Optimization Lecture 3+4

Qingfu Zhang

Dept of CS
CityUHK

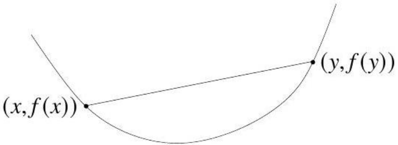
Outline

Outline

Definition

- $f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and for all $x, y \in \mathbf{dom}\, f, 0 \le \theta \le 1$,

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



- f is concave if $-f$ is convex
- f is **strictly** convex if $\mathbf{dom}\, f$ is convex and for $x, y \in \mathbf{dom}\, f, x \neq y, 0 < \theta < 1$,
$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$$

can be generated to other spaces.

Restriction of a convex function to a line

- $f : \mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function $g : \mathbf{R} \to \mathbf{R}$,

$$g(t) = f(x + tv), \quad \mathbf{dom}\, g = \{t \mid x + tv \in \mathbf{dom}\, f\}$$

  is convex in t for any $x \in \mathbf{dom}\, f$, $v \in \mathbf{R}^n$

can be used to check convexity of f by checking convexity of functions of one variable

First-order condition

- f is **differentiable** if **dom** f is open and the gradient

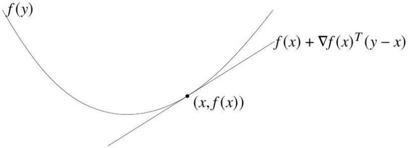
$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^T \in \mathbf{R}^n$$

  exists at each $x \in \mathbf{dom}\, f$

- **1st-order condition**: differentiable f with convex domain is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \mathbf{dom}\, f$$

- first order Taylor approximation of convex f is a **global underestimator** of f



Proof ($m = 1$)

- Necessary: For any $t \in (0, 1)$, and $x, y \in domf$. By def:

$$f(x + t(y - x)) \leq (1 - t)f(x) + tf(y)$$

$$f(y) \geq f(x) + \frac{f(x + t(y - x)) - f(x)}{t}$$

  Let $t \to 0$, we have $f(y) \leq f(x) + f^{'}(x)(y - x)$

- Sufficiency: Let $z = \theta x + (1 - \theta)y$:

$$f(x) \geq f(z) + f^{'}(z)(x - z)$$

  and

$$f(y) \geq f(z) + f^{'}(z)(y - z)$$

$$\Rightarrow ??$$

Second-order conditions

- f is **twice differentiable** if $\mathbf{dom}\, f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \ldots, n$$

  exists at each $x \in \mathbf{dom}\, f$

- **2nd-order conditions**: for twice differentiable f with convex domain
  - f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbf{dom}\, f$
  - if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

Extended-value extension

- suppose f is convex on $\mathbf{R}^n$, with domain $\mathbf{dom}\, f$
- its extended-value extension $\tilde{f}$ is function $\tilde{f} : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathbf{dom}\, f \\ \infty & x \notin \mathbf{dom}\, f \end{cases}$$

- often simplifies notation; for example, the condition

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad \tilde{f}(\theta x + (1-\theta)y) \leq \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y)$$

  (as an inequality in $\mathbf{R} \cup \{\infty\}$), means the same as the two conditions
  - $\mathbf{dom}\, f$ is convex
  -

$$x, y \in \mathbf{dom}\, f, 0 \leq \theta \leq 1 \quad \Longrightarrow$$

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

Examples on $\mathbf{R}$

convex functions:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- exponential: e^{ax}, for any $a \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on $\mathbf{R}$, for $p \geq 1$
- positive part (relu): $\max\{0, x\}$

concave functions:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$ (by def)
- powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$
- entropy: $-x \log x$ on $\mathbf{R}_{++}$
- negative part: $\min\{0, x\}$ (by def)

Examples on $\mathbf{R}^n$

convex functions:

- affine functions: $f(x) = a^T x + b$
- any norm, e.g., the ℓ_p norms
- $\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$ for $p \geq 1$
- $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$
- sum of squares: $\|x\|_2^2 = x_1^2 + \cdots + x_n^2$
- max function: $\max(x) = \max\{x_1, x_2, \ldots, x_n\}$
- softmax or log-sum-exp function: $\log(\exp x_1 + \cdots + \exp x_n)$ (ex in class)

