Conjugate Functions, Optimality Conditions, and Duality

Pontus Giselsson

Outline

- **Conjugate function – Definition and basic properties**
- Examples
- Biconjugate
- Fenchel-Young's inequality
- Duality correspondence
- Moreau decomposition
- Duality and optimality conditions
- Weak and strong duality

Conjugate Functions

Conjugate function – Definition

- The conjugate function of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined as

$$f^*(s) := \sup_x \left(s^T x - f(x) \right)$$

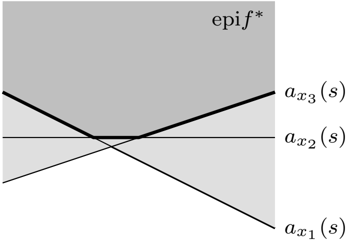
- Implicit definition via optimization problem

Conjugate function properties

- Let $a_x(s) := s^T x - f(x)$ be affine function parameterized by x:

$$f^*(s) = \sup_x a_x(s)$$

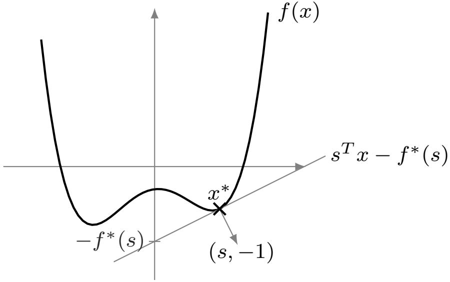
  is supremum of family of affine functions
- Epigraph of f^* is intersection of epigraphs of (below three) a_x



- f^* convex: epigraph intersection of convex halfspaces $\operatorname{epi} a_x$
- f^* closed: epigraph intersection of closed halfspaces $\operatorname{epi} a_x$

Conjugate interpretation

- Conjugate $f^*(s)$ defines affine minorizer to f with slope s:



where $-f^*(s)$ decides constant offset to get support

- Why?

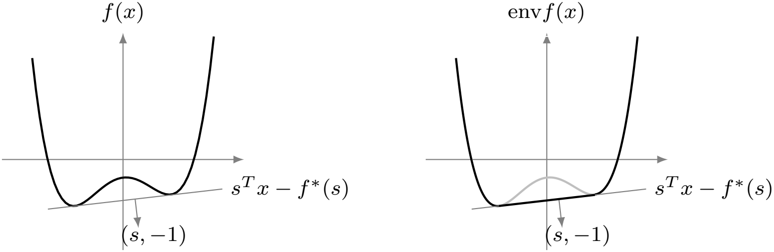
$$f^*(s) = \sup_x \left(s^Tx - f(x)\right) \quad \Leftrightarrow \quad f^*(s) \geq s^Tx - f(x) \text{ for all } x$$

$$\Leftrightarrow \quad f(x) \geq s^Tx - f^*(s) \text{ for all } x$$

- Maximizing argument x^* gives support: $f(x^*) = s^Tx^* - f^*(s)$
- We have $f(x^*) = s^Tx^* - f^*(s)$ if and only if $s \in \partial f(x^*)$

Consequence

- Conjugate of f and $\mathrm{env} f$ are the same, i.e., $f^* = (\mathrm{env} f)^*$



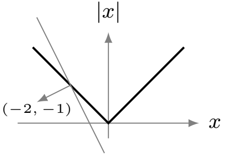
- Functions have same supporting affine functions
- Epigraphs have same supporting hyperplanes

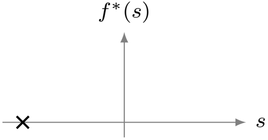
Outline

- Conjugate function – Definition and basic properties
- **Examples**
- Biconjugate
- Fenchel-Young's inequality
- Duality correspondence
- Moreau decomposition
- Duality and optimality conditions
- Weak and strong duality

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

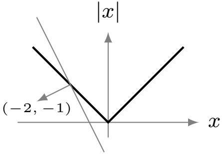


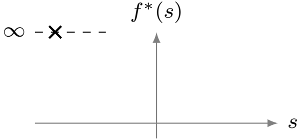


Slope, $s = -2$ $\quad f^*(s)$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis





Slope, $s = -2$ $\qquad$ $f^*(s) \to \infty$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

Slope, $s = -1$ $\quad$ $f^*(s)$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

Slope, $s = -1$ $\qquad f^*(s) = 0$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

Slope, $s = -0.5$ $\quad f^*(s)$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

