2.$$

Hence, $g \in \operatorname{dom} \psi_{Q_2}$ and $\psi_{Q_1}(g) > \psi_{Q_2}(g)$.

2. In view of 1), $Q_1 \subseteq Q_2$ and $Q_2 \subseteq Q_1 \Rightarrow Q_1 \equiv Q_2$. □

Theorem 4.12 *Let Q be a closed convex set and $x_0 \in \partial Q$.*

Then there exists a hyperplane $\mathcal{H}(g, \gamma)$:

- *$\mathcal{H}(g, \gamma)$ is supporting to Q.*
- *$x_0 \in \mathcal{H}(g, \gamma)$.*

(Such vector g is called *supporting to Q at x_0*.)

Proof. Consider a sequence:

$$\{y_k\} : \quad y_k \notin Q, \quad y_k \to x_0.$$

Denote

$$g_k = \tfrac{y_k - \pi_Q(y_k)}{\|y_k - \pi_Q(y_k)\|}, \quad \gamma_k = \langle g_k, \pi_Q(y_k) \rangle.$$

Note that $\forall x \in Q$ we have (see Theorem 4.11):

$$\langle g_k, x \rangle \leq \gamma_k \leq \langle g_k, y_k \rangle. \tag{4.9}$$

But $\| g_k \| = 1$ and (see Lemma 4.5)

$$\begin{aligned} | \gamma_k | &= | \langle g_k, \pi_Q(y_k) - x_0 \rangle + \langle g_k, x_0 \rangle | \\ &\leq \| \pi_Q(y_k) - x_0 \| + \| x_0 \| \\ &\leq \| y_k - x_0 \| + \| x_0 \| . \end{aligned}$$

Thus $\{g_k\}$ and $\{\gamma_k\}$ are bounded.

Therefore, without loss of generality

$$\exists g^*, \gamma^* : \quad g^* = \lim_{k \to \infty} g_k, \quad \gamma^* = \lim_{k \to \infty} \gamma_k.$$

It remains to take the limit in (4.9). □

Subgradients

Definition. Let f be a convex function.

A vector g is called the *subgradient* of f at $x_0 \in \operatorname{dom} f$

if for any $x \in \operatorname{dom} f$ we have:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle. \qquad (4.10)$$

Note: The subgradient sometimes is not unique!

Consider the function $f(x) = \mid x \mid$, $x \in R$.

Then $\forall y \in R$ and $g \in [-1, 1]$ we have:

$$f(y) = \mid y \mid \geq g \cdot y = f(0) + g \cdot (y - 0).$$

Definition.

The set of all subgradients of f at x_0,

$$\partial f(x_0),$$

is called the *subdifferential* of f at x_0.

Note: $\partial f(x_0)$ is a *closed convex* set (= intersection of half-spaces, the closed convex sets).

In our example $\partial f(0) = [-1, 1]$.

Theorem 4.13 *Let f be a closed convex function and $x_0 \in \mathrm{int}\,(\mathrm{dom}\, f)$.*

Then $\partial f(x_0)$ is a nonempty bounded set.

Proof. Note that $(f(x_0), x_0) \in \partial(\mathrm{epi}\,(f))$. Hence, $\exists$ a hyperplane supporting to epi (f) at $(f(x_0), x_0)$:

$-\alpha\tau + \langle d, x\rangle \leq -\alpha f(x_0) + \langle d, x_0\rangle, \quad \forall(\tau, x) \in \mathrm{epi}\,(f).$

Note that we can take

$$\| d \|^2 + \alpha^2 = 1. \qquad (4.11)$$

1. If $\tau \geq f(x_0) \Rightarrow (\tau, x_0) \in \mathrm{epi}\,(f) \Rightarrow \alpha \geq 0$.

2. $\exists \epsilon > 0$ and $L > 0 : B_2(x_0, \epsilon) \subseteq \mathrm{dom}\, f$ and

$$f(x) - f(x_0) \leq L \| x - x_0 \|, \quad \forall x \in B_2(x_0, \epsilon)$$

(Lemma 4.8). Therefore $\forall x \in B_2(x_0, \epsilon)$

$$\langle d, x - x_0\rangle \leq \alpha(f(x) - f(x_0) \leq \alpha L \| x - x_0 \| .$$

Choosing $x = x_0 + \epsilon d$ we get: $\| d \|^2 \leq L\alpha \| d \|$.

