Optimization Theory and Algorithms

Instructor: Prof. LIAO, Guocheng （廖国成）

Email: liaogch6@mail.sysu.edu.cn

School of Software Engineering
Sun Yat-sen University

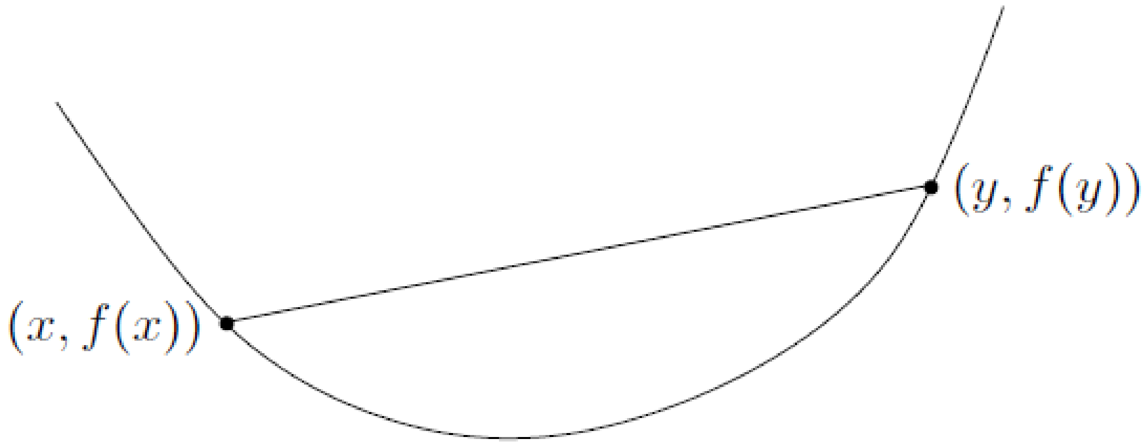
Outline

- Convex function
- Equivalent characterization
- Examples of convex function
- Properties of convex function
- Convexity-preserving operations

Convex function

A function f: $\mathbb{R}^n \to \mathbb{R}$ is convex if **dom** f is a convex set and for all x, y, and $0 \leq \theta \leq 1$:

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



Function lies below the line segment between $(x, f(x))$ and $(y, f(y))$.

Strict convex : for all $x \neq y$, and $0 < \theta < 1$:

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

A function f is concave if $-f$ is convex.

