Convex Functions (I)

Lecture 3, Convex Optimization

National Taiwan University

March 11, 2021

Table of contents

Open sets and closed sets

Open and closed can hold simultaneously e.g. $\{\emptyset\}$

Open Sets

A set $C \subseteq \mathbb{R}^n$ is said to be **open** if every element in C is an interior point.

Closed Sets

A set $C \subseteq \mathbb{R}^n$ is said to be **closed** if the complement of C, (i.e., $\mathbb{R}^n \setminus C$), is open.

- An alternative definition for closed sets is: a set $C \subseteq \mathbb{R}^n$ is said to be closed if every convergent sequence converges to a point in C. $(1,2)$ is open set. $\Rightarrow 1.1, 1.01, 1.001, \cdots, 1 \Rightarrow$ converge out of C

Examples of Open and Closed Sets

- The set $(1, 2) = \{x \in \mathrm{R} \mid 1 < x < 2\}$ is open.
- The set $[1, 2] = \{x \in \mathrm{R} \mid 1 \leq x \leq 2\}$ is closed.
- The empty set $\varnothing$ is open. It is also closed.
- The set R^n is open. It is also closed.

Properties of Closed Sets

Properties of Closed Sets

A set $C \subseteq \mathrm{R}^n$ is closed if and only if the limit point of every convergent sequence is in C. $\Leftrightarrow$ R^n/C is open

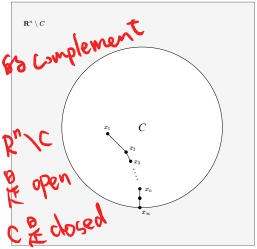
Proof: "only if": Suppose $\{x_k\}, k = 1, 2, ...$ is a convergent sequence in C. Then, for any $\epsilon > 0$, $\exists N_0 \in \mathbb{N}$ such that $\forall n_1, n_2 > N_0$, $||x_{n_1} - x_{n_2}||_2 < \epsilon$. Then, there exists uniquely a limit point $x_\infty \in \mathrm{R}^n$, such that $\forall \epsilon > 0, \exists n_0 \in \mathbb{N}$ s.t. $\forall n_1 > n_0$,

$$||x_{n_1} - x_\infty||_2 < \epsilon. \quad (1)$$

表 x∞ 在 C 的 complement
⇓
表 $R^n \setminus C$
是 open
⇒ C 是 closed

Suppose x_∞ is NOT in C. Then $x_\infty \in \mathrm{R}^n \setminus C$ instead. Then x_∞ must be an interior point in $\mathrm{R}^n \setminus C$ (since it is open). And this means $\exists \epsilon > 0$, the ball $B(x_\infty, \epsilon) \subseteq \mathrm{R}^n \setminus C$. But this is contradictory to (1).

"if": Suppose C is not closed. Then $\mathrm{R}^n \setminus C$ is not open, and $\exists x_0 \in \mathrm{R}^n \setminus C$ that is not in int $(\mathrm{R}^n \setminus C)$. Thus, $\forall \epsilon > 0$, $B(x_0, \epsilon) \nsubseteq \mathrm{R}^n \setminus C$, i.e., $B(x_0, \epsilon) \bigcap C \neq \varnothing, \forall \epsilon > 0$. Construct a sequence $\{x_n\}$ and let the nth point $x_n \in C$ be chosen as any point in $B(x_0, \frac{1}{n}) \bigcap C$. Then, $\lim_{n \to \infty} x_n = x_0 \notin C$.



Closure

Closure

The **closure** of a set $C \subseteq \mathbb{R}^n$ is defined as

把它擴張成 min close set

不一定 open ∵ C不一定是 close

$$\textbf{cl } C = \big(\mathbb{R}^n \setminus \underbrace{\text{int } (\mathbb{R}^n \setminus C)}_{\text{open}}\big)$$

close

- A point $x \in \textbf{cl } C$ if $\forall \epsilon > 0$, $\exists y \in C$ such that $||x - y||_2 \leq \epsilon$.
- It can be shown that the closure of C is the set of all the limit points of convergent sequences in C.

Boundary

Boundary

The **boundary** of the set $C \subseteq \mathrm{R}^n$ is defined as

$$\textbf{bd } C = \textbf{cl } C \setminus \text{int } C.$$

- $\forall \epsilon > 0$, $\exists y \in C$ and $z \notin C$ s.t.

$$||y - x||_2 \leq \epsilon, \quad ||z - x||_2 \leq \epsilon.$$

- A set $C \in \mathrm{R}^n$ is closed if it contains its boundary: $\textbf{bd } C \subseteq C$.
- A set $C \in \mathrm{R}^n$ is open if it contains no boundary points: $C \bigcap \textbf{bd } C = \varnothing$.

Infimum and Supremum

↳ min in set ↳ max in set ∵ Set不一定有 max/min (e.g. (0.9, 0.99, ⋯)), ∴ 用 sup/inf 来處理 max/min

Supremum

Suppose $C \subseteq \mathrm{R}$. A number a is an **upper bound** on C if for each $x \in C, x \leq a$. The number b is called the **least upper bound** or **supremum** of the set C, and is denoted $\sup C$.