Thus, $(4.11) \Rightarrow \alpha \geq [1 + L^2]^{-1/2}$.

Hence, choosing $g = d/\alpha$ we get:

$$f(x) \geq f(x_0) + \langle g, x - x_0\rangle, \quad \forall x \in \mathrm{dom}\, f.$$

3. Finally, if $g \in \partial f(x_0)$, $g \neq 0$, then for $x = x_0 + \epsilon g / \| g \|$ we have:

$$\epsilon \| g \| = \langle g, x - x_0\rangle \leq f(x) - f(x_0) \leq L \| x - x_0 \| = L\epsilon. \quad \square$$

Theorem 4.14 *For any* $x_0 \in \operatorname{int}(\operatorname{dom} f)$ *and* $p \in R^n$ *we have:*

$$f'(x_0; p) = \max\{\langle g, p\rangle \mid \ g \in \partial f(x_0)\}.$$

Proof. 1. Since $f'(x_0, p)$ is convex, $\forall y \in \operatorname{dom} f$ we have:

$$f(y) \geq f(x_0) + f'(x_0; y - x_0) \geq f(x_0) + \langle g, y - x_0\rangle,$$

where $g \in \partial_p f'(x_0; 0)$. Thus, $\partial_p f'(x_0; 0) \subseteq \partial f(x_0)$.

On the other hand, $\forall g \in \partial f(x_0)$

$$f'(x_0; p) = \lim_{\alpha \downarrow 0} \tfrac{1}{\alpha}[f(x_0 + \alpha p) - f(x_0)] \geq \langle g, p\rangle. \tag{4.12}$$

Thus, $\partial f(x_0) \subseteq \partial_p f'(x_0; 0) \Rightarrow \partial f(x_0) \equiv \partial_p f'(x_0; 0)$.

2. Let $g_p \in \partial_p f'(x_0; p)$. Then $\forall v \in R^n$, $\tau > 0$, we have:

$$\tau f'(x_0; v) = f'(x_0; \tau v) \geq f'(x_0; p) + \langle g_p, \tau v - p\rangle.$$

This is possible for any $\tau > 0$ only if

$$f'(x_0; v) \geq \langle g_p, v\rangle, \tag{4.13}$$

and

$$f'(x_0; p) - \langle g_p, p\rangle \leq 0. \tag{4.14}$$

However, (4.13) $\Rightarrow g_p \in \partial_p f'(x_0; 0)$.

Hence, (4.12) + (4.14) $\Rightarrow \langle g_p, p\rangle = f'(x_0; p)$.

□

Some properties

Theorem 4.15 *We have*

$$f(x^*) = \min_{x \in \operatorname{dom} f} f(x).$$

if and only if $0 \in \partial f(x^*)$.

Proof. 1. If $0 \in \partial f(x^*)$ then

$$f(x) \geq f(x^*) + \langle 0, x - x^* \rangle = f(x^*) \quad \forall x \in \operatorname{dom} f.$$

2. If $f(x) \geq f(x^*)$ for all $x \in \operatorname{dom} f$ then $0 \in \partial f(x^*)$. □

Theorem 4.16 *For any* $x_0 \in \operatorname{dom} f$ *all vectors* $g \in \partial f(x_0)$ *are supporting to the sublevel set* $\mathcal{L}_f(f(x_0))$*:*

$$\langle g, x_0 - x \rangle \geq 0 \quad \forall x \in \mathcal{L}_f(f(x_0)),$$

$$\mathcal{L}_f(f(x_0)) = \{x \in \operatorname{dom} f : \ f(x) \leq f(x_0)\}.$$

Proof. Indeed, if $f(x) \leq f(x_0)$ and $g \in \partial f(x_0)$ then

$$f(x_0) + \langle g, x - x_0 \rangle \leq f(x) \leq f(x_0). \quad \square$$

Corollary 4.4 *Let* $Q \subseteq \operatorname{dom} f$ *be a closed convex set,* $x_0 \in Q$ *and*

$$x^* = \arg\min\{f(x) \mid \ x \in Q\}.$$

Then for any $g \in \partial f(x_0)$ *we have:*

$$\langle g, x_0 - x^* \rangle \geq 0.$$

The proof is evident.

