Part III Convex and Conjugate Functions, Computable Problems

Wenxun Xing

Department of Mathematical Sciences
Tsinghua University
Office hour: 4:00-5:00pm, Thursday
Office: The New Science Building, A416
Tel: 62787945
Email: wxing@tsinghua.edu.cn

Oct., 2019

Convex and conjugate functions, computable problems

Content

- Continuous, differential functions
- Convex functions
- Conjugate functions
- Computable problems

Functions

- Continuous: $f : \mathcal{X} \subset \mathbb{R}^n$ is continuous at x^0
  (i) $x^0 \in \mathcal{X}$
  (ii) $\lim_{x \to x^0} f(x) = f(x^0)$
- Continuous function: $f \in C^0(\mathcal{X})$ means f is continuous at all points in $\mathcal{X} \subset \mathbb{R}^n$.
- Gradient: For $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$

$$\nabla f(x) = [\frac{\partial f(x)}{\partial x_1}, \cdots, \frac{\partial f(x)}{\partial x_n}]_{1 \times n}$$

- Hessian: For $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$

$$F(x) = [\frac{\partial^2 f(x)}{\partial x_i \partial x_j}]_{n \times n}$$

- Continuously differentiable function: $f \in C^p(\mathcal{X})$ $(p = 1, 2, \cdots)$ means f is p-th continuously differentiable over $\mathcal{X} \subset \mathbb{R}^n$.

Functions

Theorem (Taylor theorem)

Let $\mathcal{X}$ be open, $f \in C^p(\mathcal{X})$, $x^1, x^2 \in \mathcal{X}$, $x^1 \neq x^2$ and

$$x(\theta) = \theta x^1 + (1-\theta)x^2 \in \mathcal{X}, \ \forall\, 0 \leq \theta \leq 1.$$

Then $\exists\ \bar{x} = \bar{\theta}x^1 + (1-\bar{\theta})x^2 \in \mathcal{X}$, $0 < \bar{\theta} < 1$, s.t.

$$f(x^2) = f(x^1) + \sum_{k=1}^{p-1} \frac{1}{k!} d^k f(x^1; x^2 - x^1) + \frac{1}{p!} d^p f(\bar{x}; x^2 - x^1)$$

where $d^k f(x; h)$ is the k-th order differential of function f along h.

Functions: big O and small o

Let $g(\cdot)$ be a real-valued function on $\mathbb{R}$.

- $g(x) = O(x)$

  $\exists c \geq 0$ such that

$$\left|\frac{g(x)}{x}\right| \leq c \text{ as } x \to 0 \text{ (or } +\infty)$$

- $g(x) = o(x)$

$$\left|\frac{g(x)}{x}\right| = 0 \text{ as } x \to 0 \text{ (or } +\infty)$$

Functions

Taylor theorem in small o formulation:

- $p = 1$
$$f(x+h) = f(x) + \nabla f(x)h + o(\|h\|)$$
- $p = 2$
$$f(x+h) = f(x) + \nabla f(x)h + \frac{1}{2}h^T F(x)h + o(\|h\|^2)$$
- $p \geq 3$. Tensor expressions.

Convex functions and properties

- Epigraph of a function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$

$$\operatorname{epi} f = \{(x, \lambda) \in \mathbb{R}^{n+1} | \lambda \geq f(x), x \in \mathcal{X}\}$$

- Closed function: if $\operatorname{epi} f$ is a closed set.
- Convex function: if $\operatorname{epi} f$ is a convex set.
- Concave function: if $-f$ is a convex function.
- Convex hull function $\operatorname{conv}(f)$ of a function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ is a function on $\mathcal{X}$ such that $\operatorname{epi}(\operatorname{conv}(f)) = \operatorname{conv}(\operatorname{epi}(f))$.
- 'Proper convex function'='convex function' in this course.

