IE 521 Convex Optimization

Lecture 7: Convex Conjugate

Niao He

14th February 2019

Outline

Recap: Subgradient

- Subgradient and subdifferential
  - $g \in \partial f(x_0)$ if $f(x) \geq f(x_0) + g^T(x - x_0), \forall x$.
- Properties
  - Subdifferential is closed and convex.
  - Subgradient exists and is bounded at interior.
  - Subdifferential is a monotone operator.
- Directional derivative
  - $f'(x; d) = \max_{g \in \partial f(x)} g^T d$
- Calculus of Subgradients
  - Conic combination
  - Affine transformation
  - Point maximum/supreme
  - Taking minimization
  - Composition

Simple Examples

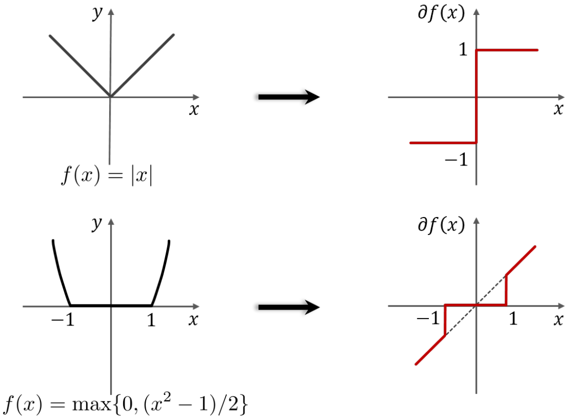


Figure: Examples of subdifferential sets

Example: Piecewise Linear Function

Example . Consider a single period inventory system. The cost $f(x)$ at inventory level x given demand d is

$$f(x) = h \cdot \max(x - d, 0) + p \cdot \max(d - x, 0).$$

Example: Piecewise Linear Function

Example . Consider a single period inventory system. The cost $f(x)$ at inventory level x given demand d is

$$f(x) = h \cdot \max(x - d, 0) + p \cdot \max(d - x, 0).$$

The subdifferential of $f(x)$ is

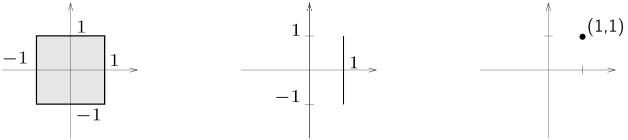
$$\partial f(x) = \begin{cases} h, & x > d \\ [-p, h], & x = d \\ -p, & x < d \end{cases}.$$

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$



$\partial f(x)$ at $x = (0, 0)$ | at $x = (1, 0)$ | at $x = (1, 1)$

Figure: Subgradient of $f(x) = \|x\|_1$ on $\mathbb{R}^2 (d = 2)$

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$

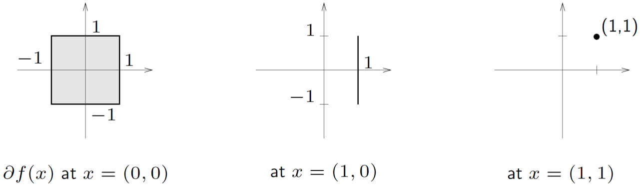


Figure: Subgradient of $f(x) = \|x\|_1$ on $\mathbb{R}^2 (d = 2)$

Example . $f(x) = \|x\|_2 = \max_{s:\|s\|_2\leq 1}\{s^T x\}$

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s \in \{-1,1\}^d} \{s^T x\}$

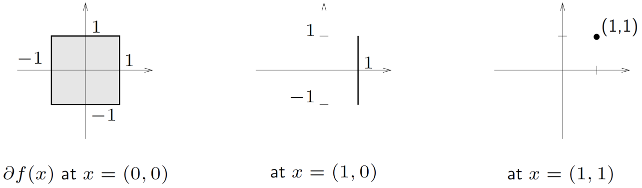


Figure: Subgradient of $f(x) = \|x\|_1$ on $\mathbb{R}^2 (d = 2)$

Example . $f(x) = \|x\|_2 = \max_{s: \|s\|_2 \le 1} \{s^T x\}$

$$\partial f(x) = \begin{cases} \frac{x}{\|x\|_2}, & x \neq 0 \\ \{s : \|s\|_2 \le 1\}, & x = 0 \end{cases}.$$

Question

Which function below is different from others?

