Convex Functions

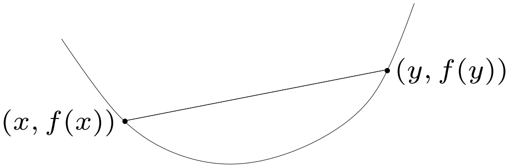
November 14, 2023

Definition

$f : \mathbf{R}^n \to \mathbf{R}$ 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 $x, y \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$



Interpretation. Geometrically, the line segment connecting $(x, f(x))$ to $(y, f(y))$ must sit above the graph of f.

- f is concave if $-f$ is convex

Example

convex:

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

concave:

- affine: $ax + b$ on **R**, for any $a, b \in$ **R**
- powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$

Question Can we prove convexity of e^{ax} by definition?

Convex Function: Example

Question Can we prove convexity of $f(x) = e^x$ by definition?
From the AM-GM inequality

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq (x_1 x_2 \cdots x_n)^{1/n}$$

let $x_1 = \cdots = x_k = a$ and $x_{k+1} = \cdots = x_n = b$.

$$\frac{k}{n}a + \left(1 - \frac{k}{n}\right) b \geq a^{k/n} b^{1-k/n}$$

For an arbitrarry number $0 \leq \theta \leq 1$, let $m(k)$ be a sequence s.t. $m(k)/n \to \theta$ as $n \to \infty$.

$$\theta a + (1 - \theta)\, b \geq a^{\theta} b^{1-\theta}$$

For any x and $y > 0$, let $a = e^x$ and $b = e^y$, $f(x)$ is convex since

$$\theta e^x + (1 - \theta) e^y \geq e^{\theta x} e^{(1-\theta) y} = e^{\theta x + (1-\theta) y}$$

Convex Function

- If f is differentiable, then it is convex if and only if f' is non-decreasing.
- If f is twice differentiable, it is convex if and only if $f''(x) \geq 0$ for all $x \in \mathbf{dom} f$

Convex Function

If $f(x, y)$ is convex in (x, y), then it is convex in x and convex in y, i.e.,

- $g(x) = f(x, y)$ is convex
- $h(y) = f(x, y)$ is convex

The converse is not necessarily true.
$g(x, y) = x^2 y$ is convex in x and convex in y where $x \in \mathbb{R}$ $y \in \mathbb{R}_+$

$$\theta g(x_1, y) + (1-\theta)g(x_2, y) - g(\theta x_1 + (1-\theta)x_2, y) = \theta(1-\theta)(x_1 - x_2)^2 y \geq 0$$
$$g(x, \theta y_1 + (1-\theta)y_2) = \theta g(x, y_1) + (1-\theta)g(x, y_2)$$

where $x, x_1, x_2 \in \mathbb{R}$, $y, y_1, y_2 \in \mathbb{R}_+$. However, $g(x, y)$ is not convex in (x, y)

$$g(0.5(0,2) + 0.5(2,0)) = 1 \nleq 0.5g(0,2) + 0.5g(2,0) = 0$$

Strict and strong convexity I

- A function is strictly convex if it is convex with strict inequality

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

  for all $x \neq y$ and $0 < \theta < 1$
- A function is σ-strongly convex if $\exists \sigma > 0$ such that

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) - \frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2$$

  - strongly convex functions are strictly convex
  - a function $f(x)$ is σ-strongly convex iff $f(x) - \frac{\sigma}{2}\|x\|^2$ is convex
  - its curvature is lower bounded by the curvature of the quadratic $\frac{\sigma}{2}\|x\|^2$

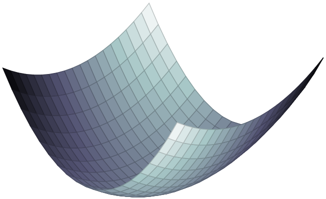
$$\kappa(x) = \frac{|f''(x)|}{(1 + (f'(x))^2)^{3/2}}$$

Strong Convexity

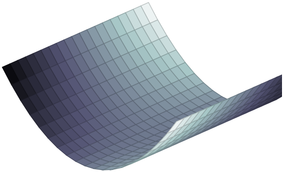
- f is strongly convex function and g is a convex function, then $h = f + g$ is strongly convex function

$$
\begin{aligned}
h(\theta x + (1-\theta)y) &= f(\theta x + (1-\theta)y) + g(\theta x + (1-\theta)y) \\
&\le \theta f(x) + (1-\theta) f(y) - \frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2 + g(\theta x + (1-\theta)y) \\
&\le \theta h(x) + (1-\theta) h(y) - \frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2
\end{aligned}
$$

- L_2-regularized problem of the form $h(x) = f(x) + \frac{\lambda}{2}\|x\|^2$, where f is convex and $\lambda > 0$, is strongly convex



$f(x)$



$f(x) - \|x\|_2^2$

Examples of Strict/strongly convex functions

Strictly convex

- $f(x) = -\log(x) + \iota_{>0}(x)$
- $f(x) = 1/x + \iota_{>0}(x)$
- $f(x) = e^x$
- $f(x) = e^{-x}$

Strongly convex

- $f(x) = \frac{\lambda}{2}\|x\|_2^2$
- $f(x) = \frac{1}{2}x^T Q x$ where Q positive definite
- $f(x) = f_1(x) + f_2(x)$ where f_1 strongly convex and f_2 convex
- $f(x) = \frac{1}{2}x^T Q x + \iota_C(x)$ where Q positive definite and C convex

Uniqueness of minimizers

- if a function is strictly (strongly) convex the minimizers are unique
- proof: assume that $x_1 \neq x_2$ and that both satisfy

  $$x_2 = x_1 = \operatorname*{argmin}_x f(x)$$

  i.e., $f(x_1) = f(x_2) = \inf_x f(x)$, then

  $$f(\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2) < \tfrac{1}{2}(f(x_1) + f(x_2)) = \inf_x f(x)$$

  contradiction!
- (minimizer might not exist for strictly convex, but always for strongly convex)

- $\exp^x$ has no minimiser

Examples on $\mathbb{R}^n$

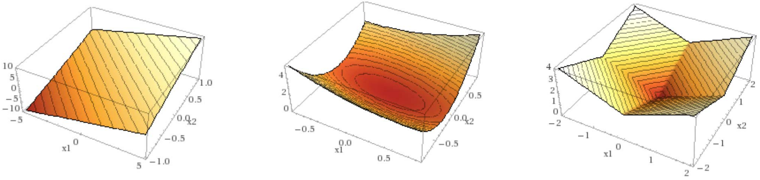
- Affine function $f(x) = a^T x + b$

Affine functions are convex (but no strictly convex) and also concave.

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

All norms are convex.

- Quadratic functions $f(x) = x^T Qx + c^T x + d$
  - Convex if and only if $Q \succeq 0$
  - Strictly convex if and only if $Q \succ 0$
  - Concave if and only if $Q \preceq 0$; strictly concave if and only if $Q \prec 0$.



