Fenchel-Rockafellar Duality

Wednesday, March 3, 2021 10:19 AM

Refs: Bauschke + Combettes, 2nd ed '17

(P) $p^* = \min_x f(x) + g(A\cdot x)$ ← linear operator; not our dual fcn

$f, g \in \Gamma_0$, allow $+\infty$ values (encode constraints)

A a $m \times n$ matrix

(D) $d^* = \min_v f^*(A^* v) + g^*(-v)$

$\Gamma_0(\mathbb{R}^n) = \{$ all lsc, proper, cvx functions on $\mathbb{R}^n \}$

$f(x) \in (-\infty, +\infty]$

$\mathrm{dom}(f) = \{x : f(x) < \infty\}$

Connections to Lagrangian Duality

take (P), recast as

$$\left.\begin{array}{l} \min_{x,z} f(x) + g(z) \\ \text{s.t. } z = A\cdot x \end{array}\right\} \mathcal{L}(\underbrace{x, z}_{\text{primal}}, \underbrace{v}_{\text{dual}}) = f(x) + g(z) + \langle z - Ax, v\rangle$$

$$[A, -I]\cdot\begin{bmatrix} x \\ z\end{bmatrix} = 0$$

$$\text{dualFcn}(v) = \inf_{x,z} \mathcal{L}(x,z,v) = \inf_x f(x) - \langle Ax, v\rangle + \inf_z g(z) + \langle z, v\rangle$$

$$\inf_x f(x) - \langle Ax, v\rangle = -\sup_x \langle x, A^* v\rangle - f(x) = -f^*(A^* v)$$

$$\inf_z g(z) + \langle z, v\rangle = -\sup_z -\langle z, v\rangle - g(z) = -g^*(-v)$$

(Recall: $f^*(y) = \sup_x \langle x, y\rangle - f(x)$)

Trick $\min f(x) = -\max -f(x)$

So... $\text{dualFcn}(v) = -f^*(A^* v) - g^*(-v)$

$(\tilde{D})$ is $\tilde{d}^* = \max_v -f^*(A^* v) - g^*(-v)$ via Lagrangian Duality

$d^* = \tilde{d}^*$ via

Saddle-point interpretation: if $g \in \Gamma_0$ then $g = g^{**}$, $g(x) = \sup_y \langle x, y\rangle - g^*(y)$

(P) $\min_x f(x) + g(Ax)$ $= \sup_v \langle Ax, v\rangle - g^*(v)$

(P) $\min_x \sup_v f(x) + \langle Ax, v\rangle - g^*(v)$

if $\exists$ a saddle-pt., then strong min-max principle and we can interchange

$$\Rightarrow \sup_v \underbrace{\min_x f(x) + \langle x, A^* v\rangle}_{-f^*(-A^* v)} - g^*(v)$$

$$\sup_v -f^*(-A^* v) - g^*(v)$$

$$\sup_{v} \; -f^*(A^*v) - g^*(-v) \qquad (D)$$

Why?

Differentiability Facts

Prop 18.9 $f \in \Gamma_0(\mathcal{H})$, f^* is strictly convex $\Rightarrow$ f is (Gateaux) differentiable on $\operatorname{int}(\operatorname{dom}(f))$

Prop 18.15 If f is continuous and convex, then

$$\left(\begin{array}{l} f \text{ (Frechet) differentiable} \\ \text{and } \nabla f \text{ is } L\text{-Lipschitz} \end{array} \right) \iff \left(\begin{array}{c} f^* \text{ is } L^{-1} \text{ strongly} \\ \text{convex} \end{array} \right)$$

and $f \in \Gamma_0$, $f = f^{**}$

Algorithms:

use:
1) gradients (or Hessians ...)
2) projections
2') proximity operators
3) log-barriers

1) gradients: if I know ∇g, can I find $\nabla(g \circ A)$?

$$= A^*(\nabla g \circ A)$$

(chain rule) ✓

$$\nabla(f+g) = \nabla f + \nabla g$$

2) projections / proximity operators

$$C = \{ x : \|x\|_2 \le 1 \}$$

$$C \circ A = \{ x : \|Ax\|_2 \le 1 \}$$

In general, if I know prox_g, I don't know $\operatorname{prox}_{g \circ A}$.