Slope, $s = -0.5$ $\quad f^*(s) = 0$

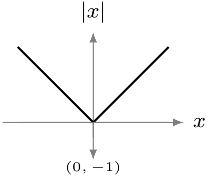
Example – Absolute value

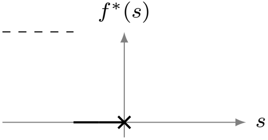
- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

Slope, $s = 0$ $f^*(s)$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

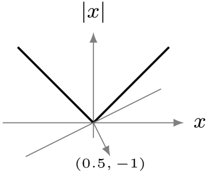


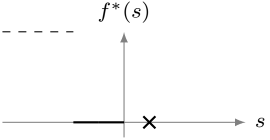


Slope, $s = 0$ $\qquad$ $f^*(s) = 0$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

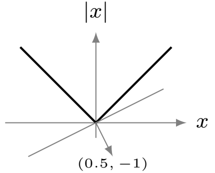


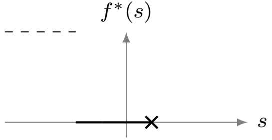


Slope, $s = 0.5$ $\qquad f^*(s)$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis





Slope, $s =0.5$ $\qquad f^*(s) = 0$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

Slope, $s = 1$ $\qquad$ $f^*(s)$

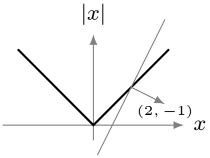
Example – Absolute value

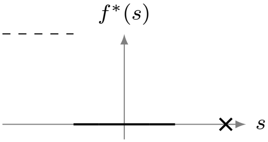
- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

Slope, s =1 $f^*(s) = 0$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis

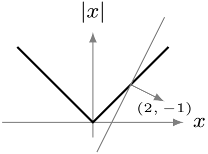




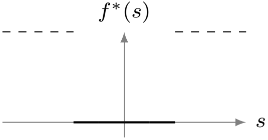
Slope, $s = 2$ $f^*(s)$

Example – Absolute value

- Compute conjugate of $f(x) = |x|$
- For given slope s: $-f^*(s)$ is point that crosses $|x|$-axis



Slope, $s = 2$

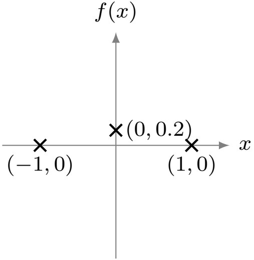


$f^*(s) \to \infty$

- Conjugate is $f^*(s) = \iota_{[-1,1]}(s)$

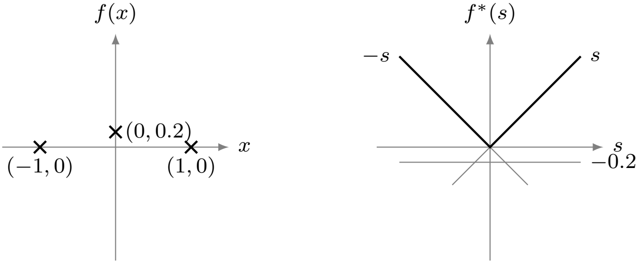
A nonconvex example

- Draw conjugate of f ($f(x) = \infty$ outside points)



A nonconvex example

- Draw conjugate of f ($f(x) = \infty$ outside points)



- Draw all affine $a_x(s)$ and select for each s the max to get $f^*(s)$

$$
\begin{aligned}
f^*(s) &= \sup_x (sx - f(x)) = \max(-s - 0, 0s - 0.2, s - 0) \\
&= \max(-s, -0.2, s) = |s|
\end{aligned}
$$

Example – Quadratic functions

Let $g(x) = \frac{1}{2}x^TQx + p^Tx$ with Q positive definite (invertible)

- Gradient satisfies $\nabla g(x) = Qx + p$
- Fermat's rule for $g^*(s) = \sup_x(s^Tx - \frac{1}{2}x^TQx - p^Tx)$:

$$0 = s - Qx - p \quad \Leftrightarrow \quad x = Q^{-1}(s-p)$$

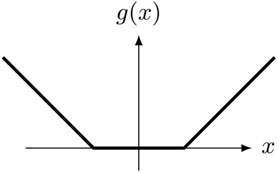
- So

$$\begin{aligned} g^*(s) &= s^TQ^{-1}(s-p) - \tfrac{1}{2}(s-p)^TQ^{-1}QQ^{-1}(s-p) + p^TQ^{-1}(s-p) \\ &= \tfrac{1}{2}(s-p)^TQ^{-1}(s-p) \end{aligned}$$