Note: The set of upper bounds on a set C is either empty (in which case we say C is **unbounded above**), all of R (only when $C = \varnothing$), or a closed infinite interval $[b, \infty)$.
We take $\sup \varnothing = -\infty$, and $\sup C = \infty$ if C is unbounded above. When $\sup C \in C$, we say the supremum of C is **attained** or **achieved**.

Lower Bound and Infimum

A number a is a **lower bound** on $C \subseteq \mathrm{R}$ if for each $x \in C, a \leq x$. The **infimum** (or **greatest lower bound**) of a set $C \subseteq \mathrm{R}$ is defined as $\inf C = -\sup(-C)$. When C is finite, the infimum is the minimum of its elements. We take $\inf \varnothing = \infty$, and $\inf C = -\infty$ if C is unbounded below, i.e., has no lower bound.

Derivative

Let $z = x + \varepsilon$, 希望能够近似 $f(z) \approx f(x) + Df(x)\varepsilon$, $Df(x) = R^{m\times n}$

Derivative

Suppose $f : \mathrm{R}^n \to \mathrm{R}^m$ and $x \in \mathbf{int\ dom}\ f$. The function f is **differentiable** at x if there exists a matrix $Df(x) \in \mathrm{R}^{m\times n}$ that satisfies

$$\lim_{z \in \mathbf{dom}\ f, z \neq x, z \to x} \frac{||f(z) - f(x) - Df(x)(z - x)||_2}{||z - x||_2} = 0.$$

The matrix $Df(x)$ is called the **derivative** (or **Jacobian**) of f at x.

- $Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}$, $\quad i = 1, \cdots, m, \quad j = 1, \cdots, n$, where $f_i : \mathrm{R}^n \to \mathrm{R}$ is the function such that $f_i(x)$ is the ith component of $f(x)$ for all $x \in \mathrm{R}^n$.

Gradient

Gradient

Row Vector

If f is real-valued (i.e., $f : \mathbf{R}^n \to \mathbf{R}$), the derivative $Df(x)$ is a $1 \times n$ matrix, i.e., it is a row vector. Its transpose, as a column vector in $\mathbf{R}^n$, is called the **gradient** of the function:

转成 column vector 比较好看

$$\nabla f(x) = Df(x)^T = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

The first-order approximation of f at a point $x \in \mathbf{int\,dom}\, f$ can be expressed as (the affine function of z)

$$f(z) \approx f(x) + \nabla f(x)^T (z - x).$$

Examples for Gradient

As a simple example, consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$,

$$f(x) = (1/2)x^T Px + q^T x + r,$$

where $P \in \mathbf{S}^n, q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. Then, its gradient is

$$\nabla f(x) = Px + q.$$

Gradient of Functions Defined on Symmetric Matrices (1/3)

Gradient of functions on symmetric matrices

If $f : \mathbf{S}^n \to \mathbf{R}$, then the **gradient** of f is defined as

$$\nabla f(X) = \mathrm{svec}^{-1}\left(\nabla \tilde{f}(\mathrm{svec}(X))\right)$$

where $\tilde{f} : \mathbf{R}^{n(n+1)/2} \to \mathbf{R}$ is defined from f as $\tilde{f}(x) \triangleq f\left(\mathrm{svec}^{-1}(x)\right)$.

one-to-one & onto linear mapping

svec^{-1} ⇒ isomorphism

$$\begin{tikzcd} \mathbf{R}^{n(n+1)/2} \arrow[r, shift left, "\mathrm{svec}^{-1}"] \arrow[rr, bend right, "\tilde{f}"'] & \mathbf{S}^n \arrow[l, shift left, "\mathrm{svec}"] \arrow[r, "f"] & \mathbf{R} \end{tikzcd}$$

$\parallel$

S^n

norm-preserving

⇕

inner-product preserving

Gradient of Functions Defined on Symmetric Matrices (2/3)

Gradient of functions on symmetric matrices

If $f : \mathbf{S}^n \to \mathbf{R}$, then the **gradient** of f is defined as

$$\nabla f(X) = \mathrm{svec}^{-1}\left(\nabla \tilde{f}(\mathrm{svec}(X))\right)$$

where $\tilde{f} : \mathbf{R}^{n(n+1)/2} \to \mathbf{R}$ is defined from f as $\tilde{f}(x) \triangleq f\left(\mathrm{svec}^{-1}(x)\right)$.

Gradient w.r.t. Symmetric Matrices

It can be shown that

still symmetric, 且 $x_{12} = x_{21}$

$$\nabla f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & & & \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} & \cdots & \frac{\partial f(X)}{\partial x_{nn}} \end{bmatrix} \in \mathbf{S}^n.$$

Gradient of Functions Defined on Symmetric Matrices (3/3)

$$\nabla f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & & & \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} & \cdots & \frac{\partial f(X)}{\partial x_{nn}} \end{bmatrix} \in \mathbf{S}^n.$$

- E.g. Let $f : \mathbf{S}^2 \to \mathbf{R}$ defined as $f\left(\begin{bmatrix} x & y \\ y & z \end{bmatrix}\right) = 2x + 2y + 3z$, then

affine function

∂x ∂y

$$\nabla f\left(\begin{bmatrix} x & y \\ y & z \end{bmatrix}\right) = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