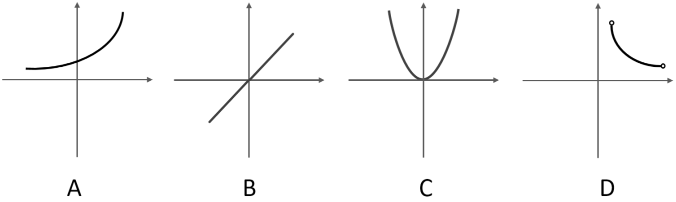


Figure: Convex functions

Existence of Global Minimizer

Definition. x^* is a *global minimizer* of $f(x)$ if

$$f(x^*) \leq f(x), \forall x.$$

Existence of Global Minimizer

Definition. x^* is a *global minimizer* of $f(x)$ if

$$f(x^*) \leq f(x), \forall x.$$

Definition. f is called *coercive* if all level sets are bounded, i.e., $f(x_k) \to \infty$ if $\|x_k\|_2 \to \infty$.

Existence of Global Minimizer

Definition. x^* is a *global minimizer* of $f(x)$ if

$$f(x^*) \leq f(x), \forall x.$$

Definition. f is called *coercive* if all level sets are bounded, i.e., $f(x_k) \to \infty$ if $\|x_k\|_2 \to \infty$.

Theorem. If f is closed (l.s.c.) and coercive, then it has a global minimizer.

Uniqueness of Global Minimizer

Recall f is strictly convex if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y), \forall \lambda \in (0,1), x \neq y.$$

- (sufficient condition): $\nabla^2 f(x) \succ 0$ (why?)

Uniqueness of Global Minimizer

Recall f is strictly convex if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y), \forall \lambda \in (0,1), x \neq y.$$

- (sufficient condition): $\nabla^2 f(x) \succ 0$ (why?)

Theorem. If f is strictly convex, then the global minimizer (if exists) must be unique.

Finding Global Minimizer

Theorem. Let f be convex. Then x^* is a global minimizer if and only if

$$0 \in \partial f(x^*).$$

Finding Global Minimizer

Theorem. Let f be convex. Then x^* is a global minimizer if and only if

$$0 \in \partial f(x^*).$$

- If f is convex and differentiable, x^* is a global minimizer iff $\nabla f(x^*) = 0$.

Finding Global Minimizer

Theorem. Let f be convex. Then x^* is a global minimizer if and only if

$$0 \in \partial f(x^*).$$

- If f is convex and differentiable, x^* is a global minimizer iff $\nabla f(x^*) = 0$.

Proof.

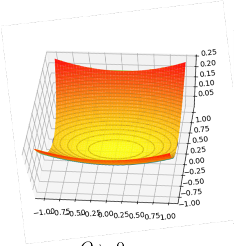
$$0 \in \partial f(x^*) \Leftrightarrow f(x) \geq f(x^*) + \langle 0, x - x^* \rangle = f(x^*), \forall x.$$

Example: Quadratic functions

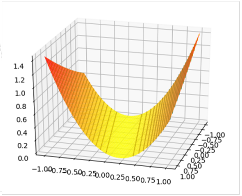
$$f(x) = \frac{1}{2}x^T Qx + b^T x + c$$

Example: Quadratic functions

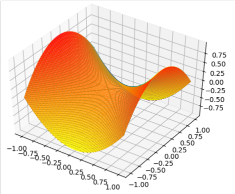
$$f(x) = \frac{1}{2}x^T Qx + b^T x + c$$



$Q \succ 0$



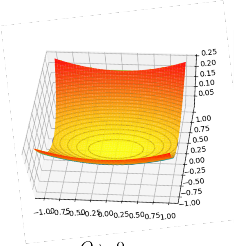
$Q \succeq 0$



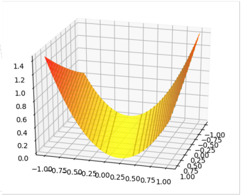
$Q \not\succeq 0$

Example: Quadratic functions

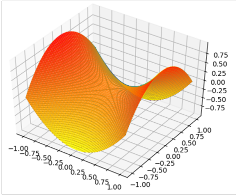
$$f(x) = \frac{1}{2}x^T Qx + b^T x + c$$



$Q \succ 0$



$Q \succeq 0$

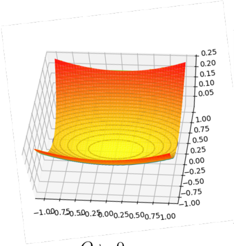


$Q \not\succeq 0$

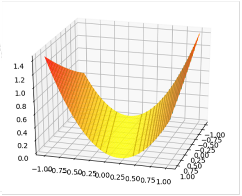
- $Q \succ 0$: strictly convex, unique minimizer $x^* = -Q^{-1}b$.