Example – A piece-wise linear function

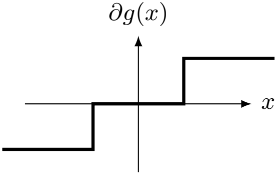
- Consider

$$g(x) = \begin{cases} -x-1 & \text{if } x \le -1 \\ 0 & \text{if } x \in [-1,1] \\ x-1 & \text{if } x \ge 1 \end{cases}$$



- Subdifferential satisfies

$$\partial g(x) = \begin{cases} -1 & \text{if } x < -1 \\ [-1,0] & \text{if } x = -1 \\ 0 & \text{if } x \in (-1,1) \\ [0,1] & \text{if } x = 1 \\ 1 & \text{if } x > 1 \end{cases}$$



Example cont'd

- We use $g^*(s) = sx - g(x)$ if $s \in \partial g(x)$:
  - $x < -1$: $s = -1$, hence $g^*(-1) = -1x - (-x - 1) = 1$
  - $x = -1$: $s \in [-1, 0]$ hence $g^*(s) = -s - 0 = -s$
  - $x \in (-1, 1)$: $s = 0$ hence $g^*(0) = 0x - 0 = 0$
  - $x = 1$: $s \in [0, 1]$ hence $g^*(s) = s - 0 = s$
  - $x > 1$: $s = 1$ hence $g^*(1) = x - (x - 1) = 1$
- That is

$$g^*(s) = \begin{cases} -s & \text{if } s \in [-1, 0] \\ s & \text{if } s \in [0, 1] \end{cases}$$

- For $s < -1$ and $s > 1$, $g^*(s) = \infty$:
  - $s < -1$: let $x = t \to -\infty$ and $g^*(s) \geq ((s + 1)t + 1) \to \infty$
  - $s > 1$: let $x = t \to \infty$ and $g^*(s) \geq ((s - 1)t + 1) \to \infty$

Example – Separable functions

- Let $f(x) = \sum_{i=1}^{n} f_i(x_i)$ be a separable function, then

$$f^*(s) = \sum_{i=1}^{n} f_i^*(s_i)$$

  is also separable
- Proof:

$$\begin{aligned} f^*(s) &= \sup_x (s^T x - \sum_{i=1}^{n} f_i(x_i)) \\ &= \sup_x (\sum_{i=1}^{n} (s_i x_i - f_i(x_i))) \\ &= \sum_{i=1}^{n} \sup_{x_i} (s_i x_i - f_i(x_i)) \\ &= \sum_{i=1}^{n} f_i^*(s_i) \end{aligned}$$

Example – 1-norm

- Let $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$ be the 1-norm
- It is a separable sum of absolute values
- Use separable sum formula and that $|\cdot|^* = \iota_{[-1,1]}$:

$$f^*(s) = \sum_{i=1}^n f_i^*(s_i) = \sum_{i=1}^n \iota_{[-1,1]}(s_i) = \begin{cases} 0 & \text{if } \max_i(|s_i|) \leq 1 \\ \infty & \text{else} \end{cases}$$

- We have $\max_i(|s_i|) = \|s\|_\infty$, let

$$B_\infty(r) = \{s : \|s\|_\infty \leq r\}$$

be the infinity norm ball of radius r, then

$$f^*(s) = \iota_{B_\infty(1)}(s)$$

is the indicator function for the unit infinity norm ball

Outline

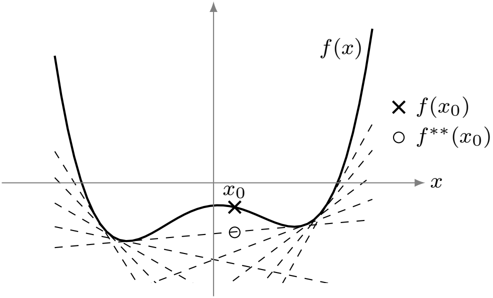
- Conjugate function – Definition and basic properties
- Examples
- **Biconjugate**
- Fenchel-Young's inequality
- Duality correspondence
- Moreau decomposition
- Duality and optimality conditions
- Weak and strong duality

Biconjugate

- Biconjuate $f^{**} := (f^*)^*$ is conjugate of conjugate

$$f^{**}(x) = \sup_s (x^T s - f^*(s))$$

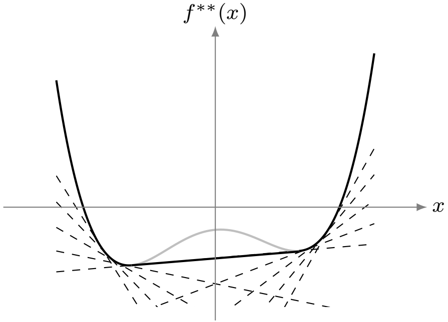
- For every x, it is largest value of all affine minorizers



