SDSC6015 - Semester A, 2025

Stochastic Optimization for Machine Learning

1. Convex Sets & Convex Functions

1 Why Convexity? An Example

Theorem 1 (Fermat's Last Theorem (1637)). *For $n \geq 3$, then equation $x^n + y^n = z^n$ has no solution over positive integers.*

- Examples: $x^3 + y^3 = z^3$? $x^4 + y^4 = z^4$?
- Proved in 1994 by Andrew Wiles ($\approx$ 350 years)
- **Question:** Can we formulate this as an (continuous) optimization problem?

Small $\neq$ Easy!! Convexity is useful to identify optimization problems that are easy to solve.

- Most applied math:
- Optimization:

Key Property: Any locally optimal point of a convex optimization problem is also (globally) optimal.

2 Convex Sets

2.1 Affine & Convex Sets

— Affine Sets

Definition 1 (Line)**.** *Given two points $\boldsymbol{x}_1,\ \boldsymbol{x}_2 \in \mathbb{R}^n$ ($\boldsymbol{x}_1 \neq \boldsymbol{x}_2$). All the points on the line through $\boldsymbol{x}_1, \boldsymbol{x}_2$ can be represented as*

$$\boldsymbol{x} = \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2, \qquad \theta \in \mathbb{R}.$$

Definition 2 (Line Segment)**.** *Given two points $\boldsymbol{x}_1,\ \boldsymbol{x}_2 \in \mathbb{R}^n$ ($\boldsymbol{x}_1 \neq \boldsymbol{x}_2$), the set,*

$$\Big\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2,\ 0 \leq \theta \leq 1\Big\}$$

is called the line segment between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Definition 3 (Affine Set)**.** *A set $S \subseteq \mathbb{R}^n$ is an affine set if for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$, $\theta \in \mathbb{R}$,*

$$\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

Example (Solution set of linear equations): Consider $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Show that the set,

$$S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$$

is affine.

— Convex Sets

Definition 4 (Convex Set)**.** *A set $S \subseteq \mathbb{R}^n$ is* **convex** *if every point on the line segment between any two points $\boldsymbol{x}, \boldsymbol{y}$ in S is itself in S,*

$$\theta\, \boldsymbol{x} + (1-\theta)\, \boldsymbol{y} \in S, \quad \forall \theta \in [0,1].$$

Examples:

- Non-convex cases:

- Non-negative Orthant: $\mathbb{R}^n_+ = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \geq \mathbf{0}\}$

- Hyperplane: $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^\top \boldsymbol{x} = \boldsymbol{b}\} \quad (\boldsymbol{a} \neq \mathbf{0})$

- Halfspace: $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^\top \boldsymbol{x} \leq \boldsymbol{b}\} \quad (\boldsymbol{a} \neq \mathbf{0})$

Questions: Are these convex sets also affine?

Examples:

- Euclidean Ball: $B(\bar{\boldsymbol{x}}, r) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2 \leq r\}$

- Ellipsoid: $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \leq 1\}$, where $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is a symmetric, positive definite matrix.

- Polyhedron: $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\}$

  Polyhedron is intersection of finite number of halfspaces and hyperplanes.

— **Convex Cone**

Definition 5 (Conic combination)**.** *A conic combination of* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$ *has the form*

$$\boldsymbol{x} = \sum_{i=1}^{k} \theta_i \boldsymbol{x}_i,$$

where $\theta_i \geq 0$ *for* $i \in [k]$.

Definition 6 (Convex cone)**.** *A set K is a convex cone if it contains all conic combination of points in the set. That is, for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in K$ and $\theta_1, \theta_2 \geq 0$,*

$$\boldsymbol{x} = \theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 \in K.$$

Examples:

- Denote $\mathbb{S}^n$ to be the set of symmetric $n \times n$ matrices (**Question:** Is $\mathbb{S}^n$ affine/convex?).

  The *positive semidefinite cone* is the set of symmetric and positive semidefinite $n \times n$ matrices, i.e.

$$\begin{aligned} \mathbb{S}^n_+ &= \{\boldsymbol{X} \in \mathbb{S}^n \mid \boldsymbol{z}^\top \boldsymbol{X} \boldsymbol{z} \geq 0, \forall \boldsymbol{z} \in \mathbb{R}^n\} \\ &= \{\boldsymbol{X} \in \mathbb{S}^n \mid \boldsymbol{X} \succcurlyeq 0\}, \end{aligned}$$

  is a convex cone.