Computing the subgradients

Lemma 4.6 *Let f be a convex function differentiable on its domain*
Then $\partial f(x) = \{f'(x)\}$ for any $x \in \text{int}\,(\text{dom}\, f)$.

Proof. Let $x_0 \in \text{int}\,(\text{dom}\, f)$. Clearly, $f'(x) \in \partial f(x)$.

2. Let us fix a direction $p \in R^n$. Then, in view of Theorem 4.14,

$$\langle f'(x_0), p \rangle = f'(x_0; p) \geq \langle g, p \rangle, \quad \forall g \in \partial f(x_0).$$

Changing the sign of p, we get:

$$\langle f'(x_0), p \rangle = \langle g, p \rangle \quad \forall g \in \partial f(x_0).$$

Taking $p = e_k$, $k = 1 \dots n$, we get $g = f'(x_0)$. □

Lemma 4.7 *Let $f(y)$ be convex,* $\text{dom}\, f \subseteq R^m$*, and*

$$\mathcal{A}(x) = Ax + b : \quad R^n \to R^m.$$

Then the function $\phi(x) = f(\mathcal{A}(x))$ is convex and

$$\text{dom}\,\phi = \{x \mid \ \mathcal{A}(x) \in \text{dom}\, f\},$$

$$\partial \phi(x) = A^T \partial f(\mathcal{A}(x)), \quad x \in \text{int}\,(\text{dom}\,\phi).$$

Proof. Indeed, let $y_0 = \mathcal{A}(x_0)$. Then $\forall p \in R^n$

$$\phi'(x_0, p) \ = f'(y_0; Ap) = \max\{\langle g, Ap \rangle \mid g \in \partial f(y_0)\}$$

$$= \max\{\langle \bar{g}, p \rangle \mid \bar{g} \in A^T \partial f(y_0)\}.$$

Using Theorem 4.14 and Corollary 4.3, we get
$\partial \phi(x_0) = A^T \partial f(\mathcal{A}(x_0))$. □

Lemma 4.8 *Let $f_1(x)$ and $f_2(x)$ are convex functions and $\alpha_1,\ \alpha_2 \geq 0$.*
Then the function

$$f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x)$$

is convex and

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x) \tag{4.15}$$

for $x \in \operatorname{int}(\operatorname{dom} f) = \operatorname{int}(\operatorname{dom} f_1) \cap \operatorname{int}(\operatorname{dom} f_2)$.

Proof. Indeed, let $x_0 \in \operatorname{int}(\operatorname{dom} f_1) \cap \operatorname{int}(\operatorname{dom} f_2)$. Then, for any $p \in R^n$ we have:

$$\begin{aligned}
f'(x_0;p) = \ & \alpha_1 f_1'(x_0;p) + \alpha_2 f_2'(x_0;p) \\
= \ & \max\{\langle g_1, \alpha_1 p\rangle \mid \ g_1 \in \partial f_1(x_0)\} \\
& + \max\{\langle g_2, \alpha_2 p\rangle \mid \ g_2 \in \partial f_2(x_0)\} \\
= \ & \max\{\langle \alpha_1 g_1 + \alpha_2 g_2, p\rangle \mid \\
& g_1 \in \partial f_1(x_0),\ g_2 \in \partial f_2(x_0)\} \\
= \ & \max\{\langle g, p\rangle \mid \\
& g \in \alpha_1 \partial f_1(x_0) + \alpha_2 \partial f_2(x_0)\}.
\end{aligned}$$

Using Theorem 4.14 and Corollary 4.3, we get (4.15). □

Lemma 4.9 *Let* $f_i(x)$, $i = 1 \ldots m$, *are convex. Then the function* $f(x) = \max\limits_{1 \le i \le m} f_i(x)$ *is convex. For* $x \in \mathrm{int}\,(\mathrm{dom}\, f) = \bigcap\limits_{i=1}^{m} \mathrm{int}\,(\mathrm{dom}\, f_i)$ *we have:*

$$\partial f(x) = \mathrm{Conv}\,\{\partial f_i(x) \mid i \in I(x)\}, \tag{4.16}$$

where $I(x) = \{i : f_i(x) = f(x)\}$.

