Gauge and Perspective Duality

CHEN Li

IORA, NUS

Seminar
August 28th, 2020

Gauge optimization framework

$$\begin{aligned} \min_{x} \quad & \kappa(x) \\ \text{s.t.} \quad & \rho(b - Ax) \leq \sigma \end{aligned} \tag{Gp}$$

- κ and ρ are closed gauge functions on $\mathbb{R}^n$ and $\mathbb{R}^m$.

Definition 1 (Gauge)

A convex function $\kappa : \mathcal{X} \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ a gauge if it is non-negative, positive honogeneous and vanishes at the origin.

- $\sigma < \rho(b)$, otherwise $x = 0$ is optimal.
- Examples of gauges:
  - norm $\|\cdot\|$
  - indicator function of convex cone $\delta_{\mathcal{K}}$

Preliminaries: polar

- The polar of a gauge κ defined as

$$\kappa^\circ(y) = \inf\{\mu > 0 \mid \langle x, y\rangle \le \mu\kappa(x), \forall x\}$$

  is a closed gauge. It is the generalization of dual norm.
- $\kappa^{\circ\circ} = \operatorname{cl}\kappa = \kappa^{**}$.
- For a convex set $\mathcal{C}$, its polar is

$$\mathcal{C}^\circ = \{y \mid \langle x, y\rangle \le 1, \forall x \in \mathcal{C}\}$$

  and its antipolar is

$$\mathcal{C}' = \{y \mid \langle x, y\rangle \ge 1, \forall x \in \mathcal{C}\}$$

- $\operatorname{epi}\kappa^\circ = \{(y, -\lambda) | (y, \lambda) \in (\operatorname{epi}\kappa)^\circ\}$
- $\kappa^\circ(y) = \delta^*_{\mathcal{U}_\kappa}(y)$ where $\mathcal{U}_\kappa = \{x \mid \kappa(x) \le 1\}$
- $\kappa^*(y) = \delta_{\mathcal{U}_{\kappa^\circ}}(y)$
- Hölder like inequality

$$\langle x, y\rangle \le \kappa(x)\kappa^\circ(y), \forall x \in \operatorname{dom}\kappa, y \in \operatorname{dom}\kappa^\circ$$

Preliminaries

- Representation of gauge function: the gauge of a nonempty convex set $\mathcal{C}$ is

$$\gamma_{\mathcal{C}}(x) = \inf\{\lambda > 0 \mid x \in \lambda\mathcal{C}\}$$

  For a gauge κ, we have

$$\kappa = \gamma_{\mathcal{U}_\kappa}$$

- Results to deal with case of $\sigma = 0$: let $\mathcal{H}_\kappa = \{x \mid \kappa(x) = 0\}$,

$$\mathcal{U}_\kappa^\circ = \mathcal{U}_{\kappa^\circ},\ \mathcal{U}_\kappa^\infty = \mathcal{H}_\kappa,\ (\operatorname{dom}\kappa)^\circ = \mathcal{H}_\kappa,\ \mathcal{H}_\kappa^\circ = \operatorname{cl}\operatorname{dom}\kappa^\circ$$

- Let κ_1 and κ_2 be gauges. Then $\kappa(x_1, x_2) := \kappa_1(x_1) + \kappa_2(x_2)$ is a gauge and its polar is

$$\kappa^\circ(y_1, y_2) = \max\{\kappa_1^\circ(y_1), \kappa_2^\circ(y_2)\}$$

Gauge duality

Different from Lagrange dual

$$\begin{aligned} \max_{y} \quad & \langle b, y\rangle - \sigma\rho^\circ(y) \\ \text{s.t.} \quad & \kappa^\circ(A^T y) \le 1 \end{aligned} \tag{Ld}$$

Gauge dual has the form

$$\begin{aligned} \min_{y} \quad & \kappa^\circ(A^T y) \\ \text{s.t.} \quad & \langle b, y\rangle - \sigma\rho^\circ(y) \ge 1 \end{aligned} \tag{Gd}$$

Goal:

- Derive related duality theory (weak and strong duality)
- Derive related optimal condition (like KKT)

Deriving gauge dual: a quick intuition

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & \kappa(x) \\ \text{s.t.} \quad & x \in \mathcal{C} \end{aligned} \tag{1}$$

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & \kappa^\circ(y) \\ \text{s.t.} \quad & y \in \mathcal{C}' \end{aligned} \tag{2}$$

Then for any $x \in \mathcal{C}, y \in \mathcal{C}'$, we have

$$\kappa(x)\kappa^\circ(y) \geq \langle x, y \rangle \geq 1$$

as weak duality. Hopefully we can have strong duality

$$\kappa(x)\kappa^\circ(y) = 1.$$

When $\mathcal{C} = \{x \mid \rho(b - Ax) \leq \sigma\}$, we need to derive corresponding $\mathcal{C}'$. It turns out (2) becomes Gd.

