Convex Functions

Cong Ma

University of Chicago, Winter 2026

Definition and Subclasses

Definition 1

$f : \mathcal{C} \to \mathbb{R}$ is **convex** if $\mathcal{C}$ is a convex set and

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

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ and $0 \leq \theta \leq 1$.

Strict Convexity: The inequality is strict $(<)$ for $\boldsymbol{x} \neq \boldsymbol{y}$ and $\theta \in (0,1)$, implying the graph lies strictly below the chord.

Strong Convexity: f is strongly convex $(m > 0)$ if $f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2$ is convex. Geometrically, f is lower-bounded by a quadratic bowl.

Catalogue of Convexity

On $\mathbb{R}$: e^{ax}, x^p ($p \geq 1$ or $p \leq 0$), and $x \log x$ (negative entropy).

On $\mathbb{R}^n$:

- **Affine:** $f(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} + b$ (both convex and concave).
- **Norms:** $\|\boldsymbol{x}\|_p$ ($p \geq 1$) and $\|\boldsymbol{x}\|_\infty = \max_k |x_k|$.
- **Least Squares:** $\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$ is convex ($\nabla^2 f = \boldsymbol{A}^\top \boldsymbol{A} \succeq \boldsymbol{0}$).

On $\mathbb{R}^{m \times n}$:

- **Affine:** $f(\boldsymbol{X}) = \mathsf{tr}(\boldsymbol{A}^\top \boldsymbol{X}) + b$.
- **Matrix Norms:** Spectral norm $\|\boldsymbol{X}\|_2$ and Frobenius norm $\|\boldsymbol{X}\|_\mathsf{F}$.
- **Log-Det:** $f(\boldsymbol{X}) = \log \det \boldsymbol{X}$ is **concave** on $\mathbb{S}^n_{++}$.

Convexity Restricted to a Line

Theorem 2

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if it is convex when restricted to any line that intersects its domain.

Specifically, f is convex if and only if for all $\boldsymbol{x} \in \mathrm{dom}\ f$ and all vectors $\boldsymbol{v} \in \mathbb{R}^n$, the univariate function

$$g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$$

is convex on its domain $\{t : \boldsymbol{x} + t\boldsymbol{v} \in \mathrm{dom}\ f\}$.

Restriction of a convex function to a line

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

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

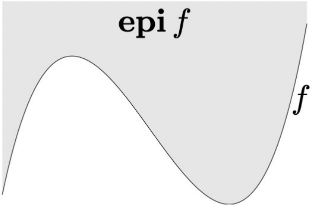
where λ_i are the eigenvalues of $X^{-1/2}VX^{-1/2}$

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

Epigraph

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

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



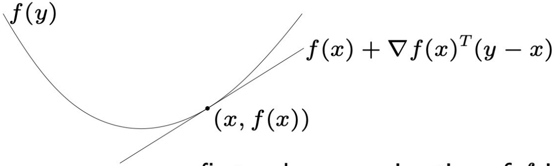
Theorem 3

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

First-order condition

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

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



first-order approximation of f is global underestimator

Proof

- Suppose first that f is convex. Let $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\lambda \in (0, 1]$. If $\boldsymbol{x} = \boldsymbol{y}$, then (1) trivially holds. We will therefore assume that $\boldsymbol{x} \neq \boldsymbol{y}$.
- $\frac{f(\boldsymbol{x}+\lambda(\boldsymbol{y}-\boldsymbol{x}))-f(\boldsymbol{x})}{\lambda} \leq f(\boldsymbol{y}) - f(\boldsymbol{x})$.
- Taking $\lambda \to 0^+$, we obtain

$$f'(\boldsymbol{x}; \boldsymbol{y} - \boldsymbol{x}) \leq f(\boldsymbol{y}) - f(\boldsymbol{x}).$$

- Since f is continuously differentiable, $f'(\boldsymbol{x}; \boldsymbol{y} - \boldsymbol{x}) = \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x})$, and (1) follows.

Proof Contd.

- To prove the reverse direction, assume that that the gradient inequality holds.
- Let $\boldsymbol{z}, \boldsymbol{w} \in C$, and let $\lambda \in (0,1)$. We will show that $f(\lambda \boldsymbol{z} + (1-\lambda)\boldsymbol{w}) \leq \lambda f(\boldsymbol{z}) + (1-\lambda) f(\boldsymbol{w})$.
- Let $\boldsymbol{u} = \lambda \boldsymbol{z} + (1-\lambda)\boldsymbol{w} \in C$. Then

$$\boldsymbol{z} - \boldsymbol{u} = \frac{\boldsymbol{u} - (1-\lambda)\boldsymbol{w}}{\lambda} - \boldsymbol{u} = -\frac{1-\lambda}{\lambda}(\boldsymbol{w} - \boldsymbol{u}).$$

- We have

$$\begin{aligned} f(\boldsymbol{u}) + \nabla f(\boldsymbol{u})^T(\boldsymbol{z} - \boldsymbol{u}) &\leq f(\boldsymbol{z}), \\ f(\boldsymbol{u}) - \frac{\lambda}{1-\lambda} \nabla f(\boldsymbol{u})^T(\boldsymbol{z} - \boldsymbol{u}) &\leq f(\boldsymbol{w}). \end{aligned}$$

- Thus,

$$f(\boldsymbol{u}) \leq \lambda f(\boldsymbol{z}) + (1-\lambda) f(\boldsymbol{w}).$$

Second-order conditions

2nd-order conditions: for twice differentiable f

- f is convex if and only if

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

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

Proof

Suppose that $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}\ \forall \boldsymbol{x} \in C$. Let $\boldsymbol{x}, \boldsymbol{y} \in C$, then $\exists \boldsymbol{z} \in [\boldsymbol{x}, \boldsymbol{y}] \in C$:

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2} (\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{x}).$$

$(\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{x}) \geq 0 \Rightarrow f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) \Rightarrow f$ convex.

