Convex Functions

Ye Shi

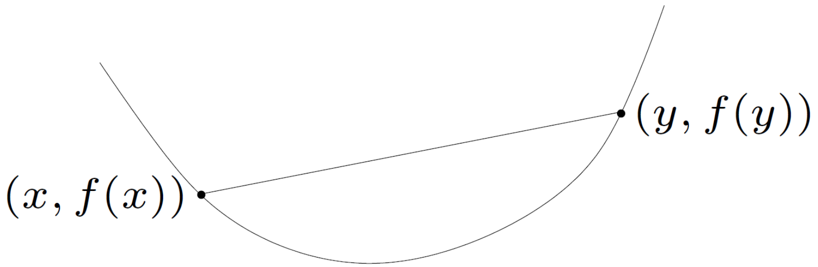
ShanghaiTech University

Outline

Definition of Convex Function

- A function $f : \mathbb{R}^n \Rightarrow \mathbb{R}$ is said to be **convex** if the domain, $\mathsf{dom} f$, is convex and for any $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{dom} f$ and $0 \leq \theta \leq 1$,

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$



- f is **strictly convex** if the inequality is strict for $0 < \theta < 1$

- f is **concave** if $-f$ is convex

Examples on $\mathbb{R}$

Convex functions:

- affine: $ax + b$ on $\mathbb{R}$
- powers of absolute value: $|x|^p$ on $\mathbb{R}$, for $p \geq 1$ (e.g., $|x|$)
- powers: x^p on $\mathbb{R}_{++}$, for $p \geq 1$ or $p \leq 0$ (e.g., x^2)
- exponential: e^{ax} on $\mathbb{R}$
- negative entropy: $x \log x$ on $\mathbb{R}_{++}$

Concave functions:

- affine: $ax + b$ on $\mathbb{R}$
- powers: x^p on $\mathbb{R}_{++}$, for $0 \leq p \leq 1$
- logarithm: $\log x$ on $\mathbb{R}_{++}$

$f(x) = x \log x$

$f'(x) = \log x + 1$

$f''(x) = \frac{1}{x} > 0, \quad x > 0$

Examples on $\mathbb{R}^n$

- **Affine functions** $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$ are convex and concave on $\mathbb{R}^n$

- **Norms** $\|\boldsymbol{x}\|$ are convex on $\mathbb{R}^n$ (e.g., $\|\boldsymbol{x}\|_\infty, \|\boldsymbol{x}\|_1, \|\boldsymbol{x}\|_2$)

- **Quadratic functions** $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} + 2\boldsymbol{q}^T\boldsymbol{x} + r$ are convex $\mathbb{R}^n$ if and only if $\boldsymbol{P} \succeq \mathbf{0}$

  $f'(x) = 2Px + 2q$, $f''(x) = P$

- The **geometric mean** $f(\boldsymbol{x}) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbb{R}^n_{++}$

- The **log-sum-exp** $f(\boldsymbol{x}) = \log \sum_i e^{x_i}$ is convex on $\mathbb{R}^n$ (it can be used to approximate $\max_{i=1,\cdots,n} x_i$)

- **Quadratic over linear:** $f(\boldsymbol{x}, y) = \boldsymbol{x}^T\boldsymbol{x}/y$ is convex on $\mathbb{R}^n \times \mathbb{R}_{++}$

$$\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

Examples on $\mathbb{R}^{n\times n}$

- **Affine functions:** (prove it!)

$$f(\boldsymbol{X}) = \mathrm{Tr}(\boldsymbol{A}\boldsymbol{X}) + b$$

are convex and concave on $\mathbb{R}^{n\times n}$

- **Logarithmic determinant function:** (prove it!)

$$f(\boldsymbol{X}) = \mathrm{logdet}(\boldsymbol{X})$$

is concave on $\mathbb{S}^n = \{\boldsymbol{X} \in \mathbb{R}^{n\times n} \mid \boldsymbol{X} \succeq \boldsymbol{0}\})$