Convex functions: examples

- **Support function** of any set is convex

$$h_A(x) = \sup\{\langle x, z\rangle | z \in A\}$$

Convex functions: examples

- **Support function** of any set is convex

$$h_A(x) = \sup\{\langle x, z\rangle | z \in A\}$$

Proof Let $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$.

$$\begin{aligned} h_A(\theta x + (1-\theta)y) &= \sup_{z\in A}\langle \theta x + (1-\theta)y, z\rangle \\ &= \sup_{z\in A}(\theta\langle x, z\rangle + (1-\theta)\langle y, z\rangle) \\ &\le \theta \sup_{z\in A}\langle x, z\rangle + (1-\theta)\sup_{z\in A}\langle y, z\rangle \\ &= \theta h_A(x) + (1-\theta)h_A(y), \end{aligned}$$

which shows that h_A is convex.

Example: Support function of Unit Ball

Support function of unit ball $B = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2 \leq 1\}$

$$h_B(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in B} \boldsymbol{y}^T \boldsymbol{x}$$

Since

$$\boldsymbol{y}^T \boldsymbol{x} \leq \|\boldsymbol{y}\|_2 \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_2$$

we have

$$h_B(\boldsymbol{x}) \leq \|\boldsymbol{x}\|_2$$

For any $\boldsymbol{x}$, choose $\boldsymbol{y} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}$, i.e., $\boldsymbol{y} \in B$.

$$h_B(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in B} \boldsymbol{y}^T \boldsymbol{x} \geq \frac{\boldsymbol{x}^T \boldsymbol{x}}{\|\boldsymbol{x}\|_2} = \|\boldsymbol{x}\|_2$$

We conclude that $h_B(\boldsymbol{x}) = \|\boldsymbol{x}\|_2$.

Convex functions: examples

• **Indicator function** of a set is convex if and only if the set is convex.

$$\delta_C(x) = \begin{cases} 0 & \text{, if } x \in C \\ +\infty & \text{, if } x \notin C \end{cases}$$

• **Max function**: $f(x) = \max\{x_1, \dots, x_n\}$ is convex.
Hint Let $x^* = f(x)$ and $y^* = f(y)$.

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max\{\theta x_1 + (1-\theta)y_1, \dots, \theta x_n + (1-\theta)y_n\} \\ &\leq \max\{\theta x^* + (1-\theta)y^*, \dots, \theta x^* + (1-\theta)y^*\} \\ &= \theta f(x) + (1-\theta)f(y)\,. \end{aligned}$$

Convex functions: examples

examples on $\mathbf{R}^{m \times n}$ ($m \times n$ matrices)

- 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

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

(See this function as $\sup_{\|y\|_2=1} y^T X y$)

Restriction of a convex function to a line I

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

$$g(t) = f(x + tv), \qquad \textbf{dom}\, g = \{t \mid x + t\, v \in \textbf{dom}\, f\}$$

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

Proof

- **Necessity:** Assume $g(t)$ is nonconvex for some x and v. There exist t_1, t_2 in **dom**g and $x + t_1 v$, $x + t_2 \in$ **dom**f such that

  $$g(\theta t_1 + (1-\theta)t_2) > \theta g(t_1) + (1-\theta)g(t_2), \quad 0 \le \theta \le 1$$

  i.e.,

  $$f(\theta(x + t_1 v) + (1-\theta)(x + t_2 v)) > \theta f(x + t_1 v) + (1-\theta) f(x + t_2 v)$$

  therefore f is nonconvex (contradiction). It means g is convex.

Restriction of a convex function to a line II

- **Sufficiency:** Assume that $g(t)$ is convex and $f(x)$ is nonconvex. Then there exist $x_1, x_2 \in \textbf{dom} f$ and some $0 < \theta < 1$ such that

$$f(\theta x_1 + (1-\theta)x_2) > \theta f(x_1) + (1-\theta)f(x_2)$$

Let $x = x_1$ and $v = x_2 - x_1$. $[0, 1] \subset \textbf{dom} g$

$$\begin{aligned} g(1-\theta) &= f(\theta x_1 + (1-\theta)x_2) \\ &> \theta f(x_1) + (1-\theta)f(x_2) \\ &= \theta g(0) + (1-\theta)g(1). \end{aligned}$$

Therefore $g(t)$ is nonconvex (contradiction). Thus $f(x)$ must be convex.

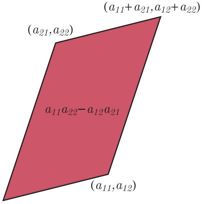
Restriction of a convex function to a line III

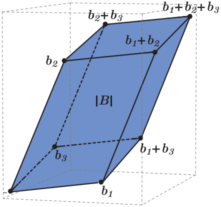
• Can check convexity of f by checking convexity of functions of one variable
Example $f : \mathbf{S}^n_{++} \to \mathbf{R}$ with $f(\mathbf{X}) = \log\det(\mathbf{X})$
Note that $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ for $\mathbf{A}$ and $\mathbf{B} \in \mathbf{S}^n$.

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

where λ_i are the eigenvalues of $\mathbf{X}^{-1/2}\mathbf{V}\mathbf{X}^{-1/2}$.
g is concave in t (for any choice of $\mathbf{X} \succ 0$ and $\mathbf{V}$). Hence f is concave.

Volume Minimization





- Volume of n-simplex with vertices $(\boldsymbol{v}_0, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$:

$$\frac{1}{n!} | \det(\boldsymbol{v}_1 - \boldsymbol{v}_0, \boldsymbol{v}_2 - \boldsymbol{v}_0, \ldots, \boldsymbol{v}_n - \boldsymbol{v}_0)|$$

- Volume of a closed ellipsoid: $\{\boldsymbol{x} \in \mathbb{R}^n \,||\, (\boldsymbol{x} - \boldsymbol{x}_c)^T \mathbf{Q}^{-1} (\boldsymbol{x} - \boldsymbol{x}_c) \leq 1\}$, $\mathbf{Q} \succ 0$

$$\frac{\pi^{n/2}}{\Gamma(n/2+1)} \sqrt{\det(\mathbf{Q})}$$

- $\det()$ is not convex.

First-order condition I

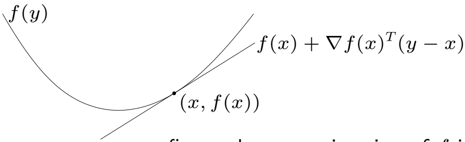
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

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

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

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

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



first-order approximation of f is global underestimator

Note: $tf(x) + (1-t)f(y) \geq f(tx + (1-t)y)$. Hence
$f(y) \geq f(x) + \lim_{t \to 0} \frac{f(x+t(y-x)) - f(x)}{t}$

First-order condition II

Proof

- Sufficiency: For $x, y, z \in \textbf{dom}\, f$ and $0 < \theta < 1$

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

we have

$$\theta f(y) + (1 - \theta) f(z) \geq f(x) + \nabla f(x)^T (\theta y + (1 - \theta) z - x)$$

Let $x = \theta y + (1 - \theta) z \in \textbf{dom} f$, we obtain

$$\theta f(y) + (1 - \theta) f(z) \geq f(\theta y + (1 - \theta) z)$$

Therefore f is convex.

