Convex Optimization

Part 4: Duality V

Namhoon Lee

POSTECH

14 Nov 2022

Conjugate function

the conjugate of a function f is

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

f^* is closed and convex (even when f is not)

Fenchel's inequality: the definition implies that

$$f(x) + f^*(y) \geq x^\top y \quad \text{for all } x, y$$

this is an extension to non-quadratic convex f of the inequality

$$\frac{1}{2}x^\top x + \frac{1}{2}y^\top y \geq x^\top y$$

Calculus rules

Separable sum

$$f(x_1, x_2) = g(x_1) + h(x_2) \qquad f^*(y_1, y_2) = g^*(y_1) + h^*(y_2)$$

Scalar multiplication ($\alpha > 0$)

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

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

- the operation $f(x) = \alpha g(x/\alpha)$ is sometimes called "right scalar multiplication"
- a convenient notation is $f = g\alpha$ for the function $(g\alpha)(x) = \alpha g(x/\alpha)$
- conjugates can be written concisely as $(g\alpha)^* = \alpha g^*$ and $(\alpha g)^* = g^*\alpha$

Calculus rules

Addition to affine function

$$f(x) = g(x) + a^\top x + b \qquad f^*(y) = g^*(y-a) - b$$

Translation of argument

$$f(x) = g(x-b) \qquad f^*(y) = b^\top y + g^*(y)$$

Composition with invertible linear mapping: if A is square and nonsingular

$$f(x) = g(Ax) \qquad f^*(y) = g^*(A^{-\top} y)$$

Infimal convolution

$$f(x) = \inf_{u+v=x} (g(u) + h(v)) \qquad f^*(y) = g^*(y) + h^*(y)$$

Conjugates and subgradients

if f is closed and convex, then

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

Proof.

if $y \in \partial f(x)$, then $f^*(y) = \sup_u (y^\top u - f(u)) = y^\top x - f(x)$; hence

$$\begin{aligned}
f^*(v) &= \sup_u (v^\top u - f(u)) \\
&\geq v^\top x - f(x) \\
&= x^\top (v - y) - f(x) + y^\top x \\
&= f^*(y) + x^\top (v - y)
\end{aligned}$$

this holds for all v; therefore, $x \in \partial f^*(y)$

reverse implication $x \in \partial f^*(y) \Longrightarrow y \in \partial f(x)$ follows from $f^{**} = f$ □

Conjugate of strongly convex function

assume f is closed and strongly convex with parameter $\mu > 0$ for the norm $\|\cdot\|$

- f^* is defined for all y (*i.e.*, $\operatorname{dom} f^* = \mathbb{R}^n$)
- f^* is differentiable everywhere, with gradient

$$\nabla f^*(y) = \operatorname*{arg\,max}_x \left(y^\top x - f(x)\right)$$

- ∇f^* is Lipschitz continuous with constant $1/\mu$ for the dual norm $\|\cdot\|_*$:

$$\|\nabla f^*(y) - \nabla f^*(y')\| \le \frac{1}{\mu}\|y - y'\|_* \qquad \text{for all } y \text{ and } y'$$

Proof.

if f is strongly convex and closed

- $y^\top x - f(x)$ has a unique maximizer x for every y
- x maximizes $y^\top x - f(x)$ if and only if $y \in \partial f(x)$;

$$y \in \partial f(x) \quad \Longleftrightarrow \quad x \in \partial f^*(y)$$

  hence $\nabla f^*(y) = \arg\max_x (y^\top x - f(x))$

- from first-order condition: if $y \in \partial f(x)$, $y' \in \partial f(x')$:

$$f(x') \geq f(x) + y^\top (x' - x) + \frac{\mu}{2}\|x' - x\|^2$$
$$f(x) \geq f(x') + (y')^\top (x - x') + \frac{\mu}{2}\|x' - x\|^2$$

  combining these inequalities shows

$$\mu\|x - x'\|^2 \leq (y - y')^\top (x - x') \leq \|y - y'\|_* \|x - x'\|$$

- now substitute $x = \nabla f^*(y)$ and $x' = \nabla f^*(y')$

Moreau decomposition

$$x = \mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x) \qquad \text{for all } x$$