Example: Quadratic functions

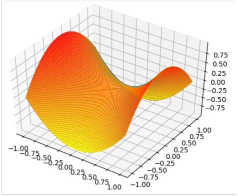
$$f(x) = \frac{1}{2}x^TQx + b^Tx + c$$



$Q \succ 0$



$Q \succeq 0$

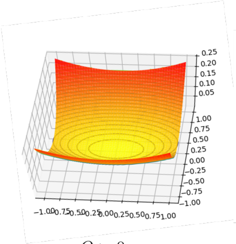


$Q \not\succeq 0$

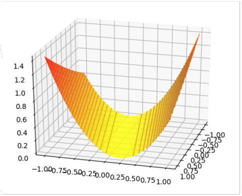
- $Q \succ 0$: strictly convex, unique minimizer $x^* = -Q^{-1}b$.
- $Q \succeq 0$ and $Q \not\succ 0$: convex but not strictly convex

Example: Quadratic functions

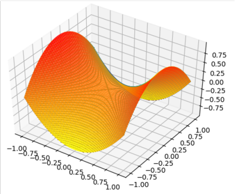
$$f(x) = \frac{1}{2}x^T Qx + b^T x + c$$



$Q \succ 0$



$Q \succeq 0$

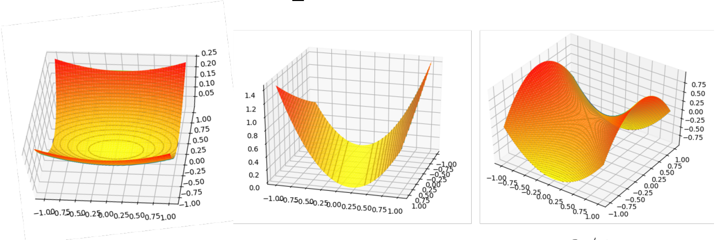


$Q \nsucceq 0$

- $Q \succ 0$: strictly convex, unique minimizer $x^* = -Q^{-1}b$.
- $Q \succeq 0$ and $Q \nsucc 0$: convex but not strictly convex
  - $b \in \text{range}(Q)$: infinitely many global minimizers

Example: Quadratic functions

$$f(x) = \frac{1}{2}x^TQx + b^Tx + c$$

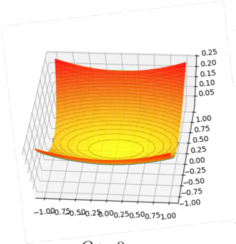


$Q \succ 0$ $\qquad$ $Q \succeq 0$ $\qquad$ $Q \not\succeq 0$

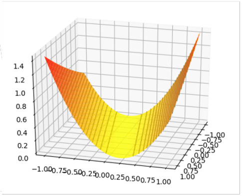
- $Q \succ 0$: strictly convex, unique minimizer $x^* = -Q^{-1}b$.
- $Q \succeq 0$ and $Q \not\succ 0$: convex but not strictly convex
  - $b \in \text{range}(Q)$: infinitely many global minimizers
  - $b \notin \text{range}(Q)$: unbounded and no global minimizer

Example: Quadratic functions

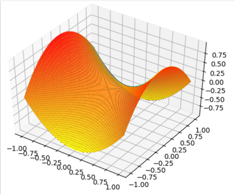
$$f(x) = \frac{1}{2}x^T Qx + b^T x + c$$



$Q \succ 0$



$Q \succeq 0$

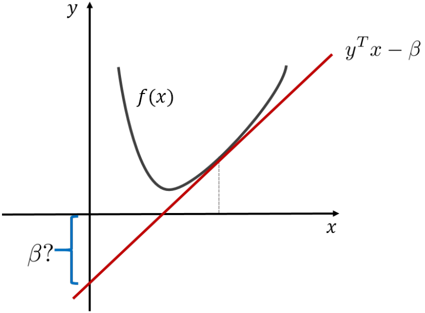


$Q \not\succeq 0$

- $Q \succ 0$: strictly convex, unique minimizer $x^* = -Q^{-1}b$.
- $Q \succeq 0$ and $Q \not\succ 0$: convex but not strictly convex
  - $b \in \text{range}(Q)$: infinitely many global minimizers
  - $b \notin \text{range}(Q)$: unbounded and no global minimizer
- $Q \not\succeq 0$: nonconvex, unbounded below