- A function $\|\cdot\|$ is a norm if
  - $\|\boldsymbol{x}\| \geq 0$; $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$
  - $\|\theta \boldsymbol{x}\| = |\theta| \cdot \|\boldsymbol{x}\|$
  - $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

  Norm ball: $\{\boldsymbol{x} \mid \|\boldsymbol{x} - \bar{\boldsymbol{x}}\| \leq r\}$ (convex set but not a convex cone.)

  Norm cone: $\{(\boldsymbol{x}, t) \mid \|\boldsymbol{x}\| \leq t\}$

Question: How about $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \leq (\boldsymbol{c}^\top \boldsymbol{x})^2, \boldsymbol{c}^\top \boldsymbol{x} \geq 0\}$, where $\boldsymbol{Q} \in \mathbb{S}^n_+$ and $\boldsymbol{c} \in \mathbb{R}^n$?

2.2 Convexity-Preserving Operations

So far, we have been using the definition to check convexity. Some complicated convex sets is obtained from simple convex sets and operations that preserve convexity!

— Intersection

Proposition 1. *If S_1 and S_2 are convex sets, $S_1 \cap S_2$ is also a convex set. Therefore, the intersection of convex sets is convex. (This holds for infinite intersections.)*

Examples:

- $\left\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\right\} = \bigcap_{i=1}^{m} \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i\right\}$
- $\left\{\boldsymbol{X} \in \mathbb{S}^n \mid \boldsymbol{z}^\top \boldsymbol{X} \boldsymbol{z} \geq 0, \forall \boldsymbol{z} \in \mathbb{R}^n\right\} = \bigcap_{\boldsymbol{z} \in \mathbb{R}^n} \left\{\boldsymbol{X} \in \mathbb{S}^n \mid \boldsymbol{z}^\top \boldsymbol{X} \boldsymbol{z} \geq 0\right\}$

— Affine Function

Definition 7 (Affine function)**.** *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function if*

$$f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b},$$

for some $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$.

Proposition 2. *Consider an affine function $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ and a convex set $S \subseteq \mathbb{R}^n$. Then the image $f(S) = \{f(\boldsymbol{x}) \mid \boldsymbol{x} \in S\}$ is convex. Similarly, if T is a convex set, then the inverse image $f^{-1}(T) = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \in T\}$ is also convex.*

Examples:

- Hyperbolic cone: $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \le (\boldsymbol{c}^\top \boldsymbol{x})^2, \boldsymbol{c}^\top \boldsymbol{x} \ge 0\}$, where $\boldsymbol{Q} \in \mathbb{S}^n_+$ and $\boldsymbol{c} \in \mathbb{R}^n$.

- Ellipsoid: $\{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \boldsymbol{Q}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \le 1\}$, where $\boldsymbol{Q} \in \mathbb{S}^n_{++}$ and $\bar{\boldsymbol{x}} \in \mathbb{R}^n$.

— **Perspective and Linear-fractional Function**

Definition 8 (Perspective function)**.** *A function $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ is a perspective function if $P(\boldsymbol{x}, t) = \frac{1}{t}\boldsymbol{x}$*

Definition 9 (Linear-fractional Function)**.** *A function $g : \mathbb{R}^n \to \mathbb{R}^m$ is a linear-fractional function if*

$$g(\boldsymbol{x}) = \frac{1}{\boldsymbol{c}^\top \boldsymbol{x} + d}\left(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\right), \qquad \boldsymbol{dom}(g) = \{\boldsymbol{x} \mid \boldsymbol{c}^\top \boldsymbol{x} + d > 0\}$$

Proposition 3. *Consider a perspective/linear-fractional function f and a convex set S. Then the image $f(S) = \{f(\boldsymbol{x}) \mid \boldsymbol{x} \in S\}$ is convex. Similarly, if T is a convex set, then the inverse image $f^{-1}(T) = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \in T\}$ is also convex.*

2.3 Projection onto Convex Sets

- The concept of projection is important to the theory and computation of convex optimization. Often it is used as a subroutine in many optimization algorithms.
- **Goal:** Finding the closest point in a convex set to a given point in $\mathbb{R}^n$.

Definition 10. *Let $C \subseteq \mathbb{R}^n$ be nonempty closed and convex set. The orthogonal projection of $\boldsymbol{x}$ onto C is defined as*

$$P_C(\boldsymbol{x}) \equiv \arg\min_{\boldsymbol{y} \in C} \|\boldsymbol{y} - \boldsymbol{x}\|.$$

That is, $P_C(\boldsymbol{x})$ is the closest point in C to $\boldsymbol{x}$.

Note: Requires C to be closed to guarantee the existence of the projection; otherwise, it's not well defined (e.g., projection of point $\{2\}$ onto $(0, 1)$).

