DATA8004: Optimization for Statistical Learning

Subject Lecturer: Man-Chung YUE

Lecture 4
Convex Optimization
Convex sets and functions

Convex set

Definition: A set $C \subseteq \mathbb{R}^n$ is said to be convex if for any $x \in C$, $y \in C$, and $\lambda \in (0, 1)$, it holds that

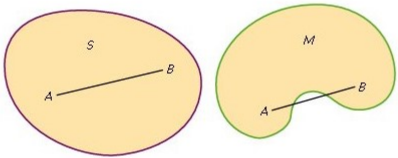
$$\lambda x + (1 - \lambda) y \in C.$$

Convex set

Definition: A set $C \subseteq \mathbb{R}^n$ is said to be convex if for any $x \in C$, $y \in C$, and $\lambda \in (0, 1)$, it holds that

$$\lambda x + (1 - \lambda) y \in C.$$

Geometrically, $\{\lambda x + (1 - \lambda) y : \; \lambda \in [0, 1]\}$ is the line segment joining x and y. Thus, convex sets are sets that contain all line segments that connect points in it.



Picture downloaded from https://www.britannica.com/science/convex-set.

Projections

Theorem 4.1:
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $y \in \mathbb{R}^n$. Then there exists a unique solution to the following optimization problem:

$$\text{Minimize } \frac{1}{2}\|x - y\|_2^2 \text{ subject to } x \in C. \tag{1}$$

Remark: The unique solution of (1) is called the projection of y onto C, denoted by $P_C(y)$.

Projections

Theorem 4.1:
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $y \in \mathbb{R}^n$. Then there exists a unique solution to the following optimization problem:

$$\text{Minimize } \frac{1}{2}\|x - y\|_2^2 \text{ subject to } x \in C. \tag{1}$$

Remark: The unique solution of (1) is called the projection of y onto C, denoted by $P_C(y)$.

Proof: If $y \in C$, the claim clearly holds.

Suppose that $y \notin C$. Fix any $w \in C$. Then it is not hard to see (Really?) that the following problem has the same optimal value and the set of optimal solutions (if nonempty) as (1):

$$\text{Minimize } \frac{1}{2}\|x - y\|_2^2 \text{ subject to } x \in C, \|x - y\|_2 \leq \|w - y\|_2. \tag{2}$$

Projections cont.

Proof of Theorem 4.1 cont.: Since the feasible set of (2) is compact, its set of optimal solutions is nonempty by Theorem 1.4. Thus, minimizers for (2) (and hence for (1)) exist.

To prove uniqueness, let x^* and $\tilde{x}$ be minimizers of (1) and denote the optimal value of (1) by v. Then $\frac{x^*+\tilde{x}}{2} \in C$ by convexity and

$$\sqrt{2v} \leq \left\| \frac{x^* + \tilde{x}}{2} - y \right\|_2 \leq \frac{1}{2}\|x^* - y\|_2 + \frac{1}{2}\|\tilde{x} - y\|_2 = \sqrt{2v}.$$

Thus, $\frac{x^*+\tilde{x}}{2}$ is also an optimal solution of (1).

Let $u^* := x^* - y$ and $\tilde{u} := \tilde{x} - y$. Then

$$\|0.5(u^* + \tilde{u})\|_2^2 = 0.5\|u^*\|_2^2 + 0.5\|\tilde{u}\|_2^2 \ (= 2v)$$
$$\Rightarrow \|0.5(u^* - \tilde{u})\|_2^2 = 0.$$

Hence, $u^* = \tilde{u}$, i.e., $x^* = \tilde{x}$.

Projection property

Theorem 4.2:
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set, $y \in \mathbb{R}^n$ and $u \in C$. Then

$$(y - P_C(y))^T (u - P_C(y)) \leq 0.$$

Projection property

Theorem 4.2:
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set, $y \in \mathbb{R}^n$ and $u \in C$. Then

$$(y - P_C(y))^T(u - P_C(y)) \leq 0.$$

Proof: For any $t \in (0, 1)$, $(1 - t)P_C(y) + tu \in C$ by convexity. Then

$$\begin{aligned} \|y - P_C(y)\|_2^2 &\leq \|y - P_C(y) + t[P_C(y) - u]\|_2^2 \\ &= \|y - P_C(y)\|_2^2 + 2t(y - P_C(y))^T(P_C(y) - u) + t^2\|P_C(y) - u\|_2^2. \end{aligned}$$

This implies (upon cancelling $\|y - P_C(y)\|_2^2$ and dividing by $t > 0$)

$$0 \leq 2(y - P_C(y))^T(P_C(y) - u) + t\|P_C(y) - u\|_2^2.$$

Passing to the limit as $t \downarrow 0$ gives the desired inequality.

Separation theorem

Theorem 4.3: (Separation theorem)
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $y \in \mathbb{R}^n \setminus C$. Then there exists $v \in \mathbb{R}^n \setminus \{0\}$ and $\alpha \in \mathbb{R}$ so that

$$v^T y > \alpha > v^T u$$

for all $u \in C$.

Separation theorem

Theorem 4.3: (Separation theorem)
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $y \in \mathbb{R}^n \backslash C$. Then there exists $v \in \mathbb{R}^n \backslash \{0\}$ and $\alpha \in \mathbb{R}$ so that

$$v^T y > \alpha > v^T u$$

for all $u \in C$.

Proof: Since $y \in \mathbb{R}^n \backslash C$, $v := y - P_C(y) \neq 0$.

Let $\alpha := \frac{1}{2} v^T [y + P_C(y)]$. Then

$$\begin{aligned} v^T y - \alpha &= [y - P_C(y)]^T \left[y - \frac{1}{2}[y + P_C(y)] \right] \\ &= \frac{1}{2}[y - P_C(y)]^T [y - P_C(y)] > 0. \end{aligned}$$

Separation theorem cont.

