Convex Optimization II

Lecture 5: Convex Functions

Hamed Shah-Mansouri

Department of Electrical Engineering

Sharif University of Technology

1400-2

Motivations

- Convex and concave functions have many special and important properties.
- Example: Any local minimum of a convex function over a convex set is also a global minimum.
- In this lecture, we introduce some of the important topics of convex functions and their properties.
- These properties can be used to develop suitable optimality conditions and computational schemes for convex optimization problems.
- All materials and figures in this lecture are from [1].

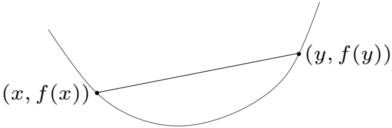
  [1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- Thanks to Prof. Vincent Wong and Prof. Stephen Boyd for all the slides used in this lecture.

Convex Functions

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



- f is concave if $-f$ is convex
- f is strictly convex if $\mathbf{dom}\, f$ is convex and

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

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

Extended-value Extension

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

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

often simplifies notation; for example, the condition

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

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

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

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

First-order Condition

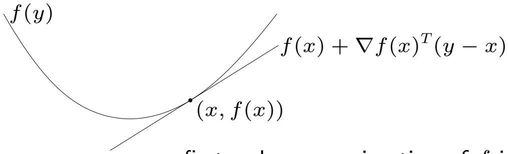
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

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

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

Symmetric Positive (Semi)-Definite Matrices

- The set of symmetric $n \times n$ matrices

$$\mathbf{S}^n = \{X \in \mathbf{R}^{n\times n} \mid X = X^T\}.$$

- The set of symmetric positive semidefinite matrices

$$\begin{aligned}\mathbf{S}^n_+ &= \{X \in \mathbf{S}^n \mid X \succeq 0\} \\ &= \{X \in \mathbf{R}^{n\times n} \mid X = X^T, \forall\, v \in \mathbf{R}^n, v^T X v \geq 0\}.\end{aligned}$$

- Equivalently, matrix X has non-negative eigenvalues.
- The set of symmetric positive definite matrices

$$\begin{aligned}\mathbf{S}^n_{++} &= \{X \in \mathbf{S}^n \mid X \succ 0\} \\ &= \{X \in \mathbf{R}^{n\times n} \mid X = X^T, \forall\, v \in \mathbf{R}^n, v \neq \mathbf{0}, v^T X v > 0\}.\end{aligned}$$

- Equivalently, matrix X has positive eigenvalues.

Second-order Conditions

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 on R

convex:

- affine: $ax + b$ on **R**, for any $a, b \in$ **R**
- exponential: e^{ax}, for any $a \in$ **R**
- powers: x^α on **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 **R**$_{++}$

concave:

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

Examples on $\mathbf{R}^n$ and $\mathbf{R}^{n\times m}$

affine functions are convex and concave; all norms are convex

examples on $\mathbf{R}^n$

- 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 $\mathbf{R}^{m\times n}$ ($m \times n$ 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

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

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

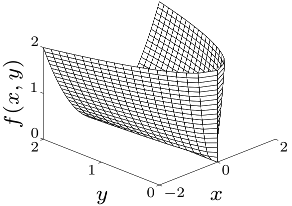
$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

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: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbf{R}^n$.

The Hessian of the log-sum-exp function is

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

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)

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)

Sublevel Set and Epigraph

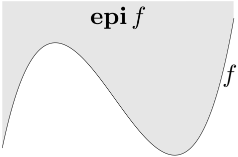
α-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$:

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

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

Epigraph

- Many results for convex functions can be proven (or interpreted) geometrically using epigraphs, and applying results for convex sets.
- e.g., consider the first-order condition for convexity

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

where f is a differentiable convex function and $x, y \in \mathbf{dom}\, f$.

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

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

توی epi　　　مرز epi

- The hyperplane defined by $(\nabla f(x), -1)$ supports $\mathbf{epi}\ f$ at the boundary point $(x, f(x))$.

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

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

How to Investigate the Convexity of a Function?

Practical methods for establishing convexity of a function

- Verify convexity definition
- First-order condition of convexity for differentiable functions
- Second-order condition of convexity for twice differentiable functions
- Operations that preserve convexity (show that the function is obtained from simple convex functions)
  - nonnegative weighted sum
  - composition with affine function
  - pointwise maximum and supremum
  - composition
  - minimization
  - perspective

Positive Weighted Sum and 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 functions, then their pointwise maximum f

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

  with $\mathbf{dom}\, f = \mathbf{dom}\, f_1 \cap \cdots \cap \mathbf{dom}\, f_m$, is also convex.