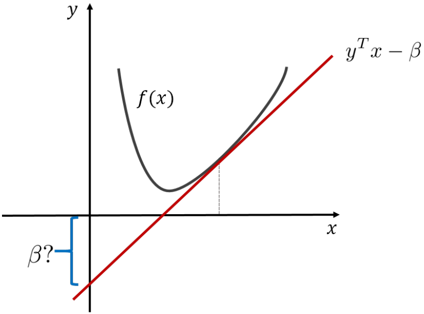
Question

For a given slope y, when is it an affine minorant?



Question

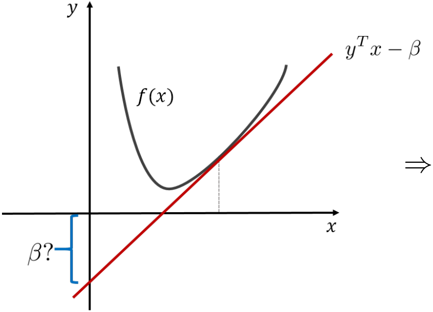
For a given slope y, when is it an affine minorant?



$$f(x) \geq y^T x - \beta, \forall x$$

Question

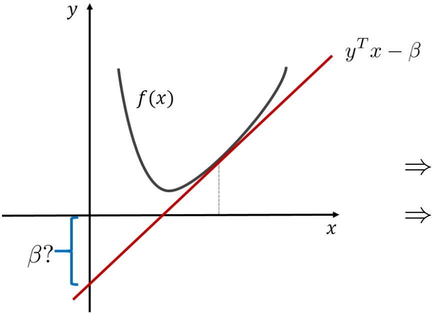
For a given slope y, when is it an affine minorant?



$$f(x) \geq y^T x - \beta, \forall x$$
$$\Rightarrow \beta \geq y^T x - f(x), \forall x$$

Question

For a given slope y, when is it an affine minorant?



$$
\begin{aligned}
& f(x) \geq y^T x - \beta, \forall x \\
\Rightarrow & \beta \geq y^T x - f(x), \forall x \\
\Rightarrow & \beta \geq \sup_x \{y^T x - f(x)\}
\end{aligned}
$$

Conjugate Function

Definition. The conjugate function of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\} = \sup_{x \in dom(f)} \left\{ y^T x - f(x) \right\}$$

Also called Legendre-Fenchel transformation.


A. Legendre (1752-1833)


Werner Fenchel (1905-1988)

Conjugate Function

Definition. The conjugate function of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\} = \sup_{x \in dom(f)} \left\{ y^T x - f(x) \right\}$$

Also called Legendre-Fenchel transformation.

Remark.

- **Fenchel's inequality:**

$$f(x) + f^*(y) \geq x^T y, \forall x, y$$

A. Legendre (1752-1833)

Werner Fenchel (1905-1988)

Conjugate Function

Definition. The conjugate function of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\} = \sup_{x \in dom(f)} \left\{ y^T x - f(x) \right\}$$

Also called Legendre-Fenchel transformation.

Remark.

- **Fenchel's inequality:**

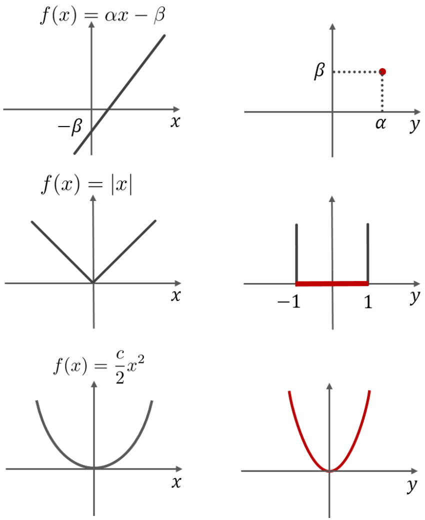
$$f(x) + f^*(y) \geq x^T y, \forall x, y$$

- f^* is convex and closed.