Proof of Theorem 4.3 cont.: Moreover, for any $u \in C$, we have

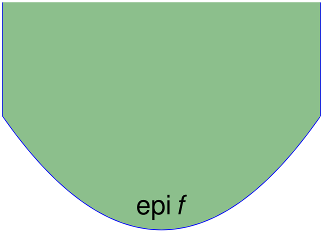
$$
\begin{aligned}
v^T u - \alpha &= [y - P_C(y)]^T \left[u - \frac{1}{2}[y + P_C(y)] \right] \\
&= [y - P_C(y)]^T \left[u - P_C(y) - \frac{1}{2}[y - P_C(y)] \right] \\
&= [y - P_C(y)]^T [u - P_C(y)] - \frac{1}{2}\|y - P_C(y)\|_2^2 \\
&< [y - P_C(y)]^T [u - P_C(y)] \le 0,
\end{aligned}
$$

where the last inequality follows from Theorem 4.2.

Extended-real-valued functions

Definition: A function $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ is called

- proper if $f(x) \neq -\infty$ for all x and $\operatorname{dom} f := \{x : \ f(x) < \infty\} \neq \emptyset$;
- convex if $\operatorname{epi} f := \{(x, r) \in \mathbb{R}^{n+1} : \ r \geq f(x)\}$ is a convex set;
- closed if epi f is closed.



Closed functions

Theorem 4.5:
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Then f is closed if and only if $\liminf_{x \to \hat{x}} f(x) \geq f(\hat{x})$ for all $\hat{x} \in \mathbb{R}^n$.

Closed functions

Theorem 4.5:
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Then f is closed if and only if $\liminf_{x \to \hat{x}} f(x) \geq f(\hat{x})$ for all $\hat{x} \in \mathbb{R}^n$.

Proof: First, suppose that $\liminf_{x \to \hat{x}} f(x) \geq f(\hat{x})$ for all $\hat{x} \in \mathbb{R}^n$.
Let $\{(x^k, r_k)\} \subseteq \operatorname{epi} f$ and suppose that $\lim_{k \to \infty} (x^k, r_k) = (x^*, r_*)$ for some $x^* \in \mathbb{R}^n$ and $r_* \in \mathbb{R}$. Then

$$r_* = \lim_{k \to \infty} r_k \geq \liminf_{k \to \infty} f(x^k) \geq f(x^*),$$

showing that $(x^*, r_*) \in \operatorname{epi} f$. Thus, $\operatorname{epi} f$ is closed.

Closed functions cont.

Proof of Theorem 4.5 cont.: Conversely, suppose that f is closed. Fix any $\hat{x} \in \mathbb{R}^n$.

If $\liminf_{x \to \hat{x}} f(x) = \infty$, then clearly $\liminf_{x \to \hat{x}} f(x) \geq f(\hat{x})$.
Now, suppose that $\liminf_{x \to \hat{x}} f(x) = c$ for some $c \in [-\infty, \infty)$. By the definition of liminf, there exists $\{x^k\}$ with $x^k \to \hat{x}$ such that

$$\lim_{k \to \infty} f(x^k) = c.$$

Pick any $r > c$, $r \in \mathbb{R}$. Then, for all large k, we have $r > f(x^k)$ and hence $(x^k, r) \in \operatorname{epi} f$. As $\operatorname{epi} f$ is closed, this further implies that $(\hat{x}, r) \in \operatorname{epi} f$, i.e., $r \geq f(\hat{x})$. Since this is true for any $r > c$, we have

$$f(\hat{x}) \leq c = \liminf_{x \to \hat{x}} f(x).$$

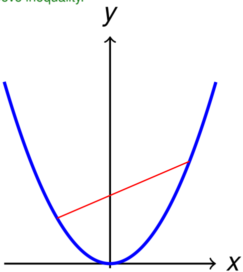
This completes the proof.

Convex functions

Theorem 4.6:
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper. It is convex if and only if for any u, $v \in \mathbb{R}^n$ and $\lambda \in (0, 1)$, it holds that

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda)f(v).$$

Note: Infinity can appear on both sides of the above inequality.



On global min

Theorem 4.7: (First-order necessary condition is sufficient under convexity.)
Let $f \in C^1(\mathbb{R}^n)$. If f is convex and $\nabla f(x^*) = 0$, then x^* is a global minimizer of f.

On global min

Theorem 4.7: (First-order necessary condition is sufficient under convexity.)
Let $f \in C^1(\mathbb{R}^n)$. If f is convex and $\nabla f(x^*) = 0$, then x^* is a global minimizer of f.

Proof: Let $\lambda \in (0, 1)$ and $x \in \mathbb{R}^n$. By convexity, we have

$$\begin{aligned}
\lambda f(x) + (1-\lambda) f(x^*) &\geq f(x^* + \lambda(x - x^*)) \\
\lambda(f(x) - f(x^*)) &\geq f(x^* + \lambda(x - x^*)) - f(x^*) \\
f(x) - f(x^*) &\geq \lambda^{-1}[f(x^* + \lambda(x - x^*)) - f(x^*)].
\end{aligned}$$

Passing to the limit as $\lambda \downarrow 0$, we see that

$$f(x) - f(x^*) \geq \nabla f(x^*)^T (x - x^*).$$

Since $\nabla f(x^*) = 0$, we conclude that $f(x) \geq f(x^*)$.

2nd order characterization

Theorem 4.8:
Suppose that $f \in C^2(\mathbb{R}^n)$. Then f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^n$.

2nd order characterization

Theorem 4.8:
Suppose that $f \in C^2(\mathbb{R}^n)$. Then f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^n$.