Examples

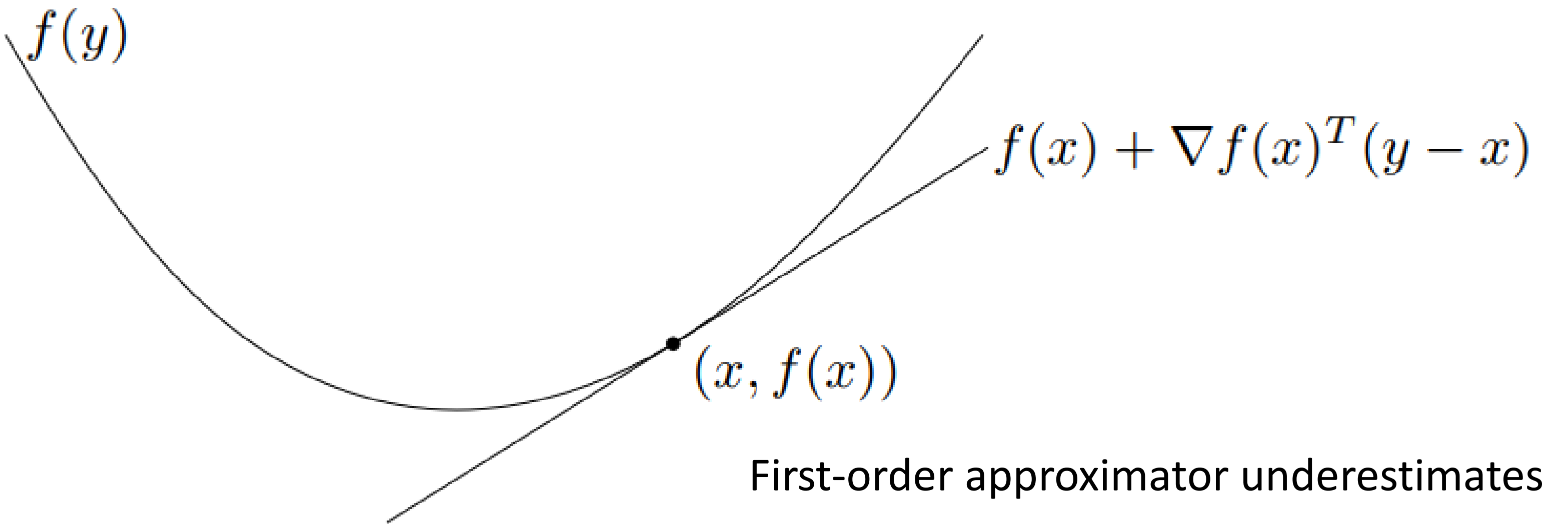
- Univariate function:
  - Exponential function $f(x) = e^{ax}$, for any a
  - Logarithmic function $f(x) = \log x$ over $\mathbb{R}_{++}$
  - Quadratic function $f(x) = ax^2 + bx + c$, $a > 0$
  - Power function $f(x) = x^a$ is convex for $a \geq 1$ or $a \leq 0$, and is concave for $0 \leq a \leq 1$.
- Affine function: $f(x) = Ax + b$
- Quadratic function $f(x) = x^T Q x + q^T x$, where $Q \succcurlyeq 0$
- Squared loss $f(x) = ||Ax - b||_2^2$ ($A^T A$ is always positive semidefinite)
- Norm: $\|x\|_p \triangleq (\sum_{i=1}^{n} |x_i|^p)^{1/p}$, where $p \geq 1$

First-order characterization

Suppose f is differentiable. Then f is convex *if and only if* **dom** f is convex and

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

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n$$



First-order approximator underestimates f

Second-order characterization

Suppose f is twice differentiable. Then f is convex *if and only if* **dom** f is convex and its Hessian matrix is positive semidefinite:

$$\nabla^2 f(x) \succcurlyeq 0, \text{for all } x \in \textbf{dom}\, f.$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & & \frac{\partial^2 f}{\partial^2 x_n} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Strict convex $\Leftrightarrow \nabla^2 f(x) \succ 0$

Concave $\Leftrightarrow \nabla^2 f(x) \preccurlyeq 0$

Quadratic function: $f(x) = \frac{1}{2} x^T Q x + q^T x$

$\nabla f(x) = Qx + q$ $\qquad \nabla^2 f(x) = Q$ $\qquad f(x)$ is convex $\Leftrightarrow Q \succcurlyeq 0$

Squared loss: $f(x) = ||Ax - b||_2^2$

$\nabla f(x) = 2A^T(Ax - b)$ $\qquad \nabla^2 f(x) = 2A^T A$ $\qquad f(x)$ is convex

Examples

Univariate function:

- Exponential function $f(x) = e^{ax}$, for any a. $f''(x) = a^2 e^{ax} > 0$
- Logarithmic function $f(x) = \log x$ over $\mathbb{R}_{++}$. $f''(x) = -\frac{1}{x^2} < 0$
- Quadratic function $f(x) = ax^2 + bx + c$, $a > 0$. $f''(x) = a > 0$
- Power function $f(x) = x^a$ is convex for $a \geq 1$ or $a \leq 0$, and is concave for $0 \leq a \leq 1$.

  $f''(x) = a(a-1)x^{a-2}$
- Negative entropy $f(x) = x \log x$. $f'(x) = \log x + 1$ $f''(x) = \frac{1}{x} > 0$

Examples

- Quadratic-over-linear function: $f(x,y) = x^2/y$, where $y > 0$.

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

- Norm: $f(x) = \|x\|_p \triangleq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, where $p \geq 1$

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

- Max: $f(x) = \max_i x_i$

$$\begin{aligned} f(\theta x + (1-\theta)y) = \max_i(\theta x_i + (1-\theta)y_i) &\leq \theta \max_i x_i + (1-\theta)\max_i y_i \\ &= \theta f(x) + (1-\theta)f(x) \end{aligned}$$

Restriction to a line

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

$$g(t) = f(x + ty), \textbf{dom}\ g = \{t | x + ty \in \textbf{dom}\ f\}$$

is convex in t for any $x \in \textbf{dom}\ f$.