First-order condition III

- Necessity: For $x, y \in \textbf{dom}\, f$ and $0 < t \leq 1$, consider restricted line of f defined as

$$g(t) = f(ty + (1-t)x), \quad \textbf{dom}\, g = \{t \mid ty + (1-t)x \in \textbf{dom} f\}$$

which has

$$g'(t) = \nabla f(ty + (1-t)x)^T (y - x)$$

Since $f(x)$ is convex, $g(t)$ is also convex. Hence

$$g(1) \geq g(0) + g'(0)(1-0)$$

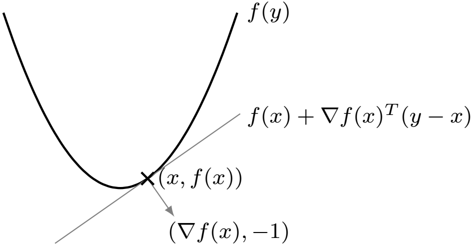
this leads to $f(y) \geq f(x) + \nabla f(x)^T (y - x)$.

First-order condition for strict convexity

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable
- f is strictly convex if and only if

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

for all $x, y \in \mathbb{R}^n$ where $x \neq y$



- Function f has for all $x \in \mathbb{R}^n$ an affine minorizer that:
  - has slope s defined by ∇f
  - coincides with function f *only* at x
  - is supporting hyperplane to epigraph of f

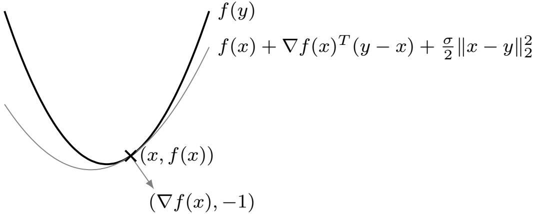
Question: compare $|x|$ and x^2

First-order condition for strong convexity I

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable
- f is σ-strongly convex with $\sigma > 0$ if and only if

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\sigma}{2}\|x - y\|_2^2$$

  for all $x, y \in \mathbb{R}^n$



- Function f has for all $x \in \mathbb{R}^n$ a quadratic minorizer that:
  - curvature defined by σ
  - coincides with function f at x
  - defines normal $(\nabla f(x), -1)$ to epigraph of f

First-order condition for strong convexity II

Normal or supporting hyperplane to epigraph of a convex function f
(y, t) in epigraph of $f(x)$ at the point $(x, f(x))$ means that

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

Then supporting hyperplane for the epigraph of f at the point $(x, f(x))$ is given by

$$\nabla f(x)^T y - t \leq -f(x) + \nabla f(x)^T x$$

or

$$(\nabla f(x), -1)^T (y, t) \leq -f(x) + \nabla f(x)^T x$$

i.e., the vector $(\nabla f(x), -1)$ defines normal to epigraph of f

Example I

Prove that $f(\mathbf{X}) = \log\det(\mathbf{X})$ is a concave function where $\mathbf{X} \in \mathbf{S}^n_+$. Given that 1st derivative of $f(\mathbf{X})$

$$\nabla_X \log\det(\mathbf{X}) = \mathbf{X}^{-1}$$

Following the first-order condition we need to prove that

$$\begin{aligned} f(\mathbf{Y}) &\le f(\mathbf{X}) + \langle \nabla_X f(\mathbf{X}), \mathbf{Y} - \mathbf{X} \rangle \\ \log\det(\mathbf{Y}) &\le \log\det(\mathbf{X}) + \langle \mathbf{X}^{-1}, \mathbf{Y} - \mathbf{X} \rangle \\ &= \log\det(\mathbf{X}) + \langle \mathbf{X}^{-1}, \mathbf{Y} \rangle - n \end{aligned}$$

or the following inequality

$$\log\det(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2}) - \mathrm{tr}(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2}) + n \le 0.$$

Let $\lambda_k \ge 0$ be eigenvectors of $\mathbf{Q} = \mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2}$. Taking into account that $\det(\mathbf{Q}) = \prod_{k=1}^n \lambda_k, \qquad \mathrm{tr}(\mathbf{Q}) = \sum_{k=1}^n \lambda_k$.

Example II

$$\sum_{k=1}^{n} (\log(\lambda_k) - \lambda_k + 1) \leq 0.$$

or $h(x) = \log(x) - x + 1 \leq 0$ for $x \geq 0$. (Note that $h'(x) = \frac{1}{x} - 1$)

Example III

Derivative of $\log\det(X^{-1})$

$$\log\det(X^{-1}) = \log\det(X)^{-1} = -\log\det(X)$$

$$\frac{\partial \log\det(X^{-1})}{\partial X_{i,j}} = \frac{-1}{\det(X)}\frac{\partial \det X}{\partial X_{i,j}} = \frac{-1}{\det(X)}\mathsf{adj}(X)_{j,i} = -(X^{-1})_{ji}$$

Alternative proof
Let $X \in \mathbf{S}^n_{++}$ and $H \in \mathbf{S}^n$ such that $X + H \in \mathbf{S}^n_{++}$

$$\begin{aligned} f(X+H) - f(X) &= \log\det(X+H) - \log\det(X) = \log\det(X^{-1}(X+H)) \\ &= \log\det(I + X^{-1/2}HX^{-1/2}) \end{aligned}$$

Example IV

Applying arithmetic-geometric inequality to the eigenvalues of $X^{-1/2}HX^{-1/2}$ we have

$$\begin{aligned}\log\det(I + X^{-1/2}HX^{-1/2}) &\le \log\left(\frac{1}{n}\,\mathrm{tr}(I + X^{-1/2}HX^{-1/2})\right)^n \\ &= n\log\left(\frac{1}{n}\,\mathrm{tr}(I + X^{-1/2}HX^{-1/2})\right) \\ &= n\log\left(1 + \frac{1}{n}\,\mathrm{tr}(X^{-1/2}HX^{-1/2})\right)\end{aligned}$$

Since $\log(1+t) \le t$ we arrive at

$$f(X+H) - f(X) \le \mathrm{tr}(X^{-1/2}HX^{-1/2}) = \mathrm{tr}(X^{-1}H)$$

This shows X^{-1} is a subgradient of f at X. Since f is differentiable, the subgradient is unique and equals the gradient $\nabla f(X)$.
Note: g is subgradient of $f(x)$ if for all $z \in \mathbf{dom} f$: $f(z) \ge f(x) + g^T(z - x)$

Second-order condition I

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

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

Second-order condition II

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

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

Examples

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

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

convex if $P \succeq 0$

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

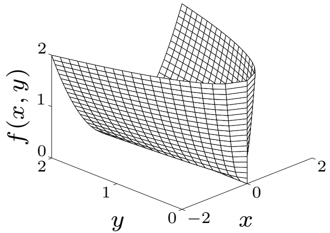
$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

convex (for any A)

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

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

convex for $y > 0$



Examples

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

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) = \begin{cases} \frac{-e^{x_i} e^{x_j}}{\left(\sum_j e^{x_j}\right)^2}, & i \neq j \\ \frac{-e^{x_i} e^{x_j}}{\left(\sum_j e^{x_j}\right)^2} + \frac{e^{x_j}}{\sum_j e^{x_j}}, & i = j \end{cases}$$

