03. Convex functions

By Yang Lin[1] (2023 秋季, @NJU)

Contents

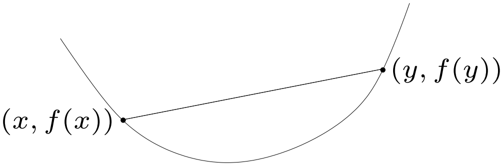
[1]Institute: Nanjing University. Email: linyang@nju.edu.cn.

1 Basic properties and examples

Definition $f : \boldsymbol{R}^n \to \boldsymbol{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

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

for all $x, y \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$



f is concave if $-f$ is convex

f is strictly convex if $\mathbf{dom}\, f$ is convex and

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

for all $x, y \in \mathbf{dom}\, f$, $x \neq y$ and $0 \leq \theta \leq 1$

1 Basic properties and examples

Convex:

Affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in R$

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 $p \geq 1$

Negative entropy: $-x \log x$ on $\mathbf{R}_{++}$

Concave:

Affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in R$

Powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$

Logarithmic: $\log x$ on $\mathbf{R}_{++}$

1 Basic properties and examples

Affine functions are convex and concave; all norms are convex

Examples on $\mathbf{R}^n$

Affine function: $f(x) = a^T x + b$

Norms: $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$

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

Affine function

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

Spectral(maximum singular value) norm

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

1 Basic properties and examples

Restriction of a convex function to a line

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

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

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

Can check convexity of f by checking convexity of functions of one variable

Example $f : \mathbf{S}^n \to \mathbf{R}$ with $f(X) = \log\det X$, $\mathbf{dom}\ f = \mathbf{S}^n_{++}$

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

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

g is concave in t (for any choice of $X \succ 0$, V); hence f is concave

1 Basic properties and examples

Extended-value extension $\tilde{f}$ of f is

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

often simplifies notation; for example, the condition

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

(as an inequality in $R \bigcup \{\infty\}$), means the same as the two conditions

- $\mathbf{dom}\ f$ is convex
- For $x, y \in \mathbf{dom}\ f$,

$$0 \leq \theta \leq 1 \Rightarrow f(\theta x + (1-\theta y)) \leq \theta f(x) + (1-\theta)f(y)$$

1 Basic properties and examples

First-order condition

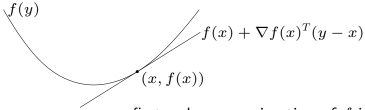
f is differentiable if **dom** f is open and the gradient

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

exists at each $x \in \textbf{dom}\ f$

1st-order condition: differentiable f with convex domain is convex iff

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



first-order approximation of f is global underestimator

1 Basic properties and examples

Second-order condition

f is twice differentiable if **dom** f is open and the Hassian $\nabla^2 f(x) \in \boldsymbol{S}^n$

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

exists at each $x \in$ **dom** f

2nd-order condition: for twice differentiable f with convex domain

- f is convex iff
$$\nabla^2 f(x) \succeq 0, \;\; \text{for all } x \in \textbf{dom } f$$
- If $\nabla^2 f(x) \succ 0$, for all $x \in$ **dom** f, then f is stricly convex

1 Basic properties and examples

Examples

Quadratic function: $f(x) = (1/2)x^T Px + q^T x + r$, ($P \in \mathbf{S}^n$, convex if $P \succeq 0$)

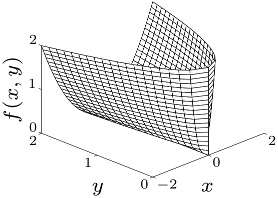
$$\nabla f(x) = Px + q, \;\; \nabla^2 f(x) = P$$

Least-squares objective: $f(x) = \|Ax - b\|_2^2$ (convex for any A)

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

Quadratic-over-linear: $f(x, y) = x^2/y$ (convex for $y > 0$)

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



1 Basic properties and examples

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

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

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

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

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

Geometric mean: $f(x) = (\prod_{k=1}^{n} x_k)^{1/n}$ on $\mathbf{R}_{++}^n$ is concave

(similar proof as for log-sum-exp)

1 Basic properties and examples

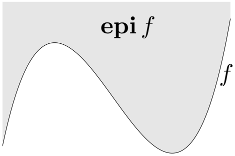
α-sublevel set of f: $\mathbf{R}^n \to \mathbf{R}$

$$C_\alpha = \{x \in \mathbf{dom}\ f | f(x) \leq \alpha\}$$

Sublevel sets of convex functions are convex (converse is false)