Assumptions on feasibility

- Feasible region

$$\mathcal{F}_p := \{u \mid \rho(b-u) \le \sigma\}, \quad , \mathcal{F}_d = \{y \mid \langle b, y\rangle - \sigma\rho^\circ(y) \ge 1\}$$

- Convention: when $\sigma = 0$, do the replacement

$$(\rho, \sigma) \to (\delta_{\mathcal{H}_\rho}, 1), \quad \sigma\rho^\circ := \delta \mathrm{cl}\,\mathrm{dom}\,\rho^\circ \equiv \delta^\circ_{\mathcal{H}_\rho}$$

- Feasibility of Gp and Gd.

$$A^{-1}\mathcal{F}_p \cap \mathrm{dom}\,\kappa \ne \emptyset, \quad A^T\mathcal{F}_d \cap \mathrm{dom}\,\kappa^\circ \ne \emptyset$$

- Relatively strictly feasibility (strict: ri replaced by int)

$$A^{-1}(\mathrm{ri}\,\mathcal{F}_p) \cap (\mathrm{ri}\,\mathrm{dom}\,\kappa) \ne \emptyset, \quad A^T\mathrm{ri}\,(\mathcal{F}_d) \cap \mathrm{ri}\,(\mathrm{dom}\,\kappa^\circ) \ne \emptyset$$

where

$$\mathrm{ri}\,\mathcal{F}_p = \begin{array}{ll} \{u \mid b-u \in \mathrm{ri}\,\mathrm{dom}\,\rho, \rho(b-u) < \sigma\} & \text{if } \sigma > 0 \\ \{u \mid b-u \in \mathrm{ri}\,\mathcal{H}_\rho\} & \text{if } \sigma = 0 \end{array}$$

$$\mathrm{ri}\,\mathcal{F}_d = \begin{array}{ll} \{y \mid y \in \mathrm{ri}\,\mathrm{dom}\,\rho^\circ, \langle b, y\rangle - \sigma\rho^\circ(y) > 1\} & \text{if } \sigma > 0 \\ \{y \mid y \in \mathrm{ri}\,\mathcal{H}^\circ_\rho, \langle b, y\rangle > 1\} & \text{if } \sigma = 0 \end{array}$$

General perturbation framework

Theorem 1 ([1])

Consider a proper, closed, convex perturbation function F on $\mathbb{R}^n \times \mathbb{R}^m$, let $p(u) = \inf_x F(x, u)$ and $q(v) = \inf_y F^*(v, y)$. The optimal value of primal-dual pair is $p(0)$ and $p^{**}(0) = -q(0)$.

(a) $p(0) \geq -q(0)$

(b) *If $0 \in \operatorname{ri} \operatorname{dom} p$, then $p(0) = -q(0)$ and if finite, the infimal $q(0)$ is attained with $\partial p(0) = \arg\max_y -F^*(0, y)$. If $0 \in \operatorname{ri} \operatorname{dom} q$, then $p(0) = -q(0)$ and if finite, the infimal $p(0)$ is attained with $\partial q(0) = \arg\min_x F(x, 0)$.*

(c) *The set $\arg\max_y -F^*(0, y)$ is nonempty and bounded if $0 \in \operatorname{int} \operatorname{dom} p$ and $p(0)$ is finite. The set $\arg\min_x F(x, 0)$ is nonempty and bounded if $0 \in \operatorname{int} \operatorname{dom} q$ and $q(0)$ is finite.*