- follows from properties of conjugates and subgradients:

$$\begin{aligned} u = \mathrm{prox}_f(x) \quad &\Longleftrightarrow \quad x - u \in \partial f(u) \\ &\Longleftrightarrow \quad u \in \partial f^*(x - u) \\ &\Longleftrightarrow \quad x - u = \mathrm{prox}_{f^*}(x) \end{aligned}$$

- generalizes decomposition by orthogonal projection on subspaces:

$$x = \mathrm{P}_L(x) + \mathrm{P}_{L^\perp}(x)$$

  if L is a subspace, $L^\perp$ its orthogonal complement
  (this is the Moreau decomposition with $f = \delta_L$, $f^* = \delta_{L^\perp}$)

Extended Moreau decomposition

for $\lambda > 0$,

$$x = \operatorname{prox}_{\lambda f}(x) + \lambda \operatorname{prox}_{\lambda^{-1} f^*}(x/\lambda) \quad \text{for all } x$$

Proof.

apply Moreau decomposition to λf

$$\begin{aligned} x &= \operatorname{prox}_{\lambda f}(x) + \operatorname{prox}_{(\lambda f)^*}(x) \\ &= \operatorname{prox}_{\lambda f}(x) + \lambda \operatorname{prox}_{\lambda^{-1} f^*}(x/\lambda) \end{aligned}$$

second line uses $(\lambda f)^*(y) = \lambda f^*(y/\lambda)$ and that for $f(x) = \lambda g(x/\lambda)$,
$\operatorname{prox}_f(x) = \lambda \operatorname{prox}_{\lambda^{-1} g}(x/\lambda)$ □

Duality

primal: minimize $f(x) + g(Ax)$

dual: maximize $-g^*(z) - f^*(-A^\top z)$

- follows from Lagrange duality applied to reformulated primal

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax = y \end{array}$$

dual function for the formulated problem is:

$$\inf_{x,y} \left(f(x) + z^\top Ax + g(y) - z^\top y\right) = -f^*(-A^\top z) - g^*(z)$$

- Slater's condition (for convex f, g): strong duality holds if there exists an $\tilde{x}$ with

$$\tilde{x} \in \operatorname{int} \operatorname{dom} f, \qquad A\tilde{x} \in \operatorname{int} \operatorname{dom} g$$

this also guarantees that the dual optimum is attained if optimal value is finite

Set constraint

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax - b \in C \end{array}$$

Primal and dual problem

primal: $\quad \text{minimize} \quad f(x) + \delta_C(Ax - b)$

dual: $\quad \text{maximize} \quad -b^\top z - \delta_C^*(z) - f^*(-A^\top z)$

Examples

	constraint	set C	support function $\delta_C^*(z)$
equality	$Ax = b$	$\{0\}$	0
norm inequality	$\\|Ax - b\\| \leq 1$	unit $\\| \cdot \\|$-ball	$\\|z\\|_*$
conic inequality	$Ax \preceq_K b$	$-K$	$\delta_{K^*}(z)$

Norm regularization

$$\text{minimize} \quad f(x) + \|Ax - b\|$$

- take $g(y) = \|y - b\|$ in general problem

$$\text{minimize} \quad f(x) + g(Ax)$$

- conjugate of $\|\cdot\|$ is indicator of unit ball for dual norm

$$g^*(z) = b^\top z + \delta_B(z) \quad \text{where } B = \{z \mid \|z\|_* \leq 1\}$$

- hence, dual problem can be written as

$$\begin{array}{ll} \text{maximize} & -b^\top z - f^*(-A^\top z) \\ \text{subject to} & \|z\|_* \leq 1 \end{array}$$

Optimality conditions

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax = y \end{array}$$

assume f, g are convex and Slater's condition holds

Optimality conditions: x is optimal if and only if there exists a z such that

1. primal feasibility: $x \in \operatorname{dom} f$ and $y = Ax \in \operatorname{dom} g$
2. x and $y = Ax$ are minimizers of the Lagrangian $f(x) + z^\top Ax + g(y) - z^\top y$:

$$-A^\top z \in \partial f(x), \qquad z \in \partial g(Ax)$$

if g is closed, this can be written symmetrically as

$$-A^\top z \in \partial f(x), \qquad Ax \in \partial g^*(z)$$

Any questions?