- **Maximum eigenvalue function**: (prove it!)

$$f(\boldsymbol{x}) = \lambda_{\max}(\boldsymbol{X}) = \sup_{\boldsymbol{y}\neq\boldsymbol{0}} \frac{\boldsymbol{y}^T\boldsymbol{X}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{y}}$$

is convex on $\mathbb{S}^n$

$\lambda_1, \lambda_2 \cdots \lambda_n$

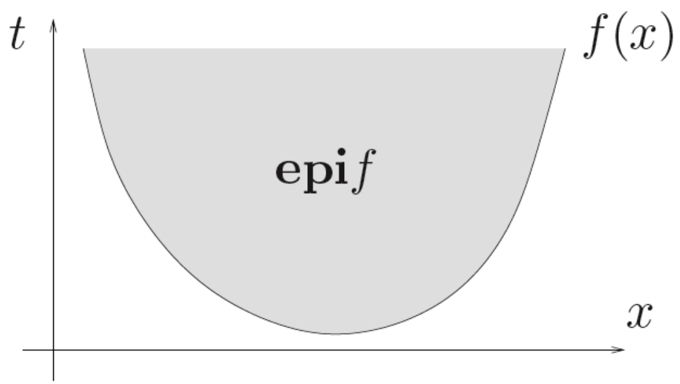
Epigraph

- The **epigraph** of f if the set

$$\operatorname{epi} f = \{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in \operatorname{dom} f,\ f(\boldsymbol{x}) \leq t\}$$

- Relation between convexity in sets and convexity in functions:

$$f \text{ is convex} \iff \operatorname{epi} f \text{ is convex}$$



$$\min_{x \in \mathbb{R}^n} \|x\|_1 \qquad \underbrace{\sum_{i=1}^{n} |x_i|}_{\partial}$$

$$|x_i| < t_i$$

$$\min \sum_{i=1}^{n} t_i$$

$$-t_i \leq x_i \leq t_i$$

$$\begin{cases} (1) & \min\ t_i \\ & -t_i \leq x_i \leq t_i \\ & \vdots \\ (n) & \min\ t_n \\ & -t_n \leq x_n \leq t_n \end{cases}$$

Outline

Restriction of a Convex Function to a Line

- $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex if and only if the function $g\colon \mathbb{R} \longrightarrow \mathbb{R}$

$$g(t) = f(\boldsymbol{x} + t\boldsymbol{v}), \quad \operatorname{dom} g = \{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \operatorname{dom} f\}$$

  is convex for any $\boldsymbol{x} \in \operatorname{dom} f$, $\boldsymbol{v} \in \mathbb{R}^n$

- In words: a function is convex if and only if it is convex when restricted to an arbitrary line.

- Implication: we can check convexity of f by checking convexity of functions of one variable!

- Example: concavity of $\log\det(\boldsymbol{X})$ follows from concavity of $\log(x)$

Example

Example: concavity of $\operatorname{logdet}(\boldsymbol{X})$:

$$\det\left(X^{\frac{1}{2}}\left(I+tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}\right)X^{\frac{1}{2}}\right)$$

$$g(t) = \operatorname{logdet}(\boldsymbol{X} + t\boldsymbol{V}) = \operatorname{logdet}(\boldsymbol{X}) + \operatorname{logdet}(\boldsymbol{I} + t\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2})$$

$$= \operatorname{logdet}(\boldsymbol{X}) + \sum_{i=1}^{n} \log(1 + t\lambda_i)$$

where λ_i's are the eigenvalues of $\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2}$.

$$\prod_{i=1}^{n}(1+t\lambda_i)$$

The function g is concave in t for any choice of $\boldsymbol{X} \succ \boldsymbol{0}$ and $\boldsymbol{V}$; therefore, f is concave.