Proof: Let $\lambda \in (0,1)$ and $x, y \in \mathbb{R}^n$. Consider

$$\begin{aligned}
&\lambda f(x) + (1-\lambda) f(y) - f(\underbrace{\lambda x + (1-\lambda) y}_{x_\lambda}) \\
=&\lambda [f(x) - f(x_\lambda)] + (1-\lambda)[f(y) - f(x_\lambda)] \\
=&\lambda \nabla f(\xi)^T (x - x_\lambda) + (1-\lambda) \nabla f(\eta)^T (y - x_\lambda),
\end{aligned}$$

for some

$$\xi \in \{\alpha x + (1-\alpha) y : \ \alpha \in (\lambda, 1)\}, \quad \eta \in \{\alpha x + (1-\alpha) y : \ \alpha \in (0, \lambda)\}$$

2nd order characterization cont.

Proof of Theorem 4.8 cont.: Hence

$$\begin{aligned}&\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y)\\ =&\lambda(1-\lambda)[\nabla f(\xi) - \nabla f(\eta)]^T (x-y)\\ =&\lambda(1-\lambda)(\xi-\eta)^T \nabla^2 f(\mu)(x-y),\end{aligned}$$

where $\mu \in \{t\xi + (1-t)\eta : \ t \in [0,1]\}$.

Since there exist $\beta > \lambda > \gamma$ such that

$$\xi - \eta = x_\beta - x_\gamma = (\beta-\gamma)(x-y),$$

we conclude that,

$$\begin{aligned}&\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y)\\ =&\lambda(1-\lambda)(\beta-\gamma)(x-y)^T \nabla^2 f(\mu)(x-y).\end{aligned} \tag{3}$$

2nd order characterization cont.

Proof of Theorem 4.8 cont.: Now, suppose that $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^n$. Then we see immediately from (3) that

$$\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda)y)$$

showing that f is convex.

Conversely, suppose that f is convex. Then we have from (3) that for any x, y, there exists $\mu \in \{tx + (1-t)y : \ t \in [0,1]\}$ such that

$$(y-x)^T \nabla^2 f(\mu)(y-x) \geq 0.$$

Fix any $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ and set $y = x + \nu d$ for $\nu > 0$. Then we have

$$d^T \nabla^2 f(\mu_\nu) d \geq 0$$

for some $\mu_\nu \in \{x + \alpha d : \ \alpha \in [0, \nu]\}$. Letting $\nu \downarrow 0$, we conclude that $d^T \nabla^2 f(x) d \geq 0$. Since d is arbitrary, we have shown that $\nabla^2 f(x) \succeq 0$.

Example 1

Example: Show that the function

$$f(x_1, x_2) = x_1^2 + 3x_1x_2 + 6x_2^2 + 6x_2 + 3x_1 + 7$$

is convex.

Example 1

Example: Show that the function

$$f(x_1, x_2) = x_1^2 + 3x_1x_2 + 6x_2^2 + 6x_2 + 3x_1 + 7$$

is convex.

Solution: Note that

$$\nabla^2 f(x) = \begin{bmatrix} 2 & 3 \\ 3 & 12 \end{bmatrix}.$$

The eigenvalues of $\nabla^2 f(x)$ are $7 \pm \sqrt{34}$, both positive. Thus, we have $\nabla^2 f(x) \succ 0$, showing that f is convex.

Example 2

Example: Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\mu \geq 0$. Show that the function

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2$$

is convex.

Example 2

Example: Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\mu \geq 0$. Show that the function

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2$$

is convex.

Solution: Recall from Lecture 3 that

$$\nabla^2 f(x) = A^T A + \mu I.$$

The above matrix is positive semidefinite (as the sum of a positive semidefinite matrix $A^T A$ and a positive semidefinite matrix μI) for all $x \in \mathbb{R}^n$. Hence, f is convex.

Calculus of convex functions

Proposition 4.1:
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ both be proper and convex, $A \in \mathbb{R}^{n \times p}$, and $\alpha > 0$. Then the following functions are convex:

- $f + g$;
- αf;
- $f \circ A$;
- $\max\{f, g\}$.

Proof: Verify using the inequality characterization. Left as exercise.

Calculus of convex functions

Proposition 4.1:
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ both be proper and convex, $A \in \mathbb{R}^{n \times p}$, and $\alpha > 0$. Then the following functions are convex:

- $f + g$;
- αf;
- $f \circ A$;
- $\max\{f, g\}$.

Proof: Verify using the inequality characterization. Left as exercise.

Proposition 4.2: Norms are convex.

Proof: Verify using the inequality characterization. Left as exercise.

Calculus of convex functions

Proposition 4.1:
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ both be proper and convex, $A \in \mathbb{R}^{n \times p}$, and $\alpha > 0$. Then the following functions are convex:

- $f + g$;
- αf;
- $f \circ A$;
- $\max\{f, g\}$.

Proof: Verify using the inequality characterization. Left as exercise.

Proposition 4.2: Norms are convex.

Proof: Verify using the inequality characterization. Left as exercise.

Proposition 4.3:
Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be convex and $g : \mathbb{R}_+ \to \mathbb{R}$ be convex and non-decreasing. Then $g \circ f$ is convex.

Proof: Verify using the inequality characterization. Left as exercise.

Example 1

Example: Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\nu > 0$ $\mu > 0$. Argue that

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \nu\|x\|_1 + \mu\|x\|_2^2$$

is convex. This model includes the elastic net as a special case.

Example 1

Example: Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\nu > 0$ $\mu > 0$. Argue that

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \nu\|x\|_1 + \mu\|x\|_2^2$$

is convex. This model includes the elastic net as a special case.

Solution: Let $f_1(x) = \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_2^2$ and $f_2(x) = \nu\|x\|_1$.

