Convex Optimization Problem IV

Jong-June Jeon

Fall, 2023

Department of Statistics, University of Seoul

Things to know

- Operation preserving convexity
- Examples of convex functions
- Quasiconvex function

Operations preserving convexity

Nonnegative weighted sums

If f_i for $i = 1, \cdots, m$ are convex and $w_i \geq 0$, then $g(x) = \sum_{i=1}^m w_i f_i(x)$ is convex.

If $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, and $w(y) \geq 0$ for each $y \in \mathcal{A}$, then

$$g(x) = \int_{\mathcal{A}} w(y) f(x, y) dy$$

is convex in x.

Let $w_i \geq 0$ then $w_i f_i(\lambda x + (1-\lambda)y) \leq \lambda w_i f_i(x) + (1-\lambda) w_i f_i(y)$ for all i and $\lambda \in [0,1]$.

$$
\begin{aligned}
g(\lambda x + (1-\lambda)y) &= \sum_i w_i f_i(\lambda x + (1-\lambda)y) \\
&\leq \sum_i \lambda w_i f_i(x) + (1-\lambda) w_i f_i(y) \\
&= \lambda g(x) + (1-\lambda) g(y).
\end{aligned}
$$

Composition with an affine mapping

Let $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{n\times m}$, and $b \in \mathbb{R}^n$. If f is convex, then

$$g(x) = f(Ax + b)$$

is also convex.

Pointwise maximum function and supremum

If f_1 and f_2 are convex, then $f(x) = \max(f_1(x), f_2(x))$ with $\text{dom}(f) = \text{dom}(f_1) \cap \text{dom}(f_2))$ is convex. Generally, if $f_1, \cdots, f_m$ are convex, then

$$f(x) = \max(f_1(x), \cdots, f_m(x))$$

is convex.

(proof) $f_i(\lambda x + (1-\lambda)y) \le \lambda f_i(x) + (1-\lambda) f_i(y)$ for all i and $\lambda \in [0,1]$.

$$
\begin{aligned}
f(\lambda x + (1-\lambda)y) &= \max_i f_i(\lambda x + (1-\lambda)y) \\
&\le \max_i \lambda f_i(x) + (1-\lambda) f_i(y) \\
&\le \max_i \lambda f_i(x) + \max_i (1-\lambda) f_i(y) \\
&= \lambda \max_i f_i(x) + (1-\lambda) \max_i f_i(y) \\
&= \lambda f(x) + (1-\lambda) f(y),
\end{aligned}
$$

Pointwise maximum and supremum

If $f(x, y)$ is convex in x for each $y \in \mathcal{A}$,

$$g(x) = \mathsf{sup}_{y \in \mathcal{A}} f(x, y)$$

is convex in x.

Example 1

- $f : x \mapsto \max\{a_1^\top x + b_1, \cdots, a_m^\top x + b_m\}$ is convex.
- Let $x \in \mathbb{R}^n$ and $x_{[i]}$ be the ith largest component of x. Then $f(x) = \sum_{i=1}^r x_{[i]}$ is convex.

(proof)

$$f(x) = \max\{x_{i_1} + \cdots + x_{i_r} : 1 \le i_1 < \cdots < i_r \le n\}$$

Since $f(x)$ is the maximum of affine functions, f is convex.

Check the convexity of $f(x) = \sum_{i=1}^r w_i x_{[i]}$.

Example 2 (Distance to the farthest point of a set)

Let $C \subset \mathbb{R}^n$.

$$f(x) = \sup_{y \in C} \|x - y\|$$

is convex.

(proof) $g(x, y) = \|x - y\|$ is convex. Thus, $\sup_{y \in C} g(x, y)$ is convex.