- Why?:
  - $x^T s - f^*(s)$: supporting affine minorizer to f with slope s
  - $f^{**}(x)$ picks largest over all these affine minorizers evaluated at x

Biconjugate and convex envelope

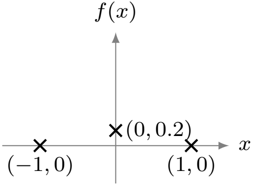
- Biconjugate is closed convex envelope of f



- $f^{**} \leq f$ and $f^{**} = f$ if and only if f (closed and) convex

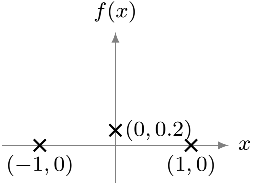
Biconjugate – Example

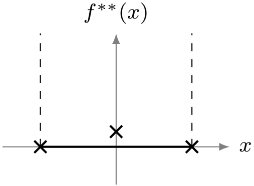
- Draw the biconjugate of f $\big(f(x) = \infty$ outside points$\big)$



Biconjugate – Example

- Draw the biconjugate of f $(f(x) = \infty$ outside points)





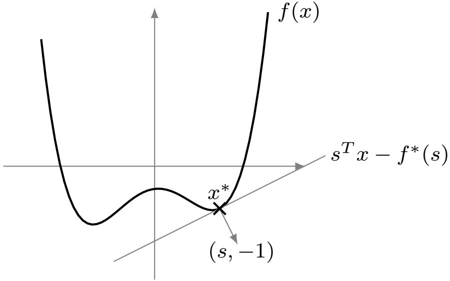
- Biconjugate is convex envelope of f
- We found before $f^*(s) = |s|$, and now $(f^*)^*(x) = \iota_{[-1,1]}(x)$
- Therefore also $\iota^*_{[-1,1]}(s) = |s|$
  (since $f^* = (\text{env} f)^* = (f^{**})^* =: f^{***}$)

Outline

- Conjugate function – Definition and basic properties
- Examples
- Biconjugate
- **Fenchel-Young's inequality**
- Duality correspondence
- Moreau decomposition
- Duality and optimality conditions
- Weak and strong duality

Fenchel-Young's inequality

- Going back to conjugate interpretation:

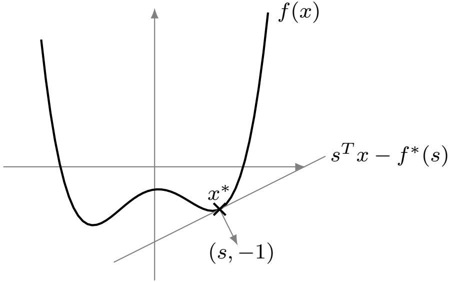


- Fenchel-Youngs's inequality: $f(x) \geq s^T x - f^*(s)$ for all x, s
- Follows immediately from definition: $f^*(s) = \sup_x (s^T x - f(x))$

Fenchel-Young's equality

- When is do we have equality in Fenchel-Young?

$$f(x) = s^T x - f^*(s)$$



- Fenchel-Young's equality and equivalence:

$$\boxed{f(x^*) = s^T x^* - f^*(s) \text{ holds if and only if } s \in \partial f(x^*)}$$

Proof – Fenchel-Young's equality

$$\boxed{f(x) = s^T x - f^*(s) \text{ holds if and only if } s \in \partial f(x)}$$

- $s \in \partial f(x)$ if and only if (by defintion of subgradient)

$$\begin{aligned} & f(y) \geq f(x) + s^T(y - x) \text{ for all } y \\ \Leftrightarrow \quad & s^T x - f(x) \geq s^T y - f(y) \text{ for all } y \\ \Leftrightarrow \quad & s^T x - f(x) \geq \sup_y \big(s^T y - f(y)\big) \\ \Leftrightarrow \quad & s^T x - f(x) \geq f^*(s) \end{aligned}$$

  which is Fenchel-Young's inequality with inequality reversed

- Fenchel-Young's inequality always holds:

$$f^*(s) \geq s^T x - f(x)$$

  so we have equality if and only if $s \in \partial f(x)$

A subdifferential formula for convex f

$\boxed{\text{Assume } f \text{ closed convex, then } \partial f(x) = \operatorname{Argmax}_s(s^T x - f^*(s))}$