Theorem 2 (The Weierstrass Theorem)**.** *Let $C \subset \mathbb{R}^n$ be compact (closed and bounded), and let $f : C \to \mathbb{R}$ be a continuous function on C. Then f attains a maximum and a minimum on C, i.e., there exist points $\boldsymbol{x}_{low}, \boldsymbol{x}_{up} \in C$ such that*

$$f(\boldsymbol{x}_{low}) \leq f(\boldsymbol{x}) \leq f(\boldsymbol{x}_{up}), \quad \forall \boldsymbol{x} \in C.$$

Theorem 3. *Let $C \subseteq \mathbb{R}^n$ be nonempty, closed, and convex. Then, for every $\boldsymbol{x}_0 \in \mathbb{R}^n$, there exists a unique point $\boldsymbol{x}^\star \in C$ that is closest (in the Euclidean norm) to $\boldsymbol{x}_0$.*

Examples

When C is simple, we can compute $P_C(\boldsymbol{x})$ explicitly (analytically).

1. If $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{w}^\top \boldsymbol{x} + b = 0\}$, $P_H(\boldsymbol{x}) = \boldsymbol{x} - \frac{\boldsymbol{w}^\top \boldsymbol{x} + b}{\|\boldsymbol{w}\|_2^2} \boldsymbol{w}$.

2. If $B = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 \leq 1\}$, $P_B(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}$.

2.4 Separation Theorem

Theorem 4 (Strict Separating Hyperplane Theorem)**.** *Consider a nonempty closed convex set* $C \subset \mathbb{R}^n$ *and a point* $\boldsymbol{x}_0 \in \mathbb{R}^n \backslash C$. *Then, there exists* $\boldsymbol{a} \neq \mathbf{0}$ *in* $\mathbb{R}^n$ *and* $b \in \mathbb{R}$ *with*

$$\boldsymbol{a}^\top \boldsymbol{x} < b, \ \forall \boldsymbol{x} \in C, \qquad \textit{and} \qquad \boldsymbol{a}^\top \boldsymbol{x}_0 > b.$$

Definition 11 (Supporting hyperplane)**.** *Consider a nonempty set* $C \subseteq \mathbb{R}^n$ *and a boundary point* $\boldsymbol{x}_0 \in \boldsymbol{bd}(C)$*. If* $\boldsymbol{w} \neq \mathbf{0}$ *in* $\mathbb{R}^n$ *satisfies* $\langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle$, $\forall \boldsymbol{x} \in C$*, then* $\{\boldsymbol{x} \mid \boldsymbol{w}^\top \boldsymbol{x} = \boldsymbol{w}^\top \boldsymbol{x}_0\}$ *is called a supporting hyperplane to* C *at* $\boldsymbol{x}_0$*.*

Theorem 5 (Weak Separating Hyperplane Theorem)**.** *Consider any convex set* $C \subseteq \mathbb{R}^n$ *and a point* $\boldsymbol{x}_0 \in \mathbb{R}^n \backslash C$*. Then, there exist* $\boldsymbol{w} \neq \mathbf{0}$ *(in* $\mathbb{R}^n$*) and* $b \in \mathbb{R}$ *with*

$$\langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq b,\ \forall \boldsymbol{x} \in C, \qquad \textit{and} \qquad \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle \geq b.$$

Theorem 6. *For every convex set* $C \subseteq \mathbb{R}^n$ *and any* $\boldsymbol{x}_0 \in \boldsymbol{bd}(C)$ *there is a supporting hyperplane to* C *at* $\boldsymbol{x}_0$*.*

3 Convex Functions

3.1 Definitions & Basics

Definition 12 (domain). *For a function f, the set $\boldsymbol{dom}(f)$ is its domain. If $f : \mathbb{R}^n \to (-\infty, +\infty]$ is convex, we let $\boldsymbol{dom}(f) = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) < +\infty\}$.*

Definition 13 (convex function). *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain $\boldsymbol{dom}(f)$ is convex and the line segment connecting $f(\boldsymbol{x}_1)$ and $f(\boldsymbol{x}_2)$ at any two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{dom}(f)$ lies above the function between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$,*

$$f(\alpha\,\boldsymbol{x}_1 + (1-\alpha)\,\boldsymbol{x}_2) \le \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2), \quad \forall \alpha \in [0,1]$$

- **Strict** convexity when the inequality is strict:

$$f\left(\alpha\,\boldsymbol{x}_1 + (1-\alpha)\,\boldsymbol{x}_2\right) < \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2), \quad \forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathbf{dom}(f), \forall \alpha \in (0,1)$$

- f is (strictly) **concave** if $-f$ is (strictly) convex.