Examples on $\mathbf{R}^{m \times n}$

- $X \in \mathbf{R}^{m \times n}$ ($m \times n$ matrices) is the variable
- general affine function has form

$$f(X) = \mathrm{tr}\left(A^T X\right) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$$

  for some $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}$
- spectral norm (maximum singular value) is convex

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = \left(\lambda_{\max}\left(X^T X\right)\right)^{1/2}$$

- log-determinant: for $X \in \mathbf{S}^n_{++}$, $f(X) = \log\det X$ is concave

Example

- $f : \mathbf{S}^n \to \mathbf{R}$ with $f(X) = \log\det X$, $\mathbf{dom}\, f = \mathbf{S}^n_{++}$
- consider line in $\mathbf{S}^n$ given by $X + tV$, $X \in \mathbf{S}^n_{++}$, $V \in \mathbf{S}^n$, $t \in \mathbf{R}$

$$
\begin{aligned}
g(t) &= \log\det(X + tV) \\
&= \log\det\left(X^{1/2}\left(I + tX^{-1/2}VX^{-1/2}\right)X^{1/2}\right) \\
&= \log\det X + \log\det\left(I + tX^{-1/2}VX^{-1/2}\right) \\
&= \log\det X + \sum_{i=1}^{n} \log\left(1 + t\lambda_i\right)
\end{aligned}
$$

where λ_i are the eigenvalues of $X^{-1/2}VX^{-1/2}$

- g is concave in t (for any choice of $X \in \mathbf{S}^n_{++}$, $V \in \mathbf{S}^n$); hence f is concave

Examples

- **quadratic function**: $f(x) = (1/2)x^T Px + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P$$

  convex if $P \succeq 0$ (concave if $P \preceq 0$)

- **least-squares objective**: $f(x) = \|Ax - b\|_2^2$

$$\nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^T A$$

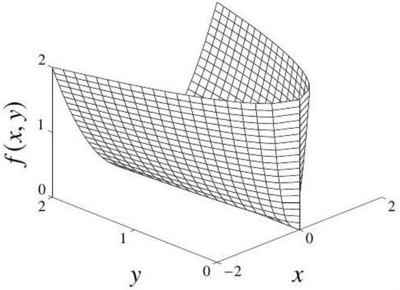
  convex (for any A)

Examples (continued)

- **quadratic-over-linear**: $f(x, y) = x^2/y$, $y > 0$

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

convex for $y > 0$



More examples

- **log-sum-exp**: $f(x) = \log \sum_{k=1}^{n} \exp x_k$ is convex

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z} \,\mathbf{diag}(z) - \frac{1}{(\mathbf{1}^T z)^2} z z^T \qquad (z_k = \exp x_k)$$

- to show $\nabla^2 f(x) \succeq 0$, we must verify that $v^T \nabla^2 f(x) v \geq 0$ for all v :

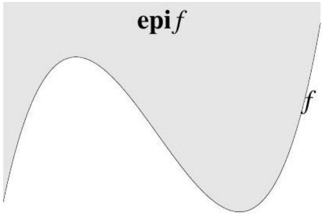
$$v^T \nabla^2 f(x) v = \frac{\left(\sum_k z_k v_k^2\right)\left(\sum_k z_k\right) - \left(\sum_k v_k z_k\right)^2}{\left(\sum_k z_k\right)^2} \geq 0$$

since $\left(\sum_k v_k z_k\right)^2 \leq \left(\sum_k z_k v_k^2\right)\left(\sum_k z_k\right)$ (from Cauchy-Schwarz inequality)

- **geometric mean**: $f(x) = \left(\prod_{k=1}^{n} x_k\right)^{1/n}$ on $\mathbf{R}_{++}^n$ is concave (similar proof as above)

Epigraph and sublevel set

- α-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$ is
  $C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \le \alpha\}$
- sublevel sets of convex functions are convex sets (but converse is false)
- **epigraph** of $f : \mathbf{R}^n \to \mathbf{R}$ is
  $\mathbf{epi}\, f = \left\{(x,t) \in \mathbf{R}^{n+1} \mid x \in \mathbf{dom}\, f, f(x) \le t\right\}$