① $\det(XY) = \det(X)\cdot\det(Y)$

② $Z = X^{-\frac{1}{2}}VX^{-\frac{1}{2}} = Q\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}Q^T$, $I+tZ = Q\begin{bmatrix}1+t\lambda_1 & & \\ & \ddots & \\ & & 1+t\lambda_n\end{bmatrix}$

$$g'(t) = \sum_{i=1}^{n} \frac{\lambda_i}{\underline{1 + t\lambda_i}}$$

$$g''(t) \leq 0$$

Outline

① $f(\theta x+(1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$, $\theta\in[0,1]$, $\forall x, y$

② $f(y) \geq f(x) + \nabla f^T(x)(y-x)$, $\forall x, y \in \mathrm{dom} f$

③ $\nabla^2 f(x) \succeq 0$. f is twice differentiable.

Proof:

$$\nabla f^T(x)(y-x) = \lim_{\theta\to 0} \frac{f(x+\theta)-f(x)}{\theta(y-x)}(y-x)$$

$$= \lim_{\theta\to 0} \frac{f(x+\theta(y-x))-f(x)}{\theta(y-x)}(y-x)$$

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

$$f(x) \geq f(y) + \nabla f^T(y)(x-y), \quad \forall x, y$$

$$d^T\nabla^2 f(x) d \geq 0, \quad \forall d$$

① ⇒ ②: $\frac{②}{\theta}$: $f(x) \geq f(y) + \frac{f(y+\theta(x-y))}{\theta} - \frac{f(y)}{\theta}$

$\Rightarrow f(x) \geq f(y) + \frac{f(y+\theta(x-y)) - f(y)}{\theta(x-y)} \cdot (x-y)$

$\lim_{\theta \to 0}$ ( ) $\Rightarrow$

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

② ⇒ ① Suppose $z = \theta x + (1-\theta) y$.

$$\Rightarrow \begin{cases} f(x) \geq f(z) + \nabla f^T(z)(x-z) & ⓐ \\ f(y) \geq f(z) + \nabla f^T(z)(y-z) & ⓑ \end{cases}$$

$\theta \cdot ⓐ + (1-\theta) \cdot ⓑ \Rightarrow$

$\theta f(x) + (1-\theta) f(y) \geq f(z) + \nabla f^T(z)(\theta x + (1-\theta) y - z)$

0

② ⇒ ③:

$$f(x+\tau d) = f(x) + \nabla^T f(x)\tau d + \frac{\tau^2}{2} d^T \nabla^2 f(x) d + o(\|\tau d\|^2)$$

By ②, we have

$$\frac{\tau^2}{2} d^T \nabla^2 f(x) d + o(\|\tau d\|^2) \geq 0$$

$$\frac{1}{2} d^T \nabla^2 f(x) d + \frac{o(\|\tau d\|^2)}{\tau^2} \geq 0$$

$$\lim_{\tau \to 0} (\quad) = d^T \nabla^2 f(x) d \geq 0$$

③ $\Rightarrow$ ② $\forall x, y.\ \exists z$

$$f(y) = \underbrace{f(x) + \nabla^T f(x)(y-x)} + \underbrace{(y-x)^T \nabla^2 f(z)(y-x)}_{\geq 0}$$

First and Second Order Conditions I

- **Gradient** (for differentiable f):

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}^T \in \mathbb{R}^n$$

- **Hessian** (for twice differentiable f):

$$\nabla^2 f(\boldsymbol{x}) = \left(\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \right)_{ij} \in \mathbb{R}^{n \times n}$$

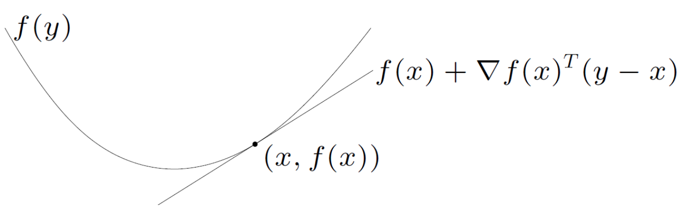
- **Taylor series**:

$$f(\boldsymbol{x} + \boldsymbol{\delta}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{\delta} + o\left(\|\boldsymbol{\delta}\|^2\right)$$

