Convex functions

Definition

$f : R^n \to R$ is convex if dom f is a convex set and

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

for all $x, y \in$ dom f, and $\theta \in [0, 1]$.

Convex functions

Definition

$f : R^n \rightarrow R$ is convex if dom f is a convex set and

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

for all $x, y \in$ dom f, and $\theta \in [0, 1]$.

- f is concave if $-f$ is convex
- f is strictly convex if dom f is convex and

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

  for all $x, y \in$ dom f, $x \neq y$, $\theta \in (0, 1)$.

Convex function: examples on R

- affine: $ax + b$
- exponential: $e^{\alpha x}$, for any $\alpha \in R$
- powers:
  - x^a on R_{++}, for $a \geq 1$ or $a \leq 0$
  - $-x^a$ on R_{++}, for $0 \leq a \leq 1$
- powers of absolute value: $|x|^p$ on R, for $p \geq 1$
- negative logarithm: $-\log x$ on R_{++}
- $x \log x$ on R_{++}

Convex function: affine functions

Affine functions are both convex and concave

- affine function in R^n:

$$f(x) = a^T x + b$$

- affine function in $R^{m \times n}$:

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

Convex function: norms

All norms are convex

- norms in R^n:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

  for all $p \geq 1$
- norms in $R^{m \times n}$:
  - Frobenius norm: $\|X\|_F = \langle X, X \rangle^{1/2}$
  - spectral norm: $\|X\|_2 = \sigma_{max}(X) = (\lambda_{\max}(X^T X))^{1/2}$

Operator norms

Let $\|\cdot\|_a$ and $\|\cdot\|_b$ be norms on R^m and R^n, respectively. The operator norm of $X \in R^{m\times n}$, induced by $\|\cdot\|_a$ and $\|\cdot\|_b$, is defined to be

$$\|X\|_{a,b} = \sup \ \{\|Xu\|_a \mid \|u\|_b \leq 1\}$$

- Spectral norm (ℓ_2-norm):

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

- Max-row-sum norm:

$$\|X\|_\infty = \|X\|_{\infty,\infty} = \max_{i=1,\cdots,m} \sum_{j=1}^{n} |X_{ij}|$$

- Max-column-sum norm:

$$\|X\|_1 = \|X\|_{1,1} = \max_{j=1,\cdots,n} \sum_{i=1}^{m} |X_{ij}|$$

Dual norm

Let $\|\cdot\|$ be a norm on R^n. The associated dual norm, denoted $\|\cdot\|_*$, is defined as

$$\|z\|_* = \sup \ \{z^T x \mid \|x\| \leq 1\}.$$

- $z^T x \leq \|x\| \ \|z\|_*$ for all $x, z \in R^n$
- The dual of the ℓ_p-norm is the ℓ_q-norm, where $1/p + 1/q = 1$
- The dual of the ℓ_2-norm on $R^{m \times n}$ is the nuclear norm,

$$\begin{aligned} \|Z\|_{2*} &= \sup \ \{\mathrm{tr}(Z^T X) \mid \|X\|_2 \leq 1\} \\ &= \sigma_1(Z) + \cdots + \sigma_r(Z) = \mathrm{tr}(Z^T Z)^{1/2}, \end{aligned}$$

where $r = \textbf{rank} \ Z$.

Restriction of a convex function to a line

$f : R^n \rightarrow R$ is convex iff $g : R \rightarrow R$,

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

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

So we can check the convexity of a function with multiple variables by checking the convexity of functions of one variable

Restriction of a convex function to a line: example

Show that $f : S^n_{++} \to R$ with $f(X) = \log\det X$ is concave.

Proof.

Define $g : R \to R$, $g(t) = f(X + tV)$ with $\mathbf{dom}\, g = \{t \mid X + tV \succ 0\}$, for any $X \succ 0$ and $V \in S^n$.