- f is convex if and only if $\mathbf{epi}\, f$ is a convex set

Jensen's inequality

- **basic inequality**: if f is convex, then for $x, y \in \textbf{dom}\, f, 0 \le \theta \le 1$,

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

- **extension**: if f is convex and z is a random variable on $\textbf{dom}\, f$,

$$f(\mathbf{E}z) \le \mathbf{E}f(z)$$

- basic inequality is special case with discrete distribution

$$\textbf{prob}(z = x) = \theta, \quad \textbf{prob}(z = y) = 1 - \theta$$

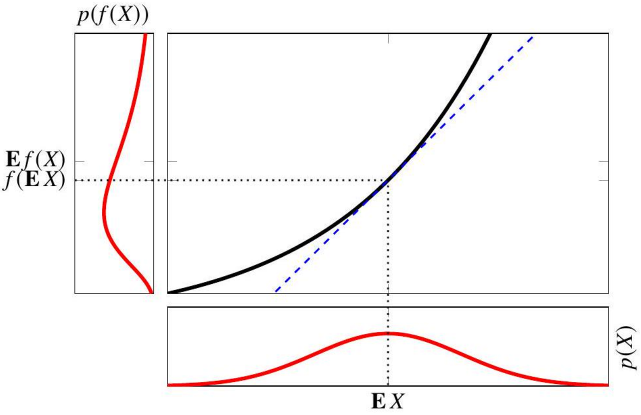
Example: log-normal random variable

- suppose $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$
- with $f(u) = \exp u$, $Y = f(X)$ is log-normal
- we have $\mathbf{E}f(X) = \exp\left(\mu + \sigma^2/2\right)$
- Jensen's inequality is

$$f(\mathbf{E}X) = \exp\mu \leq \mathbf{E}f(X) = \exp\left(\mu + \sigma^2/2\right)$$

which indeed holds since $\exp\sigma^2/2 > 1$

Example: log-normal random variable



Outline

Showing a function is convex

methods for establishing convexity of a function f

1. verify definition (often simplified by restricting to a line)
2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
   - recommended only for **very simple** functions
3. show that f is obtained from simple convex functions by operations that preserve convexity
   - nonnegative weighted sum
   - composition with affine function
   - pointwise maximum and supremum
   - composition
   - minimization
   - perspective

you'll mostly use methods 2 and 3

Nonnegative scaling, sum, and integral

- **nonnegative multiple**: αf is convex if f is convex, $\alpha \geq 0$
- **sum**: $f_1 + f_2$ convex if f_1, f_2 convex
- **infinite sum**: if $f_1, f_2, \ldots$ are convex functions, infinite sum $\sum_{i=1}^{\infty} f_i$ is convex
- **integral**: if $f(x, \alpha)$ is convex in x for each $\alpha \in \mathcal{A}$, then $\int_{\alpha \in \mathcal{A}} f(x, \alpha) d\alpha$ is convex

- there are analogous rules for concave functions

Composition with affine function

(pre-)composition with affine function: $f(Ax+b)$ is convex if f is convex

examples

- log barrier for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log\left(b_i - a_i^T x\right),$$

$$\mathbf{dom}\, f = \left\{x \mid a_i^T x < b_i, i = 1, \ldots, m\right\}$$

- norm approximation error: $f(x) = \|Ax - b\|$ (any norm)

Pointwise maximum

if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

examples:

- piecewise-linear function: $f(x) = \max_{i=1,\ldots,m}\left(a_i^T x + b_i\right)$
- sum of r largest components of $x \in \mathbf{R}^n$:

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

$\left(x_{[i]}\right.$ is i th largest component of $\left.x\right)$
Proof:

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$$

(all the possible combinations of r different components of x)

Pointwise supremum

if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, then $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex

examples

- distance to farthest point in a set C : $f(x) = \sup_{y \in C} \|x - y\|$
- maximum eigenvalue of symmetric matrix: for $X \in \mathbf{S}^n, \lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^T X y$ is convex
- support function of a set C : $S_C(x) = \sup_{y \in C} y^T x$ is convex