First and Second Order Conditions II

- **First-order condition:** a differentiable f with convex domain is convex if and only if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$$



- Interpretation: first-order approximation is a global under estimator

- **Second-order condition:** a twice differentiable f with convex domain is convex if and only if

$$\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0} \quad \forall \boldsymbol{x} \in \operatorname{dom} f$$

Examples

- **Quadratic function:** $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^T\boldsymbol{x} + r$(with $\boldsymbol{P} \in \mathbb{S}^n$)

$$\nabla f(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}, \qquad \nabla^2 f(\boldsymbol{x}) = \boldsymbol{P}$$

  is convex if $\boldsymbol{P} \succeq \mathbf{0}$.

- **Least-squares objective:** $f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}), \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}^T\boldsymbol{A}$$

  is convex.

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

$$\nabla^2 f(x, y) = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y & -x \end{bmatrix} \succeq \mathbf{0}$$

  is convex for $y > 0$.

Outline

Operations that Preserve Convexity I

How to establish the convexity of a given function?

- Applying the definition
- With first- or second-order conditions
- By restricting to a line
- Showing that the functions can be obtained from simple functions by operations that preserve convexity:
  - nonnegative weighted sum
  - composition with affine function (and other compositions)
  - pointwise maximum and supremum, minimization
  - perspective

Operations that Preserve Convexity II

- **Nonnegative weighted sum:** if f_1, f_2 are convex, then $\alpha_1 f_1 + \alpha_2 f_2$ is convex, with $\alpha_1, \alpha_2 \geq 0$.
- **Composition with affine functions:** if f is convex, then $f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is convex (e.g., $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|$ is convex, $\operatorname{logdet}(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{X}\boldsymbol{H}^T)$ is concave).
- **Pointwise maximum:** $f := \max\{f_1, \cdots, f_m\}$ is convex, if $f_1, \cdots, f_m$ are convex

Example: sum of r largest components of $\boldsymbol{x} \in \mathbb{R}^n$:

$$f(\boldsymbol{x}) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

where $x_{[i]}$ is the ith largest component of $\boldsymbol{x}$.

Proof: $f(\boldsymbol{x}) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} | 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$.

Operations that Preserve Convexity III

- **Pointwise supremum:** if $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $\boldsymbol{x}$ for each $\boldsymbol{y} \in \mathcal{A}$, then

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

  is convex.

Example: distance to farthest point in a set C:

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

Example: maximum eigenvalue of symmetric matrix: for $\boldsymbol{X} \in \mathbb{S}^n$,

$$\lambda_{\max}(\boldsymbol{X}) = \sup_{\boldsymbol{y} \neq \mathbf{0}} \frac{\boldsymbol{y}^T \boldsymbol{X} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{y}}$$

Operations that Preserve Convexity IV

- **Composition with scalar functions:** let $g : \mathbb{R}^n \longrightarrow \mathbb{R}$, $h : \mathbb{R} \longrightarrow \mathbb{R}$, then the function $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$ satisfies:

$$f(\boldsymbol{x}) \text{ is convex if } \begin{array}{l} g \text{ convex, } h \text{ convex nondecreasing} \\ g \text{ concave, } h \text{ convex nonincreasing} \end{array}$$

- **Minimization:** if $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $(\boldsymbol{x}, \boldsymbol{y})$ and C is a convex set, then

$$g(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in C} f(\boldsymbol{x}, \boldsymbol{y})$$

  is convex (e.g., distance to a convex set).

