Convex Optimization Problem III

Jong-June Jeon

Fall, 2023

Department of Statistics, University of Seoul

Things to know

- Convex function
- Gradient vector, Hessian matrix
- Sublevel set, Epigraph
- Equivalent definitions of convex function.

Note

In the previous chapter, we discussed the definition of a convex set as a feasible set to consider in optimization problems and explored several commonly used examples. In this chapter, we will learn about Convex functions as the objective functions in optimization problems. Similar to the previous chapter, it is essential for us to determine whether the objective function is a Convex function in a given problem. To do so, we will study various sufficient conditions for convex functions.

Definition 1 (Convex function)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and denote the domain of f by dom(f).

We call f convex function if

- dom(f) is convex
- for all $x, y \in$ dom(f) and $\theta \in [0, 1]$,

$$f(\theta x + (1 - \theta)y) \le \theta f(x) + (1 - \theta)f(y)$$

If the strict inequality holds whenever $0 < \theta < 1$, we call f strictly convex.

Note that a convex function f is continuous on the relative interior of its domain.

Example of convex functions

- Let $f : x \in \mathbb{R}^n \mapsto Ax + b \in \mathbb{R}^m$
- Let $f : x \in \mathbb{R}^n \mapsto \|x\| \in \mathbb{R}$
- $f(x) = (1/2)x^\top Px + q^\top x + r$ with $P \in \mathcal{S}^n_+$
- $f : \mathbb{R}^n \times \mathcal{S}^n \to \mathbb{R}$ defined as $f(x, Y) = x^\top Y^{-1} x$ is convex on $\mathrm{dom}(f) = \mathbb{R} \times \mathcal{S}^n_{++}$.

Some examples are proved by the definition of convexity. For some cases, the equivalent definition of a convex function is helpful.

Proposition 1

f *is convex if and only if for all* $x \in$ *dom*(f) *and all* ν, $g(t) = f(x + t\nu)$ *with* *dom*$(g) = \{t : x + t\nu \in$ *dom*$(f)\}$ *is convex.*

(proof)

($\leftarrow$) For arbitrary x_1 and x_2, we can choose $t_1, t_2 \neq 0$ such that $x_1 = x + t_1 v$ and $x_2 = x + t_2 v$

$$
\begin{aligned}
f(\lambda x_1 + (1-\lambda)x_2) &= f(x + (\lambda t_1 + (1-\lambda)t_2)v) = g(\lambda t_1 + (1-\lambda)t_2) \\
&\leq \lambda g(t_1) + (1-\lambda)g(t_2) \\
&= \lambda f(x_1) + (1-\lambda)f(x_2).
\end{aligned}
$$

$f(x) = (1/2)x^\top Px + q^\top x + r$ with $P \in \mathcal{S}_+^n$. Prove f is a convex function.

(proof) For an arbitrary $v \in \mathbb{R}^n$ let

$$g(t) = f(x + tv) = \frac{1}{2}(v^\top Pv)t^2 + (v^\top Px)t + \frac{1}{2}x^\top Px$$

Because of $P \in \mathcal{S}_+^n$, $(v^\top Pv) \geq 0$, which completes the proof.

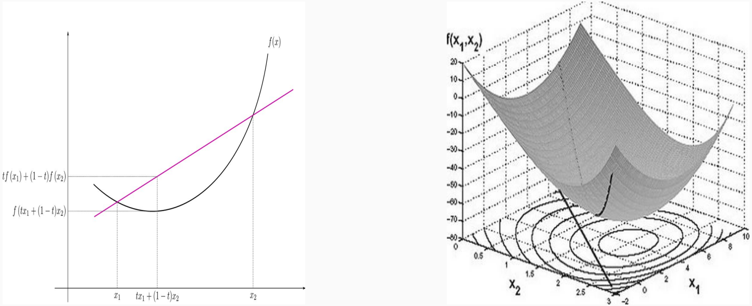

Figure 1: Convex functions on $\mathbb{R}$ and $\mathbb{R}^2$

Definition 2 (Gradient vector)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. The gradient vector of f on x is defined by

$$\nabla f(x) = \frac{\partial f(x)}{\partial x} = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2} \cdots, \frac{\partial f(x)}{\partial x_n} \right)^\top$$

Examples

- $f(x) = \frac{1}{2} x^\top A x + b^\top x + c$,
- $f(x) = -\log(1 + \exp(b^\top x + c))$

Proposition 2 (Gradient as a directional derivative)

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ *and suppose that*

$$u = \textit{argmax}_{u \in \mathbb{R}^n : \|u\| = 1} \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h}$$

exists. Then, u *is proportional to* $\nabla f(x)$*.*

(proof) For a fixed u with $\|u\| = 1$

$$\begin{aligned}\lim_{h\to 0}\frac{f(x+hu)-f(x)}{h} &= \frac{\partial}{\partial h}\left(f(x+hu)\right)\Big|_{h=0}\\ &= \nabla f(x)^\top u.\end{aligned}$$

$|\nabla f(x)^\top u| \leq \|\nabla f(x)\|\|u\| = \|\nabla f(x)\|$ by Cauchy-Schwartz inequality and the equality above holds when $u \propto \nabla f(x)$. So, the u, the direction maximizing the infinitesimal change, is proportional to $\nabla f(x)$.