The Hessian matrix is then written as

$$\nabla^2 f(\boldsymbol{x}) = \mathrm{diag}(\mathbf{z}) - \mathbf{z}\,\mathbf{z}^T, \tag{1}$$

where $\mathbf{z} = [z_i]$, $z_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$ and $\boldsymbol{z}^T \mathbf{1} = 1$

Need to show that $\mathbf{v}^T \nabla^2 f(\boldsymbol{x})\, \mathbf{v} \geq 0$ for all $\mathbf{v}$.

$$\mathbf{v}^T \nabla^2 f(\boldsymbol{x})\, \mathbf{v} = \sum_i z_i v_i^2 - \left(\sum_i z_i v_i\right)^2$$

Following the Cauchy-Schwarz inequality we have

$$\left(\sum_i z_i v_i\right)^2 \leq \left(\sum_i z_i\right)\left(\sum_i z_i v_i^2\right) = \sum_i z_i v_i^2$$

Convex functions: examples

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)

Convex functions: examples I

- **Log-sum-exp** is convex

$$f(x) = \log(\exp^{x_1} + \exp^{x_2} + \cdots + \exp^{x_n})$$

and is approximation to the maximum $\max_i x_i$

$$\max\{x_1, \ldots, x_n\} \leq f(x) \leq \max\{x_1, \ldots, x_n\} + \log n$$

Hint Applying the mid-point rule to prove $f(x)$ is convex

$$\frac{1}{2}\log\left(\sum_i \exp^{x_i}\right) + \frac{1}{2}\log\left(\sum_i \exp^{y_i}\right) \geq \log\left(\sum_i \exp^{x_i/2+y_i/2}\right)$$

and substitute $\exp^{x_i/2} = a_i$ and $\exp^{y_i/2} = b_i$

$$\left(\sum_i a_i^2\right)^{1/2}\left(\sum_i b_i^2\right)^{1/2} \geq \sum a_i b_i$$

$$f(x) - y = \log(\exp^{x_1-y} + \exp^{x_2-y} + \cdots + \exp^{x_n-y})$$

Hence $f(x) - \max\{x_i\} \leq \log(\exp(0) + \cdots + \exp(0)) = \log n$.

Convex functions: examples I

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2 + \boxed{e^{x_1}}$$

$$\begin{aligned} f_1(x_1, x_2) &= x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2 \\ &= [x_1, x_2]\left[\begin{array}{cc} 1 & 1 \\ 1 & 3 \end{array}\right][x_1, x_2]^T + [2, -3][x_1, x_2]^T \end{aligned}$$

$f_1(x_1, x_2)$ is convex because its Hessian is positive definite

$$\mathbf{H} = \left[\begin{array}{cc} 1 & 1 \\ 1 & 3 \end{array}\right] > 0$$

$f(x_1, x_2)$ is sum of two convex functions, hence it is convex.

Convex functions: examples II

$$f(x, y, z) = x^2 + y^2 + 3z^2 - xy + 2xz + yz$$
$$= \frac{1}{2}[x, y, z]\begin{bmatrix} 2 & -1 & 2 \\ -1 & 2 & 1 \\ 2 & 1 & 6 \end{bmatrix}[x, y, z]^T$$

The Hessian $\mathbf{H} > \mathbf{0}$ is positive definitive because its leading principal minors are $2 > 0$, $3 > 0$, and $4 > 0$. Hence the function is strictly convex.

$$f(x_1, x_2) = x_1^3 + 2x_1^2 + 2x_1x_2 + (1/2)x_2^2 - 8x_1 - 2x_2 - 8$$

$$f(x_1, x_2, x_3) = e^{x_1 - x_2 + x_3} + e^{2x_2} + x_1$$

$$f(x_1, x_2) = -\log(x_1x_2) \qquad \text{over } R^2_{++},$$

Monotonicity of the Gradient I

Suppose that f is a continuously differentiable function over a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^T(\boldsymbol{x} - \boldsymbol{y}) \geq 0 \qquad \text{for any } \boldsymbol{x}, \boldsymbol{y} \in C$$

Proof

- Sufficiency:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x})$$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y})$$

$$f(\boldsymbol{x}) + f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + f(\boldsymbol{y}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y})$$

or

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^T(\boldsymbol{x} - \boldsymbol{y}) \geq 0$$

Monotonicity of the Gradient II

- Necessity : if ∇f is monotone, let $g(t) = f(x + t(y - x))$, and $g'(t) \geq g'(0)$ for all $t \geq 0$ and $t \in \operatorname{dom} g$

$$g'(t) = \nabla f(x + t(y - x))^T (y - x)$$

Hence

$$\begin{aligned} f(y) = g(1) &= g(0) + \int_0^1 g'(t)dt \geq g(0) + g'(0) \\ &= f(x) + \nabla f(x)^T (y - x) \end{aligned}$$

Monotonicity of the Gradient

σ-Strongly Convex Function:

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^T (\boldsymbol{x} - \boldsymbol{y}) \geq \sigma \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \qquad \text{for any } \boldsymbol{x}, \boldsymbol{y} \in \textbf{dom} f$$

Optimality conditions

Consider an unconstrained optimization problem

$$\min \quad f(x)$$

where f is convex and differentiable.
Then, any point x^* that satisfies $\nabla f(x^*) = 0$ is a global minimum.
Proof: From the 1st order characterization of convexity, since $\nabla f(x^*) = 0$, we get

$$f(y) \geq f(x^*) + \nabla f(x^*)(y - x^*) = f(x^*), \quad \forall y.$$

In absence of convexity, $\nabla f(x) = 0$ is not sufficient even for local optimality.
For example, $f(x) = x^3$ and and $x = 0$.

Optimality conditions

Consider an optimization problem

$$\min \quad f(x) \quad \text{s.t.} \quad x \in \Omega$$

where $f : \mathbb{R}^n \to R$ is convex and differentiable and Ω is convex. Then a point $x^\star$ is optimal if and only if $x^\star \in \Omega$ and

$$\nabla f(x^\star)^T (y - x^\star) \geq 0, \quad \forall y \in \Omega.$$

Proof

(Sufficiency) Suppose $x \in \Omega$ satisfies

$$\nabla f(x)^T (y - x) \geq 0, \quad \forall y \in \Omega$$

By the 1st order characterization of convexity

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \forall y \in \Omega$$

it is obvious that $f(y) \geq f(x)$ for all $y \in \Omega$. Hence x is optimal.

Optimality conditions

Consider an optimization problem

$$\min \quad f(x) \quad \text{s.t.} \quad x \in \Omega$$

where $f : \mathbb{R}^n \to R$ is convex and differentiable and Ω is convex. Then a point x is optimal if and only if $x \in \Omega$ and

$$\nabla f(x)^T (y - x) \geq 0, \quad \forall y \in \Omega.$$

Proof

(Necessity) Suppose x is optimal but for some $y \in \Omega$ we had

$$\nabla f(x)^T (y - x) < 0\,.$$

For $\theta \in [0, 1]$, $x + \theta(y - x) \in \Omega$. Let $g(\theta) = f(x + \theta(y - x))$.

$$g'(\theta) = (y - x)^T \nabla f(x + \theta(y - x))$$

Hence $g'(0) = (y - x)^T \nabla f(x) < 0$, implies that $\exists \delta > 0$ such that $g(\theta) < g(0)$, $\forall \theta \in (0, \delta)$.

$$f(x + \theta(y - x)) < f(x), \quad \forall \theta \in (0, \delta)$$

This contradicts the optimality of x.