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

where $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of $X^{-1/2}VX^{-1/2}$. Hence g is concave for any $X \succ 0$ and $V \in S^n$, so is f. □

Extended-value extensions

Definition

If $f : R^n \to R$ is convex, we define its extended-value extension $\tilde{f} : R^n \to R \cup \{\infty\}$ by

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

By this notation, the condition

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

is equivalent to the two conditions:

- **dom** f is convex
- $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \quad \forall \theta \in [0,1]$

First-order conditions

Theorem (first-order condition)

If $f : R^n \to R$ is differentiable, then f is convex if and only if **dom** *f is convex and*

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

First-order conditions

Theorem (first-order condition)

If $f : R^n \to R$ is differentiable, then f is convex if and only if **dom** *f is convex and*

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

- local information (gradient) leads to global information (convexity)
- f is strictly convex if and only if **dom** f is convex and

$$f(y) > f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in \mathrm{dom}\, f, x \neq y$$

First-order condition: proof

Proof.

- Suppose $f(x)$ is convex. Then

$$\begin{aligned} f(x+\theta(y-x)) - f(x) &\le \theta(f(y)-f(x)), \quad \forall \theta \in [0,1] \\ \lim_{\theta \to 0} \frac{f(x+\theta(y-x)) - f(x)}{\theta} &\le f(y) - f(x) \\ \implies \nabla f(x)^T (y-x) &\le f(y) - f(x) \end{aligned}$$

- Suppose the first-order condition holds. Let $z = \theta x + (1-\theta) y$. Then

$$\begin{aligned} & f(x) \ge f(z) + \nabla f(z)^T (x-z) \\ & f(y) \ge f(z) + \nabla f(z)^T (y-z) \\ & \implies \theta f(x) + (1-\theta) f(y) \ge f(z) \end{aligned}$$

which is true for any $\theta \in [0,1]$, so f is convex.

□

Second-order conditions

Theorem (second-order condition)

If $f : R^n \rightarrow R$ is twice differentiable, then f is convex if and only if **dom** f *is convex and its Hessian is positive semidefinite, i.e.,*

$$\nabla f(x) \succeq 0 \quad \forall x \in \mathrm{dom}\, f$$

Second-order conditions

Theorem (second-order condition)

If $f : R^n \rightarrow R$ is twice differentiable, then f is convex if and only if **dom** *f is convex and its Hessian is positive semidefinite, i.e.,*

$$\nabla f(x) \succeq 0 \quad \forall x \in \mathrm{dom}\, f$$

- if $\nabla f(x) \succ 0$ for all $x \in \mathrm{dom}\, f$, then f is strictly convex

Second-order conditions: proof

Proof.

- Suppose f is convex. Because f is twice differentiable, we have

$$f(x+\delta x) = f(x) + \nabla f(x)^T \delta x + \frac{1}{2}\delta x^T \nabla^2 f(x)\delta x + R(x;\delta x)\|\delta x\|^2$$

  where $R(x;\delta x) \to 0$ as $\delta x \to 0$. Because f is convex, by the first-order condition, $f(x+\delta x) \geq f(x) + \nabla f(x)^T \delta x$. Hence

$$\delta x^T \nabla^2 f(x)\delta x + R(x;\delta x)\|\delta x\|^2 \geq 0$$

  for any δx. Let $\delta x = \epsilon d$. Taking $\epsilon \to 0$ yields $d^T \nabla^2 f(x) d \geq 0$ for any d, thus $\nabla^2 f(x) \succeq 0$.

- Suppose $\nabla f(x) \succeq 0\ \forall x \in \mathbf{dom}\ f$. Then for any $x, y \in \mathbf{dom}\ f$ and some $z = \theta x + (1-\theta) y$ with $\theta \in [0,1]$,

$$\begin{aligned} f(y) &= f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x) \\ &\geq f(x) + \nabla f(x)^T (y-x) \end{aligned}$$

  By the first-order condition, f is convex.

Some special cases

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

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

  convex if $P \succeq 0$

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

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

  convex for any A

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

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

  convex for $y > 0$

Some special cases: log-sum-exp

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

$\max\{x_1, \cdots, x_n\} \leq f(x) \leq \max\{x_1, \cdots, x_n\} + \log n$, so f can be viewed as a differentiable approximation of the max function.

Proof.

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

$$\Longrightarrow \; v^T \nabla^2 f(x) v = \frac{1}{(1^T z)^2}\left(\sum_{i=1}^{n} z_i \sum_{i=1}^{n} z_i v_i^2 - (\sum_{i=1}^{n} z_i v_i)^2\right)$$

$$= \frac{1}{(1^T z)^2}(\|a\|_2^2 \|b\|_2^2 - \langle a, b\rangle^2) \geq 0$$

where $a = (\sqrt{z_1}, \cdots, \sqrt{z_n})$, $b = (\sqrt{z_1} v_1, \cdots, \sqrt{z_n} v_n)$, for any v. □

Some special cases: geometric mean

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

Proof.