The result of the directional derivatives shows that the direction of the acute angle to $\nabla f(x)$ increases the value function.

Property of the gradient vector

- Tangent space at $x^{(k)}$: $\mathcal{T}(x^{(k)}) = \{x \in \mathbb{R}^p : \nabla f(x^{(k)})^\top x = 0\}$
- Gradient vector at $x^{(k)}$ satisfies that

$$\nabla f(x^{(k)})^\top x = 0$$

  for $x \in \mathcal{T}(x^{(k)})$ by definition (orthogonal to tangent space).
- The gradient vector is the fastest direction to increase the value of the objective function at the point $x^{(k)}$.
- To decrease the value of an objective function, it is possible to move the current solution $x^{(k)}$ in the opposite direction of the gradient vector.

(fig)

Illustration of gradient vectors

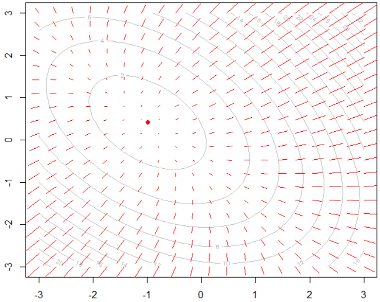

Figure 2: Contour and gradient vector of an $L(w)$: gray curve denotes the contour; the red point denotes w^*; the red lines denote the gradient vectors

Definition 3 (Hessian matrix)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. The Hessian matrix of f on x is defined by

$$\nabla^2 f(x) = \frac{\partial^2 f(x)}{\partial x^\top \partial x} = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

Hessian matrix as a linear map

The directional derivatives of $\nabla f(x)$ implies that

$$\lim_{h \to 0} \frac{\nabla f(x + hv) - \nabla f(x)}{h} = \nabla^2 f(x) v$$

- Hessian matrix represents the map to produce the directional change of $\nabla f(x)$.
- $\nabla^2 f(x) v$ is approximated by

$$\frac{\nabla f(x + hv) - \nabla f(x)}{h}$$

  for small h.
- The positive definiteness of $\nabla^2 f(x)$ implies that all angles between directional derivative of $\nabla f(x)$ and v are acute.

Property of Hessian matrix

- When f is twice differentiable, the Hessian matrix is well defined.
- Hessian matrix is a symmetric matrix.
- The Hessian matrix is always non-negative definite?
  ex) $f(x_1, x_2) = x_1^2/2 + x_2^2/2 - 4x_1x_2$
- Cf) Talyor's expansion of a twice differentiable $f : \mathbb{R}^n \mapsto \mathbb{R}$.

$$f(x) = f(x_0) + \nabla f(x_0)^\top (x - x_0) + \frac{1}{2}(x - x_0)^\top \nabla^2 f(x^*)(x - x_0),$$

where $x^* = hx + (1 - h)x_0$ is given by some $h \in [0, 1]$.

Proposition 3 (Characterizations of convex function)

All following statements are the characterization of a convex function.

1. f *is convex on* $\mathbb{R}^n$.
2. f *is convex on any restricted line of* $\mathbb{R}^n$.
3. *(First order condition): If* f *is differentiable,*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

   for all $x, y \in$ *dom*(f).

4. *(Second order condition): If* f *is twice differentiable,* $\nabla^2 f(x)$ *is non-negative definite.*

(proof) $(1) \Leftrightarrow (2)$ is shown by Proposition 1. We will show $(1) \to (3)$. Let f be a convex and differentiable function.

$f((1-h)x + hy) = f(x + h(y-x))$ for $h \in (0,1)$ and

$$\lim_{h \to 0} \frac{f(x + h(y-x)) - f(x)}{h} = \nabla f(x)^\top (y - x).$$

In addition,

$$\frac{f((1-h)x + hy)}{h} \leq \frac{(1-h)f(x) + hf(y)}{h}.$$

Thus,

$$f(y) - f(x) \geq \nabla f(x)^\top (y - x).$$

(proof) $(3) \to (1)$ Let $z = (1-h)x + hy$.

$$f(y) \geq f(z) + \nabla f(z)^\top (y - z)$$
$$f(x) \geq f(z) + \nabla f(z)^\top (x - z)$$

By multiplying h and $(1-h)$ to above inequalities respectively and adding them,

$$hf(y) + (1-h)f(x) \geq f(z),$$

which completes the proof of $(1) \Leftrightarrow (3)$.

(proof) (1) + (twice diff) $\rightarrow (4)$

Let $g(h) = f(x + hv)$ for $h \in \mathbb{R}$. Since f is convex, $g(h)$ is convex for all x and v.

$$g'(h) = \nabla f(x + hv)v$$
$$g''(h) = v^\top \nabla^2 f(x + hv)v$$

Because of $g''(0) \geq 0$, $v^\top \nabla^2 f(x)v \geq 0$

(proof) $(3) \leftarrow \nabla^2 f(x)$ is nnd.