Optimization problem with linear constraints

Consider the optimization problem

$$\min f(\boldsymbol{x}) \qquad \text{s.t.} \quad \mathbf{A}\boldsymbol{x} = \boldsymbol{b}$$

where f is a convex function and $\mathbf{A} \in \mathbb{R}^{m \times n}$.
A point $\boldsymbol{x} \in \mathbb{R}^n$ is optimal if and only if it is feasible and $\exists \boldsymbol{\mu} \in \mathbb{R}^m$ s.t.

$$\nabla f(\boldsymbol{x}) = \mathbf{A}^T \boldsymbol{\mu}\,.$$

Optimization problem with linear constraints

Consider the optimization problem

$$\min f(\boldsymbol{x}) \qquad \text{s.t.} \quad \mathbf{A}\boldsymbol{x} = \boldsymbol{b}$$

where f is a convex function and $\mathbf{A} \in \mathbb{R}^{m \times n}$.
A point $\boldsymbol{x} \in \mathbb{R}^n$ is optimal if and only if it is feasible and $\exists \boldsymbol{\mu} \in \mathbb{R}^m$ s.t.

$$\nabla f(\boldsymbol{x}) = \mathbf{A}^T \boldsymbol{\mu}\,.$$

Proof: Assume that $\boldsymbol{x}^\star$ is a solution, then from the optimality condition

$$\nabla f(\boldsymbol{x}^\star)^T (\boldsymbol{y} - \boldsymbol{x}^\star) \geq 0 \quad \forall\, \boldsymbol{y} : \mathbf{A}\boldsymbol{y} = \boldsymbol{b}.$$

Let $\boldsymbol{y} = \boldsymbol{x}^\star + \boldsymbol{v}$. Since $\boldsymbol{b} = \mathbf{A}\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^\star + \mathbf{A}\boldsymbol{v}$, we have $\mathbf{A}\boldsymbol{v} = 0$.

$$\nabla f(\boldsymbol{x}^\star)^T \boldsymbol{v} \geq 0 \quad \forall\, \boldsymbol{v} \text{ and } -\boldsymbol{v} : \mathbf{A}\boldsymbol{v} = 0.$$

Hence $\nabla f(\boldsymbol{x}^\star)^T \boldsymbol{v} = 0$. In other words,

$$\nabla f(\boldsymbol{x}^\star) = \mathbf{A}^T \boldsymbol{\mu}\,.$$

Optimization problem with nonnegativity constraints

Consider the optimization problem

$$\min f(\boldsymbol{x}) \qquad \text{s.t.} \quad \boldsymbol{x} \geq 0$$

where f is a convex function.
$\boldsymbol{x}^\star$ is a stationary point iff $\nabla f(\boldsymbol{x}^\star)^T(\boldsymbol{x} - \boldsymbol{x}^\star) \geq 0$ for all $\boldsymbol{x} \geq 0$
$\iff$ $\nabla f(\boldsymbol{x}^\star) \geq 0$ and $\nabla f(\boldsymbol{x}^\star)^T x^\star \leq 0$
$\iff$ $\nabla f(\boldsymbol{x}^\star) \geq 0$ and $x_i^\star \frac{\partial f}{\partial x_i}(\boldsymbol{x}^\star) = 0$
$\iff$ $\frac{\partial f}{\partial x_i}(x^\star) \begin{cases} = 0 & x_i^\star > 0 \\ \geq 0 & x_i^\star = 0, \end{cases}$

Optimization problem over unit Ball

Consider the optimization problem

$$\min f(\boldsymbol{x}) \qquad \text{s.t.} \quad \|\boldsymbol{x}\| \le 1$$

where f is a continuously differentiable convex function.
The point $\boldsymbol{x}^\star \in B[0,1]$ is a stationary point iff for all $\boldsymbol{x} \in B$

$$\nabla f(\boldsymbol{x}^\star)^T(\boldsymbol{x} - \boldsymbol{x}^\star) \ge 0 \quad \Longleftrightarrow \quad \boxed{\nabla f(\boldsymbol{x}^\star)^T \boldsymbol{x}} \ge \nabla f(\boldsymbol{x}^\star)^T \boldsymbol{x}^\star$$

Assume $\|\nabla f(\boldsymbol{x}^\star)\| \ne 0$, LHS attains minimum when $\boldsymbol{x} = -\frac{\nabla f(\boldsymbol{x}^\star)}{\|\nabla f(\boldsymbol{x}^\star)\|}$, implying that

$$-\|\nabla f(\boldsymbol{x}^\star)\| = \nabla f(\boldsymbol{x}^\star)^T \boldsymbol{x}^\star \quad \Longleftrightarrow \quad \|\nabla f(\boldsymbol{x}^\star)\| \left(\frac{\nabla f(\boldsymbol{x}^\star)^T}{\|\nabla f(\boldsymbol{x}^\star)\|} \boldsymbol{x}^\star + 1 \right) = 0$$

which leads to $\boldsymbol{x}^\star = -\frac{\nabla f(\boldsymbol{x}^\star)}{\|\nabla f(\boldsymbol{x}^\star)\|}$, i.e., $\|\boldsymbol{x}^\star\|_2 = 1$ and $\nabla f(\boldsymbol{x}^\star) = \lambda \boldsymbol{x}^\star$ for $\lambda < 0$.
Finally, the optimality condition is

- $\nabla f(\boldsymbol{x}^\star) = 0$ or
- $\|\boldsymbol{x}^\star\| = 1$ and $\nabla f(\boldsymbol{x}^\star) = \lambda \boldsymbol{x}^\star$ where $\lambda < 0$

Stationary Condition

Feasible set	Stationary condition
$\mathbb{R}^n$	$\nabla f(\boldsymbol{x}^\star) = 0$
$\mathbb{R}^n_+$	$\frac{\partial}{\partial x_i^*} f(\boldsymbol{x}^\star) \begin{cases} = 0, & x_i^* > 0, \\ \geq 0, & x_i^* = 0 \end{cases}$
$\{\boldsymbol{x} \in \mathbb{R}^n : \mathbf{A}\boldsymbol{x} = \boldsymbol{b}\}$	$\nabla f(\boldsymbol{x}^\star) = \mathbf{A}^T \nu$
$\{\boldsymbol{x} \in \mathbb{R}^n : \mathbf{1}_n^T \boldsymbol{x} = 1\}$	$\frac{\partial}{\partial x_1^*} f(\boldsymbol{x}^\star) = \cdots = \frac{\partial}{\partial x_n^*} f(\boldsymbol{x}^\star)$
$B[0, 1]$	$\nabla f(\boldsymbol{x}^\star) = 0$ or $\lVert\boldsymbol{x}^\star\rVert = 1$ and $\nabla f(\boldsymbol{x}^\star) = \lambda \boldsymbol{x}^\star$ where $\lambda < 0$