- Since $f^{**} = f$, we have $f(x) = \sup_s(x^T s - f^*(s))$ and

$$\begin{aligned} s^* \in \operatorname*{Argmax}_s(x^T s - f^*(s)) \quad &\iff \quad f(x) = x^T s^* - f^*(s^*) \\ &\iff \quad s^* \in \partial f(x) \end{aligned}$$

- The last equivalence is from previous slide

Subdifferential formulas for f^*

- For general f, we have that

$$\partial f^*(s) = \operatorname*{Argmax}_x (s^T x - f^{**}(x))$$

  by previous formula and since f^* closed and convex

- For closed convex f, we have, since $f = f^{**}$, that

$$\partial f^*(s) = \operatorname*{Argmax}_x (s^T x - f(x))$$

Relation between ∂f and ∂f^* – General case

$$\boxed{s \in \partial f(x) \text{ implies that } x \in \partial f^*(s)}$$

- Since $f^{**} \leq f$ and $s \in \partial f(x)$, Fenchel-Young's equality gives:

$$0 = f^*(s) + f(x) - s^T x \geq f^*(s) + f^{**}(x) - s^T x \geq 0$$

  where last step is Fenchel-Young's inequality

- Hence $f^*(s) + f^{**}(x) - s^T x = 0$ and FY $\Rightarrow x \in \partial f^*(s)$

Inverse relation between ∂f and ∂f^* – Convex case

$$\boxed{\text{Suppose } f \text{ closed convex, then } s \in \partial f(x) \iff x \in \partial f^*(s)}$$

- Using implication on previous slide twice and $f^{**} = f$:

$$s \in \partial f(x) \Rightarrow x \in \partial f^*(s) \Rightarrow s \in \partial f^{**}(x) \Rightarrow s \in \partial f(x)$$

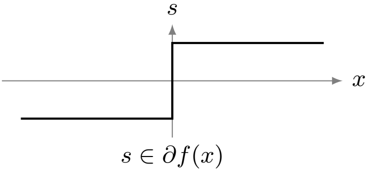
- Another way to write the result is that for closed convex f:

$$\partial f^* = (\partial f)^{-1}$$

(Definition of inverse of set-valued A: $x \in A^{-1}u \iff u \in Ax$)

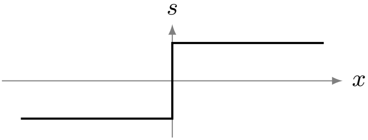
Example 1 – Relation between ∂f and ∂f^*

- What is ∂f^* for below ∂f?

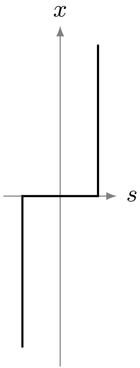


Example 1 – Relation between ∂f and ∂f^*

- What is ∂f^* for below ∂f?



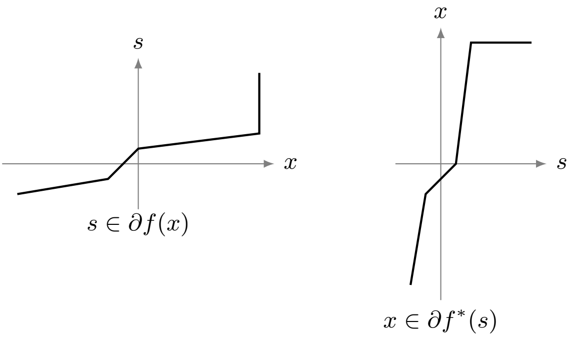
$s \in \partial f(x)$



$x \in \partial f^*(s)$

- Since $\partial f^* = (\partial f)^{-1}$, we flip the figure

Example 2 – Relation between ∂f and ∂f^*



- region with slope σ in $\partial f(x)$ $\Leftrightarrow$ region with slope $\frac{1}{\sigma}$ in $\partial f^*(s)$
- Implication: ∂f σ-strong monotone $\Leftrightarrow$ $\partial f^*(s)$ σ-cocoercive?
  (Recall: σ-cocoercivity $\Leftrightarrow$ $\frac{1}{\sigma}$-Lipschitz and monotone)

Outline

- Conjugate function – Definition and basic properties
- Examples
- Biconjugate
- Fenchel-Young's inequality
- **Duality correspondence**
- Moreau decomposition
- Duality and optimality conditions
- Weak and strong duality

Cocoercivity and strong monotonicity

$$\boxed{\begin{array}{c}\partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n} \text{ maximal monotone and } \sigma\text{-strongly monotone} \\ \iff \\ \partial f^* = \nabla f^* : \mathbb{R}^n \to \mathbb{R}^n \text{ single-valued and } \sigma\text{-cocoercive}\end{array}}$$