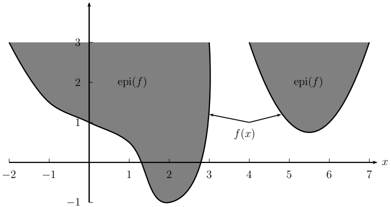

Figure: Figure of $\mathrm{epi}(f)$

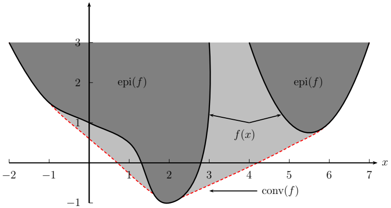

Figure: Figure of $\mathrm{conv}(f)$

Theorem

$f : \mathcal{X}$ is a convex function if and only if $\mathcal{X}$ is a nonempty convex set and for any $x^1,\ x^2 \in \mathcal{X}$ and $0 \leq \alpha, \beta \leq 1, \alpha + \beta = 1$, we have

$$f(\alpha x^1 + \beta x^2) \leq \alpha f(x^1) + \beta f(x^2).$$

Theorem

$f_1 : \mathcal{X}$ and $f_2 : \mathcal{X}$ are two convex functions, then $f_1 + f_2 : \mathcal{X}$, $\max\{f_1, f_2\} : \mathcal{X}$ are convex functions.

Theorem

Suppose $f : \mathcal{X}$ be a convex function. We have

$$\mathrm{ri}(\mathrm{epi}(f)) = \left\{ \begin{pmatrix} x \\ \lambda \end{pmatrix} \mid x \in \mathrm{ri}(\mathcal{X}) \ and \ f(x) < \lambda \right\}.$$

Theorem

Given a convex set $\mathcal{X} \subseteq \mathbb{R}^n$ and a convex function $f : \mathcal{X}$, there exists a $d \in \mathbb{R}^n$ for a given $\bar{x} \in \mathrm{ri}(\mathcal{X})$, such that

$$f(x) \geq f(\bar{x}) + d^T(x - \bar{x})$$

for any $x \in \mathcal{X}$.

Subgradient

For a $f(x) : \mathcal{X} \subseteq \mathbb{R}^n$ and an $\bar{x} \in \mathcal{X}$, a $d \in \mathbb{R}^n$ is called a subgradient if

$$f(x) \geq f(\bar{x}) + d^T(x - \bar{x}) \text{ for any } x \in \mathcal{X}.$$

$$\partial f(\bar{x}) = \{d \in \mathbb{R}^n \mid d \text{ is a subgradient of } f(x) \text{ at } \bar{x}\}.$$

Theorem

If the subgradient set of $f : \mathcal{X}$ at $\bar{x}$ is nonempty, then it is closed and convex.

Geometry explanation of subgradient

If $f(x)$ is a convex function, a subgradient d determines a hyperplane

$$\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{n+1} \mid y - d^T x = f(\bar{x}) - d^T \bar{x} \right\},$$

with the supporting point $\begin{pmatrix} \bar{x} \\ f(\bar{x}) \end{pmatrix}$ to support $\mathrm{epi}(f)$.

- When $f(x)$ is differential at $\bar{x}$, $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ having a unique point.
- For any relative interior of $\mathcal{X}$, there exists a subgradient.
- The subgradient set may be empty at the boundary of $\mathcal{X}$ or $f(x)$ is not convex.

$$f(x) = \begin{cases} e^x, & -1 \leq x < 0, \\ 2, & x = 0 \end{cases}$$

It is convex on $[-1, 0]$, but the subgradient set is empty at $x = 0$.