∂y ∂z

- It can be shown that

$$\lim_{Z \in \mathbf{dom}\ f, Z \neq X, Z \to X} \frac{|f(Z) - f(X) - \mathbf{tr}(\nabla f(X)(Z - X))|}{||Z - X||_F} = 0.$$

Inner-product explain : $\underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{A} \cdot \begin{bmatrix} x & y \\ z & w \end{bmatrix} = ax+by+cz+dw$

$= tr\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix}\right) = tr\begin{bmatrix} ax+cz & ay+cw \\ bx+dz & by+dw \end{bmatrix}$

A^T

$= ax+cz + by+dw$

Gradient

linear affine function

- Let $f : \mathbf{S}^n \to \mathbf{R}$ be defined as $f(X) = \mathbf{tr}(AX)$ where $A \in \mathbf{R}^{n \times n}$. Then,

$$\nabla f(X) = \frac{A + A^T}{2}.$$

why ?

代 P.13 的公式

If $A \in \mathbf{S}^n$, then $\nabla f(X) = A$.

- The first-order approximation of f at a point $X \in \mathbf{int\,dom}\, f$ can be expressed as (the affine function of Z)

$$f(X) + \mathbf{tr}\left(\nabla f(X)(Z - X)\right).$$

Gradient of the Log-Determinant Function (1/2)

Gradient of the Log-Determinant Function

Consider the function $f : \mathbf{S}^n \to \mathbf{R}$ given by

$$f(X) = \log\det X, \quad \operatorname{dom} f = \mathbf{S}^n_{++}.$$

We will show that

$$\nabla f(X) = X^{-1}.$$

$n=1$. $(\log x)' = x^{-1}$ 很合理

且 dom logarithm $\in R_+$

Gradient of the Log-Determinant Function (2/2)

Notice that

$\Delta X = Z - X$, $\Delta X \in S^n$, $Z, X \in S^n_{++}$, $X = UDU^T$

$\Rightarrow X^{\frac{1}{2}} = U D^{\frac{1}{2}} U^T$

$X + \Delta X = X^{\frac{1}{2}} \cdot I \cdot X^{\frac{1}{2}} + X^{\frac{1}{2}} X^{-\frac{1}{2}} \cdot \Delta X \cdot X^{-\frac{1}{2}} \cdot X^{\frac{1}{2}}$

$$
\begin{aligned}
\log\det Z &= \log\det(X + \Delta X) \\
&= \log\det(X^{1/2}(I + X^{-1/2}\Delta X X^{-1/2})X^{1/2}) \quad \text{det } ABC = \det A \cdot \det B \cdot \det C \\
&= \log\det X + \log\det(I + X^{-1/2}\Delta X X^{-1/2}) \quad \text{assume 知道 eig. vals} \\
&= \log\det X + \sum_{i=1}^{n} \log(1 + \lambda_i) \quad \text{(where } \lambda_i\text{'s are eig.vals. of } X^{-1/2}\Delta X X^{-1/2}\text{)} \quad \text{加了 } I\text{, 所有 eig. vals } +1 \\
&\approx \log\det X + \sum_{i=1}^{n} \lambda_i = \text{trace} \\
&= \log\det X + \mathbf{tr}(X^{-1/2}\Delta X X^{-1/2}) \quad \mathrm{tr}(AB) = \mathrm{tr}(BA) \\
&= \log\det X + \mathbf{tr}(X^{-1}(Z - X)).
\end{aligned}
$$

$\because \Delta X$ 很小

So, $f(Z) \approx f(X) + \mathbf{tr}(X^{-1}(Z - X))$.
Therefore,

$$\nabla f(X) = X^{-1}.$$

Second derivative and Hessian matrix

Hessian matrix

Let $f : \mathrm{R}^n \to \mathrm{R}$. Then **second derivative** or **Hessian matrix** of f at $x \in \operatorname{int} \operatorname{dom} f$, denoted $\nabla^2 f(x)$, is given by

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

- The Hessian matrix is a symmetric matrix as long as we assume that any second derivatives of f are continuous.
- The second-order approximation of f, at or near x, is the quadratic function of z defined by

$$f(z) \approx f(x) + \nabla f(x)^T (z - x) + (1/2)(z - x)^T \nabla^2 f(x)(z - x).$$

Second derivative and Hessian matrix

- Note that $D\nabla f(x) = \nabla^2 f(x) = (\nabla^2 f(x))^T$.
- Consider again the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$,

$$f(x) = (1/2)x^T Px + q^T x + r,$$

where $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. Then, its gradient is

$$\nabla f(x) = Px + q.$$

So, its Hessian matrix is

$$\nabla^2 f(x) = P.$$

Chain Rules

- Let $f : \mathrm{R}^n \to \mathrm{R}^m$ and $g : \mathrm{R}^m \to \mathrm{R}^p$. Consider the composite function $h : \mathrm{R}^n \to \mathrm{R}^p$, with $\mathrm{dom}\ h = \mathrm{dom}\ f \cap f^{-1}(\mathrm{dom}\ g)$, defined as $h(z) = g(f(z))$ for any $z \in \mathrm{dom}\ h$.
- Then,

$$Dh(x) = Dg(f(x)) \cdot Df(x).$$

To see this, note that

$$\begin{aligned} h(z) &= h(x + \Delta x) \\ &= g(f(x + \Delta x)) \\ &\approx g(f(x) + Df(x)\Delta x) \\ &\approx g(f(x)) + Dg(f(x))Df(x)\Delta x \\ &= h(x) + Dh(x)(z - x) \end{aligned}$$

for any Δx with a sufficiently small $\|\Delta x\|$.