- Since $\nabla^2 f_1(x) = A^T A + 2\mu I \succeq 0$ for all $x \in \mathbb{R}^n$, f_1 is convex.
- Since norm functions are convex and positive multiple of convex functions are convex, f_2 is convex.

Hence, f is convex as the sum of convex functions.

Example 2

Example: Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\nu > 0$ $\mu > 0$. Argue that

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \nu\|x\|_1 + \mu\sum_{i=1}^{n-1}|x_{i+1} - x_i|$$

is convex. This model corresponds to fused LASSO.

Example 2

Example: Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\nu > 0$ $\mu > 0$. Argue that

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \nu\|x\|_1 + \mu \sum_{i=1}^{n-1} |x_{i+1} - x_i|$$

is convex. This model corresponds to fused LASSO.

Solution:Let $f_1(x) = \frac{1}{2}\|Ax - b\|_2^2$, $f_2(x) = \nu\|x\|_1$, $f_3(x) = \mu\sum_{i=1}^{n-1}|x_{i+1} - x_i|$.

- Since $\nabla^2 f_1(x) = A^T A \succeq 0$ for all $x \in \mathbb{R}^n$, f_1 is convex.
- Since norm functions are convex and positive multiple of convex functions are convex, f_2 is convex.

Example 2

Example: Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\nu > 0$ $\mu > 0$. Argue that

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \nu\|x\|_1 + \mu\sum_{i=1}^{n-1}|x_{i+1} - x_i|$$

is convex. This model corresponds to fused LASSO.

Solution:Let $f_1(x) = \frac{1}{2}\|Ax - b\|_2^2$, $f_2(x) = \nu\|x\|_1$, $f_3(x) = \mu\sum_{i=1}^{n-1}|x_{i+1} - x_i|$.

- Since $\nabla^2 f_1(x) = A^T A \succeq 0$ for all $x \in \mathbb{R}^n$, f_1 is convex.
- Since norm functions are convex and positive multiple of convex functions are convex, f_2 is convex.
- Let $D \in \mathbb{R}^{(n-1)\times n}$ with $d_{ii} = -1$, $d_{i,i+1} = 1$ for $i = 1, \ldots, n-1$ and $d_{ij} = 0$ otherwise. Then $f_3(x) = \mu\|Dx\|_1$. Since norm functions are convex, composition of convex function with a linear map is convex, and positive multiple of convex functions are convex, f_3 is convex.

Hence, f is convex as the sum of convex functions.

Indicator functions

Definition: Let C be a nonempty closed set. The indicator function is defined as

$$\delta_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{otherwise.} \end{cases}$$

This function is proper and closed.

Indicator functions

Definition: Let C be a nonempty closed set. The indicator function is defined as

$$\delta_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{otherwise.} \end{cases}$$

This function is proper and closed.

Example: The optimization problem

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & \frac{1}{2}\|Ax - b\|_2^2 \\ \text{s.t.} & \|x\|_1 \leq \sigma \end{array}$$

can be written as $\min\limits_{x \in \mathbb{R}^n} f(x) + g(x)$, where $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ and $g(x) = \delta_{\|\cdot\|_1 \leq \sigma}(x)$.

Subdifferentials

Definition: Let f be a proper convex function. The subdifferential of f at x is defined as

$$\partial f(x) := \{\xi \in \mathbb{R}^n : \; f(y) - f(x) \geq \xi^T(y - x) \;\; \forall y \in \mathbb{R}^n\}.$$

An element of $\partial f(x)$ is called a subgradient.

Subdifferentials

Definition: Let f be a proper convex function. The subdifferential of f at x is defined as

$$\partial f(x) := \{\xi \in \mathbb{R}^n : \ f(y) - f(x) \geq \xi^T(y - x) \ \ \forall y \in \mathbb{R}^n\}.$$

An element of $\partial f(x)$ is called a subgradient.

Remark: Following the arguments on Slide 11, one can show that if f is convex and differentiable at x^*, then $\partial f(x^*) = \{\nabla f(x^*)\}$.

Subdifferentials

Definition: Let f be a proper convex function. The subdifferential of f at x is defined as

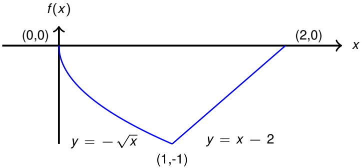
$$\partial f(x) := \{\xi \in \mathbb{R}^n : \; f(y) - f(x) \geq \xi^T(y - x) \;\; \forall y \in \mathbb{R}^n\}.$$

An element of $\partial f(x)$ is called a subgradient.

Remark: Following the arguments on Slide 11, one can show that if f is convex and differentiable at x^*, then $\partial f(x^*) = \{\nabla f(x^*)\}$.

Example: Consider

$$f(x) = \begin{cases} -\sqrt{x} & \text{if } 0 \leq x \leq 1, \\ x - 2 & \text{if } 1 < x \leq 2, \\ \infty & \text{otherwise.} \end{cases}$$



Then $\partial f(0) = \emptyset$, $\partial f(1) = [-0.5, 1]$, $\partial f(1.5) = \{1\}$, $\partial f(2) = [1, \infty)$.

Subdifferentials cont.

Proposition 4.4:
Let f be proper and convex. Then the following statements hold:

(i) $\partial f(x) = \emptyset$ whenever $x \notin \operatorname{dom} f$.

(ii) An $\bar{x}$ minimizes f if and only if $0 \in \partial f(\bar{x})$.

(iii) (Monotonicity) If $\xi_1 \in \partial f(x_1)$ and $\xi_2 \in \partial f(x_2)$, then

$$(\xi_1 - \xi_2)^T (x_1 - x_2) \geq 0.$$

(iv) (Closedness) Suppose f is also closed. If $\xi^k \in \partial f(x^k)$ for all k and $(x^k, \xi^k) \to (x^*, \xi^*)$ for some x^* and $\xi^* \in \mathbb{R}^n$, then

$$\xi^* \in \partial f(x^*).$$

Subdifferentials cont.