Epigraphs

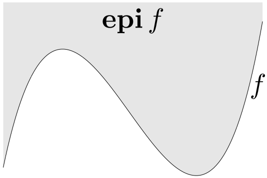
α-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$:

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

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

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

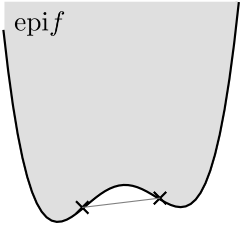
$$\mathbf{epi}\, f = \{(x, t) \in \mathbf{R}^{n+1} \mid x \in \mathbf{dom}\, f,\ f(x) \leq t\}$$

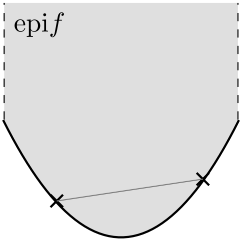


f is convex if and only if $\mathbf{epi}\, f$ is a convex set

Epigraphs and Convexity

- Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$
- Then f is convex if and only $\mathrm{epi} f$ is a convex set in $\mathbb{R}^n \times \mathbb{R}$





- f is called closed (lower semi-continuous) if $\mathrm{epi} f$ is closed set

Epigraphs and Convexity

A function f is convex if and only if its epigraph is a convex set.

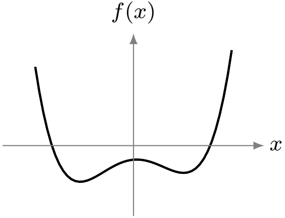
- Suppose f is convex and let $(x_1, t_1), (x_2, t_2) \in \mathrm{epi} f$.
  For $0 \leq \theta \leq 1$, the point $(x, t) = \theta(x_1, t_1) + (1 - \theta)(x_2, t_2)$ has

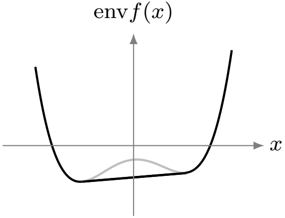
$$\begin{aligned} t &= \theta t_1 + (1 - \theta)t_2 \geq \theta f(x_1) + (1 - \theta) f(x_2) \\ &\geq f(\theta x_1 + (1 - \theta)x_2) = f(x) \end{aligned}$$

  Hence $(x, t) \in \mathrm{epi} f$, and $\mathrm{epi} f$ is convex.
- The converse is similar.

Convex Envelope

- Convex envelope of f is largest convex minorizer



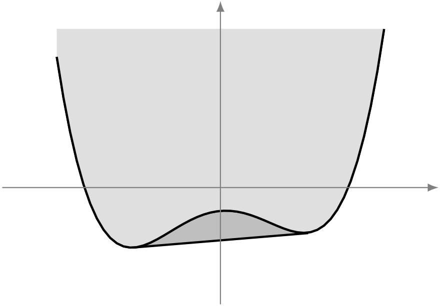


- Definition: The convex envelope $\mathrm{env} f$ satisfies: $\mathrm{env} f$ convex,

$$\mathrm{env} f \le f \qquad \text{and} \qquad \mathrm{env} f \ge g \text{ for all convex } g \le f$$

Convex Envelope

- Epigraph of convex envelope of f is convex hull of $\mathrm{epi} f$



- $\mathrm{epi} f$ in light gray, $\mathrm{epi}\,\mathrm{env} f$ includes dark gray

Jensen's inequality

basic 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

$$\mathbf{prob}(z = x) = \theta, \qquad \mathbf{prob}(z = y) = 1 - \theta$$

Jensen's inequality

- By convexity of $-\log x$ and Jensen's inequality with general θ

$$-\log(\theta a + (1-\theta)b) \le -\theta \log(a) - (1-\theta)\log(b)$$

$$a^\theta b^{1-\theta} \le \theta a + (1-\theta)b$$

- **Example** Prove that the set $C = \{x \in \mathbb{R}^n \mid x_1 x_2 \cdots x_n \ge a\}$ is convex.

$$\begin{aligned}
f(\theta x + (1-\theta)y) &= \prod_{k=1}^{n}(\theta x_k + (1-\theta)y_k) \\
&\ge \prod_{k=1}^{n} x_k^\theta y_k^{(1-\theta)} \\
&= \left(\prod_{k=1}^{n} x_k\right)^\theta \left(\prod_{k=1}^{n} y_k\right)^{(1-\theta)} \\
&\ge a^\theta a^{1-\theta} = a\,.
\end{aligned}$$

Arithmetic-geometric mean inequality

For all $x \in \mathbb{R}^n_{++}$

$$\left(\prod_i x_i\right)^{1/n} \leq \frac{1}{n}\left(\sum_i x_i\right)$$

Note: $\log(x)$ is concave on $\mathbb{R}_{++}$. Hence

$$\log\left(\sum_i \theta_i x_i\right) \geq \sum_i \theta_i \log(x_i) = \log\left(\prod_i x_i^{\theta_i}\right)$$

Jensen's inequality: example

• **Maximum of a convex function over a polyhedron.**
Show that the maximum of a convex function f over the polyhedron $P = \text{conv}\{v_1, \ldots, v_k\}$ is achieved at one of its vertices, i.e.,

$$\max_{x \in P} f(x) = \max_{i=1,\ldots,k} f(v_i)$$

Hint: Assume the statement is false, and $f(x^\star \in P)$ has a global maximum at a point $x^\star$.
Since $x^\star \in P$, $x^\star = \sum_{n=1}^{k} \theta_n v_n$, $\theta^T \mathbf{1}_k = 1$.
Denote $f* = \max_i f(v_i)$.

$$f(x^\star) \leq \sum_n \theta_n f(v_n) \leq \left(\sum_n \theta_n \right) f* = \max_{i=1,\ldots,k} f(v_i)\,.$$

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

Power of a nonnegative function

• If f is convex and nonnegative, i.e., $f(x) \geq 0, \forall x$, and $k \geq 0$, then f^k is convex.
Proof Consider the case when f is twice differentiable. Let $g = f^k$.

$$\begin{aligned} \nabla g(x) &= k f^{k-1} \nabla f(x) \\ \nabla^2 g(x) &= k((k-1) f^{k-2} \nabla f(x) \, \nabla^T f(x) + f^{k-1} \nabla^2 f(x)). \end{aligned}$$

$\nabla^2 g(x) > 0$

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

examples

- log barrier for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \qquad \mathbf{dom}\, f = \{x \mid a_i^T x < b_i, i = 1, \ldots, m\}$$

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

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: $f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$ 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 ith largest component of x)

  proof:

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

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

- 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 $X \in \mathbf{S}^n$,

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

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 $\begin{array}{l} g \text{ convex, } h \text{ convex, } \tilde{h} \text{ nondecreasing} \\ g \text{ concave, } h \text{ convex, } \tilde{h} \text{ nonincreasing} \end{array}$

- 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 $\tilde{h}$

examples

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

Vector composition

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

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

f is convex if g_i convex, h convex, $\tilde{h}$ nondecreasing in each argument
g_i concave, h convex, $\tilde{h}$ nonincreasing in each argument

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

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

examples

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

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

examples

- $f(x, y) = x^T A x + 2x^T B y + y^T C y$ with

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

  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

Perspective

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

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

g is convex if f is convex

examples

- $f(x) = x^T x$ is convex; 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

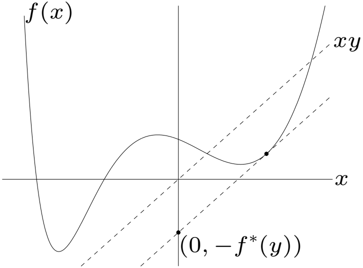
$$g(x) = (c^T x + d) f\left((Ax+b)/(c^T x + d)\right)$$

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

Conjugate function

the **conjugate** of a function f is

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



• $f^*(y)$ tells how far to shift a supporting hyperplane with directional slopes y so that it barely touches the graph of f
• $f^*(y)$ is a convex funtion.