Definitions of Convex Functions

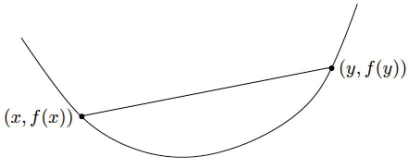
Convex functions

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

其實只是在看 line segment

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

- The line segment between $(x, f(x))$ and $(y, f(y))$, which is the **chord** from x to y, lies above the graph[1]



[1] The term graph will be formally defined in a later section.

Definitions of Convex Functions

Convex functions

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

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y). \tag{2}$$

- A function f is **strictly convex** if strict inequality holds in (2) whenever $x \neq y$ and $0 < \theta < 1$:

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

- We say f is **concave** if $-f$ is convex, and **strictly concave** if $-f$ is strictly convex.

Affine Functions

Affine Functions

For an affine function we always have equality in (2), i.e.,

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

so all affine functions are both convex and concave.

- Conversely, any function that is convex and concave is affine.

Convexity

需任意怎麼切
都 convex

*
- A function is convex if and only if it is convex when restricted to any line that intersects its domain. 只看 intersection lines的 dom 是否 convex
- That is, f is convex if and only if $\forall x \in \operatorname{dom} f, v \in \mathrm{R}^n$, the function $g(t) = f(x + tv)$ is convex on $\{t \mid x + tv \in \operatorname{dom} f\}$.
- A convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

Proof for "A convex function is continuous on the relative interior of its domain"

自己看

Suppose $f : \mathbb{R}^n \to \mathbb{R}$, with domain dom f, is convex. Then, we claim that f is continuous at every point in the interior of dom f:
Claim:

$$\forall x_0 \in \text{int dom } f, \forall \epsilon > 0, \exists \delta > 0 \text{ s.t. }, \forall x \in B(x_0, \delta), |f(x) - f(x_0)| < \epsilon$$

Proof: Since $x_0 \in$ int dom f, there exists $r > 0$ such that $B(x_0, r) \subseteq$ int dom f. Consider any line that passes x_0, i.e., $\mathcal{L}_a = \{x_0 + ta \mid t \in \mathbb{R}, x_0 + ta \in \text{int dom } f\}$ for any $a \in \mathbb{R}^n$ with $a \neq 0$. Without loss of generality, assume $||a||_2 = r$. We consider the function values at the two intersection points of $\mathcal{L}_a$ and **bd** $B(x_0, r)$(i.e., $x_0 \pm a$). Due to convexity of f, we have $f(x_0) \leq \frac{1}{2}(f(x_0 + a) + f(x_0 - a))$. Further, it can be shown that $\forall s \in [0, 1]$,

$$-s(f(x_0 - a) - f(x_0)) \leq f(x_0 + sa) - f(x_0) \leq s(f(x_0 + a) - f(x_0)).$$

Let

$$E = \sup_{x \in B(x_0, r)} |f(x) - f(x_0)|.$$

Then it is obvious that $|f(x_0 + a) - f(x_0)| \leq E$ for any $a \in \mathbb{R}^n$ with $||a||_2 = r$. Now, let us choose $\delta = \frac{r}{E} \cdot \epsilon$. Then, $\forall x \in B(x_0, \delta)$, we can show that $|f(x) - f(x_0)| \leq \epsilon$ by letting $s = ||x - x_0||_2/r$ (note that $s \leq \delta/r$) and $a = r \cdot \frac{x - x_0}{||x - x_0||_2}$, with which we obtain
$|f(x) - f(x_0)| = |f(x_0 + sa) - f(x_0)| \leq \max\{|s||f(x_0 + a) - f(x)|, |-s||f(x_0 - a) - f(x_0)|\} \leq sE \leq \frac{\delta}{r} \cdot E = \epsilon$.
So, the proof is complete.

Extended-Value Extensions

Extended-Value Extensions

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

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

∵開口向上，到某點就可能 over dom f
∴若 over dom f,
set to ∞

First-Order Conditions

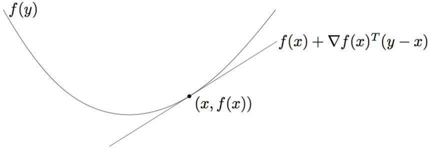
First-Order Conditions

Suppose f is differentiable (implying that dom f is open). Then f is convex if and only if dom f is convex and

$$f(y) \geq \underbrace{f(x) + \nabla f(x)^T (y - x)}_{P_1}.$$

若f是convex, 原先 f(y)≈-- 會使得 P_1 為 lower bound (underestimator)

- Observation: the first-order Taylor approximation is a global underestimator of the function.
- Conversely, if the first-order Taylor approximation of a function is always a global underestimator of the function, then the function is convex.