A. Legendre (1752-1833)

Werner Fenchel (1905-1988)

Examples



$$f^*(y) = \begin{cases} \beta, & y = \alpha \\ +\infty, & y \neq \alpha \end{cases}$$

$$f^*(y) = \begin{cases} 0, & |y| < 1 \\ +\infty, & |y| > 1 \end{cases}$$

$$f^*(y) = \frac{1}{2c}y^2$$

Figure: Examples of conjugate functions

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^T Qx + b^T x + c \quad (Q \succ 0)$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^T Qx + b^T x + c \quad (Q \succ 0)$

$$f^*(y) = \frac{1}{2}(x - b)^T Q^{-1}(x - b) - c$$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^T Qx + b^T x + c \quad (Q \succ 0)$

$$f^*(y) = \frac{1}{2}(x-b)^T Q^{-1}(x-b) - c$$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

$$f^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^TQx + b^Tx + c \quad (Q \succ 0)$

$$f^*(y) = \frac{1}{2}(x-b)^TQ^{-1}(x-b) - c$$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

$$f^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

$$f^*(y) = -\sum_{i=1}^n \log(-y_i) - n$$

More Examples

Example 4. Indicator function: $I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$

Example 5. Norm: $f(x) = \|x\|$

More Examples

Example 4. Indicator function: $I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$

$$I_C^*(y) = \sup_{x \in C} y^T x$$

Example 5. Norm: $f(x) = \|x\|$

More Examples

Example 4. Indicator function: $I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$

$$I_C^*(y) = \sup_{x \in C} y^T x$$

Example 5. Norm: $f(x) = \|x\|$

$$f^*(y) = \begin{cases} 0, & \|y\|_* \leq 1 \\ +\infty, & \|y\|_* > 1 \end{cases}$$

Calculus of Conjugate Functions

- Separable sum: If $g(x_1, x_2) = f_1(x_1) + f_2(x_2)$, then

$$g^*(y_1, y_2) = f_1^*(y_1) + f_2^*(y_2).$$

- Scaling: If $g(x) = \alpha f(x)$ with $\alpha > 0$, then

$$g^*(y) = \alpha f^*(y/\alpha).$$

- Summation: If $g(x) = f_1(x) + f_2(x)$, then

$$g^*(y) = \inf_z \{f_1^*(z) + f_2^*(y - z)\}$$

Biconjugate Function

- The conjugate of f is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\}$$

- The conjugate of the conjugate function $f^*(y)$,

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \left\{ x^T y - f^*(y) \right\}$$

Biconjugate Function

- The conjugate of f is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\}$$

- The conjugate of the conjugate function $f^*(y)$,

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \left\{ x^T y - f^*(y) \right\}$$

Q: is it true that $f^{**} = f$?

Conjugate Theory

Theorem.

(a) $f \geq f^{**}$.

(b) If f is closed and convex, then $f^{**} = f$.

Conjugate Theory

Theorem.

(a) $f \geq f^{**}$.

(b) If f is closed and convex, then $f^{**} = f$.

Proof.

(a) By definition of f^*, $f^*(y) \geq y^T x - f(x), \forall y$.

Conjugate Theory

Theorem.

(a) $f \geq f^{**}$.

(b) If f is closed and convex, then $f^{**} = f$.

Proof.

(a) By definition of f^*, $f^*(y) \geq y^T x - f(x), \forall y$.

$$\Rightarrow f(x) \geq y^T x - f^*(y), \forall y.$$

Conjugate Theory

Theorem.

(a) $f \geq f^{**}$.

(b) If f is closed and convex, then $f^{**} = f$.

Proof.

(a) By definition of f^*, $f^*(y) \geq y^T x - f(x), \forall y$.

$$\Rightarrow f(x) \geq y^T x - f^*(y), \forall y.$$

$$\Rightarrow f(x) \geq \sup_y \left\{ y^T x - f^*(y) \right\} = f^{**}(x).$$

Conjugate Theory

Proof (continued).

(b) Suppose $f^{**} \neq f$, then $\mathrm{epi}(f) \subsetneq \mathrm{epi}(f^{**})$

$$\exists (x_0, f^{**}(x_0)), \text{ s.t. } (x_0, t_0) \in \mathrm{epi}(f^{**}) \text{ and } (x_0, t_0) \notin \mathrm{epi}(f)$$

Conjugate Theory

Proof (continued).

(b) Suppose $f^{**} \neq f$, then $\text{epi}(f) \subsetneq \text{epi}(f^{**})$

$$\exists (x_0, f^{**}(x_0)), \text{ s.t. } (x_0, t_0) \in \text{epi}(f^{**}) \text{ and } (x_0, t_0) \notin \text{epi}(f)$$

- f is convex and closed, then $\text{epi}(f)$ is closed and convex.