Example

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

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

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

$$\begin{aligned} f^*(y) &= \sup_x (y^T x - (1/2)x^T Q x) \\ &= \frac{1}{2} y^T Q^{-1} y \end{aligned}$$

Conjugate of norm

Conjugate of norm is the indicator function of dual norm ball.
For a norm $||x||$, its dual norm $||x||_* = \sup_{||u|| \le 1} u^T x$.
• Let $x = ||x||u$ where $||u|| = 1$.

$$
\begin{aligned}
f^*(y) &= \sup_x (y^T x - ||x||) = \sup_{||x||} ||x|| \sup_{||u||=1} (y^T u - ||u||) \\
&= \sup_{||x||} ||x||(||y||_* - 1) \\
&= \begin{cases} 0, & ||y||_* \le 1 \\ \infty, & ||y||_* > 1 \end{cases}
\end{aligned}
$$

Let $||\cdot||$ be a norm on $\mathbb{R}^n$. The associated dual norm
$||z||_* = \sup\{z^T x : ||x|| \le 1\}$.
(support function of the unit ball of the original norm)

Example

	Function	Conjugate
Quadratic	$\frac{1}{2}x^2$	$\frac{1}{2}y^2$
Exponential	$\exp(x)$	$y(\log(y) - 1)$
Log	$-\log(x)$	$-\log(-y) - 1$
Log-exponential	$\log(1 + \exp(x))$	$[1 - y]\log(1 - y) + y\log y$
Cross entropy	$x\log(x)$	$\exp(y - 1)$
Affine	$a^T x - b$	b
Norm	$\|x\|$	$I_{\|\cdot\|_* \leq 1}(y)$

Conjugate of Indicator function I

I_C is the indicator function for the set C. Conjugate of I_C is the support function of the same set

$$I_C^\star(y) = \sup_{x \in C} \langle x, y \rangle \tag{2}$$

Let $y \in \mathbf{dom} I_C^\star$. For any $x \in C$

$$\langle x, y \rangle - I_C(x) = x^T y$$

For any $x \notin C$

$$\langle x, y \rangle - I_C(x) = -\infty$$

Hence

$$\sup\{\langle x, y \rangle - f(x)\} = \sup_{x \in C} \langle x, y \rangle$$

Convex Quadratic I

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \mathbf{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$$

where $\mathbf{A} \in \mathbb{S}^n_+$, $\boldsymbol{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$
Conjugate of $f(\boldsymbol{x})$ is

$$f^\star(\boldsymbol{y}) = \begin{cases} \frac{1}{2}(\boldsymbol{y}-\boldsymbol{b})^T\mathbf{A}^\dagger(\boldsymbol{y}-\boldsymbol{b}) - c\,, & \boldsymbol{y} \in \boldsymbol{b} + \text{range}(\mathbf{A}), \\ \infty, & \text{otherwise} \end{cases}$$

Negative entropy over unit simplex I

$$f(\boldsymbol{x}) \triangleq \begin{cases} \sum_{i=1}^{n} x_i \ln x_i & \boldsymbol{x} \in \Delta_n \\ \infty & \text{otherwise} \end{cases}.$$

where $\Delta_n = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}^T \mathbf{1} = 1, \boldsymbol{x} \geq \mathbf{0}\}$.

$$f^*(\boldsymbol{y}) = \ln\left(\sum_{j=1}^{n} e^{y_j}\right)$$

Fenchen's Inequality

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

Proof: By definition of conjugate function

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

Thus

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

- $f(x) = \frac{1}{2}\|x\|_2^2 \quad \rightarrow \quad f^*(y) = \frac{1}{2}\|y\|_2^2$

$$x^T x + y^T y \geq 2x^T y$$

- $f(x) = \frac{1}{2}x^T Q x \quad \rightarrow \quad f^*(y) = \frac{1}{2}y^T Q^{-1} y$

$$x^T Q x + y^T Q^{-1} y \geq 2x^T y$$

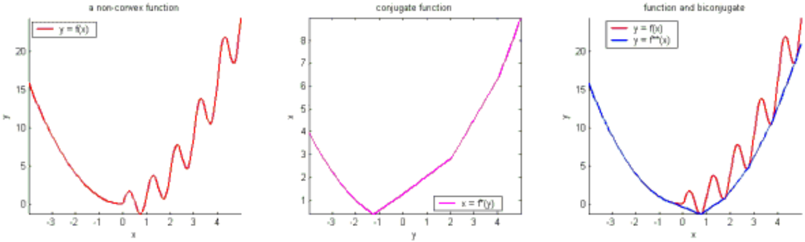
Bi-Conjugate function I

- The bi-conjugate function $f^{**}(x) = \sup_y (x^T y - f^*(y))$ is the maximal convex function that bounds the original function from below, $f^{**}(x) \leq f(x)$

Proof: From Fenchen's inequality $f(x) + f^*(y) \geq x^T y$,

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

- If $f(x)$ is convex, then it is its own bi-conjugate, $f(x) = f^{**}(x)$



(red line) a non-convex function,(middle) its convex conjugate (purple line), and (right) the original function (red) is shown together with its biconjugate (blue).

Bi-Conjugate function II

$f(x)$ is closed and convex, then $f^{**}(x) = f(x), \forall x, \quad \Longleftrightarrow \quad$ epi$f$ = epif^{**}

Proof: Assume that epi$f \neq$ epif^{**} and exists $(x, f^{**}(x)) \notin$ epi$f$.
Implying that there a hyperplane can separate $(x, f^{**}(x))$ and epi$f$, i.e.,

$$\begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} z \\ s \end{bmatrix} \leq 0 < \begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} x \\ f^{**}(x) \end{bmatrix}, \quad \forall [z, s]^T \in \text{epi} f(x).$$

Thus $\exists c$ s.t .

$$\begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} \leq c < 0.$$

Note that $b \leq 0$, otherwise, we can choose a significant large s.

- If $b < 0$. Let $y = \frac{a}{-b}$, then $(y^T z - s) - y^T x + f^{**}(x) \leq \frac{-c}{b} < 0$ for all $[z, s]^T \in$ epi$f(x)$.
  Following Fenchen's inequality $\frac{-c}{b} \geq f^*(y) + f^{**}(x) - y^T x \geq 0$.
  Contradiction!