Remark: In this course, we do not define the derivative of a function at a boundary point.

First-Order Conditions

- A convex function f satisfies

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

  for all $x, y \in \operatorname{dom} f$.

只要有$f(x)$ and $\nabla f(x)^T$,就能保証 $f(y) \geq$ ——

- This shows that from local information about a convex function (i.e., $f(x), \nabla f(x)$), we can derive global information (i.e., a global underestimator).
- Example: if $\nabla f(x) = 0$, then for all $y \in \operatorname{dom} f$, $f(y) \geq f(x)$. (x is the global minimizer of f.)

  if f is convex

First-Order Conditions – Strict Convexity, Concavity

First-Order Conditions for strict convexity

f is strictly convex if and only if dom f is convex and for $x, y \in \text{dom } f, x \neq y$, we have

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

First-Order Conditions for (strict) concavity

f is concave if and only if dom f is convex and for $x, y \in \text{dom } f$, we have

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

f is strictly concave if and only if dom f is convex and for $x, y \in \text{dom } f, x \neq y$, we have

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

Proof of First-Order Conditions

Proof ideas:

- Consider the special case $n = 1$ first.
  - Then we only need to prove that f is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x).$$

- For the general case $f : \mathrm{R}^n \to \mathrm{R}$, with dom f convex, consider the line passing by any two points $x, y \in \operatorname{dom} f, x \neq y$, and define a function $g : \mathrm{R} \to \mathrm{R}$ with $g(t) = f(ty + (1 - t)x)$ and $\operatorname{dom} g = \{t \in \mathrm{R} \mid ty + (1 - t)x \in \operatorname{dom} f\}$.

if f convex $\Rightarrow$ g convex, then satisfy

$g(y) \geq g(x) + \cdots$

Proof of first-order convexity condition (1/2)

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

Proof. We first consider the case $n = 1$: a differentiable function $f : \mathrm{R} \to \mathrm{R}$ is convex if and only if $f(y) \geq f(x) + f'(x)(y - x)$ for all x and y in dom f. Assume first that f is convex and $x, y \in$ dom f. Since dom f is convex (i.e., an interval), we conclude that for all $0 < t \leq 1, x + t(y - x) \in$ dom f, and by convexity of f, $f(x + t(y - x)) \leq (1 - t)f(x) + tf(y)$. Dividing both sides by t, we obtain

$$f(y) \geq f(x) + \frac{f(x + t(y - x)) - f(x)}{t},$$

and taking the limit as $t \to 0$ yields (1).

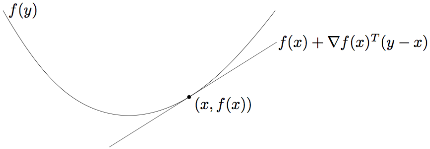
To show sufficiency, assume the function satisfies (1) for all x and y in dom f (which is an interval). Choose any $x \neq y$, and $0 \leq \theta \leq 1$, and let $z = \theta x + (1 - \theta)y$. Applying (1) twice yields $f(x) \geq f(z) + f'(z)(x - z), f(y) \geq f(z) + f'(z)(y - z)$. Multiplying the first inequality by θ, the second by $1 - \theta$, and adding them yields $\theta f(x) + (1 - \theta)f(y) \geq f(z)$, which proves that f is convex.

Proof of first-order convexity condition (2/2)

Now we can prove the general case, with $f : \mathrm{R}^n \to \mathrm{R}$. Let $x, y \in \mathrm{R}^n$ and consider f restricted to the line passing through them, i.e., the function defined by $g(t) = f(ty + (1-t)x)$, so $g'(t) = \nabla f(ty + (1-t)x)^T(y-x)$. First assume f is convex, which implies g is convex, so by the argument above we have $g(1) \geq g(0) + g'(0)$, which means $f(y) \geq f(x) + \nabla f(x)^T(y-x)$. Now assume that this inequality holds for any x and y, so if $ty + (1-t)x \in \mathrm{dom}\, f$ and $\tilde{t}y + (1-\tilde{t})x \in \mathrm{dom}\, f$, we have $f(ty + (1-t)x) \geq f(\tilde{t}y + (1-\tilde{t})x) + \nabla f(\tilde{t}y + (1-\tilde{t})x)^T(y-x)(t-\tilde{t})$, i.e., $g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t-\tilde{t})$. We have seen that this implies that g is convex.

Second-Order Conditions

- Assume that $f : \mathbf{R} \to \mathbf{R}$ is twice differentiable with $\mathbf{dom}\, f = \mathbf{R}$, then it is convex if and only if its second derivative is nonnegative.



Second-Order Conditions

$R^n \to R$

- Assume that f is twice differentiable, that is, its **Hessian** or second derivative $\nabla^2 f$ exists at each point in dom f (open).

Second-Order Conditions

Then, f is convex if and only if dom f is convex and its Hessian is positive semidefinite:

定義 matrix ≥ 0 ⇒ 只關心 eig.vals 是否都≥ 0

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \text{dom } f.$$

- For a function on R, this means $f''(x) \geq 0$, and dom f is convex.