- σ-strong monotonicity: for all $u \in \partial f(x)$ and $v \in \partial f(y)$

$$(u - v)^T (x - y) \geq \sigma \|x - y\|_2^2 \tag{1}$$

  or equivalently for all $x \in \partial f^*(u)$ and $y \in \partial f^*(v)$
- ∂f^* is single-valued:
  - Assume $x \in \partial f^*(u)$ and $y \in \partial f^*(u)$, then lhs of (1) 0 and $x = y$
- ∇f^* is σ-cocoercive: plug $x = \nabla f^*(u)$ and $y = \nabla f^*(v)$ into (1)
- That ∂f^* has full domain follows from Minty's theorem

Duality correspondance

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Then the following are equivalent:

(i) f is closed and σ-strongly convex
(ii) ∂f is maximally monotone and σ-strongly monotone
(iii) ∇f^* is σ-cocoercive
(iv) ∇f^* is maximally monotone and $\frac{1}{\sigma}$-Lipschitz continuous
(v) f^* is closed convex and satisfies descent lemma (is $\frac{1}{\sigma}$-smooth)

where $\nabla f^* : \mathbb{R}^n \to \mathbb{R}^n$ and $f^* : \mathbb{R}^n \to \mathbb{R}$

Comments:

- (i) $\Leftrightarrow$ (ii) and (iii) $\Leftrightarrow$ (iv) $\Leftrightarrow$ (v): Previous lecture
- (ii) $\Leftrightarrow$ (iii): This lecture
- Since $f = f^{**}$ the result holds with f and f^* interchanged
- Full proof available on course webpage

Example – Proximal operator is 1-cocoercive

$$\boxed{\text{Assume } g \text{ closed convex, then } \mathrm{prox}_{\gamma g} \text{ is 1-cocoercive}}$$

- Prox definition $\mathrm{prox}_{\gamma g}(z) = \mathrm{argmin}_x(g(x) + \frac{1}{2\gamma}\|x - z\|_2^2)$
- Let $r = \gamma g + \frac{1}{2}\|\cdot\|_2^2$, then

$$\begin{aligned}
\mathrm{prox}_{\gamma g}(z) &= \underset{x}{\mathrm{argmin}}(g(x) + \tfrac{1}{2\gamma}\|x - z\|_2^2) \\
&= \underset{x}{\mathrm{argmax}}(-\gamma g(x) - \tfrac{1}{2}\|x - z\|_2^2) \\
&= \underset{x}{\mathrm{argmax}}(z^T x - (\tfrac{1}{2}\|x\|_2^2 + \gamma g(x))) \\
&= \underset{x}{\mathrm{argmax}}(z^T x - r(x)) \\
&= \nabla r^*(z)
\end{aligned}$$

  where last step is subdifferential formula for r^* for convex r
- Now, r is 1-strongly convex and $\nabla r^* = \mathrm{prox}_{\gamma g}$ is 1-cocoercive

Example – Proximal operator for strongly convex g

$\boxed{\text{Assume } g \text{ is } \sigma\text{-strongly convex, then } \mathrm{prox}_{\gamma g} \text{ is } (1+\gamma\sigma)\text{-cocoercive}}$

- Let $r = \gamma g + \frac{1}{2}\|\cdot\|_2^2$, and use $\mathrm{prox}_{\gamma g}(z) = \nabla r^*(z)$
- r is $(1+\gamma\sigma)$-strongly convex and ∇r^* is $(1+\gamma\sigma)$-cocoercive

Outline

- Conjugate function – Definition and basic properties
- Examples
- Biconjugate
- Fenchel-Young's inequality
- Duality correspondence
- **Moreau decomposition**
- Duality and optimality conditions
- Weak and strong duality

Moreau decomposition – Statement

$$\boxed{\text{Assume } g \text{ closed convex, then } \mathrm{prox}_g(z) + \mathrm{prox}_{g^*}(z) = z}$$

- When g scaled by $\gamma > 0$, Moreau decomposition is

  $$z = \mathrm{prox}_{\gamma g}(z) + \mathrm{prox}_{(\gamma g)^*}(z) = \mathrm{prox}_{\gamma g}(z) + \gamma \mathrm{prox}_{\gamma^{-1} g^*}(\gamma^{-1} z)$$