Convex functions and properties

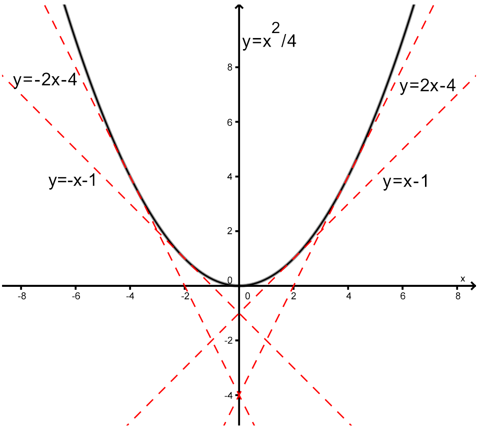

Figure: $(x,\ f(x)) \leftrightarrow (y,\ \lambda)$

Conjugate functions

- The negative of λ-intercept: $\lambda - d^T y = f(x) - d^T x, y \in \mathbb{R}^n$.

$$f^*(d) = \sup_{x \in \mathcal{X}} \{d \bullet x - f(x)\}$$

- Conjugate of $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$:

$$f^*(y) = \sup_{x \in \mathcal{X}} \{y \bullet x - f(x)\}$$

with f^* being defined on $\mathcal{Y} = \{y \in \mathbb{R}^n | f^*(y) < +\infty\}$.

Conjugate functions

Lemma

If $f^* : \mathcal{Y}$ exists then $\mathcal{Y}$ is a convex set and $f^* : \mathcal{Y}$ is a convex function.

Lemma (Fenchel's inequality)

Given $f : \mathcal{X}$ and its conjugate $f^* : \mathcal{Y}$, then

$$x \bullet y \leq f(x) + f^*(y), \; \forall \, x \in \mathcal{X} \text{ and } y \in \mathcal{Y}.$$

Moreover,

$$x \bullet y = f(x) + f^*(y) \iff y \in \partial f(x)$$

Conjugate functions–examples

Example 1: $f(x) = x^2$.

$$f^*(y) = \sup_{x \in \mathbb{R}} (xy - x^2) = \frac{y^2}{4},$$

$$\mathcal{Y} = \mathbb{R}.$$

Example 2: $f(x) = x^3$.

$$f^*(y) = \sup_{x \in \mathbb{R}} (xy - x^3) = +\infty,$$

$$\mathcal{Y} = \emptyset.$$

Example 3: $f(x) = 2x^2, x \geq 1$.

$$f^*(y) = \sup_{x \geq 1} (xy - 2x^2) = \begin{cases} \frac{y^2}{8}, & y \geq 4 \\ y - 2, & y < 4 \end{cases}$$

$$\mathcal{Y} = \mathbb{R}.$$

Lemma

Given $f : \mathcal{X}$, if $f^* : \mathcal{Y}$ is well defined, then $\mathcal{Y}$ is a nonempty convex set and $f^*(y)$ is convex on $\mathcal{Y}$.

Theorem

Suppose $\mathcal{X} \neq \emptyset$, $f : \mathcal{X}$ and $f^* : \mathcal{Y}$ be well-defined. Then

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

satisfies

$$f^{**}(x) = \mathrm{cl}(\mathrm{conv}(f))(x), \forall\, x \in \mathrm{ri}(\mathrm{conv}(\mathcal{X}))$$

and

$$f^{**}(x) = +\infty, \forall\, x \notin \mathrm{cl}(\mathrm{conv}(\mathcal{X})),$$

where $\mathrm{cl}(\mathrm{conv}(f))$ is defined by $\mathrm{epi}(\mathrm{cl}(\mathrm{conv}(f))) = \mathrm{cl}(\mathrm{conv}(\mathrm{epi}(f)))$. Specially, if $f : \mathcal{X}$ is a convex and continuous function and $\mathcal{X}$ is a closed convex set, then $f(x) = f^{**}(x), \forall\, x \in \mathcal{X}$.