Conjugate Theory

Proof (continued).

(b) Suppose $f^{**} \neq f$, then $\text{epi}(f) \subsetneq \text{epi}(f^{**})$

$$\exists (x_0, f^{**}(x_0)), \text{ s.t. } (x_0, t_0) \in \text{epi}(f^{**}) \text{ and } (x_0, t_0) \notin \text{epi}(f)$$

- f is convex and closed, then $\text{epi}(f)$ is closed and convex.
- By separation theorem, $\exists$ hyperplane with $(y, \beta) \neq 0$ separates $\text{epi}(f)$ and $(x_0, f^{**}(x_0))$:

$$y^T x + \beta t < y^T x_0 + \beta f^{**}(x_0), \forall (x, t) \in \text{epi}(f)$$

Conjugate Theory

Proof (continued).

(b) Suppose $f^{**} \neq f$, then $\text{epi}(f) \subsetneq \text{epi}(f^{**})$

$$\exists (x_0, f^{**}(x_0)), \text{ s.t. } (x_0, t_0) \in \text{epi}(f^{**}) \text{ and } (x_0, t_0) \notin \text{epi}(f)$$

- f is convex and closed, then $\text{epi}(f)$ is closed and convex.
- By separation theorem, $\exists$ hyperplane with $(y, \beta) \neq 0$ separates $\text{epi}(f)$ and $(x_0, f^{**}(x_0))$:

$$y^T x + \beta t < y^T x_0 + \beta f^{**}(x_0), \forall (x, t) \in \text{epi}(f)$$

- Note $\beta \leq 0$. W.l.o.g. let $\beta = -1$ (why?), and

$$y^T x - f(x) < y^T x_0 - f^{**}(x_0), \forall x$$

Conjugate Theory

Proof (continued).

(b) Suppose $f^{**} \neq f$, then $\text{epi}(f) \subsetneq \text{epi}(f^{**})$

$$\exists (x_0, f^{**}(x_0)), \text{ s.t. } (x_0, t_0) \in \text{epi}(f^{**}) \text{ and } (x_0, t_0) \notin \text{epi}(f)$$

- f is convex and closed, then $\text{epi}(f)$ is closed and convex.
- By separation theorem, $\exists$ hyperplane with $(y, \beta) \neq 0$ separates $\text{epi}(f)$ and $(x_0, f^{**}(x_0))$:

$$y^T x + \beta t < y^T x_0 + \beta f^{**}(x_0), \forall (x, t) \in \text{epi}(f)$$

- Note $\beta \leq 0$. W.l.o.g. let $\beta = -1$ (why?), and

$$y^T x - f(x) < y^T x_0 - f^{**}(x_0), \forall x$$

- This implies that $f^*(y) < y^T x_0 - f^{**}(x_0)$. Contradicts with Fenchel's inequality!

Conjugate Subgradient Theorem

Fenchel's inequality: $x^T y \leq f(x) + f^*(y)$.

Conjugate Subgradient Theorem

Fenchel's inequality: $x^T y \leq f(x) + f^*(y)$.

Theorem. Let f be closed and convex. Then

$$x^T y = f(x) + f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y).$$

Conjugate Subgradient Theorem

Fenchel's inequality: $x^T y \leq f(x) + f^*(y)$.

Theorem. Let f be closed and convex. Then

$$x^T y = f(x) + f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y).$$

Proof.

$$\begin{aligned} x^T y = f(x) + f^*(y) &\Leftrightarrow x^T y - f(x) = \sup_z \{z^T y - f(z)\} \\ &\Leftrightarrow x^T y - f(x) \geq z^T y - f(z), \forall z \\ &\Leftrightarrow f(z) \geq f(x) + y^T (z - x), \forall z \\ &\Leftrightarrow y \in \partial f(x) \end{aligned}$$

Implications on Global Minimizers

Corollary. Let f be closed and convex. Denote X^* as the set of global minimizers. Then

(a) $X^* = \partial f^*(0)$.

(b) X^* is nonempty if $0 \in \text{rint}(\text{dom}(f^*))$.

(c) X^* is nonempty and compact if $0 \in \text{int}(\text{dom}(f^*))$.

References

- Ben-Tal & Nemirovski, Chapter 2.5-2.6
- Bertsekas, Nedich, & Ozdaglar, Chapter 7