Bi-Conjugate function III

- If $b = 0$, select $\varepsilon > 0$. Since $f^\star(y)$ is the supremum of $\tilde{y}^T x - s$ over all $\tilde{y} \in \mathbf{dom} f(y)$, we can choose $\tilde{y} \in \mathbf{dom} f^*(y)$: $f^*(y) \geq \tilde{y}^T z - s$ where $s = f(z)$

$$\begin{aligned}
\begin{bmatrix} a + \varepsilon\tilde{y} \\ -\varepsilon \end{bmatrix}^T \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} &\leq c + \varepsilon(\tilde{y}^T(z - x) - s + f^{**}(x)) \\
&\leq c + \varepsilon(f^*(y) - x^T y + f^{**}(x))
\end{aligned}$$

Since $c < 0$, we can always select sufficiently small $\varepsilon > 0$ such that

$$c + \varepsilon(f^*(y) - x^T y + f^{**}(x)) < 0$$

i.e.,

$$\begin{bmatrix} a + \varepsilon\tilde{y} \\ -\varepsilon \end{bmatrix}^T \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} < 0$$

To the previous case with $\tilde{a} = a + \varepsilon\tilde{y}$ and $\tilde{b} = -\varepsilon < 0$.

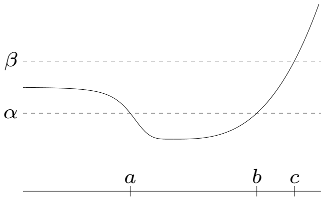
Quasiconvex functions I

$f : \mathbb{R}^n \to \mathbb{R}$ 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. In other words, f is quasiconvex if for all $x, y \in \mathbf{dom} f$ and $\theta \in [0, 1]$ we have

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



Suppose that $f(x)$ is quasiconvex, and let $x, y \in \mathbf{dom} f$. Then, for any $\theta \in [0, 1]$, we have:

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

Quasiconvex functions II

It means that the line segment joining x and y is contained in the level set S_a, i.e., S_a is convex, since it contains any line segment joining any two points in it.
Conversely, suppose that the level sets S_α are convex for all α, and let x and $y \in \mathbf{dom} f$. Then, for any $\theta \in [0, 1]$, we have:

$$\theta x + (1 - \theta)y \in S_{\max(f(x), f(y))}$$

This means that the point on the line segment joining x and y is in the level set $S_{\max(f(x),f(y))}$, which implies that $f(\theta x + (1 - (\theta)y) \leq \max(f(x), f(y))$.

Quasiconvex functions: example

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

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

  is quasilinear
- distance ratio

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

  is quasiconvex

Continuity of Convex Functions I

Let f be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$ and $\boldsymbol{x}_0 \in \text{int}(C)$. Then there exist $\varepsilon > 0$ and $L > 0$ s.t. the ball $B[\boldsymbol{x}_0, \varepsilon] \subseteq C$ and

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| \leq L\|\boldsymbol{x} - \boldsymbol{x}_0\| \qquad \text{for any } \boldsymbol{x} \in B[\boldsymbol{x}_0, \varepsilon]$$

Proof

- Take a sufficient small $\varepsilon > 0$ such that

$$B_\infty[\boldsymbol{x}_0, \varepsilon] = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_0\|_\infty \leq \varepsilon\} \subseteq C$$

- Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{2^n}$ be 2^n extreme points of $B_\infty[\boldsymbol{x}_0, \varepsilon]$.
- Any $\boldsymbol{x} \in B_\infty[\boldsymbol{x}_0, \varepsilon]$ can be represented by a convex combination $\boldsymbol{x} = \sum_i \lambda_i \mathbf{v}_i$.

$$f(\boldsymbol{x}) \leq \sum_i \lambda_i f(\mathbf{v}_i) \leq M$$

where $M = \max f(\mathbf{v}_i)$.

Continuity of Convex Functions II

- Since $B[\boldsymbol{x}_0, \varepsilon] = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 \leq \varepsilon\} \subseteq B_\infty[\boldsymbol{x}_0, \varepsilon]$, we conclude that $f(\boldsymbol{x}) \leq M$ for any $\boldsymbol{x} \in B[\boldsymbol{x}_0, \varepsilon]$
- Let $\boldsymbol{x} \in B[\boldsymbol{x}_0, \varepsilon]$, $\boldsymbol{x} \neq \boldsymbol{x}_0$ and $\alpha = \frac{1}{\varepsilon}\|\boldsymbol{x} - \boldsymbol{x}_0\|$.
- Since $\boldsymbol{z} = \boldsymbol{x}_0 + \frac{1}{\alpha}(\boldsymbol{x} - \boldsymbol{x}_0) \in B[\boldsymbol{x}_0, \varepsilon]$, then $f(\boldsymbol{z}) \leq M$.
- $\boldsymbol{x} = \alpha\boldsymbol{z} + (1-\alpha)\boldsymbol{x}_0$, $L = \frac{M - f(\boldsymbol{x}_0)}{\varepsilon}$

$$
\begin{aligned}
f(\boldsymbol{x}) &\leq \alpha f(\boldsymbol{z}) + (1-\alpha) f(\boldsymbol{x}_0) = f(\boldsymbol{x}_0) + \alpha(f(\boldsymbol{z}) - f(\boldsymbol{x}_0)) \\
&\leq f(\boldsymbol{x}_0) + \alpha(M - f(\boldsymbol{x}_0)) \\
&= f(\boldsymbol{x}_0) + \frac{M - f(\boldsymbol{x}_0)}{\varepsilon}\|\boldsymbol{x} - \boldsymbol{x}_0\| \\
&= f(\boldsymbol{x}_0) + L\|\boldsymbol{x} - \boldsymbol{x}_0\|
\end{aligned}
$$

Continuity of Convex Functions III

- Define $\boldsymbol{u} = \boldsymbol{x}_0 + \frac{1}{\alpha}(\boldsymbol{x}_0 - \boldsymbol{x})$. Since $\boldsymbol{u} \in B[\boldsymbol{x}_0, \varepsilon]$, then $f(\boldsymbol{u}) \leq M$. Note that $\boldsymbol{x}_0 = \frac{1}{\alpha+1}\boldsymbol{x} + \frac{\alpha}{\alpha+1}\boldsymbol{u} \implies$

$$f(\boldsymbol{x}_0) \leq \frac{1}{\alpha + 1} f(\boldsymbol{x}) + \frac{\alpha}{\alpha + 1} f(\boldsymbol{u})$$

  Therefore,

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{x}_0 + \alpha(\boldsymbol{x}_0 - \boldsymbol{u})) \geq f(\boldsymbol{x}_0) + \alpha(f(\boldsymbol{x}_0) - f(\boldsymbol{u})) \\ &\geq f(\boldsymbol{x}_0) - \frac{M - f(\boldsymbol{x}_0)}{\varepsilon} \|\boldsymbol{x} - \boldsymbol{x}_0\| = f(\boldsymbol{x}_0) - L\|\boldsymbol{x} - \boldsymbol{x}_0\| \end{aligned}$$

Log-concave and log-convex functions

a positive function f is log-concave if $\log f$ is concave:

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

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

- powers: x^a on $\mathbf{R}_{++}$ is log-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

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

Log-concave function: properties

- 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)

Convexity w.r.t. generalized inequalities

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

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

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

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.*,

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

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

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