  (since $\mathrm{prox}_{(\gamma g)^*} = \gamma \mathrm{prox}_{\gamma^{-1} g^*} \circ \gamma^{-1} \mathrm{Id}$)
- Don't need to know g^* to compute $\mathrm{prox}_{\gamma g^*}$

Moreau decomposition – Proof

- Let $u = z - x$
- Fermat's rule: $x = \mathrm{prox}_g(z)$ if and only if

$$\begin{aligned} 0 \in \partial g(x) + x - z \quad &\Leftrightarrow \quad z - x \in \partial g(x) \\ &\Leftrightarrow \quad u \in \partial g(x) \\ &\Leftrightarrow \quad x \in \partial g^*(u) \\ &\Leftrightarrow \quad z - u \in \partial g^*(u) \\ &\Leftrightarrow \quad 0 \in \partial g^*(u) + u - z \end{aligned}$$

  if and only if $u = \mathrm{prox}_{g^*}(z)$ by Fermat's rule
- Using $z = x + u$, we get

$$z = x + u = \mathrm{prox}_g(z) + \mathrm{prox}_{g^*}(z)$$

Optimality Conditions and Duality

Outline

- Conjugate function – Definition and basic properties
- Examples
- Biconjugate
- Fenchel-Young's inequality
- Duality correspondence
- Moreau decomposition
- **Duality and optimality conditions**
- Weak and strong duality

Composite optimization problem

- Consider *primal* composite optimization problem

$$\operatorname{minimize} f(Lx) + g(x)$$

  where f, g closed convex and L is a matrix
- We will derive primal-dual optimality conditions and dual problem

Primal optimality condition

Let $f : \mathbb{R}^m \to \overline{\mathbb{R}}$, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, $L \in \mathbb{R}^{m \times n}$ with f, g closed convex and assume CQ, then:

$$\operatorname{minimize} f(Lx) + g(x)$$

is solved by $x^\star \in \mathbb{R}^n$ if and only if $x^\star$ satisfies

$$0 \in L^T \partial f(Lx^\star) + \partial g(x^\star)$$

- Optimality condition implies that vector s exists such that

$$s \in L^T \partial f(Lx^\star) \qquad \text{and} \qquad -s \in \partial g(x^\star)$$

- So CQ implies a subgradient exists for both functions at solution

Primal-dual optimality condition 1

- Introduce *dual* variable $\mu \in \partial f(Lx)$, then optimality condition

$$0 \in L^T \underbrace{\partial f(Lx)}_{\mu} + \partial g(x)$$

  is equivalent to

$$\begin{aligned} \mu &\in \partial f(Lx) \\ -L^T\mu &\in \partial g(x) \end{aligned}$$

- This is a necessary and sufficient primal-dual optimality condition
- (*Primal-dual* since involves primal x and dual μ variables)

Primal-dual optimality condition 2

- Primal-dual optimality condition

$$
\begin{aligned}
\mu &\in \partial f(Lx) \\
-L^T\mu &\in \partial g(x)
\end{aligned}
$$

- Using subdifferential inverse:

$$
\mu \in \partial f(Lx) \qquad \Longleftrightarrow \qquad Lx \in \partial f^*(\mu)
$$

gives equivalent primal dual optimality condition

$$
\begin{aligned}
Lx &\in \partial f^*(\mu) \\
-L^T\mu &\in \partial g(x)
\end{aligned}
$$

Dual optimality condition

- Using subdifferential inverse on other condition

$$-L^T\mu \in \partial g(x) \qquad \Longleftrightarrow \qquad x \in \partial g^*(-L^T\mu)$$

gives equivalent primal dual optimality condition

$$\begin{aligned} Lx &\in \partial f^*(\mu) \\ x &\in \partial g^*(-L^T\mu) \end{aligned}$$

- This is equivalent to that:

$$0 \in \partial f^*(\mu) - L\,\underbrace{\partial g^*(-L^T\mu)}_{x}$$

which is a dual optimality condition since it involves only μ

Dual problem

- The dual optimality condition

$$0 \in \partial f^*(\mu) - L\partial g^*(-L^T\mu)$$

is a sufficient condition for solving the *dual problem*

$$\text{minimize}\, f^*(\mu) + g^*(-L^T\mu)$$

- Have also necessity under CQ on dual, which is mild

Why dual problem?