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2}(n\,\mathrm{diag}^2(q) - qq^T) \quad (q_i = 1/x_i)$$

$$\implies v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2}\left(\sum_{i=1}^{n} 1 \sum_{i=1}^{n} v_i^2 q_i^2 - (\sum_{i=1}^{n} q_i v_i)^2\right)$$

$$= -\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2}(\|a\|_2^2 \|b\|_2^2 - \langle a, b\rangle^2) \le 0$$

where $a_i = 1,\ b_i = q_i v_i$ for any v, so $\nabla^2 f(x) \preceq 0$. □

Epigraph and sublevel set

α-**sublevel set** of $f : R^n \to R$:

$$C_\alpha = \{x \in \text{dom } f \mid f(x) \leq \alpha\}$$

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

epigraph of $f : R^n \to R$:

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

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

Jensen's inequality and extensions

basic inequality: if f is convex, then for any $\theta \in [0, 1]$,

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

extension: if f is convex, then

$$f(\mathrm{E}[z]) \leq \mathrm{E}[f(z)]$$

for any random variable z.

Using Jensen's inequality: deriving Holder's inequality

$$\theta \log a + (1-\theta) \log b \leq \log(\theta a + (1-\theta) b)$$
$$\implies a^{\theta} b^{1-\theta} \leq \theta a + (1-\theta) b \quad \forall a, b \geq 0, \theta \in [0,1]$$

Applying this with

$$a = \frac{|x_i|^p}{\sum_j |x_j|^p} \quad a = \frac{|y_i|^p}{\sum_j |y_j|^p} \quad \theta = 1/p$$

yields

$$\left(\frac{|x_i|^p}{\sum_j |x_j|^p}\right)^{1/p} \left(\frac{|y_i|^p}{\sum_j |y_j|^p}\right)^{1/q} \leq \frac{|x_i|^p}{p \sum_j |x_j|^p} + \frac{|y_i|^p}{q \sum_j |y_j|^p}$$

and summing over i yields Holder's inequality.

Operations preserving convexity

- nonnegative weighted sum
- composition with affine function
- pointwise maximum and supremum
- composition
- minimization
- perspective

positive weighted sum & composition with affine function

- f is convex $\implies$ αf is convex for any $\alpha \geq 0$
- f_1, f_2 are convex $\implies$ $f_1 + f_2$ is convex (extends to infinite sums, integrals)
- f is convex $\implies$ $f(Ax + b)$ is convex

Examples:

- log barrier for linear inequalities

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

- $f(x) = \|Ax + b\|$

Pointwise maximum

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

examples

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

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

  is convex ($x_{[i]}$ is ith largest component of x)

Pointwise supremum

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

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

is convex

In terms of epigraph:

$$\text{epi } g = \bigcap_{y \in A} \text{epi } f(\cdot, y)$$

Examples

- support function of a set: $S_C(x) = \sup_{y \in C} y^T x$
- $f(x) = \sup_{y \in C} \|x - y\|$
- $\lambda_{\max}(X) = \sup_{\|u\|_2 = 1} u^T X u$ where $X \in S^n$ is convex

Minimization

If $f(x, y)$ is convex in (x, y) and C is a convex nonempty set, then

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

is convex

examples

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

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

  Because

$$g(x) = \inf_{y} \ f(x, y) = x^T (A - BC^{-1}B^T)x$$

  is convex, the Schur complement $A - BC^{-1}B^T \succeq 0$
- distance to a set: $\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex

Minimization: proof

Proof.

