3. Convex functions

- basic properties and examples
- operations that preserve convexity
- the conjugate function
- quasiconvex functions
Definition

\( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex if \( \text{dom} \, f \) is a convex set and

\[
\quad f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)
\]

for all \( x, y \in \text{dom} \, f, \ 0 \leq \theta \leq 1 \)

• \( f \) is concave if \( -f \) is convex

• \( f \) is strictly convex if \( \text{dom} \, f \) is convex and

\[
\quad f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)
\]

for \( x, y \in \text{dom} \, f, \ x \neq y, \ 0 < \theta < 1 \)
Examples on R

convex:

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

concave:

- affine: $ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$
- powers: $x^\alpha$ on $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbb{R}_{++}$
Examples on $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$

affine functions are convex and concave; all norms are convex

**examples on $\mathbb{R}^n$**

- **affine function** $f(x) = a^T x + b$
- **norms:** $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_{\infty} = \max_k |x_k|$

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

- **affine function**

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

- **spectral (maximum singular value) norm**

\[
f(X) = \|X\|_2 = \sigma_{\text{max}}(X) = (\lambda_{\text{max}}(X^T X))^{1/2}
\]
Restriction of a convex function to a line

\[ f : \mathbb{R}^n \to \mathbb{R} \text{ is convex if and only if the function } g : \mathbb{R} \to \mathbb{R}, \]
\[ g(t) = f(x + tv), \quad \text{dom } g = \{ t \mid x + tv \in \text{dom } f \} \]

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

can check convexity of \( f \) by checking convexity of functions of one variable

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

Note that: \( X + tv = X^{\frac{1}{2}}(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})X^{\frac{1}{2}} \) then \( \det(X + tv) = \det(X)\det(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}) \)

\[ g(t) = \log \det(X + tv) = \log \det X + \log \det(I + tX^{-1/2}VX^{-1/2}) \]
\[ = \log \det X + \sum_{i=1}^{n} \log(1 + t\lambda_i) \]

where \( \lambda_i \) are the eigenvalues of \( X^{-1/2}VX^{-1/2} = UDU' \) then \( I + tUDU' = U(I + tD)U' \)

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

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

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

often simplifies notation; for example, the condition

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

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

- $\text{dom } f$ is convex
- for $x, y \in \text{dom } f$,

$$0 \leq \theta \leq 1 \quad \implies \quad f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$
First-order condition

\( f \) is **differentiable** if \( \text{dom} \, f \) 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 \( x \in \text{dom} \, f \)

**1st-order condition**: differentiable \( f \) with convex domain is convex iff

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

first-order approximation of \( f \) is global underestimator
Second-order conditions

\( f \) is **twice differentiable** if \( \text{dom} \ f \) is open and the Hessian \( \nabla^2 f(x) \in S^n \),

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

exists at each \( x \in \text{dom} \ f \)

**2nd-order conditions:** for twice differentiable \( f \) with convex domain

- \( f \) is convex if and only if
  \[
  \nabla^2 f(x) \succeq 0 \quad \text{for all} \ x \in \text{dom} \ f
  \]

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

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

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

convex if $P \succeq 0$

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

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

convex (for any $A$)

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

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

convex for $y > 0$
**log-sum-exp:** $f(x) = \log \sum_{k=1}^{n} \exp x_k$ is convex

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

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

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

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

More clearly: $a_k = v_k \sqrt{z_k}$, $b_k = \sqrt{z_k}$, then $\langle a, b \rangle \leq |a|_2 |b|_2$

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

(similar proof as for log-sum-exp)
Epigraph and sublevel set

α-sublevel set of \( f : \mathbb{R}^n \to \mathbb{R} \):

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

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

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

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

\( f \) is convex if and only if \( \text{epi } f \) is a convex set

(Convex functions)
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

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

for any random variable $z$

basic inequality is special case with discrete distribution

$$\text{prob}(z = x) = \theta, \quad \text{prob}(z = y) = 1 - \theta$$
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
Positive weighted sum & composition with affine function

nonnegative multiple: $\alpha f$ 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