Values at the boundary points

Let

$$f(x) = \begin{cases} \frac{1}{x}, & 0 < x < 1 \\ 2, & x = 1. \end{cases}$$

- $f : (0, 1]$ is a convex function on $(0, 1]$. Its conjugate function

$$f^*(y) = \sup_{0<x\le 1} \{xy - f(x)\} = \max\left\{ \sup_{0<x<1} \{xy - \frac{1}{x}\}, y - 2 \right\},$$

where

$$\sup_{0<x<1} \{xy - \frac{1}{x}\} = \begin{cases} -2\sqrt{-y}, & y \le -1 \\ y - 1, & y > -1. \end{cases}$$

We have

$$f^*(y) = \begin{cases} -2\sqrt{-y}, & y \le -1 \\ y - 1, & y > -1. \end{cases}$$

- The conjugate of the conjugate function

$$f^{**}(x) = \sup_{y\in\mathbb{R}}\{xy - f^*(y)\} = \max\left\{\sup_{y\le -1}\{xy + 2\sqrt{-y}\}, \sup_{-1<y}\{xy - y + 1\}\right\},$$

where

$$\sup_{y\le -1}\{xy + 2\sqrt{-y}\} = \begin{cases} \frac{1}{x}, & 0 < x \le 1 \\ +\infty, & x \le 0, \\ 2 - x & x > 1, \end{cases}$$

$$\sup_{-1<y}\{xy - y + 1\} = \begin{cases} +\infty, & x > 1 \\ 2 - x, & x \le 1 \end{cases}$$

We have

$$f^{**}(x) = \begin{cases} \frac{1}{x}, & 0 < x \le 1 \\ +\infty, & \text{otherwise.} \end{cases}$$

- $f^{**}(x) = \mathrm{cl}(\mathrm{conv}(f))(x) = f(x), \forall\, x \in (0,1)$.
  $f^{**}(0)$ and $\mathrm{cl}(\mathrm{conv}(f))(0)$ are $+\infty$.
  $f^{**}(1) = \mathrm{cl}(\mathrm{conv}(f))(1) = 1 \ne f(1) = 2$.

Theorem

Suppose $f : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}$ be convex and continuous and $\mathcal{X}$ be closed and convex. Then $f^* : \mathcal{Y}$ is well defined and the conjugate of $f^* : \mathcal{Y}$ is $f : \mathcal{X}$. There exist $\bar{x} \in \mathcal{X}$ and $\bar{y} \in \mathcal{Y}$ such that $\bar{y} \in \partial f(\bar{x})$ if and only if $\bar{x} \in \partial f^*(\bar{y})$. Then

$$\bar{x}^T \bar{y} = f(\bar{x}) + f^*(\bar{y}) \Longleftrightarrow \bar{y} \in \partial f(\bar{x}) \text{ or } \bar{x} \in \partial f^*(\bar{y}).$$

Theorem

Suppose $\mathcal{X}$ be closed and convex, $f_1^* : \mathcal{Y}_1$ and $f_2^* : \mathcal{Y}_2$ be well-defined, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be closed, $f_1^* : \mathcal{Y}_1$ and $f_2^* : \mathcal{Y}_2$ be continuous, $f_1^{**} = f_2^{**} : \mathcal{X}$ and be continuous. We have $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Y}$ and $f_1^* = f_2^* : \mathcal{Y}$.

Corollary

Suppose $\mathcal{X}$ be closed and convex, $f_1 : \mathcal{X}$ and $f_2 : \mathcal{X}$ have the same continuous convex hull function. If $f_1^* : \mathcal{Y}_1$ is well-defined, $\mathcal{Y}_1$ is closed and $f_1^* : \mathcal{Y}_1$ is continuous, then $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Y}$ and $f_1^* = f_2^* : \mathcal{Y}$.

Conjugate functions and properties

Let $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ be a function with its conjugate transform $f^* : \mathcal{Y}$.