Example: Log-determinant. $f: \mathbf{S}_{++}^n \to \mathbb{R}$, $f(X) = \log\det X$ is concave

$$\begin{aligned} g(t) = f(X + tV) &= \log\det(X + tV) \\ &= \log\det(X^{1/2} I X^{1/2} + tX^{1/2} X^{-1/2} V X^{-1/2} X^{1/2}) \\ &= \log\det(X^{1/2}(I + tX^{-1/2} V X^{-1/2}) X^{1/2}) \\ &= 2\log\det(X^{1/2}) + \log\det(I + tX^{-1/2} V X^{-1/2}) \\ &= \log \textstyle\prod_{i=1}^n (1 + t\boxed{\lambda_i}) + 2\log\det(X^{1/2}) \\ &= \textstyle\sum_{i=1}^n \log(1 + t\lambda_i) + 2\log\det(X^{1/2}) \end{aligned}$$

Eigenvalue of $X^{-1/2} V X^{-1/2}$

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

Jensen’s inequality

Basic: if f is convex, then for any $0 \le \theta \le 1$:

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

Extension to more than two points: if f is convex, then for any $\theta_1, \ldots, \theta_k \in [0,1]$ such that $\sum_{i=1}^{k} \theta_i = 1$

$$f\left(\sum_{i=1}^{k} \theta_i x_i\right) \le \sum_{i=1}^{k} \theta_i f(x_i)$$

Expression with expectation: if f is convex, then:

$$f(\mathrm{E}[x]) \le \mathrm{E}[f(x)]$$

Example: $\frac{a+b}{2} \ge \sqrt{ab}$ $\quad f(x) = -\log x$ is convex $\implies -\log \frac{a+b}{2} \le -\frac{1}{2}(\log a + \log b)$

i.e., $-\frac{a+b}{2} \ge -\sqrt{ab}$

Sublevel set and epigraph:

$t -$ sublevel set of a function $f: \mathbb{R}^n \to \mathbb{R}$:

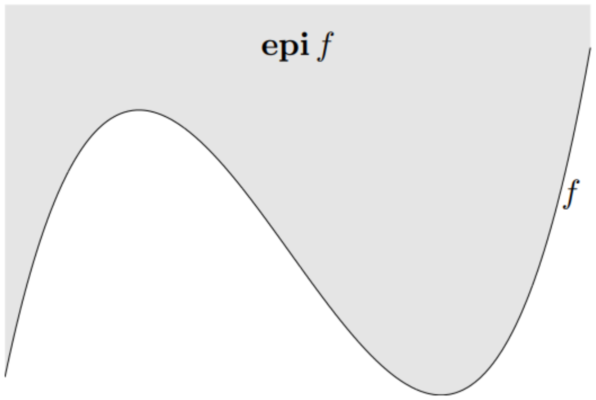
$$C_t = \{x \in \mathbf{dom}\ f | f(x) \leq t\}$$

Sublevel sets of a convex function are convex sets (inverse is false).

A function is quasi-convex if all its sublevel sets are convex.

Epigraph of a function $f: \mathbb{R}^n \to \mathbb{R}$:

$$\textbf{epi}\ f = \{(x, t) \in \mathbb{R}^{n+1} |\ x \in \mathbf{dom}\ f, f(x) \leq t\}$$



f is convex if and only if **epi** f is a convex set

Convexity-preserving operations

- Non-negative weighted sum: let $f_1,\ldots,f_m$ be convex functions and $\alpha_1,\ldots,\alpha_m \in \mathbb{R}$ be non-negative values. Then the following function is convex:

$$f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$$

- Affine composition: let f be a convex function. Then $g(x) \triangleq f(Ax + b)$ is convex

  E.g., $f(x) = ||Ax + b||_2^2$ is convex

Convexity-preserving operations

Composition: let f be the composition of $g\colon \mathbb{R}^n \to \mathbb{R}$ and $h\colon \mathbb{R} \to \mathbb{R}$,

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

- g is convex and h is non-decreasing: then f is convex.
- g is concave and h is non-increasing: then f is convex.