Proof

Suppose that f is convex over C. Let $\boldsymbol{x} \in C$ and let $\boldsymbol{y} \in \mathbb{R}^n$.

C is open $\Rightarrow \exists \varepsilon > 0$ such that $\boldsymbol{x} + \lambda \boldsymbol{y} \in C \ \forall \lambda \in (0, \varepsilon)$.

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

$f(\boldsymbol{x} + \lambda \boldsymbol{y}) = f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^T \boldsymbol{y} + \frac{\lambda^2}{2} \boldsymbol{y}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} + o(\lambda^2 \|\boldsymbol{y}\|^2)$.

Thus, $\frac{\lambda^2}{2} \boldsymbol{y}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} + o(\lambda^2 \|\boldsymbol{y}\|^2) \geq 0$ for any $\lambda \in (0, \varepsilon)$.

Dividing by λ^2, $\frac{1}{2} \boldsymbol{y}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} + \frac{o(\lambda^2 \|\boldsymbol{y}\|^2)}{\lambda^2} \geq 0$.

Taking $\lambda \to 0^+$, we have $\boldsymbol{y}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} \geq 0 \ \forall \boldsymbol{y} \in \mathbb{R}^n$.

Hence $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}$ for any $\boldsymbol{x} \in C$.

Log-sum-exp

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

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

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

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

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

Jensen's Inequality

Theorem 4

Let $f : C \to \mathbb{R}$ be a convex function where $C \subseteq \mathbb{R}^n$ is a convex set. Then for any $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in C$ and $\boldsymbol{\lambda} \in \Delta_k$, the following inequality holds:

$$f\left(\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{k} \lambda_i f(\boldsymbol{x}_i).$$

Note: The domain C must be a convex set to ensure that the convex combination $\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ remains within the domain where f is defined.

Operations Preserving Convexity

Nonnegative Combinations: $f = \sum a_i f_i$ is convex if $a_i \geq 0$ and f_i convex.

Pointwise Maximum: $f(\boldsymbol{x}) = \sup_{s \in S} f_s(\boldsymbol{x})$ is convex if each f_s is convex. Examples include the sum of r largest components and the maximum eigenvalue $\lambda_{\max}(\boldsymbol{X})$.

Composition with Affine Map: $f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is convex if f is convex.

Partial Minimization: $g(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in C} f(\boldsymbol{x}, \boldsymbol{y})$ is convex if f is jointly convex in $(\boldsymbol{x}, \boldsymbol{y})$ and C is a convex set.

Nonnegative Linear Combinations and Affine Composition

Nonnegative Linear Combination: $f = \sum_{i=1}^{m} \alpha_i f_i$ is convex if all f_i are convex and $\alpha_i \geq 0$. This property extends to infinite sums and integrals.

Composition with Affine Function: $f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is convex if f is convex.

Examples:

- **Log barrier** for linear inequalities: $f(\boldsymbol{x}) = -\sum_{i=1}^{m} \log(b_i - \boldsymbol{a}_i^\top \boldsymbol{x})$ on $\text{dom}(f) = \{\boldsymbol{x} \mid \boldsymbol{a}_i^\top \boldsymbol{x} < b_i\}$.
- **Norm of affine function:** $f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|$ for any norm $\|\cdot\|$.

Pointwise Maximum and Supremum

If f_s is convex for each $s \in S$, then $f(\boldsymbol{x}) = \sup_{s \in S} f_s(\boldsymbol{x})$ is convex.

Examples:

Piecewise-linear function: $f(\boldsymbol{x}) = \max_{i=1,\dots,m}(\boldsymbol{a}_i^\top \boldsymbol{x} + b_i)$.

Sum of r largest components: $f(\boldsymbol{x}) = \boldsymbol{x}_{[1]} + \cdots + \boldsymbol{x}_{[r]}$ is the maximum of all $\binom{n}{r}$ combinations of r components.

Support function of a set C: $S_C(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} \boldsymbol{y}^\top \boldsymbol{x}$.

Distance to farthest point in a set C: $f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|$.

Maximum eigenvalue of symmetric matrix $\boldsymbol{X}$:
$\lambda_{\max}(\boldsymbol{X}) = \sup_{\|\boldsymbol{y}\|_2 = 1} \boldsymbol{y}^\top \boldsymbol{X} \boldsymbol{y}$.

Minimization and Distance to Sets

Partial Minimization: If $f(\boldsymbol{x}, \boldsymbol{y})$ is jointly convex in $(\boldsymbol{x}, \boldsymbol{y})$ and C is a convex set, then $g(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in C} f(\boldsymbol{x}, \boldsymbol{y})$ is convex.

Examples:

Schur Complement: For $f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + 2\boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{y} + \boldsymbol{y}^\top \boldsymbol{C}\boldsymbol{y}$ with $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^\top & \boldsymbol{C} \end{bmatrix} \succeq \boldsymbol{0}$ and $\boldsymbol{C} \succ \boldsymbol{0}$, the function $g(\boldsymbol{x}) = \inf_{\boldsymbol{y}} f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^\top)\boldsymbol{x}$ is convex.

Distance to a set: $d(\boldsymbol{x}, S) = \inf_{\boldsymbol{y} \in S} \|\boldsymbol{x} - \boldsymbol{y}\|$ is convex if S is convex.