- For $\alpha \in \mathbb{R}$, the conjugate of $f + \alpha$ is $f^* - \alpha$.
- For $a \in \mathbb{R}^n$, the conjugate of $\tilde{f}(x) = f(x) + x \bullet a$ on $\mathcal{X}$ is $\tilde{f}^*(y) = f^*(y - a)$, $\forall\, y \in \mathcal{Y}$.
- For $a \in \mathbb{R}^n$, the conjugate of $\bar{f}(x) = f(x - a)$ on $\mathcal{X}$ is $\bar{f}^*(y) = f^*(y) + y \bullet a$, $\forall\, y \in \mathcal{Y}$.
- For $\lambda > 0$, the conjugate of $f_1(x) = \lambda f(x)$ on $\mathcal{X}$ is $f_1^*(y) = \lambda f^*(\frac{y}{\lambda})$, $\forall\, y \in \lambda\mathcal{Y}$.
- For $\lambda > 0$, the conjugate of $f_2(x) = f(\frac{x}{\lambda})$ on $\lambda\mathcal{X}$ is $f_2^*(y) = f^*(\lambda y)$, $\forall\, y \in \mathcal{Y}/\lambda$.

Computable problems

- Optimization problem

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \\ & x \in \mathcal{D} \subseteq \mathbb{R}^n, \end{array}$$

- Discrete (combinatorial) or continue optimization problems $\mathcal{D}$ is discrete or not.
- Computational easy or hard: computational complexity, easy<->tractable <->in polynomial time.
- Polynomial problems, NP-complete, NP-hard etc.

Complexity concepts of combinatorial problems

- Problem, instance, size of an instance, computational time.
  - A problem: a set.
  - An instance: an element of the set.
  - The size of an instance: encoding scheme, bits in Turing machine.
  - The computational time: the total basic operations of the algorithm.
- Polynomial time algorithms.
  For an algorithm A, if there exists a polynomial function $p(\cdot)$ such that

$$C_A(I) = O(p(s(I))), \ \forall I \in Q,$$

  where $C_A(I)$ is the computational time of algorithm A, Q is the problem, I is an instance, $s(I)$ is the size of I, then the algorithm is called an polynomial time algorithm.

- Polynomial time problems.
  For a problem Q, if there exists a polynomial algorithm to solve it, then it is called a polynomial problem.
- NP (Non-deterministic Problem): polynomially solved by a non-deterministic computer.
- NP-complete, NP-hard.
- Approximation ratio

$$r(A) = \sup_{I \in Q} \frac{v_A(I)}{v_{opt}(I)}, \quad \frac{v_A(I) - v_{opt}}{v_{opt}(I)} \leq r(A) - 1.$$

  where $v_A(I)$ and $v_{opt}(I)$ are values of the heuristic A and optimal of the instance I respectively.
- Polynomial time approximation scheme (PTAS).
  For any $\epsilon > 0$, if there exist a $1+\epsilon$ approximation algorithm A and a bi-variable polynomial function $g(\cdot,\cdot)$, such that

$$C_A(I) = O(g(s(I), y)), \forall\, I \in Q,$$

  where $y = r(\frac{1}{\epsilon})$ and $r(\cdot)$ is a real function.

Complexity concepts of continuous problems

- Size of an instance.
  Number of variables, coefficients and constraints.
- Basic operations
  - Black-box for some real numbers and computations.

    e, $\sin x$, $\sqrt{2}$ etc.

  - Basic operations: one black-box operation=1 operation.

    $\sin x$, $\sqrt{x}$: one unit computation.

Computable

- For an optimization problem Q and any $\epsilon > 0$,
- there exists an algorithm A and a bi-variable polynomial function $g(\cdot, \cdot)$,
- the computation time is

$$C_A(I) = O(g(d(I), y)), \forall\, I \in Q,$$

  where $d(I)$ is the size of I and $y = \log_2(\frac{1}{\epsilon})$,
- to get a solution $x_A(I)$ with its objective value $v_A(I)$ satisfying

$$\mid v_A(I) - v_{opt}(I) \mid \leq \epsilon, \ \forall\, I \in Q,$$

- and to check that the distance between $x_A(I)$ the feasible set of the instance is no great than ϵ.