— Epigraph

Definition 14 (Epigraph). *The epigraph of $f : \mathbb{R}^n \to \mathbb{R}$ is*

$$\boldsymbol{epi}(f) = \left\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in \boldsymbol{dom}(f),\ f(\boldsymbol{x}) \le t\right\}.$$

Theorem 7. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph is a convex set.*

— Jensen's Inequality

Theorem 8 (Jensen's Inequality)**.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if*

$$f\left(\sum_{i=1}^{k} \alpha_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{k} \alpha_i f(\boldsymbol{x}_i),$$

for any $\boldsymbol{x}_i \in \mathbb{R}^n$, $\boldsymbol{\alpha} \in \Delta_k$.

Note: $\Delta_k = \{\boldsymbol{\alpha} \in \mathbb{R}^k \mid \sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i \geq 0, \forall i \in [k]\}$

— Sublevel Set

Definition 15 (Sublevel Set)**.** *The α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is definied as*

$$C_\alpha = \{\boldsymbol{x} \in \boldsymbol{dom}(f) \mid f(\boldsymbol{x}) \leq \alpha\}$$

Proposition 4. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then all of its sublevel sets are convex.*

Question: Is the reverse implication true?

— Checking Convexity Along Lines

Proposition 5. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for $\forall \boldsymbol{x} \in \boldsymbol{dom}(f)$, $\forall \boldsymbol{v} \in \mathbb{R}^n$, the function $g(t) = f(\boldsymbol{x} + t \cdot \boldsymbol{v})$ is convex on its domain $\boldsymbol{dom}(g) = \{t \in \mathbb{R} \mid \boldsymbol{x} + t \cdot \boldsymbol{v} \in \boldsymbol{dom}(f)\}$.*

Question: What does this mean?

3.2 First and Second Order Conditions

— First Order Condition

Definition 16. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable if its gradient $\nabla f = [\partial f/\partial x_1, \partial f/\partial x_2, \ldots, \partial f/\partial x_n]^\top$ exists at each point in $\boldsymbol{dom}(f)$, and $\boldsymbol{dom}(f)$ is open.*

Proposition 6 (First Order Condition)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then f is convex if and only if $\boldsymbol{dom}(f)$ is convex and*

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}),$$

where $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{dom}(f)$.

— **Second Order Condition**

Definition 17. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable if its Hessian*

$$\nabla^2 f = \begin{bmatrix} \partial^2 f/\partial x_1 \partial x_1 & \cdots & \partial^2 f/\partial x_1 \partial x_n \\ \vdots & \ddots & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \cdots & \partial^2 f/\partial x_n \partial x_n \end{bmatrix}$$

exists at each point in $\boldsymbol{dom}(f)$*, and* $\boldsymbol{dom}(f)$ *is open.*

Proposition 7 (Second order condition)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Then f is convex if and only if* $\boldsymbol{dom}(f)$ *is convex and*

$$\nabla^2 f(\boldsymbol{x}) \succcurlyeq 0, \quad \forall \boldsymbol{x} \in \boldsymbol{dom}(f)$$

— **Examples**

- Exponential function: $f(x) = e^{a \cdot x}$, $\mathbf{dom}(f) = \mathbb{R}$

- Powers: $f(x) = x^{\beta}$ ($\beta \geq 1$ or $\beta \leq 0$), $\mathbf{dom}(f) = \mathbb{R}_{++}$

- Negative logarithm: $f(x) = -\log(x)$, $\mathbf{dom}(f) = \mathbb{R}_{++}$

- Negative entropy: $f(x) = x \cdot \log(x)$, $\mathbf{dom}(f) = \mathbb{R}_{++}$

3.3 Convexity-Preserving Operations

Proposition 8 (Non-negative weighted sum)**.** *Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be convex function, $\forall i \in [m]$. Suppose $\cap_{i=1}^{m} \boldsymbol{dom}(f_i) \neq \emptyset$. Then*

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i \cdot f_i(\boldsymbol{x})$$

is a convex function where $\alpha_i \in \mathbb{R}_+$, $\forall i \in [m]$.