- Sometimes easier to solve than primal
- Only useful if primal solution can be obtained from dual

Solving primal from dual

- Assume f, g closed convex and CQ holds
- Assume optimal dual μ known: $0 \in \partial f^*(\mu) - L\partial g^*(-L^T\mu)$
- Optimal primal x must satisfy any and all primal-dual conditions:

$$\begin{cases} \mu \in \partial f(Lx) \\ -L^T\mu \in \partial g(x) \end{cases} \qquad \begin{cases} Lx \in \partial f^*(\mu) \\ -L^*\mu \in \partial g(x) \end{cases}$$

$$\begin{cases} \mu \in \partial f(Lx) \\ x \in \partial g^*(-L^T\mu) \end{cases} \qquad \begin{cases} Lx \in \partial f^*(\mu) \\ x \in \partial g^*(-L^T\mu) \end{cases}$$

- If one of these uniquely characterizes x, then must be solution:
  - g^* is differentiable at $-L^T\mu$ for dual solution μ
  - f^* is differentiable at dual solution μ and L invertible
  - $\ldots$

Optimality conditions – Summary

- Assume f, g closed convex and that CQ holds
- Problem $\min_x f(Lx) + g(x)$ is solved by x if and only if

$$0 \in L^T \partial f(Lx) + \partial g(x)$$

- Primal dual necessary and sufficient optimality conditions:

$$\begin{cases} \mu \in \partial f(Lx) \\ -L^T\mu \in \partial g(x) \end{cases} \qquad \begin{cases} Lx \in \partial f^*(\mu) \\ -L^T\mu \in \partial g(x) \end{cases}$$

$$\begin{cases} \mu \in \partial f(Lx) \\ x \in \partial g^*(-L^T\mu) \end{cases} \qquad \begin{cases} Lx \in \partial f^*(\mu) \\ x \in \partial g^*(-L^T\mu) \end{cases}$$

- Dual optimality condition

$$0 \in \partial f^*(\mu) - L\partial g^*(-L^T\mu)$$

solves dual problem $\min_\mu f^*(\mu) + g^*(-L^T\mu)$

Outline

- Conjugate function – Definition and basic properties
- Examples
- Biconjugate
- Fenchel-Young's inequality
- Duality correspondence
- Moreau decomposition
- Duality and optimality conditions
- **Weak and strong duality**

Concave dual problem

- We have defined dual as convex minimization problem

$$\underset{\mu}{\text{minimize}}\, f^*(\mu) + g^*(-L^T\mu)$$

- Dual problem can be written as concave maximization problem:

$$\underset{\mu}{\text{maximize}} - f^*(\mu) - g^*(-L^T\mu)$$

- Same solutions but optimal values minus of each other
- Concave formulation gives nicer optimal value comparisons
- To compare, we let the primal and dual optimal values be

$$p^\star = \inf_x (f(Lx) + g(x)) \qquad \text{and} \qquad d^\star = \sup_\mu (-f^*(\mu) - g^*(-L^T\mu))$$

Weak duality

Weak duality always holds meaning $p^\star \geq d^\star$

- We have by Fenchel-Young's inequality for all μ and x:

$$\begin{aligned} f^*(\mu) + g^*(-L^T\mu) &\geq \mu^T Lx - f(Lx) + (-L^T\mu)^T x - g(x) \\ &= -f(Lx) - g(x) \end{aligned}$$

- Negate, maximize lhs over μ, minimize rhs over x, to get

$$d^\star = \sup_\mu(-f^*(\mu) - g^*(-L^T\mu)) \leq \inf_x(f(Lx) + g(x)) = p^\star$$

Strong duality

$$\boxed{\begin{array}{c}\text{Assume } f, g \text{ closed convex, solution } x^\star \text{ exists, and CQ} \\ \text{then } \textit{strong duality} \text{ holds meaning } p^\star = d^\star\end{array}}$$

- Dual $\mu^\star$ and primal $x^\star$ solutions exist such that

$$\mu^\star \in \partial f(Lx^\star) \qquad \text{and} \qquad -L^T\mu^\star \in \partial g(x^\star)$$

- We have by Fenchel-Young's equality:

$$\begin{aligned} p^\star &= f(Lx^\star) + g(x^\star) \\ &= (\mu^\star)^T Lx^\star - f^*(\mu^\star) + (-L^T\mu^\star)^T x^\star - g^*(-L^T\mu^\star) \\ &= -f^*(\mu^\star) - g^*(-L^T\mu^\star) = d^\star \end{aligned}$$

Dual problem gives lower bound

- Consider again concave dual problem with optimal value

$$d^\star = \sup_\mu(-f^*(\mu) - g^*(-L^T\mu))$$

- We know that for all dual variables μ

$$p^\star \geq d^\star \geq -f^*(\mu) - g^*(-L^T\mu)$$

- So can find lower bound to $p^\star$ by evaluating dual objective