Epigraph of f: $\mathbf{R}^n \to \mathbf{R}$

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



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

1 Basic properties and examples

Jensen’s inequality: if f is convex, then for $0 \leq \theta \leq 1$,

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

Extension: if f is convex, then

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

for any random variable z

Basic inequality is special case with discrete distribution $p(x) = \theta, p(y) = 1-\theta$

2 Operations that preserve convexity

Practical methods for establishing convexity of a function:

1. Verify definition (often simplified by restricting to a line)

2. For twice differentiable functions, show $\nabla^2 f(x) \succeq 0$

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

2 Operations that preserve convexity

Positive weighted sum & composition with affine function:

Nonnegative multiple: αf is convex if f is convex, $\alpha \geq 0$

Sum: f_1+f_2 convex if f_1, f_2 convex (extends to infinite sums,integrals)

Composition with affine function: $f(Ax+b)$ is convex if f is convex

Example:

- Log barrier for linear inequalities

$$f(x) = -\log(b_i - a_i^T x), \mathbf{dom}\ f = \{x | a_i^T x < b_i,\ \ i = 1, ..., m\}$$

- (Any) norm of affine function: $f(x) = \|Ax + b\|$

2 Operations that preserve convexity

Pointwise maximum:

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

Example:

- Piecewise-linear function: $f(x) = \max_{i=1,\dots,m}(a_i^T x + b_i)$ is convex
- Sum of r largest components of $x \in \mathbf{R}^n$:

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

is convex ($x_{[i]}$ is i-th largest component of x) *proof.*

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

2 Operations that preserve convexity

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

Example:

- Support function of a set C: $S_C(x) = \sup_{y \in C} y^T x$ is convex
- Distance to farthest point in a set C:

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

- Maximum eigenvalue of symmetric matrix: for $X \in \mathbf{S}^n$

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^T X y$$

2 Operations that preserve convexity

Composition with scalar functions: composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$:

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

f is convex if: (1) g convex, h convex, $\tilde{h}$ nondecreasing; (2) g concave, h convex, $\tilde{h}$ nonincreasing

proof. (for $n = 1$, differentiable g, h)

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

note: monotonicity must hold for extended-value extension $\tilde{h}$

Example:

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

2 Operations that preserve convexity

Vector composition of $g : \mathbf{R}^n \to \mathbf{R}^k$ and $h : \mathbf{R}^k \to \mathbf{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), ..., g_k(x))$$

f is convex if: (1) g_i convex, h convex, $\tilde{h}$ nondecreasing in each argument; (2) g_i concave, h convex, $\tilde{h}$ nonincreasing in each argument

proof. (for $n = 1$, differentiable g, h)

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

note: monotonicity must hold for extended-value extension $\tilde{h}$

Example:

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

2 Operations that preserve convexity

Minimization: if $f(x, y)$ is convex in (x, y) and C is a convex set, then

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

is convex

Example:

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

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

  minimizing over y gives $g(x) = \inf_y f(x, y) = x^T(A - BC^{-1}B^T)x$
  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

2 Operations that preserve convexity

Perspective of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$

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

g is convex if f is convex

Example:

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

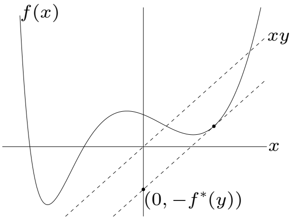
$$g(x) = (c^T x + d) f\left((Ax + b)/(c^T x + d)\right)$$

is convex on $\{x | c^T x + d > 0, (Ax + b)/(c^T x + d) \in \mathbf{dom}\ f\}$

3 The conjugate function

The conjugate of a function f is

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



f^* is convex (even if f is not)

will be useful in chapter 5

3 The conjugate function

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 $f(x) = (1/2)x^T Qx$ with $Q \in \mathbf{S}^n_{++}$

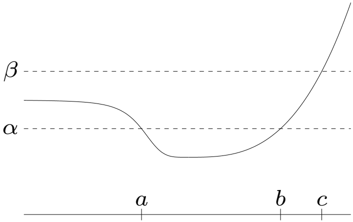
$$f^*(y) = \sup_x \left(y^T x - (1/2)x^T Qx \right) = \frac{1}{2} y^T Q^{-1} y$$

