Convex Functions

Ye Shi

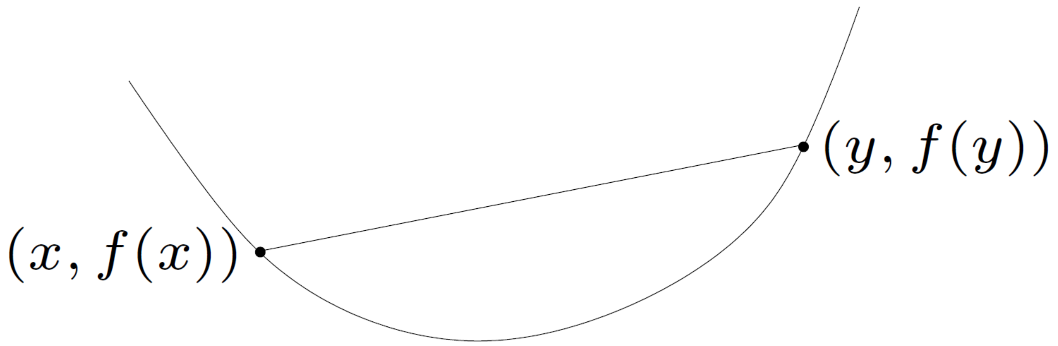
ShanghaiTech University

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 $\mathbb{R}_{++}$

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 $\boldsymbol{P} \succeq \mathbf{0}$
- 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 $\mathbb{R}^n \times \mathbb{R}_{++}$

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!)

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

$$\lambda = \frac{y^T x y}{y^T y}$$

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

$$y^T X y = y^T \lambda y = \lambda \underbrace{y^T y} \Rightarrow$$

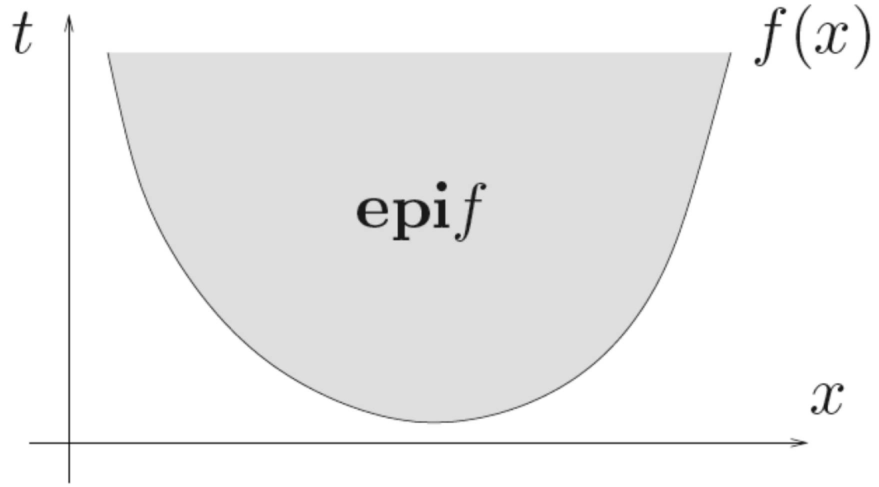
Epigraph

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

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

- Relation between convexity in sets and convexity in functions:

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



Outline

1. Definition of Convex Function
2. Restriction of a Convex Function to a Line
3. First and Second Order Conditions
4. Operations that Preserve Convexity
5. Quasi-Convexity, Log-Convexity, and Convexity w.r.t. Generalized Inequalities

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 $\operatorname{logdet}(\boldsymbol{X})$ follows from concavity of $\log(x)$

Example

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

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

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

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

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

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

$$g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1+t\lambda_i)^2} \leq 0$$

$$X + tV = X^{\frac{1}{2}}(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})X^{\frac{1}{2}}$$

① $\det(XY) = \det(X)\cdot\det(Y)$

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

$$I + tZ = Q\begin{bmatrix}1+t\lambda_1 & & \\ & \ddots & \\ & & 1+t\lambda_n\end{bmatrix}Q^T$$

$$\log\det(I + tZ) = \sum_{i=1}^{n}\log(1 + t\lambda_i)$$

Outline

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):

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

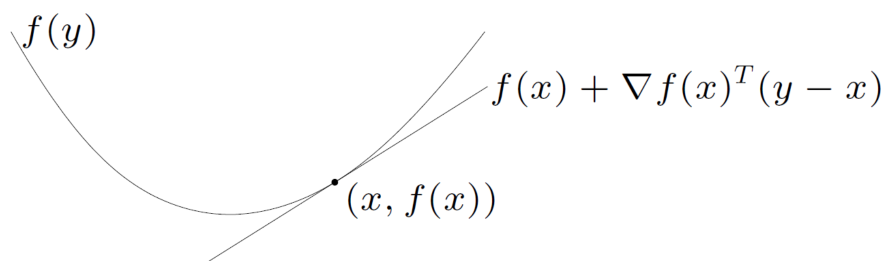
❧ **Taylor series**:

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

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

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

Convex conditions

① $f(\theta x+(1-\theta)y) \le \theta f(x)+(1-\theta)f(y)$, $0\le\theta\le 1$, $\forall x, y$.