Examples:

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

Convexity-preserving operations

Pointwise maximum: if $f_1,\ldots,f_m$ are convex function, then

$$f(x) = \max\{f_1(x),\ldots,f_m(x)\}$$

is convex

Examples:

- Piecewise-linear function: $f(x) = \max_i (a_i^T x + b_i)$ is convex.

- Sum of r largest components of $x \in \mathbb{R}^n$: let$x_{[1]} \leq x_{[2]} \leq \cdots \leq x_{[n]}$. Then

$$f(x) = \sum_{i=1}^{r} x_{[i]}$$

  is convex.

$$\max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} | 1 \leq i_1 \leq i_2 \leq \cdots \leq i_r\}$$

Sum of any r different components of x is convex function
$\Longrightarrow$ pointwise maximum of them is convex

Convexity-preserving operations

If $f(x, y)$ is convex in x for each $y \in C$, then

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

is convex

Examples:

- Distance to the farthest point of a set: $f(x) = \sup_{y \in S} ||x - y||$

- Support function of a set: $S_c(x) = \sup_{y \in C} y^T x$

How to verify convexity of a function

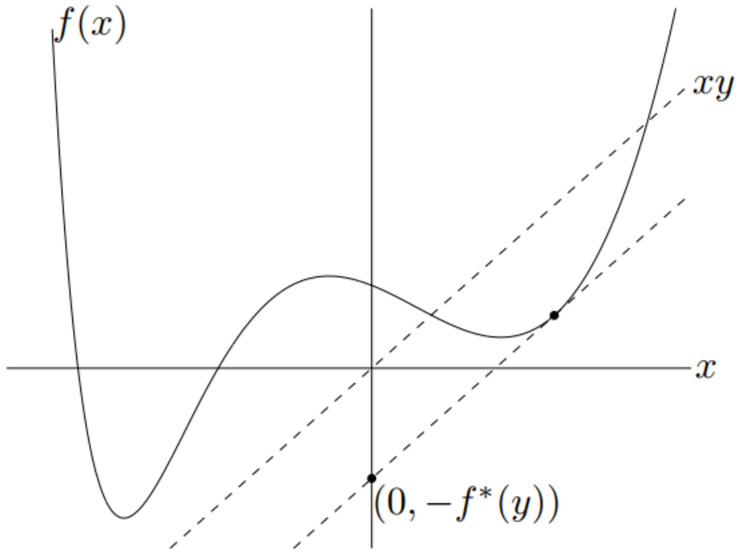
- By definition
- Hessian matrix is positive semidefinite
- By convexity-preserving operations

Conjugate function

The conjugate function of a function f is

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} y^T x - f(x)$$

The domain of conjugate function is where the supremum is finite.



In one-dimensional space, the conjugate function is the maximum gap between the linear function xy and $f(x)$.

The conjugate function is always convex (pointwise supremum of a family of affine functions)

Conjugate function: examples

- Affine function $f(x) = ax + b$

$$f^*(y) = \sup_{x \in \mathbb{R}} xy - ax - b = b,\, y = a$$

If $y \neq a$, then $xy - ax - b$ is unbounded for $x \in \mathbb{R}$

- Negative logarithmic $f(x) = -\log x$

$$f^*(y) = \sup_{x>0} xy + \log x = -1 - \log(-y)\,,\, y < 0$$

If $y > 0$, then $xy + \log x$ is unbounded as $x \to +\infty$

- Exponential function $f(x) = e^x$

$$f^*(y) = \sup_x xy - e^x = y\log(y) - y, y \geq 0$$

If $y < 0$, then $xy - e^x$ is unbounded for $x \to -\infty$

- Strictly convex quadratic $f(x) = \frac{1}{2}x^T Qx$ with $Q \succ 0$

$$f^*(y) = \sup_x y^T x - \frac{1}{2}x^T Qx = \frac{1}{2}y^T\, Q^{-1}y, \text{ for all } y$$