Example: distance to a set C:

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

is convex if C is convex.

$$f(x) = h(g(x))$$

$$f'(x) = h'(g(x)) \cdot g'(x)$$

$$f''(x) = \underbrace{h''(g(x))\,[g'(x)]^2} + \underbrace{\underset{\substack{\geq 0 \\ \leq 0}}{h'(g(x))}\;\underset{\substack{\geq 0 \\ \leq 0}}{g''(x)}}$$

$$f''(x) > 0$$

Operations that Preserve Convexity V

- **Perspective:** if $f(\boldsymbol{x})$ is convex, then its perspective

$$g(\boldsymbol{x}, t) = tf(\boldsymbol{x}/t), \quad \operatorname{dom} g = \{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} | \boldsymbol{x}/t \in \operatorname{dom} f, t > 0\}$$

  is convex.

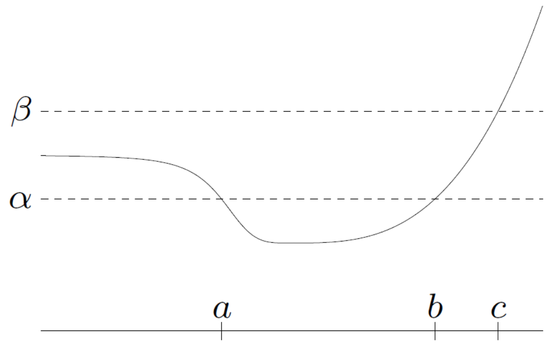
Example: $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x}$ is convex; hence $g(\boldsymbol{x}, t) = \boldsymbol{x}^T \boldsymbol{x}/t$ is convex for $t > 0$.

Example: the negative logarithm $f(\boldsymbol{x}) = -\log \boldsymbol{x}$ is convex; hence the relative entropy function $g(\boldsymbol{x}, t) = t \log t - t \log \boldsymbol{x}$ is convex on $\mathbb{R}^2_{++}$.

Outline

Quasi-Convexity Functions

- A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is quasi-convex if $\mathrm{dom}\, f$ is convex and the sublevel sets
$$S_\alpha = \{\boldsymbol{x} \in \mathrm{dom}\, f \mid f(\boldsymbol{x}) \leq \alpha\}$$
are convex for all α.



- f is quasiconcave if $-f$ is quasiconvex.

Examples

- $\sqrt{|x|}$ is quasiconvex on $\mathbb{R}$
- $\mathrm{ceil}(x) = \inf\{z \in \mathbb{Z} \mid z \geq x\}$ is quasilinear
- $\log x$ is quasilinear on $\mathbb{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbb{R}^2_{++}$
- the linear-fractional function

$$f(\boldsymbol{x}) = \frac{\boldsymbol{a}^T\boldsymbol{x} + b}{\boldsymbol{c}^T\boldsymbol{x} + d}, \qquad \mathrm{dom}\, f = \{\boldsymbol{x} \mid \boldsymbol{c}^T\boldsymbol{x} + d > 0\}$$

  is quasilinear

Log-Convexity

- A positive function f is log-concave is $\log f$ is concave:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \geq f(\boldsymbol{x})^{\theta} f(\boldsymbol{y})^{1-\theta} \quad \text{for} \quad 0 \leq \theta \leq 1$$

- f is log-convex if $\log f$ is convex.

- Example: x^a on $\mathbb{R}_{++}$ is log-convex for $a \leq 0$ and log-concave for $a \geq 0$

- Example: many common probability densities are log-concave

Convexity w.r.t. Generalized Inequalities

- $\boldsymbol{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is K-convex if $\operatorname{dom} \boldsymbol{f}$ is convex and for any $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} \boldsymbol{f}$ and $0 \leq \theta \leq 1$,

$$\boldsymbol{f}(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \preceq_K \theta \boldsymbol{f}(\boldsymbol{x}) + (1-\theta)\boldsymbol{f}(\boldsymbol{y})$$

- Example: $\boldsymbol{f} : \mathbb{S}^m \longrightarrow \mathbb{S}^m$, $\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{X}^2$ is $\mathbb{S}^m_+$-convex

Reference

Chapter 3 of:

- Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

Book:

- Petersen, Kaare Brandt, and Michael Syskind Pedersen. "The matrix cookbook." Technical University of Denmark 7 (2008): 15.