Proof. Indeed, let $x_0 \in \bigcap\limits_{i=1}^{m} \mathrm{int}\,(\mathrm{dom}\, f_i)$. Assume that $I(x_0) = 1 \ldots k$. Then, for any $p \in R^n$ we have:

$$f'(x_0; p) = \max_{1 \le i \le k} f_i'(x_0; p)$$

$$= \max_{1 \le i \le k} \max\{\langle g_i, p\rangle \mid g_i \in \partial f_i(x_0)\}.$$

Note that for any numbers $a_1 \ldots a_k$ we have:

$$\max_{1 \le i \le k} a_i = \max\{\sum_{i=1}^{k} \lambda_i a_i \mid \{\lambda_i\} \in \Delta_k\},$$

where $\Delta_k = \{\lambda_i \ge 0, \ \sum\limits_{i=1}^{k} \lambda_i = 1\}$, the k-dimensional *standard simplex.* Therefore,

$$f'(x_0; p) = \max_{\{\lambda_i\} \in \Delta_k} \{\sum_{i=1}^{k} \lambda_i \max\{\langle g_i, p\rangle \mid g_i \in \partial f_i(x_0)\}$$

$$= \max\{\langle \sum_{i=1}^{k} \lambda_i g_i, p\rangle \mid g_i \in \partial f_i(x_0), \ \{\lambda_i\} \in \Delta_k\}$$

$$= \max\{\langle g, p\rangle \mid g = \sum_{i=1}^{k} \lambda_i g_i, g_i \in \partial f_i(x_0), \{\lambda_i\} \in \Delta_k\}$$

$$= \max\{\langle g, p\rangle \mid g \in \mathrm{Conv}\,\{\partial f_i(x_0), i \in I(x)\}\ \}. \quad \square$$

Lemma 4.10 *Let* Δ *be a set and*

$$f(x) = \sup\{\phi(y,x) \mid \ y \in \Delta\}.$$

Suppose that for any fixed $y \in \Delta$ *the function* $\phi(y,x)$ *is convex in* x.

Then $f(x)$ *is convex and for any* x_0 *from*

$$\mathrm{dom}\, f = \{x \in R^n \mid \ \exists \gamma : \phi(y,x) \le \gamma \; \forall y \in \Delta\}$$

we have

$$\partial f(x_0) \supseteq \mathrm{Conv}\, \{\partial \phi_x(y,x_0) \mid y \in I(x_0)\},$$

where $I(x_0) = \{y \mid \ \phi(y,x_0) = f(x_0)\}$.

Proof. Indeed, for any $x \in \mathrm{dom}\, f$, $y \in I(x)$ and $g \in \partial \phi_x(y,x_0)$ we have:

$$\begin{aligned} f(x) \ge \phi(y,x) \ &\ge \phi(y,x_0) + \langle g, x - x_0 \rangle \\ &= f(x_0) + \langle g, x - x_0 \rangle. \end{aligned}$$

□

Examples of Subdifferentials

1. Let $f(x) = \mid x \mid$, $x \in R$. Then $\partial f(0) = [-1, 1]$ since

$$f(x) = \max_{-1 \leq g \leq 1} g \cdot x.$$

2. $f(x) = \sum_{i=1}^{m} \mid \langle a_i, x \rangle - b_i \mid$. Denote

$$\begin{aligned} I_-(x) &= \{i : \langle a_i, x \rangle - b_i < 0\}, \\ I_+(x) &= \{i : \langle a_i, x \rangle - b_i > 0\}, \\ I_0(x) &= \{i : \langle a_i, x \rangle - b_i = 0\}. \end{aligned}$$