都用 line 切法處理 ⇒ $g(t) = f(x+vt)$, $t \in \{t \mid x+vt \in \text{dom} f\}$

⇒ $x \in \text{dom} f$, $v \in R$ ⇒ 去處理 g

(Continue)

$g'(t) = \nabla f(x+vt)^T v \Rightarrow g''(t) = v^T \nabla^2 f(\underbrace{x+vt}_{y}) v$

$\because \nabla^2 f(y)$ is psd, psd matrix in quadratic form i.e. $v^T \nabla^2 f(y) v$,

will ≥ 0 (by psd's properties) $\Rightarrow g''(t) \geq 0 \Rightarrow g$ is convex

then f is convex

$f(x,y) = x^2 + 2xy + 3y + y^3$ Domain is convex

$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x + 2y \\ 2x + 3 + 3y^2 \end{bmatrix}$

$\underbrace{\nabla^2 f}_{\text{Hessian}} = \begin{bmatrix} \frac{\partial f}{\partial x \partial x} & \frac{\partial f}{\partial x \partial y} \\ \frac{\partial f}{\partial y \partial x} & \frac{\partial f}{\partial y \partial y} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6y \end{bmatrix}$

$\rightarrow$ PSD $\Rightarrow f$ is convex

eig. val

Second-Order Conditions – Strict Convexity, Concavity

Second-Order Conditions for Concavity

A function f is concave if and only if dom f is convex and $\nabla^2 f(x) \preceq 0$ for all $x \in \text{dom } f$.

Second-Order Conditions for Strict Convexity

If $\nabla^2 f(x) \succ 0$ for all $x \in \text{dom } f$ where dom f is convex, then f is strictly convex.

- If f is strictly convex, do we have $\nabla^2 f(x) \succ 0$? (e.g., think $f(x) = x^4$)

  不一定

  $f(x)=x^4$ is s-conex

  即若 convex $\Rightarrow$ R_{++}, R_{--} is domain, It's convex

- Is $f : \mathbf{R} \to \mathbf{R}$, $f(x) = 1/x^2$ a convex function? Why?

  $f''(x) = 6x^{-4} \geq 0 \Rightarrow x=0 \notin \text{dom } f$,

  but if dom $f = R \setminus \{0\}$, this function is not convex

Example – Quadratic Functions

- Consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^n$, given by

$$f(x) = (1/2)x^T Px + q^T x + r,$$

  with $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$.
    - Note that $\nabla^2 f(x) = P$.
- The function f is convex if and only if $P \succeq 0$.
- The function f is concave if and only if $P \preceq 0$.
- The function f is strictly convex if and only if $P \succ 0$.
- The function f is strictly concave if and only if $P \prec 0$.

Example Convex Functions on R

- Exponential: e^{ax} is convex on R, for any $a \in$ R.
- Powers: x^a is convex on R$_{++}$ when $a \geq 1$ or $a \leq 0$; it is concave when $0 \leq a \leq 1$.
- Powers of absolute value: $|x|^p$ with $p \geq 1$ is convex on R.
- Logarithm: $\log x$ is concave on R$_{++}$.
- Negative entropy: $x \log x$ is convex on R$_{++}$ (and also on R$_{+}$ if defined as 0 for $x = 0$).

Example Convex Functions on $\mathbb{R}^n$

- Norms. Every norm on $\mathbb{R}^n$ is convex.
- Max function. $f(x) = \max\{x_1, \cdots, x_n\}$ is convex on $\mathbb{R}^n$.
- Quadratic-over-linear function. The function $f(x, y) = x^2/y$, with $\operatorname{dom} f = \mathbb{R} \times \mathbb{R}_{++} = \left\{(x, y) \in \mathbb{R}^2 \mid y > 0\right\}$, is convex.
- Log-sum-exp. The function $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$.
  - Note that $\max\{x_1, \cdots, x_n\} \leq f(x) \leq \max\{x_1, \cdots, x_n\} + \log n$.
- Geometric mean. The geometric mean $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\operatorname{dom} f = \mathbb{R}^n_{++}$.
- Log-determinant. The function $f(X) = \log\det X$ is concave on $\operatorname{dom} f = \mathbb{S}^n_{++}$.

Norms and Max function

- If $f : \mathrm{R}^n \to \mathrm{R}$ is a norm, and $0 \leq \theta \leq 1$, then

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

since f satisfies the triangle inequality and f is homogeneous.
  - Therefore any norm is convex.
- The function $f(x) = \max_i x_i$ is convex since

$$\begin{aligned} \max_i(\theta x_i + (1-\theta)y_i) &\leq \max_i \theta x_i + \max_i (1-\theta) y_i \quad ^2 \\ &= \theta \max_i x_i + (1-\theta) \max_i y_i. \\ &= \theta f(x) + (1-\theta) f(y). \end{aligned}$$

  - In addition, $f(|x|) = \max_i |x_i|$ is a norm.

[2] It is worthy to note that the index "i" that achieves the maximum of each of the three terms in this inequality may be different in general.