Proposition 4.4:
Let f be proper and convex. Then the following statements hold:

(i) $\partial f(x) = \emptyset$ whenever $x \notin \operatorname{dom} f$.

(ii) An $\bar{x}$ minimizes f if and only if $0 \in \partial f(\bar{x})$.

(iii) (Monotonicity) If $\xi_1 \in \partial f(x_1)$ and $\xi_2 \in \partial f(x_2)$, then

$$(\xi_1 - \xi_2)^T (x_1 - x_2) \geq 0.$$

(iv) (Closedness) Suppose f is also closed. If $\xi^k \in \partial f(x^k)$ for all k and $(x^k, \xi^k) \to (x^*, \xi^*)$ for some x^* and $\xi^* \in \mathbb{R}^n$, then

$$\xi^* \in \partial f(x^*).$$

Proof sketch: Items (i) and (ii) follow directly from definition. Item (iii) follows by summing the inequalities

$$\xi_1^T (x_2 - x_1) \leq f(x_2) - f(x_1) \ \text{ and } \ \xi_2^T (x_1 - x_2) \leq f(x_1) - f(x_2).$$

Item (iv) follows from the definition of subdifferential and Theorem 4.5.

Lipschitz continuity

Theorem 4.9:
Let f be proper and convex. Let $x \in \operatorname{dom} f$ and suppose that there exists $s > 0$ so that

$$\sup\{f(w) : \; w \in B(x,s)\} < \infty.$$

Then there exists $r > 0$ so that f is Lipschitz continuous on $B(x,r)$, i.e., there exists $L > 0$ so that

$$|f(u) - f(v)| \leq L\|u - v\|_2, \qquad \forall u, v \in B(x,r).$$

Lipschitz continuity

Theorem 4.9:
Let f be proper and convex. Let $x \in \operatorname{dom} f$ and suppose that there exists $s > 0$ so that

$$\sup\{f(w) : \; w \in B(x, s)\} < \infty.$$

Then there exists $r > 0$ so that f is Lipschitz continuous on $B(x, r)$, i.e., there exists $L > 0$ so that

$$|f(u) - f(v)| \leq L\|u - v\|_2, \qquad \forall u, v \in B(x, r).$$

Proof: By applying suitable translations, we may assume that $x = 0$. We will show that f is locally Lipschitz around 0.

Lipschitz continuity

Theorem 4.9:
Let f be proper and convex. Let $x \in \operatorname{dom} f$ and suppose that there exists $s > 0$ so that

$$\sup\{f(w) : \ w \in B(x,s)\} < \infty.$$

Then there exists $r > 0$ so that f is Lipschitz continuous on $B(x,r)$, i.e., there exists $L > 0$ so that

$$|f(u) - f(v)| \leq L\|u - v\|_2, \qquad \forall u, v \in B(x,r).$$

Proof: By applying suitable translations, we may assume that $x = 0$. We will show that f is locally Lipschitz around 0.

Write $U := \sup\{f(w) : \ w \in B(0,s)\} < \infty$. Then $B(0,s) \subseteq \operatorname{dom} f$ and any $w \in B(0,s)$ satisfies

$$f(w) + f(-w) \geq 2f(0) \ \Rightarrow \ f(w) \geq 2f(0) - f(-w) \geq 2f(0) - U.$$

Thus, $|f(w)| \leq \max\{|2f(0) - U|, U\} =: U_1$ whenever $w \in B(0,s)$.

Lipschitz continuity cont.

Proof of Theorem 4.9 cont.: Consider $r = s/2$ and any $u, v \in B(0, r)$.

Lipschitz continuity cont.

Proof of Theorem 4.9 cont.: Consider $r = s/2$ and any $u, v \in B(0, r)$.

Extrapolate the line segment $[u, v]$ (in the direction from u to v) until it hits the boundary of $B(0, s)$. Let z denote the intersection. Then

$$v = \frac{\|v - z\|_2}{\|u - z\|_2} u + \frac{\|u - v\|_2}{\|u - z\|_2} z$$

and we have $\|u - z\|_2 \geq r$ and $\|v - z\|_2 + \|u - v\|_2 = \|u - z\|_2$. Hence

$$\begin{aligned} & f(v) \leq \frac{\|v - z\|_2}{\|u - z\|_2} f(u) + \frac{\|u - v\|_2}{\|u - z\|_2} f(z) \\ \Longrightarrow \quad & f(v) - f(u) \leq \frac{\|u - v\|_2}{\|u - z\|_2} (f(z) - f(u)) \leq \frac{2U_1}{r} \|u - v\|_2. \end{aligned}$$

Extrapolating in the other direction gives $f(u) - f(v) \leq \frac{2U_1}{r} \|u - v\|_2$. This completes the proof.

Lipschitz continuity cont.

Corollary 4.1:
Let f be proper and convex. Suppose that $x \in \operatorname{int}(\operatorname{dom} f)$. Then f is locally Lipschitz around x.

Lipschitz continuity cont.

Corollary 4.1:
Let f be proper and convex. Suppose that $x \in \mathrm{int}(\mathrm{dom}\, f)$. Then f is locally Lipschitz around x.

Corollary 4.2:
Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then f is locally Lipschitz.

Proof sketch: Fix any $x \in \mathbb{R}^n$. Note that $B(0,1)$ is contained in $\{y : \|y\|_1 \leq \sqrt{n}\}$. Then, for any $u \in B(x,1)$, we have

$$f(u) \leq \max\{f(x \pm \sqrt{n} e_i) : \ e_i \text{ is standard basis}, i = 1, \ldots, n\}.$$

Lipschitz continuity cont.