(d) *Optimal solutions are characterized jointly by the conditions:*

$$\left.\begin{array}{l} \bar{x} \in \arg\min_x F(x, 0) \\ \bar{y} \in \arg\max_y -F^*(0, y) \\ F(\bar{x}, 0) = -F^*(0, \bar{y}) \end{array}\right\} \Longleftrightarrow (0, \bar{y}) \in \partial F(\bar{x}, 0) \Longleftrightarrow (\bar{x}, 0) \in \partial F^*(0, \bar{y})$$

A perturbation for gauge duality

Consider the primal and dual value functions

$$
\begin{aligned}
v_p(u) \quad &= \inf_{\mu>0,x}\{\mu \mid \rho(b - Ax + \mu u) \le \sigma, \kappa(x) \le \mu\} \\
&(\lambda := 1/\mu, w := x/\mu) \\
&= \inf_{\lambda>0,w}\{1/\lambda \mid \rho(b\lambda - Aw + u) \le \sigma\lambda, \kappa(w) \le 1\} \\
v_d(t,\theta) \quad &= \inf_y\{\kappa^\circ(A^T y + t) \mid \langle b, y\rangle - \sigma\rho^\circ(y) \ge 1 + \theta\}
\end{aligned} \tag{3}
$$

Define the convex function $F(w, \lambda, u) :=$

$$
-\lambda + \delta_{\mathrm{epi}\,\rho\times\mathcal{U}_\kappa}\left(W\begin{bmatrix} w \\ \lambda \\ u \end{bmatrix}\right) \text{ where } W := \begin{bmatrix} -A & b & I_m \\ 0 & \sigma & 0 \\ I_n & 0 & 0 \end{bmatrix}
$$

- W is nonsingular
- $(0,0,0) \in \operatorname{dom} F$ and F is closed and proper
- Let $p(u) := \inf_{\lambda\ge 0,w} F(w,\lambda,u)$ and $q(t,\theta) := \inf_y F^*(t,\theta,y)$, then $v_p(u) = -1/p(u)$ and given both are finite, and $v_p(u) = 0 \iff p(u) = -\infty$, $p(u) = 0 \iff v_p(u) = +\infty$.

A perturbation for gauge duality cont.

$$
\begin{aligned}
F^*(t,\theta,y) &= \mathrm{cl}\ \inf_{z,\beta,r} \left\{ \delta^*_{\mathrm{epi}\,\rho\times\mathcal{U}_\kappa} \;\middle|\; W^T \begin{bmatrix} z \\ \beta \\ r \end{bmatrix} = \begin{bmatrix} t \\ \theta \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \\
&= \delta^*_{\mathrm{epi}\,\rho\times\mathcal{U}_\kappa} \begin{pmatrix} y \\ \sigma^{-1}(1+\theta-\langle b,y\rangle \\ t + A^T y \end{pmatrix} \\
&= \delta^*_{\mathrm{epi}\,\rho} \begin{pmatrix} y \\ \sigma^{-1}(1+\theta-\langle b,y\rangle \end{pmatrix} + \delta^*_{\mathcal{U}_\kappa}(t + A^T y) \\
&= \delta_{\mathrm{epi}\,\rho^\circ} \begin{pmatrix} y \\ -\sigma^{-1}(1+\theta-\langle b,y\rangle \end{pmatrix} + \kappa^\circ(t + A^T y)
\end{aligned}
$$

Note $F^*(0,0,y) = \left\{ \begin{array}{ll} \kappa^\circ(A^T y) & \text{if } \langle b,y\rangle - \sigma\rho^\circ(y) \geq 1 \\ +\infty & \text{otherwise} \end{array} \right\}$ and Gd is

$$
q(0,0) = \inf_y F^*(0,0,y)
$$

In summary, $-1/v_p$ and v_d plays the role of p and q in Theorem 1.

Gauge duality theorem

Theorem 2

Let $\nu_p := v(0)$ *and* $\nu_d := v_d(0,0)$. *Then for Gp and Gd, we have*

(a) $1/\nu_p \le \nu_d$ *and* $1/\nu_d \le \nu_p$. *If* $\nu_p = 0$*(resp.* $\nu_d = 0$*), then Gd(resp. Gp) is infeasible.*

(b) *(Weak duality) If* x *and* y *are primal and dual feasible, then*

$$1 \le \nu_p \nu_d \le \kappa(x)\kappa^\circ(A^T y)$$

(c) *If the dual(resp. primal) is feasible and the primal(resp. dual) is relatively strictly feasible, then* $\nu_p\nu_d = 1$ *and Gd(resp. Gp) attains its optimal value.*

Lemma 1

If the primal Gp is relatively strictly feasible, then $0 \in \operatorname{ri}\operatorname{dom} p$. *if the dual Gd is relatively strictly feasible, then* $0 \in \operatorname{ri}\operatorname{dom} q$. *If further assume strict feasibility, the* $\operatorname{ri}(\cdot)$ *can be replaced by* $\operatorname{int}(\cdot)$.