Partial minimization

the function $g(x) = \inf_{y \in C} f(x, y)$ is called the **partial minimization** of f (w.r.t. y)

- if $f(x, y)$ is convex in (x, y) and C is a convex set, then partial minimization g is convex

Pf: special case when there exists a y_x such that $g(x) = f(x, y_x)$

examples

- $f(x, y) = x^T Ax + 2x^T By + y^T Cy$ with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \quad C \succ 0$$

  minimizing over y gives
  $g(x) = \inf_y f(x, y) = x^T \left(A - BC^{-1}B^T\right) x$
  g is convex, hence Schur complement $A - BC^{-1}B^T \succeq 0$

- distance to a set: $\mathbf{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex.
  $(x - y = (I, -I)??.)$

Composition with scalar functions

- composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$ is $f(x) = h(g(x))$ (written as $f = h \circ g$)
- composition f is convex if
  - g convex, h convex, $\tilde{h}$ nondecreasing
  - or g concave, h convex, $\tilde{h}$ nonincreasing

  (monotonicity must hold for extended-value extension $\tilde{h}$)
- proof (for $n = 1$, differentiable g, h)

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

examples

- $f(x) = \exp g(x)$ is convex if g is convex
- $f(x) = 1/g(x)$ is convex if g is concave and positive

General composition rule

- composition of $g : \mathbf{R}^n \to \mathbf{R}^k$ and $h : \mathbf{R}^k \to \mathbf{R}$ is
  $f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x))$
- f is convex if h is convex and for each i one of the following holds
  - g_i convex, $\tilde{h}$ nondecreasing in its i th argument
  - g_i concave, $\tilde{h}$ nonincreasing in its i th argument
  - g_i affine
- you will use this composition rule **constantly** throughout this course
- you need to commit this rule to memory

Examples

- $\log \sum_{i=1}^{m} \exp g_i(x)$ is convex if g_i are convex
- $f(x) = p(x)^2/q(x)$ is convex if
  - p is nonnegative and convex
  - q is positive and concave

  (Noting x^2/t is convex of t, x).
- composition rule subsumes others, e.g.,
  - αf is convex if f is, and $\alpha \geq 0$
  - sum of convex (concave) functions is convex (concave)
  - max of convex functions is convex
  - min of concave functions is concave

Outline

Perspective

- the **perspective** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$,

$$g(x, t) = tf(x/t), \quad \mathbf{dom}\, g = \{(x, t) \mid x/t \in \mathbf{dom}\, f, t > 0\}$$

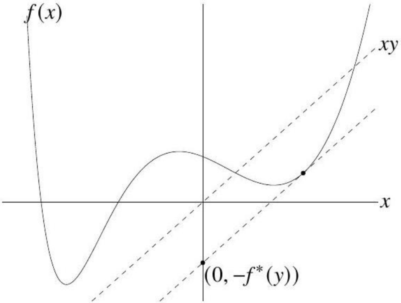
- g is convex if f is convex

examples

- $f(x) = x^T x$ is convex; so $g(x, t) = x^T x/t$ is convex for $t > 0$
- $f(x) = -\log x$ is convex; so relative entropy $g(x, t) = t \log t - t \log x$ is convex on $\mathbf{R}^2_{++}$

Conjugate function

- the **conjugate** of a function f is $f^*(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x) \right)$



- f^* is convex (even if f is not)
- very important concept.

Examples

- negative logarithm $f(x) = -\log x$

$$f^*(y) = \sup_{x>0}(xy + \log x) = \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \text{otherwise} \end{cases}$$

- strictly convex quadratic, $f(x) = (1/2)x^T Q x$ with $Q \in \mathbf{S}^n_{++}$

$$f^*(y) = \sup_x \left(y^T x - (1/2)x^T Q x \right) = \frac{1}{2} y^T Q^{-1} y$$

Basic Properties

Fenchel Inequality

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

Examples:

- $f(x) = \frac{1}{2}x^T Qx$ where $Q \in S^n_{++}$.