Corollary 4.1:
Let f be proper and convex. Suppose that $x \in \operatorname{int}(\operatorname{dom} f)$. Then f is locally Lipschitz around x.

Corollary 4.2:
Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then f is locally Lipschitz.

Proof sketch: Fix any $x \in \mathbb{R}^n$. Note that $B(0,1)$ is contained in $\{y : \|y\|_1 \leq \sqrt{n}\}$. Then, for any $u \in B(x,1)$, we have

$$f(u) \leq \max\{f(x \pm \sqrt{n}e_i) : e_i \text{ is standard basis}, i = 1, \ldots, n\}.$$

Corollary 4.3:
Let f be proper and convex and let $x \in \operatorname{int}(\operatorname{dom} f)$. Then there exists $r > 0$ so that $f + \delta_{B(x,r)}$ is closed.

Proof sketch: Apply Corollary 4.1 and Theorem 4.5.

Directional derivative

Proposition 4.5:
Let f be proper and convex, $x \in \operatorname{dom} f$ and $d \in \mathbb{R}^n$. Then the function

$$t \mapsto \frac{f(x+td) - f(x)}{t}$$

is nondecreasing on $\mathbb{R}_{++}$.

Proof: For any $s > t > 0$, we have from Theorem 4.6 that

$$\frac{t}{s} f(x+sd) + \frac{s-t}{s} f(x) \geq f(x+td)$$

The conclusion follows upon rearranging terms in the above display.

Directional derivative

Proposition 4.5:
Let f be proper and convex, $x \in \operatorname{dom} f$ and $d \in \mathbb{R}^n$. Then the function

$$t \mapsto \frac{f(x+td)-f(x)}{t}$$

is nondecreasing on $\mathbb{R}_{++}$.

Proof: For any $s > t > 0$, we have from Theorem 4.6 that

$$\frac{t}{s}f(x+sd) + \frac{s-t}{s}f(x) \geq f(x+td)$$

The conclusion follows upon rearranging terms in the above display.

Corollary 4.4:
Let f be proper and convex and let $x \in \operatorname{dom} f$. Then

$$\lim_{t \downarrow 0} \frac{f(x+td)-f(x)}{t} = \inf_{t>0} \frac{f(x+td)-f(x)}{t} \in \overline{\mathbb{R}}.$$

Subdifferential nonemptiness

Theorem 4.10:
Let f be proper and convex and $x \in \mathrm{int}(\mathrm{dom}\, f)$. Then
$$\partial f(x) \neq \emptyset.$$

Subdifferential nonemptiness

Theorem 4.10:
Let f be proper and convex and $x \in \operatorname{int}(\operatorname{dom} f)$. Then
$$\partial f(x) \neq \emptyset.$$

Proof: By Corollary 4.3, there exists $a > 0$ so that $\widehat{f} := f + \delta_{B(x,a)}$ is closed. Now, for each $k = 1, 2, \ldots$, consider the points $p^k := (x, f(x) - \frac{1}{k})$. Then $p^k \notin \operatorname{epi} \widehat{f}$ for all k.

Subdifferential nonemptiness

Theorem 4.10:
Let f be proper and convex and $x \in \text{int}(\text{dom}\, f)$. Then
$$\partial f(x) \neq \emptyset.$$

Proof: By Corollary 4.3, there exists $a > 0$ so that $\widehat{f} := f + \delta_{B(x,a)}$ is closed. Now, for each $k = 1, 2, \ldots$, consider the points $p^k := (x, f(x) - \frac{1}{k})$. Then $p^k \notin \text{epi}\, \widehat{f}$ for all k.

By Theorem 4.3, there exists (w^k, α_k) with $\|(w^k, \alpha_k)\|_2 = 1$ so that
$$(w^k)^T x + \alpha_k(f(x) - 1/k) > (w^k)^T y + \alpha_k r, \quad \forall (y, r) \in \text{epi}\, \widehat{f}.$$
Since $\|(w^k, \alpha_k)\|_2 = 1$, by passing to a convergent subsequence if necessary, we may assume that $(w^k, \alpha_k) \to (w^*, \alpha_*)$ for some $w^* \in \mathbb{R}^n$ and $\alpha_* \in \mathbb{R}$ satisfying $\|(w^*, \alpha_*)\| = 1$. The above display becomes
$$(w^*)^T x + \alpha_* f(x) \geq (w^*)^T y + \alpha_* r, \quad \forall (y, r) \in \text{epi}\, \widehat{f}. \tag{4}$$

Subdifferential nonemptiness cont.

Proof of Theorem 4.10 cont.: Since $(x, f(x) + 1) \in \operatorname{epi} \widehat{f}$, we conclude that $\alpha_* \leq 0$. We claim that $\alpha_* < 0$.

Suppose to the contrary that $\alpha_* = 0$. Then $\|w^*\|_2 \neq 0$. Moreover, since $x \in \operatorname{int}(\operatorname{dom} f)$ (and hence $x \in \operatorname{int}(\operatorname{dom} \widehat{f})$), there exist $\epsilon > 0$ so that $x + \epsilon w^* \in \operatorname{dom} \widehat{f}$. Then (4) gives

$$(w^*)^T x \geq (w^*)^T (x + \epsilon w^*) \implies \|w^*\|_2 = 0,$$

leading to a contradiction. Thus, $\alpha_* < 0$.

Subdifferential nonemptiness cont.

Proof of Theorem 4.10 cont.:
Dividing both sides of (4) by $|\alpha_*|$ and writing $u^* = w^*/|\alpha_*|$, we obtain

$$(u^*)^T x - f(x) \geq (u^*)^T y - r, \quad \forall (y, r) \in \operatorname{epi} \widehat{f}.$$

This implies in particular that

$$\widehat{f}(y) - f(x) \geq (u^*)^T (y - x), \quad \forall y \in \mathbb{R}^n.$$

Subdifferential nonemptiness cont.