Proof of Theorem 2

- Proof of Lemma 1
  - Note $u \in \operatorname{dom} p$ iff $(u, 0, 0) \in \begin{bmatrix} A & -b \\ 0 & -\sigma \\ -I & 0 \end{bmatrix} \begin{bmatrix} w \\ \lambda \end{bmatrix} + \operatorname{epi} \rho \times \mathcal{U}_\kappa$ is solvable.
  - Then everything is on explicit calculation of relative interior.
  - Same for the dual.
- Proof of Theorem 2: use Theorem 1.
  - $p(0) \geq -q(0)$ implies (a).
  - Both $p(0)$ and $q(0)$ are finite given primal and dual feasibility, then (a) implies (b).
  - Lemma 1 implies $0 \in \operatorname{ri} \operatorname{dom} p$. By part (b) of Theorem 1, strong duality holds. Same for the dual.

Gauge multipliers and sensitivity

Theorem 3

For gauge primal-dual pair Gp and Gd, we have

(a) *If the primal is relatively strictly feasible and the dual is feasible, then the set of optimal solutions for the dual is nonempty and coincides with*

$$\partial p(0) = \partial(-1/v_p)(0)$$

If it is further assumed that the primal is strictly feasible, then the set of optimal solutions to the dual is bounded.

(b) *If the dual is relatively strictly feasible and the primal is feasible, then the set of optimal solutions for the primal is nonempty with solutions $x^* = w^*/\lambda^*$ where*

$$(w^*, \lambda^*) \in \partial v_d(0,0) \text{ and } \lambda^* > 0$$

If it is further assumed that the dual is strictly feasible, then the set of optimal solutions to the primal is bounded.

Gauge optimality conditions

Theorem 4

Suppose both problems of the gauge dual pair Gp and Gd are relatively strictly feasible, and the pair (x^*, y^*) is primal-dual feasible. Then (x^*, y^*) is primal-dual optimal if and only if it satisfies the conditions:

(a)	$\rho(b - Ax^*) = \sigma$ *or* $\rho^\circ(y^*) = 0$	*(primal activity)*
(b)	$\langle b, y^* \rangle - \sigma\rho^\circ(y^*) = 1$	*(dual activity)*
(c)	$\langle x^*, A^T y^* \rangle = \kappa(x^*)\kappa^\circ(A^T y^*)$	*(objective alignment)*
(d)	$\langle b - Ax^*, y^* \rangle = \sigma\rho^\circ(y^*)$	*(constraint alignment)*

Proof of Theorem 4

- "if": note that by

$$\kappa(x^*)\kappa^\circ(A^T y^*) \overset{(c)}{=} \langle x^*, A^T y^* \rangle \overset{(b),(d)}{=} 1$$

  strong duality holds, (x^*, y^*) is optimal.

- "only if": W.l.o.g. assume $\sigma > 0$ ($\sigma = 0$ can be done with the replacement).
  - If (x^*, y^*) is optimal, since $\kappa^\circ(A^T y)$ and $\langle b, y\rangle - \sigma\rho^\circ(y)$ in Gd are positive homogeneous, (b) must hold by optimality of y^*.
  - Note $\kappa(x^*)$ and $\kappa^\circ(A^T y^*)$ are nonzero and finite. Next we will use part (d) in Theorem 1 to prove the remaining parts.
  - Let $\lambda^* := 1/\kappa(x^*)$ and $w^* := \lambda^* x^*$, we have $\kappa(w^*) = 1$. By Theorem 1 (d), $(0, 0, y^*) \in \partial F(w^*, \lambda^*, 0)$. Since the primal is relatively strictly feasible, we have

$$\partial F(w, \lambda, 0) = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} + W^T \mathcal{N}_{\mathrm{epi}\,\rho \times \mathcal{U}_\kappa}\left(\begin{bmatrix} -Aw + b\lambda \\ \sigma\lambda \\ w \end{bmatrix}\right)$$

Proof of Theorem 4 cont.