Quadratic-Over-Linear Function

- The **quadratic-over-linear** function $f : \mathrm{R}^2 \to \mathrm{R}, \mathrm{dom}\ f = \mathrm{R} \times \mathrm{R}_{++}$, $f(x, y) = x^2/y$, is convex since:

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

Log-Sum-Exp

- The **log-sum-exp** function $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbf{R}^n$ since

$$\nabla^2 f(x) = \frac{1}{(1^T z)^2}\left((1^T z)\textbf{diag}\,(z) - zz^T\right),$$

where $z = (e^{x_1}, \cdots, e^{x_n})$, and

- for all v,

$$v^T \nabla^2 f(x) v = \frac{1}{(1^T z)^2}\left(\left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n v_i^2 z_i\right) - \left(\sum_{i=1}^n v_i z_i\right)^2\right) \geq 0.$$

Geometric mean

- The **geometric mean** function $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on dom $f = \mathbf{R}_{++}^n$ since its Hessian $\nabla^2 f(x)$ can be shown to be negative semidefinite.
- Note that

$$\frac{\partial f(x)}{\partial x_k} = \frac{1}{n}\frac{(\prod_{i=1}^n x_i)^{1/n}}{\prod_{i=1}^n x_i} \cdot \prod_{i=1, i\neq k}^{n} x_i = \frac{1}{n}\frac{(\prod_{i=1}^n x_i)^{1/n}}{x_k}$$

and

$$\frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k^2}, \quad \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k x_l} (k \neq l).$$

Geometric mean

- The **geometric mean** function $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on dom $f = \mathbf{R}_{++}^n$ since its Hessian $\nabla^2 f(x)$ can be shown to be negative semidefinite.
- So,

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n \,\mathbf{diag}\,(1/x_1^2, \cdots, 1/x_n^2) - qq^T\right)$$

where $q_i = 1/x_i$.
- For any $v \in \mathbf{R}^n$, we have

$$v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n \sum_{i=1}^n v_i^2/x_i^2 - \left(\sum_{i=1}^n v_i/x_i\right)^2\right) \leq 0.$$

Log-Determinant

- The function $f : \mathbf{S}^n \to \mathbf{R}$, $f(X) = \log\det X$, with $\operatorname{dom} f = \mathbf{S}^n_{++}$ is concave.
- Proof idea: consider an arbitrary line in $\mathbf{S}^n$ (that passes through some point in $\mathbf{S}^n_{++}$) given by $X = Z + tV$, where $Z \in \mathbf{S}^n_{++}$, $V \in \mathbf{S}^n$, and define $g(t) = f(Z + tV)$, $\operatorname{dom} g = \{t \mid Z + tV \succ 0\}$.
- Then it can be shown that

$$g(t) = \sum_{i=1}^{n} \log(1 + t\lambda_i) + \log\det Z$$

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

- So,

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

Sublevel sets

Sublevel Sets

The **α-sublevel set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

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

Sublevel sets of a convex function are convex

If f is a convex function, then for any $\alpha \in \mathbb{R}$, the α-sublevel set, C_α, is convex.

- The converse is not true. A function can have all its sublevel sets convex, but not be a convex function. (e.g., $f(x) = -e^x$.)
- If f is concave, then its **α-superlevel** set, given by $\{x \in \mathrm{dom}\, f | f(x) \geq \alpha\}$, is a convex set.

Sublevel sets – Example

Example

The geometric and arithmetic means of $x \in \mathbb{R}^n_+$ are

$$G(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad A(x) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

respectively. Suppose $0 \leq \beta \leq 1$, then the set

$$\left\{x \in \mathbb{R}^n_+ \mid G(x) \geq \beta A(x)\right\}$$

is convex since it is the 0-superlevel set of the concave function $G(x) - \beta A(x)$.

- It is also a convex cone.

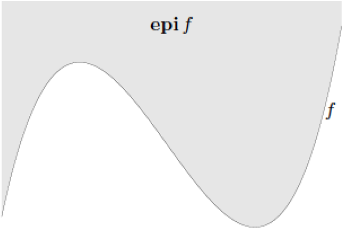
Epigraph

Graph

The **graph** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as $\{(x, f(x)) \mid x \in \text{dom } f\}$, a subset of $\mathbb{R}^{n+1}$.

Epigraph

The **epigraph** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as **epi** $f = \{(x,t)|x \in \text{dom } f, f(x) \leq t\}$, which is a subset of $\mathbb{R}^{n+1}$.



Epigraph

Graph

The **graph** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as $\{(x, f(x)) \mid x \in \text{dom } f\}$, a subset of $\mathbb{R}^{n+1}$.

Epigraph

The **epigraph** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as **epi** $f = \{(x, t) | x \in \text{dom } f, f(x) \leq t\}$, which is a subset of $\mathbb{R}^{n+1}$.

The epigraph of convex functions

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

The hypograph of concave functions

A function is concave if and only if its **hypograph**, defined as **hypo** $f = \{(x, t) \mid x \in \text{dom } f, f(x) \geq t\}$, is a convex set.

Matrix fractional function

- The function $f : \mathbf{R}^n \times \mathbf{S}^n \to \mathbf{R}$, defined as

$$f(x, Y) = x^T Y^{-1} x,$$

is called a **matrix fractional function**, and is convex on $\mathbf{dom}\, f = \mathbf{R}^n \times \mathbf{S}^n_{++}$.