Proof of Theorem 4.10 cont.:
Dividing both sides of (4) by $|\alpha_*|$ and writing $u^* = w^*/|\alpha_*|$, we obtain

$$(u^*)^T x - f(x) \geq (u^*)^T y - r, \quad \forall (y, r) \in \operatorname{epi} \widehat{f}.$$

This implies in particular that

$$\widehat{f}(y) - f(x) \geq (u^*)^T (y - x), \quad \forall y \in \mathbb{R}^n.$$

Now, for any $u \in \mathbb{R}^n$, since $x \in \operatorname{int}(\operatorname{dom} f)$ and f is convex, there exists $s \in (0, 1)$ so that $x + t(u - x) \in \operatorname{dom} f \cap B(x, a)$ whenever $t \in (0, s]$. Hence, $\widehat{f}(x + t(u - x)) = f(x + t(u - x))$ for all such t and

Subdifferential nonemptiness cont.

Proof of Theorem 4.10 cont.:
Dividing both sides of (4) by $|\alpha_*|$ and writing $u^* = w^*/|\alpha_*|$, we obtain

$$(u^*)^T x - f(x) \geq (u^*)^T y - r, \quad \forall (y, r) \in \operatorname{epi} \widehat{f}.$$

This implies in particular that

$$\widehat{f}(y) - f(x) \geq (u^*)^T (y - x), \quad \forall y \in \mathbb{R}^n.$$

Now, for any $u \in \mathbb{R}^n$, since $x \in \operatorname{int}(\operatorname{dom} f)$ and f is convex, there exists $s \in (0, 1)$ so that $x + t(u - x) \in \operatorname{dom} f \cap B(x, a)$ whenever $t \in (0, s]$. Hence, $\widehat{f}(x + t(u - x)) = f(x + t(u - x))$ for all such t and

$$(u^*)^T (u - x) \leq \frac{f(x + t(u - x)) - f(x)}{t} \leq f(u) - f(x),$$

where the last inequality follows from Proposition 4.5. Thus, we have obtained an element $u^* \in \partial f(x)$.

Relative interior

Definition: Let C be a nonempty convex set. The set $\operatorname{ri} C$ is the interior of C relative to its affine hull.

Remark: When $0 \in C$, $\operatorname{ri} C$ is just the interior relative to span C. The relative interior of a nonempty convex set is always nonempty.

Relative interior

Definition: Let C be a nonempty convex set. The set $\operatorname{ri} C$ is the interior of C relative to its affine hull.

Remark: When $0 \in C$, $\operatorname{ri} C$ is just the interior relative to span C. The relative interior of a nonempty convex set is always nonempty.

The following facts concerning relative interior can be found in Ref 5.

Proposition 4.5: Let C and D be two nonempty convex sets in $\mathbb{R}^n$, A be a linear map with suitable dimensions. Then

- $\operatorname{ri} C = \operatorname{ri} \overline{C}$ and $\overline{C} = \overline{\operatorname{ri} C}$.
- $A \operatorname{ri} C = \operatorname{ri}(AC)$.
- $A^{-1} \operatorname{ri} C = \operatorname{ri}(A^{-1}C)$ if $A^{-1} \operatorname{ri} C \neq \emptyset$.
- $\operatorname{ri} C \cap \operatorname{ri} D = \operatorname{ri}(C \cap D)$ if $\operatorname{ri} C \cap \operatorname{ri} D \neq \emptyset$.
- $\operatorname{ri} C \times \operatorname{ri} D = \operatorname{ri}(C \times D)$.
- $\operatorname{ri}(C - D) = \operatorname{ri} C - \operatorname{ri} D$.

Subdifferential nonemptiness cont.

Theorem 4.11:
Let f be proper and convex and $x \in \mathrm{ri}(\mathrm{dom}\, f)$. Then

$$\partial f(x) \neq \emptyset.$$

Subdifferential nonemptiness cont.

Theorem 4.11:
Let f be proper and convex and $x \in \text{ri}(\text{dom}\, f)$. Then

$$\partial f(x) \neq \emptyset.$$

Proof sketch: By translation, we may assume that $x = 0$.

If span(dom f) has dimension 0, then f is the indicator function of a point and $\partial f(0) = \mathbb{R}^n \neq \emptyset$. On the other hand, if span(dom f) $= \mathbb{R}^n$, then the desired conclusion follows from Theorem 4.10.

Now, suppose that span(dom f) has dimension k, where $1 \leq k < n$. By a suitable rotation, we may also assume that

$$\text{span}(\text{dom}\, f) = \{(y, 0) \in \mathbb{R}^n : \; y \in \mathbb{R}^k\}.$$

Now 0 lies in the interior of dom $f|_{\mathbb{R}^k}$. Apply Theorem 4.10 to obtain a "subgradient" in $\mathbb{R}^k$ and extend it to a subgradient of f (e.g., by appending $n - k$ zeros).

Strong convexity

Definition: Let f be a proper function. We say that f is strongly convex if there exists $\sigma > 0$ such that the following function h is convex:

$$h(x) := f(x) - \frac{\sigma}{2}\|x\|_2^2.$$

The $\sigma > 0$ is called a strong convexity modulus of f.

Strong convexity

Definition: Let f be a proper function. We say that f is strongly convex if there exists $\sigma > 0$ such that the following function h is convex:

$$h(x) := f(x) - \frac{\sigma}{2}\|x\|_2^2.$$

The $\sigma > 0$ is called a strong convexity modulus of f.