Proposition 9 (Affine Composition)**.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $f(\boldsymbol{x}) = g(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$, where $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ and $\boldsymbol{b} \in \mathbb{R}^n$. Suppose $\boldsymbol{dom}(f) \neq \emptyset$, then f is convex.*

Proposition 10 (Pointwise Maximum)**.** *If $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ are convex functions and $\boldsymbol{dom}(f_1) \cap \boldsymbol{dom}(f_2) \neq \emptyset$, then $f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ is convex.*

Proposition 11 (Pointwise Supremum)**.** *If $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $\boldsymbol{x}$ for each $\boldsymbol{y} \in \mathcal{Y}$, then*

$$g(\boldsymbol{x}) = \sup_{y \in \mathcal{Y}} f(\boldsymbol{x}, \boldsymbol{y})$$

is convex.

Examples

- Piecewise linear function: $\max_{i \in [m]} \boldsymbol{a}_i^\top \boldsymbol{x} + \boldsymbol{b}_i$

- Sum of largest k values: $x_{i_1} + x_{i_2} + \cdots + x_{i_k}$

- Support function: $S_C(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} \boldsymbol{y}^\top \boldsymbol{x}$

- Maximum eigenvalue: $\lambda_{\max}(\boldsymbol{X}) = \sup_{\|\boldsymbol{v}\|_2 = 1} \boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{v}$

— Composition

Proposition 12 (Composition)**.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be convex function and $h : \mathbb{R} \to \mathbb{R}$ be convex and non-decreasing on $\boldsymbol{dom}(h)$. Suppose $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$ and $\boldsymbol{dom}(f) \neq \emptyset$, then f is convex.*

Examples

- $e^{g(\boldsymbol{x})}$ is convex if g is convex.
- $1/g(\boldsymbol{x})$ is convex if g is concave and positive.

Proposition 13 (Composition - Generalization)**.** *Let $g : \mathbb{R}^n \to \mathbb{R}^k$ be convex in each component and $h : \mathbb{R}^k \to \mathbb{R}$ be convex and non-decreasing in each argument. Suppose $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$ and $\boldsymbol{dom}(f) \neq \emptyset$, then f is convex.*

3.4 Conjugate Functions

Definition 18 (Conjugate function)**.** *The conjugate of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as*

$$f^\star(\boldsymbol{y}) = \sup_{\boldsymbol{x} \in \boldsymbol{dom}(f)} \boldsymbol{y}^\top \boldsymbol{x} - f(\boldsymbol{x}).$$

Proposition 14. *The conjugate function $f^\star$ of function $f : \mathbb{R}^n \to \mathbb{R}$ is closed and convex.*

Theorem 9. *For any function $f : \mathbb{R}^n \to \mathbb{R}$ that is proper, convex, and closed, then $f^{\star\star} = f$.*

Examples:

- **Norm.** Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, with dual norm $\|\cdot\|_\star$. The conjugate of $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ is

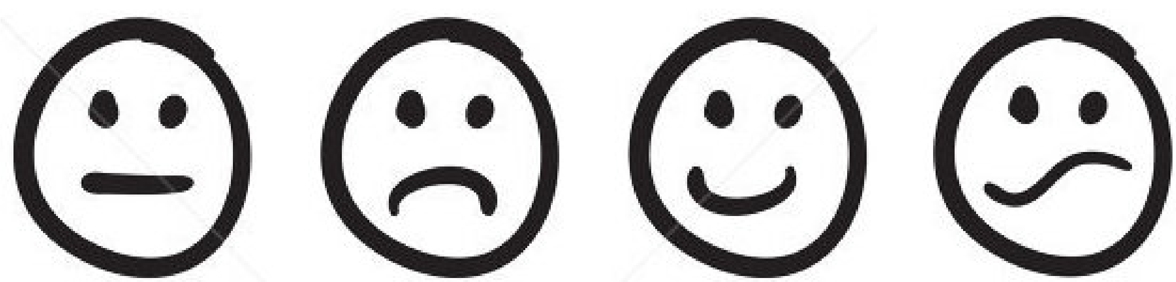
$$f^\star(\boldsymbol{y}) = \begin{cases} 0 & \text{if } \|\boldsymbol{y}\|_\star \leq 1, \\ \infty & \text{otherwise.} \end{cases}$$

  (LATER) Check lecture on Lagrangian Duality for the special case.

- **Log-determinant.** Consider $f(\boldsymbol{X}) = \log\det \boldsymbol{X}^{-1}$ with $\mathbf{dom}(f) = \mathbb{S}^n_{++}$. The conjugate is

$$f^\star(\boldsymbol{Y}) = \begin{cases} \log\det(-\boldsymbol{Y}^{-1}) - n & \text{if } \boldsymbol{Y} \in -\mathbb{S}^n_{++}, \\ \infty & \text{otherwise.} \end{cases}$$

3.5 Summary



Affine Concave Convex Neither