- Example: Piecewise-linear functions

$$f(x) = \max\{a_1^T x + b_1, \ldots, a_L^T x + b_L\}$$

Supremum and Infimum

- A number $\hat{a}$ is an upper bound on $C \subseteq \mathbf{R}$ if for each $x \in C$, $x \leq \hat{a}$.
- The number $\hat{b}$ is called the least upper bound or supremum of the set C, and is denoted $\sup C$.
- We take $\sup \emptyset = -\infty$ and $\sup C = \infty$ if C is unbounded above.
- e.g., $C = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots\}$. We have $\sup C = 1$.
- When the set C is finite, $\sup C$ is the maximum of its elements.
- Similarly, a number $\check{a}$ is a lower bound on $C \subseteq \mathbf{R}$ if for each $x \in C$, $\check{a} \leq x$.
- The number $\check{b}$ is called the greatest lower bound or infimum of the set C, and is denoted $\inf C = -\sup(-C)$.
- e.g., $C = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}$. We have $\inf C = 0$.
- When the set C is finite, $\inf C$ is the minimum of its elements.

Pointwise Supremum

- If for each $y \in \mathcal{A}$, $f(x,y)$ is convex in x, then
$$g(x) = \sup_{y \in \mathcal{A}} f(x,y)$$
is convex in x.
- Pointwise supremum of functions corresponds to intersection of epigraphs
$$\mathbf{epi}\ g = \bigcap_{y \in \mathcal{A}} \mathbf{epi}\ f(\cdot, y)$$
Examples
- Support function of a set $C \subseteq \mathbf{R}^n$, defined as $S_C(x) = \sup_{y \in C} y^T x$.
- Distance to farthest point of a set $C \subseteq \mathbf{R}^n$, defined as $f(x) = \sup_{y \in C} ||x - y||$.
- Maximum eigenvalue of symmetric matrix $X \in \mathbf{S}^n$, defined as $\lambda_{\max}(X) = \sup_{||y||_2 = 1} y^T X y$.

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

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 h is convex, $\tilde{h}$ is nondecreasing, and g is convex.
- f is convex if h is convex, $\tilde{h}$ is nonincreasing, and g is concave.
- f is concave if h is concave, $\tilde{h}$ is nondecreasing, and g is concave.
- f is concave if h is concave, $\tilde{h}$ is nonincreasing, and g is convex.
- monotonicity must hold for extended-value extension $\tilde{h}$, which assigns value ∞ ($-\infty$) to points not in $\mathbf{dom}\ h$ for h convex (concave).
- proof (for $n = 1$, differentiable g, h)

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

- $\exp g(x)$ 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 $h : \mathbf{R}^k \to \mathbf{R}$

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

- f is convex if h convex, h nondecreasing in each argument, g_i convex.
- f is convex if h convex, h nonincreasing in each argument, g_i concave.
- f is concave if h concave, h nondecreasing in each argument, g_i concave.
- f is concave if h concave, h nonincreasing in each argument, g_i convex.
- proof (for $n = 1$, twice differentiable g, h)

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

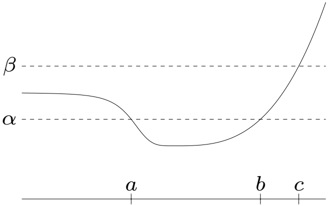
- $\sum_{i=1}^m \log g_i(x)$ is concave if g_i are concave and positive.
- $\log \sum_{i=1}^m \exp g_i(x)$ is convex if g_i are convex.

Quasiconvex Functions

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



- f is quasiconcave if $-f$ is quasiconvex
- f is quasilinear if it is quasiconvex and quasiconcave

Examples of Quasiconvex Functions

- $\sqrt{|x|}$ is quasiconvex on $\mathbf{R}$
- $\operatorname{ceil}(x) = \inf\{z \in \mathbf{Z} \mid z \geq x\}$ 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

SUMMARY

- Definition of convex function and epigraph.
- Examples of convex and concave functions in $\mathbf{R}$, $\mathbf{R}^n$, and $\mathbf{R}^{n\times m}$.
- First order and second order conditions, Jensen's inequality.
- Operations that preserve convexity.
  - Nonnegative weighted sums
  - Composition with an affine mapping
  - Pointwise maximum and supremum
  - Scalar and vector composition
  - Minimization
  - Perspective of a function
- Examples.
- Reading: Sections 3.1−3.3 in [1] by Boyd and Vandenberghe.