- Proof:

$$\begin{aligned} \mathbf{epi}\, f &= \left\{(x, Y, t) \mid Y \succ 0, x^T Y^{-1} x \leq t\right\} \\ &= \left\{(x, Y, t) \mid Y \succ 0, \begin{bmatrix} Y & x \\ x^T & t \end{bmatrix} \succeq 0\right\} \end{aligned}$$

is a convex set.

Epigraph and first-order condition for convexity

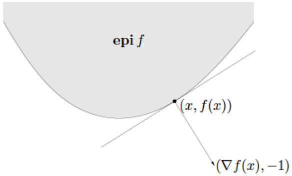
- If $(y, t) \in \textbf{epi } f$, then

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

implying

$$\begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0.$$

- This means that the hyperplane defined by $(\nabla f(x), -1)$ supports **epi** f at the boundary point $(x, f(x))$.



Jensen's Inequality

- The basic inequality for convex functions

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

  is called **Jensen's inequality**.

- Jensen's inequality can be extended to more than two points: If f is convex, $x_1, \cdots, x_k \in \mathrm{dom}\, f$, and $\theta_1, \cdots, \theta_k \geq 0$ with $\theta_1 + \cdots + \theta_k = 1$, then

$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k).$$

Jensen's Inequality

- Extension to infinite sum:

$$f\left(\int_S p(x)x dx\right) \leq \int_S f(x)p(x)dx,$$

  with $p(x) \geq 0$ on S, $\int_S p(x)dx = 1, S \subseteq \mathbf{dom}\ f$.
- If x is a random variable such that $\mathrm{Prob}\ (x \in \mathbf{dom}\ f) = 1$, then

$$f(\mathrm{E}x) \leq \mathrm{E}f(x).$$

- Suppose $x \in \mathbf{dom}\ f \subseteq \mathrm{R}^n$, $z \in \mathrm{R}^n, \mathrm{E}(z) = 0$, and assume $\Pr(x + z \in \mathbf{dom}\ f) = 1$. Then we have

$$\mathrm{E}f(x + z) \geq f(x).$$

Inequalities

- Many famous inequalities can be derived by applying Jensen's inequality to some convex functions.
- The arithmetic-geometric mean inequality: $(a+b)/2 \geq \sqrt{ab}$.
- Noting that $-\log x$ is convex, and letting $\theta = 1/2$, we obtain

$$-\log \frac{a+b}{2} \leq \frac{-\log a - \log b}{2},$$

implying the AM-GM inequality: $\sqrt{ab} \leq \dfrac{a+b}{2}$.

- Further, by taking

$$a = \frac{x_i^2}{\sum_{j=1}^n x_j^2}, b = \frac{y_i^2}{\sum_{j=1}^n y_j^2},$$

and summing over i, we get the Cauchy's inequality

$$\left(\sum_{j=1}^n x_j y_j\right)^2 \leq \left(\sum_{j=1}^n |x_j||y_j|\right)^2 \leq \left(\sum_{j=1}^n x_j^2\right)\left(\sum_{j=1}^n y_j^2\right).$$

Hölder's Inequality

- Apply the Jensen's inequality to the function $-\log x$ again, with an arbitrary θ, $0 < \theta < 1$, we get an inequality more general than the AM-GM inequality:

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

- If we take $\theta = 1/p$, where $p > 1$. Let $q = 1/(1-\theta)$, then $q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.
- By taking

$$a = \frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}, \ b = \frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q},$$

and summing over i, we obtain

$$\sum_{j=1}^n |x_j||y_j| \leq \left(\sum_{j=1}^n |x_j|^p \right)^{1/p} \left(\sum_{j=1}^n |y_j|^q \right)^{1/q},$$

which implies the Hölder's inequality

$$\sum_{j=1}^n x_j y_j \leq \left(\sum_{j=1}^n |x_j|^p \right)^{1/p} \left(\sum_{j=1}^n |y_j|^q \right)^{1/q}.$$

Minkowski Inequality

- Using Hölder's inequality, and $\frac{1}{p} + \frac{1}{q} = 1$, we have the following:

$$\begin{aligned}
||x + y||_p^p &= \sum_{i=1}^{n} |x_i + y_i|^p \\
&\leq \sum_i |x_i||x_i + y_i|^{p-1} + \sum_i |y_i||x_i + y_i|^{p-1} \\
\text{(by Hölder's ineq.)} \quad &\leq \left(\sum_i |x_i|^p\right)^{1/p} \left(\sum_i |x_i + y_i|^p\right)^{1/q} \\
&\quad + \left(\sum_i |y_i|^p\right)^{1/p} \left(\sum_i |x_i + y_i|^p\right)^{1/q} \\
&= (||x||_p + ||y||_p)\,(||x + y||_p)^{p-1}
\end{aligned}$$

- Therefore, we obtain that $||x + y||_p \leq ||x||_p + ||y||_p$, with $p > 1$.
- Note that the inequality above also holds for $p = 1$. So, we obtain the Minkowski inequality:

$$||x + y||_p \leq ||x||_p + ||y||_p,$$

with $p \geq 1$, which essentially says that $\forall p \geq 1$, ℓ_p-norm is a norm.