Example 3 (Maximum eigenvalue of a symmetric matrix)

$f : X \in \mathcal{S}^m \mapsto \lambda_{max}(X) \in \mathbb{R}$.

$$f(X) = \sup\{y^\top X y : \|y\| = 1\}$$

Let $g : (X, y) \in \mathcal{S}^n \times \mathbb{R}^m \mapsto y^\top X t \in \mathbb{R}$, then $g(X, y)$ is linear for a fixed y. Thus, f is a pointwise maximum of $g(X, y)$, and it is convex.

Definition 4 (convex optimization problem)

- $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$ is convex function.
- $C = \{x \in \mathbb{R}^n : f_i(x) \leq 0, h_j(x) = 0$ for all $i, j \geq 1\}$ is convex set where $f_i, h_j : \mathbb{R}^n \to \mathbb{R}$.

$$\begin{aligned} \text{minimize } & f_0(x) \\ \text{subject to } & f_i(x) \leq 0, \quad i = 1, \cdots, p, \\ & h_j(x) = 0 \quad j = 1, \cdots, m. \end{aligned}$$

Note that

- If f_i is convex then $\{x : f_i(x) \leq 0\}$ is convex set.
- If h_j is affine then $\{x : h_j(x) = 0\}$ is convex set.
- Any finite intersection of convex sets is convex.

Example 5 (See the proof in p.73–74)

- Quadratic over linear function (convex): $f(x, y) = x^2/y$, with

$$\operatorname{dom} f = \mathbb{R} \times \mathbb{R}_{++}$$

- Log-sum-exp (convex): $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$
- Geometric mean (concave)
- Log-determinant (concave)

proof of Geometric mean $f : x \in \mathbb{R}^n_{++} \mapsto (\prod_{i=1}^n x_i)^{1/n} \in \mathbb{R}$ is concave.

$$\frac{\partial f(x)}{\partial x_j} = \prod_{i \neq j} x_i^{1/n} \times \frac{1}{n} x_j^{1/n-1} = \left(\frac{1}{n} \prod_{i=1}^n x_i^{1/n} \right) \frac{1}{x_j}$$

$$\frac{\partial^2 f(x)}{\partial x_j^2} = -\frac{1}{n} \frac{(n-1)}{n} \prod_{i \neq j} x_i^{1/n} x_j^{1/n-2} = -\left(\frac{1}{n} \frac{(n-1)}{n} \prod_{i=1}^n x_i^{1/n} \right) \frac{1}{x_j^2}$$

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_k} = \frac{1}{n^2} \left(\prod_{i=1}^n x_i^{1/n} \right) \frac{1}{x_j} \frac{1}{x_k}$$

Let $z = (z_1, \cdots, z_n)^\top = (1/x_1, \cdots, 1/x_n)^\top$ then

$$\nabla^2 f(x) = -\frac{1}{n^2}\left(n\text{diag}(z_1^2, \cdots, z_n^2) - zz^\top\right)$$

Let $a = (a_1, \cdots, a_n)$ and $v = (a_1 z_1, \cdots, a_n z_n)$ then

$$\begin{aligned} a^\top \nabla^2 f(x) a &= -\frac{1}{n^2}\left(n\sum_{i=1}^{n} a_i^2 z_i^2 - (\sum_{i=1}^{n} a_i z_i)^2\right) \\ &= -\frac{1}{n^2}\left(\|1\|^2\|v\|^2 - < 1, v >^2\right) \leq 0 \end{aligned}$$

The inequality holds by Cauchy inequality. Therefore f is concave.

Example 6 (Linear regression)

Let $(y_i, x_i) \in \mathbb{R} \times \mathbb{R}^p$ for $i = 1, \cdots, n$ be response-covariate pairs and the objective function of linear regression is given by

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \frac{1}{2}\sum_{i=1}^{n}(y_i - x_i^\top \boldsymbol{\beta})^2 \\
&= \frac{1}{2}\boldsymbol{\beta}^\top(\sum_{i=1}^{n} x_i x_i^\top)\boldsymbol{\beta} - (\sum_{i=1}^{n} y_i x_i)^\top \boldsymbol{\beta} + \frac{1}{2}\sum_{i=1}^{n} y_i^2.
\end{aligned}
$$

Let $A = (\sum_{i=1}^{n} x_i x_i^\top)$, $b = \sum_{i=1}^{n} y_i x_i$ and $c = \frac{1}{2}\sum_{i=1}^{n} y_i^2$ then

$$
L(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\beta}^\top A\boldsymbol{\beta} - b^\top \boldsymbol{\beta} + c,
$$

a quadratic function.