Mathematical Background

Matrix Operation Cheat-sheet

Rule	Comments
$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$	order is reversed, everything is transposed
$(\mathbf{a}^T\mathbf{Bc})^T = \mathbf{c}^T\mathbf{B}^T\mathbf{a}$	as above
$\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a}$	(the result is a scalar, and the transpose of a scalar is itself)
$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$	multiplication is distributive
$(\mathbf{a}+\mathbf{b})^T\mathbf{C} = \mathbf{a}^T\mathbf{C}+\mathbf{b}^T\mathbf{C}$	as above, with vectors
$\mathbf{AB} \neq \mathbf{BA}$	multiplication is **not** commutative

For a more comprehensive reference, see

`https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`

Matrix Derivatives Cheat-sheet

Scalar derivative			Vector derivative		
$f(x)$	$\to$	$\frac{\mathrm{d}f}{\mathrm{d}x}$	$f(\mathbf{x})$	$\to$	$\frac{\mathrm{d}f}{\mathrm{d}\mathbf{x}}$
bx	$\to$	b	$\mathbf{x}^T\mathbf{B}$	$\to$	$\mathbf{B}$
bx	$\to$	b	$\mathbf{x}^T\mathbf{b}$	$\to$	$\mathbf{b}$
x^2	$\to$	$2x$	$\mathbf{x}^T\mathbf{x}$	$\to$	$2\mathbf{x}$
bx^2	$\to$	$2bx$	$\mathbf{x}^T\mathbf{B}\mathbf{x}$	$\to$	$2\mathbf{B}\mathbf{x}$

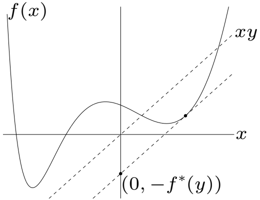
For a more comprehensive reference, see

`https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`

Conjugate Function

Given $f : \mathbf{R}^n \to \mathbf{R}$, the conjugate function $f^* : \mathbf{R}^n \to \mathbf{R}$ defined as:

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



with domain consisting of $y \in \mathbf{R}^n$ for which the supremum is finite.

- $f^*(y)$ is always convex. It is the pointwise supremum of a family of affine functions of y. This is true whether or not f is convex.
- Fenchel's inequality: $f(x) + f^*(y) \geq x^T y$ for all x, y (by definition).
- $f^{**} = f$ if f is convex and closed.

  Useful for Lagrange duality theory.

Examples of Conjugate Functions

- *Affine function.* $f(x) = ax + b$, where $x \in \mathbf{R}$.
  $f^*(a) = -b$ with $\mathbf{dom}\ f^* = \{a\}$.

- *Negative logarithm.* $f(x) = -\log x$ with $\mathbf{dom}\ f = \mathbf{R}_{++}$.
  $f^*(y) = -\log(-y) - 1$ with $\mathbf{dom}\ f^* = \{y \mid y < 0\} = -\mathbf{R}_{++}$.

- *Exponential.* $f(x) = e^x$.
  $f^*(y) = y\log(y) - y$ with $\mathbf{dom}\ f^* = \{y \mid y \geq 0\} = \mathbf{R}_{+}$.

- *Negative entropy.* $f(x) = x\log x$ with $\mathbf{dom}\ f = \mathbf{R}_{+}$ and $f(0) = 0$.
  $f^*(y) = e^{y-1}$ with $\mathbf{dom}\ f^* = \mathbf{R}$.

- *Strictly convex quadratic function.* $f(x) = \frac{1}{2}x^TQx$, where Q is positive definite.
  $f^*(y) = \frac{1}{2}y^TQ^{-1}y$ with $\mathbf{dom}\ f^* = \mathbf{R}$.

- *Log-sum-exp* function. $f(x) = \log\sum_{i=1}^n e^{x_i}$.
  $f^*(y) = \sum_{i=1}^n y_i \log y_i$ if $y \succeq 0$ and $\sum_{i=1}^n y_i = 1$. ($f^*(y) = \infty$ otherwise).