Theorem 4.12:
Let f be proper, closed, and strongly convex. Then

$$\operatorname{Arg\,min} f \neq \emptyset$$

and is a singleton. Moreover, for any $x \in \mathbb{R}^n$, it holds that

$$\frac{\sigma}{2}\|x - x^*\|_2^2 \leq f(x) - f(x^*)$$

where $\sigma > 0$ is a strong convexity modulus of f and x^* is the unique minimizer of f.

Strong convexity cont.

Proof of Theorem 4.12: Let $\hat{x} \in \text{ri}(\text{dom}\, f)$. Then Theorem 4.11 implies that $\partial f(\hat{x}) \neq \emptyset$. Define the set

$$\Omega := \{x : f(x) \leq f(\hat{x})\}.$$

Note that this set is closed in view of Theorem 4.5 and the closedness of f.

Strong convexity cont.

Proof of Theorem 4.12: Let $\hat{x} \in \mathrm{ri}(\mathrm{dom}\, f)$. Then Theorem 4.11 implies that $\partial f(\hat{x}) \neq \emptyset$. Define the set

$$\Omega := \{x : f(x) \leq f(\hat{x})\}.$$

Note that this set is closed in view of Theorem 4.5 and the closedness of f.

Next, from the definition of strong convexity, there exists $\sigma > 0$ such that $h_\sigma(x) := f(x) - \frac{\sigma}{2}\|x\|_2^2$ is convex. Then we have for any $x \in \Omega$ that

$$h_\sigma\left(\frac{x+\hat{x}}{2}\right) \leq \frac{1}{2}h_\sigma(x) + \frac{1}{2}h_\sigma(\hat{x}),$$

i.e.,

$$f\left(\frac{x+\hat{x}}{2}\right) - \frac{\sigma}{2}\left\|\frac{x+\hat{x}}{2}\right\|_2^2 \leq \frac{1}{2}f(x) - \frac{\sigma}{4}\|x\|_2^2 + \frac{1}{2}f(\hat{x}) - \frac{\sigma}{4}\|\hat{x}\|_2^2.$$

Strong convexity cont.

Proof of Theorem 4.12 cont.: Upon rearranging terms, we obtain

$$\begin{aligned}
&\frac{\sigma}{4}\|x\|_2^2 + \frac{\sigma}{4}\|\hat{x}\|_2^2 - \frac{\sigma}{8}\|x+\hat{x}\|_2^2 \left(= \frac{\sigma}{8}\|x-\hat{x}\|_2^2\right) \\
&\le \frac{1}{2}f(x) + \frac{1}{2}f(\hat{x}) - f\left(\frac{x+\hat{x}}{2}\right) \\
&= \frac{1}{2}f(x) - \frac{1}{2}f(\hat{x}) + \left[f(\hat{x}) - f\left(\frac{x+\hat{x}}{2}\right)\right] \\
&\le f(\hat{x}) - f\left(\frac{x+\hat{x}}{2}\right) \le \frac{1}{2}\xi^T(x-\hat{x}). \quad (\because \xi \in \partial f(\hat{x}))
\end{aligned}$$

Strong convexity cont.

Proof of Theorem 4.12 cont.: Upon rearranging terms, we obtain

$$\begin{aligned}
&\frac{\sigma}{4}\|x\|_2^2 + \frac{\sigma}{4}\|\hat{x}\|_2^2 - \frac{\sigma}{8}\|x+\hat{x}\|_2^2 \left(= \frac{\sigma}{8}\|x-\hat{x}\|_2^2\right) \\
&\leq \frac{1}{2}f(x) + \frac{1}{2}f(\hat{x}) - f\left(\frac{x+\hat{x}}{2}\right) \\
&= \frac{1}{2}f(x) - \frac{1}{2}f(\hat{x}) + \left[f(\hat{x}) - f\left(\frac{x+\hat{x}}{2}\right)\right] \\
&\leq f(\hat{x}) - f\left(\frac{x+\hat{x}}{2}\right) \leq \frac{1}{2}\xi^T(x-\hat{x}). \quad (\because \xi \in \partial f(\hat{x}))
\end{aligned}$$

Then,

$$\frac{\sigma}{8}\|x-\hat{x}\|_2^2 \leq \frac{1}{2}\xi^T(x-\hat{x}) \leq \|\xi\|_2 \cdot \left\|\frac{x-\hat{x}}{2}\right\|_2,$$

which further implies $\|x-\hat{x}\|_2 \leq \frac{4\|\xi\|_2}{\sigma}$, showing that Ω is bounded.
Now, Theorem 1.4 guarantees that $\operatorname{Arg\,min} f \neq \emptyset$.

Strong convexity cont.

Proof of Theorem 4.12 cont.: Next, let $\widetilde{x} \in \operatorname{Arg\,min} f$. Consider the function

$$\widetilde{h}(x) := f(x) - \frac{\sigma}{2}\|x - \widetilde{x}\|_2^2.$$

Then one can check that $\widetilde{h}$ is convex. Moreover, by a direct computation based on the definition of subdifferential, we have

$$\partial\widetilde{h}(x) + \sigma(x - \widetilde{x}) = \partial f(x), \quad \forall x \in \mathbb{R}^n.$$

In addition, since $\widetilde{x} \in \operatorname{Arg\,min} f$, we have from Proposition 4.4 and the above display that

$$0 \in \partial f(\widetilde{x}) = \partial\widetilde{h}(\widetilde{x}).$$

Hence $\widetilde{x}$ also minimizes $\widetilde{h}$ in view of Proposition 4.4. Thus,

$$f(x) - \frac{\sigma}{2}\|x - \widetilde{x}\|_2^2 = \widetilde{h}(x) \geq \widetilde{h}(\widetilde{x}) = f(\widetilde{x})$$

for all $x \in \mathbb{R}^n$. This proves the uniqueness of minimizers of f and also the desired inequality.