② $f(y) \ge f(x)+\nabla^T f(x)(y-x)$, $\forall x, y$

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

Proof: ① $\Rightarrow$ ② : $\frac{①}{\theta}$: $f(x) \ge f(y)+\frac{f(\theta x+(1-\theta)y)-f(y)}{\theta}$

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

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

$$\lim_{\theta\to 0} f(x) \ge \lim_{\theta\to 0}(\quad\quad)$$

$$\Rightarrow f(x) \ge \nabla^T f(y)(x-y)$$

② $\Rightarrow$ ① : $z=\theta x+(1-\theta)y$.

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

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

$$\theta f(x) + (1-\theta) f(y) \geq f(z) + \nabla^T f(z) (\underbrace{\theta x + (1-\theta) y - z}_{0})$$
$$= f(z)$$

② ⇒ ③. Taylor Series.

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

② $f(y) \geq f(x) + \nabla^T f(x) (y-x)$

$y = x+\tau d$

$$f(x+\tau d) \geq f(x) + \nabla^T f(x) \cdot \tau d, \quad \forall d$$

$$f(x+\tau d) - f(x) - \nabla^T f(x) \tau d \geq 0, \quad \forall d$$

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

$$\lim_{\tau \to 0} (\;) \Rightarrow d^T \nabla^2 f(x) d \geq 0 \Rightarrow \nabla^2 f(x) \succeq 0$$
$\forall d$.

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

$$f(y) = f(x) + \nabla f(x)^T (y-x) + (y-x)^T \nabla^2 f(z) (y-x).$$

since $\nabla^2 f(z) \succeq 0$

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

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 $\boldsymbol{P} \succeq \boldsymbol{0}$.

- **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. Matrix Cookbook

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

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

  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

nonconvex $f_1 - f_2$ difference of convex.

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, $\log\det(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{X}\boldsymbol{H}^T)$ 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: $f(\boldsymbol{x}) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} | 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$.

$$f(x) = \max\{f_1(x), \cdots, f_n(x)\}$$

$$f(\theta x + (1-\theta) y) = \max\{f_1(\theta x + (1-\theta) y), \cdots, f_2(\theta x + (1-\theta) y)\}$$

$$\leq \max\{\theta f_1(x) + (1-\theta) f(y), \cdots, \theta f_n(x) + (1-\theta) f_n(y)\}$$

$$\leq \theta \max\{f_1(x), \cdots, f_n(x)\} + (1-\theta) \max\{f_1(y), \cdots, f_n(y)\}$$

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

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 $\boldsymbol{X} \in \mathbb{S}^n$,

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

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).

$$z = \begin{bmatrix} x \\ y \end{bmatrix}$$

Example: distance to a set C:

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

$$\underbrace{\begin{bmatrix} 1 & -1 \end{bmatrix}}\underbrace{\begin{bmatrix} x \\ y \end{bmatrix}} = x - y$$

is convex if C is convex.

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

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

$$f''(x) = h''(g(x))[g'(x)]^2 + h'(g(x))\, g''(x)$$

Operations that Preserve Convexity V

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

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

is convex.

$$t\,(x/t)^T \cdot (x/t) = \frac{x^T x}{t}$$

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 $g(\boldsymbol{x}, t) \doteq t\log t - t\log \boldsymbol{x}$ is convex on $\mathbb{R}^2_{++}$.

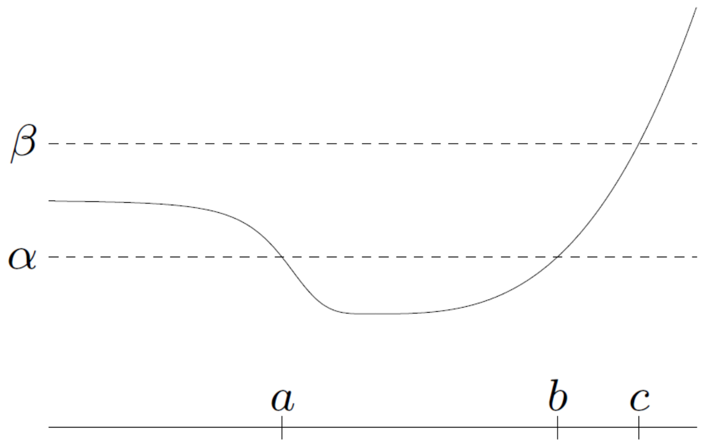
Outline

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 $\mathbb{R}$
- $\operatorname{ceil}(x) = \inf\{z \in \mathbb{Z} \mid z \geq x\}$ 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

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

  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 $\mathrm{dom}\,\boldsymbol{f}$ is convex and for any $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\,\boldsymbol{f}$ 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.