Then

$$\begin{aligned} \partial f(x) = & \sum_{i \in I_+(x)} a_i - \sum_{i \in I_-(x)} a_i \\ & + \sum_{i \in I_0(x)} [-a_i, a_i]. \end{aligned}$$

3. $f(x) = \max_{1 \leq i \leq n} x^{(i)}$. Denote

$$I(x) = \{i : \ x^{(i)} = f(x)\}.$$

Then

$$\partial f(x) = \mathrm{Conv}\, \{e_i \mid \ i \in I(x)\}.$$

For $x = 0$ we have:

$$\partial f(0) = \mathrm{Conv}\, \{e_i \mid \ 1 \leq i \leq n\}.$$

4. Euclidean norm: $f(x) = \| x \|$.

$$\partial f(0) = B_2(0,1) = \{x \in R^n \mid \| x \| \le 1\},$$
$$\partial f(x) = \{x / \| x \|\}, \quad x \neq 0.$$

5. Infinity norm:

$$f(x) = \| x \|_\infty = \max_{1 \le i \le n} | x^{(i)} | .$$

$$\partial f(0) = B_1(0,1) = \{x \in R^n \mid \sum_{i=1}^{n} | x^{(i)} | \le 1\},$$
$$\partial f(x) = \text{Conv}\,\{[-e_i, e_i] \mid i \in I(x)\},$$

where $I(x) = \{i : | x^{(i)} | = f(x)\}$.

6. l_1-norm:

$$f(x) = \| x \|_1 = \sum_{i=1}^{n} | x^{(i)} | .$$

$$\partial f(0) = B_\infty(0,1) = \{x \in R^n \mid \max_{1 \le i \le n} | x^{(i)} | \le 1\},$$
$$\partial f(x) = \sum_{i \in I_+(x)} e_i - \sum_{i \in I_-(x)} e_i + \sum_{i \in I_0(x)} [-e_i, e_i],$$

where $I_+(x) = \{i \mid x^{(i)} > 0\}$, $I_-(x) = \{i \mid x^{(i)} < 0\}$ and $I_0(x) = \{i \mid x^{(i)} = 0\}$.

Example: Optimality Conditions

Theorem 4.17 *(Kuhn-Tucker). Let f_i are differentiable convex functions, $i = 0 \dots m$.*

And let $\exists \bar{x}$: $f_i(\bar{x}) < 0$, $i = 1 \dots m$.

A point x^* is a solution to the problem

$$\min\{f_0(x) \mid \; f_i(x) \leq 0, i = 1 \dots m\} \qquad (4.17)$$

if and only if $\exists \lambda_i \geq 0$, such that

$$f_0'(x^*) + \sum_{i \in I^*} \lambda_i f_i'(x^*) = 0,$$

where $I^* = \{i \geq 1 : \; f_i(x^*) = 0\}$.

Proof. 1. Note, that x^* is a solution to (4.17) iff it is a global minimizer of the function

$$\phi(x) = \max\{f_0(x) - f^*; f_i(x), i = 1 \dots m\}.$$

2. In view of T. 4.15, this is the case iff $0 \in \partial\phi(x^*)$.

3. In view of L. 4.9, this is true iff

$$\exists \bar{\lambda}_i \geq 0 : \quad \sum_{i \in I_0} \bar{\lambda}_i f_i'(x^*) = 0, \quad \sum_{i \in I_0} \bar{\lambda}_i = 1,$$

where $I_0 = \{0\} \cup \{i \geq 1 : f_i(x^*) = 0\}$.

4. If $\bar{\lambda}_0 = 0$ then $\forall x \in R^n$

$$\sum_{i \in I^*} \bar{\lambda}_i f_i(x) \geq \sum_{i \in I^*} \bar{\lambda}_i [f_i(x^*) + \langle f_i'(x^*), x - x^* \rangle] = 0.$$

This is a contradiction. $\Rightarrow$ Take $\lambda_i = \bar{\lambda}_i / \bar{\lambda}_0$. $\square$