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

- (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: $f(x) = \max_{i=1,\ldots,m}(a_i^Tx + b_i)$ is convex

• sum of $r$ largest components of $x \in \mathbb{R}^n$:

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

is convex ($x[i]$ is $i$th largest component of $x$)

proof:

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

An index of a vector entry goes from 1 to $n$

There are $n$ choose $r$ sets of $r$ different indices

We can define $m = n$ choose $r$ functions that sum $r$ entries (See the first line of slide)

The example goes through all $n$ choose $r$ sets of indices $i_1\ldots i_r$
Pointwise supremum

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

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

is convex

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 \( X \in S^n \),

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

(Example: definition of dual norm)
Composition with scalar functions

composition of $g : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R} \rightarrow \mathbb{R}$:

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

$f$ is convex if

- $g$ convex, $h$ convex, nondecreasing
- $g$ concave, $h$ convex, nonincreasing

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

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

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 $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$:

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

$f$ is convex if $g_i$ convex, $h$ convex nondecreasing in each argument
$g_i$ concave, $h$ convex 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

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

minimizing over \( y \) gives

\[
g(x) = \inf_y f(x, y) = x^T (A - BC^{-1}B^T)x
\]

\( g \) is convex, hence Schur complement \( A - BC^{-1}B^T \succeq 0 \) (iff \([A \ B; B^T \ C] \succeq 0\))

- distance to a set: \( \text{dist}(x, S) = \inf_{y \in S} \|x - y\| \) is convex if \( S \) is convex

(Example: Lagrange dual, we will see it next week)
the **perspective** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the function $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$,

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

$g$ 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 $\mathbb{R}^2_{++}$
- if $f$ is convex, then

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

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

The conjugate of a function $f$ is

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

Properties:

- $f^*$ is convex (even if $f$ is not):
  - $y x - f(x)$ is convex in $y$
  - conjugate is pointwise supremum

- $f^{**} = f$, if $f$ is convex and epi $f$ is a closed set

- for differentiable $f$, $f^*$ is also called Fenchel conjugate or Legendre transform

(very useful in Chapter 5)
examples

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

\[
f^*(y) = \sup_{x > 0} (xy + \log x)
\]

\[
= \begin{cases} 
-1 - \log(-y) & y < 0 \\
\infty & \text{otherwise}
\end{cases}
\]

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

\[
f^*(y) = \sup_{x} (y^T x - (1/2)x^T Q x)
\]

\[
= \frac{1}{2} y^T Q^{-1} y
\]
Quasiconvex functions

\( f : \mathbb{R}^n \to \mathbb{R} \) is quasiconvex if \( \text{dom} \ f \) is convex and the sublevel sets

\[
S_\alpha = \{ x \in \text{dom} \ f \mid f(x) \leq \alpha \}
\]

are convex for all \( \alpha \)

- \( f \) is quasiconcave if \(-f\) is quasiconvex
- \( f \) is quasilinear if it is quasiconvex and quasiconcave
Examples

• $\sqrt{|x|}$ is quasiconvex on $\mathbb{R}$
• $\text{ceil}(x) = \inf\{z \in \mathbb{Z} \mid z \geq x\}$ is quasilinear
• $\log x$ is quasilinear on $\mathbb{R}_{++}$
• $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbb{R}_+^2$
• linear-fractional function
  
  $f(x) = \frac{a^T x + b}{c^T x + d}$,  \hspace{1cm} \text{dom } f = \{x \mid c^T x + d > 0\}$
  
  is quasilinear
• distance ratio
  
  $f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}$,  \hspace{1cm} \text{dom } f = \{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$
  
  is quasiconvex
Properties

**modified Jensen inequality:** for quasiconvex $f$

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

**first-order condition:** differentiable $f$ with cvx domain is quasiconvex iff

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

**sums** of quasiconvex functions are not necessarily quasiconvex