There is no chain rule!

$$\operatorname{prox}_g(y) = \operatorname*{argmin}_x \tfrac{1}{2}\|x-y\|^2 + g(x)$$

$$\operatorname{proj}_{C \cap D} = \operatorname{proj}_C \;?\; \operatorname{proj}_D$$

No! no nice rule

$$f = I_C, \quad g = I_D, \quad f+g = I_{C \cap D}$$

Thm Thm 15.23 B+C "Generalized Slater"

If $0 \in \operatorname{relint}(\operatorname{dom} g - A(\operatorname{dom} f))$ — Constraint Qualification CQ (like Slater)

then Strong duality holds

$$\inf_x f(x) + g(Ax) = -\min_v f^*(A^*v) + g^*(-v)$$

and the dual sol'n is obtained.

CQ: in finite dimensions, show $\operatorname{relint}(\operatorname{dom}(g)) \cap A(\operatorname{relint}(\operatorname{dom} f)) \neq \phi$ (Prop. 15.24 (x))

or, if f, g are polyhedral, $\operatorname{dom} g \cap A(\operatorname{dom} f) \neq \phi$ (Fact 15.25 (ii))

i.e., "strictly feasible point"

Recovering a primal solution from a dual solution

Thm 19.1 Bauschke + Combettes 2nd ed Let $f \in \Gamma_0(\mathbb{R}^n)$, $g \in \Gamma_0(\mathbb{R}^m)$, $\operatorname{dom}(g) \cap A(\operatorname{dom} f) \neq \phi$

The following are equivalent:
1. there is no duality gap and x, v are primal-dual optimal (i.e., $\exists$ saddle pts to $\mathcal{L}(x,v) = f(x) + \langle Ax, v\rangle - g^*(v)$)
2. $L^*v \in \partial f(x)$ and $-v \in \partial g(Ax)$
3. $x \in \partial f^*(A^*v)$ and $Ax \in \partial g^*(-v)$

(2) $\iff$ (3) since $\partial f^* = \partial f^{-1}$ when $f \in \Gamma_0$

In particular (Fact 19.4 B+C) under above conditions, if f^* is differentiable at (A^*v) [i.e., if f is strictly convex]

then either (A) there is no primal optimal solution

or (B) $x = \nabla f^*(A^*v)$ is primal optimal

Simple example of using duality

$$(P) \quad \min \tfrac{1}{2}\|x - x_0\|^2 \qquad f(x) = \tfrac{1}{2}\|x - x_0\|^2$$

$$\underbrace{\text{s.t. } \|Ax - b\| \leq \varepsilon}_{\text{how to solve?}} \qquad g(y) = I_{\|y-b\| \leq \varepsilon}, \qquad A = A$$

projected-gradient method isn't easy

(projection is as hard as original problem... need SVD(A))

$$f^*(v) = \sup_x \langle v, x\rangle - \tfrac{1}{2}\|x - x_0\|^2 = \langle v, x_0\rangle + \sup_{\tilde{x}} \langle v, \tilde{x}\rangle - \tfrac{1}{2}\|\tilde{x}\|^2$$

$\tilde{x} = x - x_0$

$$= \langle v, x_0\rangle + \tfrac{1}{2}\|v\|^2$$

$$g^*(v) = \sup_{\substack{y \\ \|y-b\| \leq \varepsilon}} \langle v, y\rangle = \langle v, b\rangle + \sup_{\substack{\tilde{y} \\ \|\tilde{y}\| \leq \varepsilon}} \langle v, \tilde{y}\rangle = \langle v, b\rangle + \varepsilon \cdot \|v\|_2$$

$\tilde{y} = y - b$

so

$$(D) \quad \min_{v} \underbrace{\left(\underbrace{\langle A^* v, x_0 \rangle + \tfrac{1}{2} \| A^* v \|_2^2}_{f^*(A^* v)} \right)}_{\text{differentiable}} + \underbrace{\left(\underbrace{\langle v, -b \rangle + \varepsilon \| v \|_2}_{g^*(-v)} \right)}_{\text{easy to find proximity operator}}$$