4 Quasiconvex functions

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

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

are convex for all α



f is quasiconcave if $-f$ is quasiconvex

f is quasilinear if it is quasiconvex and quasiconcave

4 Quasiconvex functions

Examples:

1. $\sqrt{|x|}$ is quasiconvex on $\mathbf{R}$
2. $\text{ceil}(x) = \inf\{z \in \mathbf{Z} | z \geq x\}$ is quasilinear
3. $\log x$ is quasilinear on $\mathbf{R}_{++}$
4. $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbf{R}^2_{++}$
5. Linear-fractional function

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

is quasilinear

6. Distance ratio

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \quad \mathbf{dom}\, f = \{x | \|x - a\|_2 \leq \|x - b\|_2\}$$

is quasiconvex

4 Quasiconvex functions

internal rate of return:

Cash flow $x = (x_0, \dots, x_n)$; x_i is payment in period i (to us if $x_i > 0$)

We assume $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$

Present value of cash flow x, for interest rate r:

$$\mathrm{PV}(x, r) = \sum_{i=0}^{n} (1+r)^{-i} x_i$$

Internal rate of return 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 halfspaces

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

4 Quasiconvex functions

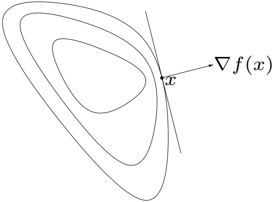
Properties:

1. Modified Jensen inequality: for quasiconvex f

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

2. First-order condition: differentiable f with cvx domain is quasiconvex iff

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



3. Sums of quasiconvex functions are not necessarily quasiconvex

5 Log-concave and log-convex functions

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

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

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

Example:

- Powers: x^a on $\mathbf{R}_{++}$ islog-convex for $a \leq 0$, log-concave for $a \geq 0$
- Many common probability densities are log-concave, e.g., normal:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})}$$

- Cumulative Gaussian distribution function Φ is log-concave

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

5 Log-concave and log-convex functions

Properties of log-concave functions

Twice differentiable f with convex domain is log-concave if and only if

$$f(x)\nabla^2 f(x) \preceq \nabla f(x)\nabla f(x)^T$$

for all $x \in \mathbf{dom} f$

Product of log-concave functions is log-concave

Sum of log-concave functions is not always log-concave

Integration: if $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x, y)\, dy$$

is log-concave (not easy to show)

5 Log-concave and log-convex functions

Consequences of integration property

Convolution $f * g$ of log-concave functions f and g is log-concave

$$(f * g)(x) = \int f(x-y)g(y)\,dy$$

if $C \subseteq \mathbf{R}^n$ convex and y is a random variable with log-concave pdf then

$$f(x) = \mathbf{prob}(x + y \in C)$$

is log-concave

proof. write $f(x)$ as integral of product of log-concave functions

$$f(x) = \int g(x+y)p(y)\,dy, \quad g(u) = \begin{cases} 1, & u \in C \\ 0, & u \notin C, \end{cases}$$

p is pdf of y

5 Log-concave and log-convex functions

Example: yield function

$$Y(x) = \mathbf{prob}(x + w \in S)$$

$x \in \mathbf{R}^n$: nominal parameter values for product

$w \in \mathbf{R}^n$: random variations of parameters in manufactured product

S: set of acceptable values

If S is convex and w has a log-concave pdf, then

Y is log-concave

Yield regions $\{x | Y(x) \geq \alpha\}$ are convex

6 Convexity with respect to generalized inequalities

$f : \mathbf{R}^n \to \mathbf{R}^m$ is K-convex if **dom** f is convex and

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

for $x, y \in$ **dom** f, $0 \le \theta \le 1$

example $f : \mathbf{S}^m \to \mathbf{S}^m$, $f(X) = X^2$ is $\mathbf{S}_+^m$-convex

proof. for fixed $z \in \mathbf{R}^m$, $z^T X^2 z = \|Xz\|_2^2$ is convex in X, i.e.,

$$z^T(\theta X + (1-\theta)Y)^2,\ z \le \theta z^T X^2 z + (1-\theta) z^T Y^2 z$$

for $X, Y \in \mathbf{S}^m$, $0 \le \theta \le 1$

Therefore $(\theta X + (1-\theta)Y)^2 \preceq \theta X^2 + (1-\theta)Y^2$