By Taylor's expansion,

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2}(y - x)^\top \nabla^2 f(x^*)(y - x)$$

for a line segment x^* between x and y.

Since $\nabla^2 f(x)$ is nnd,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

In conclusion, we prove that (1)-(4) are equivalent definitions of the convex function.

Theorem 4 (Monotone gradient condition for convexity)

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a differentiable function. Then, f is convex if and only of

$$(\nabla f(y) - \nabla f(x))^\top (y - x) \geq 0.$$

(proof) $(\rightarrow)$

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$
$$f(x) \geq f(y) + \nabla f(y)^\top (x - y)$$

By summing the inequalities,

$$0 \geq (\nabla f(x) - \nabla f(y))^\top (y - x).$$

(proof) ($\leftarrow$)

Let $g_{a,b}(t) = f(ta + (1-t)b)$. We will show that $g_{a,b}(t)$ is convex for all a and b, which implies that f is convex.

$$g`_{a,b}(t) = \frac{\partial}{\partial t} g_{a,b}(t) = \nabla f(ta + (1-t)b)^\top (a-b).$$

$g`_{a,b}(t) - g`_{a,b}(0) \geq 0$ for all $t \in [0,1]$ because

$$\begin{aligned} g`_{a,b}(t) - g`_{a,b}(0) &= (\nabla f(ta + (1-t)b) - \nabla f(b))^\top (a-b) \\ &= \frac{1}{t}(\nabla f(ta + (1-t)b) - \nabla f(b))^\top (ta - tb) \\ &= \frac{1}{t}(\nabla f(ta + (1-t)b) - \nabla f(b))^\top (ta - tb + b - b) \\ &\geq 0. \end{aligned}$$

The inequality holds for the condition of f.

$$
\begin{aligned}
f(x) &= g_{x,y}(1) = g_{x,y}(0) + \int_0^1 g`_{x,y}(t)dt \\
&\geq g_{x,y}(0) + \int_0^1 g`_{x,y}(0)dt \\
&= f(y) + \nabla f(y)^\top (x - y)
\end{aligned}
$$

Definition 5 (Sublevel set and epigraph)

- $C_\alpha = \{x \in dom(f) : f(x) \leq \alpha\}$ is called α sublevel set.
  - If f is a convex function, its sublevel set is always convex.
  - 하지만 sublevel set 이 convex라고 해서 함수 f가 convex인 것은 아니다. ($f(x) = -\exp(x)$)
  - 참고: sublevel set은 f가 closed function을 정의할 때 사용됨! (후술함)
- Epigraph: $\{(x,t) : x \in dom(f), t \geq f(x)\}$ (Epi 라는 것이 above 라는 뜻이 있음)
  - Epigraph가 convex set 인 것과 f가 convex function 인 것은 동치임.

Proposition 4

f is convex on $\mathbb{R}^n$ if and only if the Epigraph of f is convex

(see the example 3.4 in p.76)

(proof) $(\rightarrow)$ Let $(x_1, t_1), (x_2, t_2) \in$ Epi(f).

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2) \le \lambda t_1 + (1-\lambda) t_2$$

for $\lambda \in [0, 1]$, and thus $\lambda(x_1, t_1) + (1-\lambda)(x_2, t_2) \in$ Epi(f).

$(\leftarrow)$ For x_1 and x_2 let $t_1 = f(x_1)$ and $t_2 = f(x_2)$. Since Epi(f) is convex,

$$\lambda t_1 + (1-\lambda) t_2 \ge f(\lambda x_1 + (1-\lambda) x_2)$$

for $\lambda \in [0, 1]$, which completes the proof.

Example 6 (with the second order conditions)

- $f : (x, y) \in \mathbb{R} \times \mathbb{R}_{++} \mapsto x^2/y$ is convex
- $f : x \in \mathbb{R}^n \mapsto \log(\exp(x_1) + \cdots + \exp(x_n))$ is convex.
- $f : x \in \mathbb{R}^n \mapsto (\prod_{i=1}^{n} x_i)^{1/n}$ is concave.
- $f : X \in S^n_{++} \mapsto -\log(det(X))$ is convex, where S^n_{++} is a set of $n \times n$ symmetric matrices.

(Proof) See p.73-74.

Example 7 (with epigraph condition)

- $f : (x, Y) \in \mathbb{R}^n \times S^n_{++} \mapsto x^\top Y^{-1} x$ is convex.

(Proof) See p.76.

Appendix

Cauchy Schwartz inequality

$$| < x, y > | \leq \|x\|\|y\|$$

(See Ch1. Angle and inner product.)

$$x^\top y = \|x\|\|y\| \cos\theta$$

Appendix

A Hessian matrix $H \in \mathcal{S}^n_{++}$ if and only if

$$\min_{x \in \mathbb{R}^n} \frac{x'Hx}{x'x} > 0.$$

Note that $\min_{x \in \mathbb{R}^n} \frac{x'Hx}{x'x} = \min_{\|x\|=1} x'Hx$ is the minimum eigenvalue of H.

Appendix

Closed functions

A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is said to be closed if for each $\alpha \in \mathbb{R}$, the sublevel set

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

is closed.

Note that $[a, \infty)$ is closed.