A is semi-positive definite because

$$u^\top (\sum_{i=1}^{n} x_i x_i^\top) u = \sum_{i=1}^{n} (x_i^\top u)^\top (x_i^\top u) = \sum_{i=1}^{n} \|x_i^\top u\|^2 \geq 0$$

for all $u \in \mathbb{R}^p$. Thus, $L(\boldsymbol{\beta})$ is convex.

Example 7 (Logistic regression)

Let $(y_i, x_i) \in \{0, 1\} \times \mathbb{R}^p$ for $i = 1, \cdots, n$ be response-covariate pairs and the objective function (negative loglikelihood) is given by

$$L(\boldsymbol{\beta}) = -\sum_{i=1}^{n} y_i x_i^\top \boldsymbol{\beta} + \sum_{i=1}^{n} \log(1 + \exp(x_i^\top \boldsymbol{\beta})).$$

The hessian matrix of $L(\boldsymbol{\beta})$ is

$$H(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(x_i x_i^\top \frac{\exp(x_i^\top \boldsymbol{\beta})}{(1 + \exp(x_i^\top \boldsymbol{\beta}))} \frac{1}{(1 + \exp(x_i^\top \boldsymbol{\beta}))} \right).$$

Let $p(x_i;\beta) = \frac{\exp(x_i^\top \beta)}{(1+\exp(x_i^\top \beta))}$.

$$
\begin{aligned}
u^\top H(\boldsymbol{\beta}) u &= \sum_{i=1}^{n} u^\top \left(x_i x_i^\top p(x_i;\beta)(1-p(x_i;\beta)) \right) u^\top \\
&= \sum_i \| \sqrt{p(x_i;\beta)(1-p(x_i;\beta)} x_i^\top u \|^2 \\
&\geq \left(\min_i \{ p(x_i;\beta)(1-p(x_i;\beta)\} \right) \sum_i \| x_i^\top u \|^2 \geq 0,
\end{aligned}
$$

for all u.

Example 8 (Linear support vector machine)

Let $(y_i, x_i) \in \{-1, 1\} \times \mathbb{R}^p$ for $i = 1, \cdots, n$ be response-covariate pairs and the objective function is given by

$$L_\lambda(\boldsymbol{\beta}) = \sum_{i=1}^{n} \max(0, 1 - y_i x_i^\top \boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} \beta_j^2$$

Since $l(t) = \max(0, t)$ is convex, it is easily shown that $L_\lambda(\boldsymbol{\beta})$ is convex.

Definition 9 (Conjugate function)

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$. $f^* : \mathbb{R}^n \mapsto \mathbb{R}$ is defined as

$$f^*(y) = \sup_{x \in \mathsf{dom}(f)} (y^\top x - f(x))$$

- f^* is always convex.
- If f is convex and closed, then $f^{**} = f$.

Example 10

- $f(x) = ax + b$: $\mathrm{dom}(f^*) = \{a\}$ and $f^*(y) = -b$
- $f(x) = \exp(x)$: $f^*(y) = y \log y - y$ with $\mathrm{dom}(f^*) = \mathbb{R}_+$
- $f(x) = (1/2)x^\top Qx$ with $Q \in \mathcal{S}_+^n+$: $f^*(y) = (1/2)y^\top Q^{-1}y$

Theorem 11 (Fenchel inequality)

$$f(x) + f^*(y) \geq x^\top y$$

for all x, y. This is called Fenchel's inequality.