$$x^T y \leq \frac{1}{2}x^T Qx + \frac{1}{2}y^T Q^{-1} y$$

- $1/p + 1/q = 1,\ p, q > 1, a, b \in R$

$$ab \leq |a|^p/p + |b|^q/q$$

- $f(x) = a^T x + b, x \in R^n$.

$$f^*(y) = \begin{cases} -b & y = a \\ +\infty & otherwise \end{cases}$$

$$f^{**}(x) = a^T x + b \ for\ x \in R^n$$

Conjugate of the Conjugate

If $f(x)$ is convex and $dom(f) = R^n$, Then

$$f^{**} = f$$

outline of pf:

- 1: $f(x) = \sup\{g(x) | g \textit{ affine}, \ g(z) \leq f(z) \textit{ for all } z\}$
- 2: If $f(x) \geq g(x)$ for all x then $f^*(y) \leq g^*(y)$ for all y (by def) then $f^{**}(x) \geq g^{**}(x)$ for all x.
- 3: 1+2 $\Rightarrow$
  $f^{**}(x) \geq \sup\{g(x) | g \textit{ affine}, \ g(z) \leq f(z) \textit{ for all } z\} = f(x)$.
- 4. $f(x) \geq f^{**}(x)$
  (why? noting that $f(x) \geq x^T y - f^*(y) \textit{ for all } y \Rightarrow f(x) \geq f^{**}(x) = \sup_y (x^T y - f^*(y))$

1+4 $\Rightarrow f^{**}(x) = f(x)$ for all $x \in R^n$.

Differentiable Convex function

Suppose f is convex and differentiable with $dom(f) = R^n$.
$y^* = \nabla f(x^*)$ for any x^*. Then:

$$f^*(y^*) = {x^*}^T y^* - f(x^*)$$

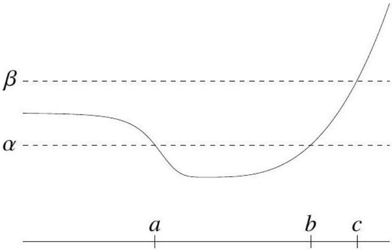
Outline

Quasiconvex functions

- $f : \mathbf{R}^n \to \mathbf{R}$ is **quasiconvex** if $\mathbf{dom}\, f$ is convex and the sublevel sets

$$S_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

are convex for all α



- f is **quasiconcave** if $-f$ is quasiconvex
- f is **quasilinear** if it is quasiconvex and quasiconcave

Examples

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

$$f(x) = \frac{a^T x + b}{c^T x + d}, \quad \mathbf{dom}\, f = \left\{x \mid c^T x + d > 0\right\}$$

is quasilinear

Example: Internal rate of return

- cash flow $x = (x_0, \ldots, x_n)$; x_i is payment in period i (to us if $x_i > 0$)
- we assume $x_0 < 0$ (i.e., an initial investment) and $x_0 + x_1 + \cdots + x_n > 0$
- **net present value** (NPV) of cash flow x, for interest rate r, is $\mathrm{PV}(x, r) = \sum_{i=0}^{n} (1+r)^{-i} x_i$
- **internal rate of return** (IRR) is smallest interest rate for which $\mathrm{PV}(x, r) = 0$:

$$\mathrm{IRR}(x) = \inf\{r \geq 0 \mid \mathrm{PV}(x, r) = 0\}$$

- IRR is quasiconcave: superlevel set is intersection of open halfspaces

$$\mathrm{IRR}(x) \geq R \quad \Longleftrightarrow \quad \sum_{i=0}^{n} (1+r)^{-i} x_i > 0 \text{ for } 0 \leq r < R$$

Properties of quasiconvex functions

- **modified Jensen inequality**: for quasiconvex f

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\}$$

  How to prove it?

- **first-order condition**: differentiable f with convex domain is quasiconvex if and only if

$$f(y) \leq f(x) \quad \Longrightarrow \quad \nabla f(x)^T(y-x) \leq 0$$

- **sum** of quasiconvex functions is not necessarily quasiconvex

Outline

- Convex (QusiConvex) functions.
- How to prove a function is convex.
- Conjugate function and its properties.