Suppose $x_1, x_2 \in \text{dom } g$. Given $\epsilon > 0$, $\exists\ y_1, y_2 \in C$ such that $f(x_i, y_i) \le g(x_i) + \epsilon$. Hence for any $\theta \in [0, 1]$,

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2) &= \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\le f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\le \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\ &\le \theta g(x_1) + (1-\theta) g(x_2) + \epsilon \end{aligned}$$

which holds for any $\epsilon > 0$. Thus

$$g(\theta x_1 + (1-\theta)x_2) \le \theta g(x_1) + (1-\theta) g(x_2)$$

□

Composition with scalar functions

Composition of $g : R^n \to R$ and $h : R \to R$:

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

Use:

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

f is convex if

- g convex, h convex, $\tilde{h}$ nondecreasing
  - example: $\exp g(x)$ is convex if g is convex
- g concave, h convex, $\tilde{h}$ nonincreasing
  - example: $1/g(x)$ is concave if g is concave and positive

Composition with vector functions

Composition of $g : R^n \to R^k$ and $h : R^k \to R$:

$$f(x) = h(g(x) = h(g_1(x), \cdots, g_k(x))$$

when $n = 1$:

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

f is convex if

- g_i convex, h convex, $\tilde{h}$ nondecreasing in each argument
  - example: $\sum_{i=1}^{m} \exp g_i(x)$ is convex if g_i is convex
- g_i concave, h convex, $\tilde{h}$ nonincreasing in each argument
  - example: $\sum_{i=1}^{m} \log g_i(x)$ is concave if g_i is concave and positive

Perspective

The perspective of a function $f : R^n \to R$ is the function $g : R^{n+1} \to R$,

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

f is convex $\implies$ g is convex.

For $t > 0$,

$$(x,t,s) \in \text{epi } g \iff f(x/t) \leq s/t \iff (x/t, s/t) \in \text{epi } f$$

Hence epi g is the inverse image of epi f under the perspective mapping

examples:

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

The conjugate function

the conjugate of a function f is

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

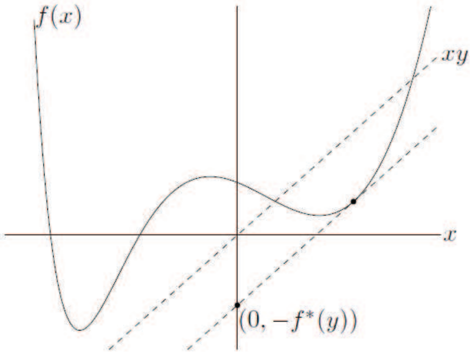

Figure: Conjugate function

The conjugate function: examples

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

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

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

$$\begin{aligned} f^*(y) &= \sup_x \; y^Tx - 1/2x^TQx \\ &= \frac{1}{2}y^TQ^{-1}y \end{aligned}$$

The conjugate function: properties

Properties:

- f^* is convex
- f is convex and closed (i.e., epi f is closed) $\implies$ $f^{**} = f$.
- Frechel's inequality: $f(x) + f^*(y) \geq x^T y$
  - Example: with $f(x) = (1/2)x^T Qx$ with $Q \in S^n_{++}$, we have

$$x^T y \leq (1/2)x^T Qx + (1/2)y^T Q^{-1} y$$

Convexity with respect to generalized inequalities

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

$$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta) f(y) \quad \forall x, y \in \mathrm{dom}\, f, \theta \in [0,1]$$

Convexity with respect to generalized inequalities

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

$$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta)f(y) \quad \forall x, y \in \mathbf{dom}\, f, \theta \in [0,1]$$

example: $f : S^m \to S^m$ with $f(X) = X^2$ is S^m_+-convex

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

$$z^T(\theta X + (1-\theta)Y)^2 z \leq \theta z^T X^2 z + (1-\theta) z^T Y^2 z \quad \forall X, Y \in S^m$$

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

□

Quasiconvex functions

Definition

A function $f(x)$ is **quasiconvex** if $\forall x, y \in \operatorname{dom} f$,

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

Theorem

$f(x)$ is quasiconvex if and only if every level set of f is convex.

Quasiconvex functions: level sets

Theorem

$f(x)$ is quasiconvex if and only if every level set of f is convex.

Proof.

1. Suppose f is quasiconvex. Suppose $x, y \in \operatorname{dom} f$ belongs to level set $S_a = \{x \mid f(x) \leq a\}$. Then

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

Thus $\theta x + (1-\theta)y \in S_a$ for all $\theta \in [0,1]$, so S_a is convex.

2. Suppose every level set of f is convex. For any $x, y \in \operatorname{dom} f$, let $a = \max\{f(x), f(y)\}$. Clearly $x, y \in S_a$. Because S_a is convex, $\theta x + (1-\theta)y \in S_a$ for any $\theta \in [0,1]$. Thus f is quasiconvex. □