Example 12

$f(x) = (1/2)x^\top Qx$ with $Q \in \mathcal{S}^n_{++}$. Then,

$$x^\top y \leq (1/2)x^\top Qx + (1/2)y^\top Q^{-1}y$$

If f is convex and differentiable. Let x^* be maximizer of $y^\top x - f(x)$ satisfying $y = \nabla f(x^*)$. Then,

$$f^*(y) = {x^*}^\top \nabla f(x^*) - f(x^*)$$

Thus, by solving $y = f(z)$ for each y, we can obtain $f^*(y) = z^\top \nabla f(z) - f(z)$.

1. For $a > 0$ and $b \in \mathbb{R}$ the conjugate function of $g(x) = af(x) + b$ is

$$g^*(y) = af^*(y/a) - b$$

2. For a nonsingular $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ let $g(x) = f(Ax + b)$.

$$g^*(y) = f^*(A^{-\top}y) - b^\top A^{-\top}y$$

3. If $f(u, v) = f_1(u) + f_2(v)$, where f_1 and f_2. Then

$$f^*(w, z) = f_1^*(w) + f_2^*(z)$$

(proof of 2)

$$
\begin{aligned}
g^*(y) &= \sup_x \left(y^\top x - f(Ax+b)\right) \\
&= \sup_x y^\top \left(A^{-1}(Ax+b) - y^\top A^{-1}b - f(Ax+b)\right) \\
&= \sup_x \left((A^{-\top}y)^\top (Ax+b) - f(Ax+b)\right) - y^\top A^{-1}b \\
&= f^*(A^{-\top}y) - b^\top A^{-\top}y
\end{aligned}
$$

Definition 13 (Quasiconvex function)

- A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is called quasiconvex if

$$S_\alpha(f) = \{x \in \mathsf{dom}(f) : f(x) \leq \alpha\}$$

  for $\alpha \in \mathbb{R}$ is convex.
- If $-f$ is quasiconvex, then f is called quasiconcave.
- If f is quasiconvex and quasiconcave as well, then f is called quasilinear.

- If f is convex, f is quasiconvex.
- f is quasiconvex if and only if $\{x : f(x) \geq \alpha\}$ is convex.
- f is quasilinear then $\{x : f(x) = \alpha\}$ is convex.

Proposition 1 (Definition of the quasiconvex function)

$S_\alpha(f)$ *is convex if and only if*

$$f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y))$$

for $\lambda \in [0,1]$.

(proof $\rightarrow$) For arbitrary x and y, let $\alpha = \max(f(x), f(y))$. By definition of α-level set, $x, y \in S_\alpha(f)$. Since $S_\alpha(f)$ is convex, $\lambda x + (1-\lambda)y) \in S_\alpha(f)$. Thus, $f(\lambda x + (1-\lambda)y) \leq \alpha = \max(f(x), f(y))$. The converse is trivial.

Example 14

- $\log x$ on $\mathbb{R}_{++}$ is quasiconvex and quasiconcave. So it is quasilinear.
- ceil$(x) = \inf\{z \in Z : z \geq z\}$ is quasiconvex and quasiconcave.
- Linear-fractional function:

$$f(x) = \frac{a^\top x + b}{c^\top x + d}$$

with dom$(f) = \{x : c^\top x + d > 0\}$ is quasiconvex and quasiconcave.

HW

Prove the following statements.

- Support function S_C associated with the set C is defined as

$$S_C(x) = \sup\{x^\top y : y \in C\}$$

  is convex.

- Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ and define

$$f^* : y \in \mathbb{R}^n \mapsto \sup\{x^\top y - f(x)\},$$

  the conjugate function of f. Then, f^* is always convex.

HW

Prob set. Ch3

- 3.4-3.7
- 3.12, 3.13
- 3.21-3.23
- 3.26, 3.28, 3.30, 3.31
- 3.42, 3.43