- "only if" (cont.):

$$\mathcal{N}_{\operatorname{epi}\rho\times\mathcal{U}_\kappa}\left(\begin{bmatrix} -Aw+b\lambda \\ \sigma\lambda \\ w \end{bmatrix}\right) = \mathcal{N}_{\operatorname{epi}\rho}\left(\begin{bmatrix} -Aw+b\lambda \\ \sigma\lambda \end{bmatrix}\right)\times\mathcal{N}_{\mathcal{U}_\kappa}(w)$$

  - If $\rho(\lambda^* b - Aw^*) = \lambda^*\sigma$ ((a) holds), we have

$$\mathcal{N}_{\operatorname{epi}\rho}\left(\begin{bmatrix} -Aw^*+b\lambda^* \\ \sigma\lambda^* \end{bmatrix}\right) = \operatorname{cone}\left(\partial\rho(b\lambda^* - Aw^*)\right)\times\{-1\}$$

    and $\mathcal{N}_{\mathcal{U}_\kappa}(w^*) = \{v \mid \kappa^\circ(v) \leq \langle v, w^*\rangle\}$. Hence

$$\partial F(w^*,\lambda^*,0) = \left\{\begin{bmatrix} v-\mu A^T z \\ \mu(\langle b,z\rangle-\sigma)-1 \\ \mu z \end{bmatrix} \middle| \begin{array}{l} \mu\geq 0, \\ z\in\partial\rho(b\lambda^*-Aw^*), \\ v\in\mathcal{N}_{\mathcal{U}_\kappa}(w^*) \end{array}\right\}$$

    We know there exists $\mu^* > 0$, $z^* \in \partial\rho(b\lambda^* - Aw^*)$ and $y^* = \mu^* z^*$ such that $v^* = A^T y^*$, $\mu^*(\langle b, z^*\rangle - \sigma) = 1$ and $\kappa^\circ(v^*) \leq \langle v^*, w^*\rangle$

Proof of Theorem 4 cont.

- "only if" (cont.):
  - If $\rho(\lambda^* b - Aw^*) = \lambda^* \sigma$ (cont.):
    - Eliminate v^*, there exists

$$\begin{cases} \mu^* > 0, z^* \in \partial\rho(b\lambda^* - Aw^*), y^* = \mu^* z^* \\ (\langle b, y^* \rangle - \sigma\mu^*) = 1 \\ \kappa(w^*)\kappa^\circ(A^T y^*) \leq \langle A^T y^*, w^* \rangle \end{cases}$$

      Hence (c) holds.
    - Note

$$\begin{cases} y^* \in \mu^* \partial\rho(b\lambda^* - Aw^*) \\ \partial\rho(u) = \arg\max_y \{\langle u, y \rangle | \rho^\circ(y) \leq 1\} \end{cases}$$

      we have $\rho(b\lambda^* - Aw^*) = \langle b\lambda^* - Aw^*, y^*/\mu^* \rangle \geq 0$.
    - If $\rho(b\lambda^* - Aw^*) > 0$, then $\rho^\circ(y^*/\mu^*) = 1$, we have

$$\rho^\circ(y^*/\mu^*)\lambda^*\sigma = \rho^\circ(y^*/\mu^*)\rho(b\lambda^* - Aw^*) = \langle b\lambda^* - Aw^*, y^*/\mu^* \rangle,$$

      hence (d) holds.
    - If $\rho(b\lambda^* - Aw^*) = 0$, then

$$0 = \rho^\circ(y^*)\rho(b\lambda^* - Aw^*) \geq \langle b\lambda^* - Aw^*, y^* \rangle \geq 0.$$

      Hence $\rho^\circ(y^*)\lambda^*\sigma = \langle b\lambda^* - Aw^*, y^* \rangle$ and (d) holds.

Proof of Theorem 4 cont.

- "only if" (cont.):
  - If $\rho(\lambda^* b - Aw^*) < \lambda^* \sigma$, then

$$\mathcal{N}_{\mathrm{epi}\,\rho}\left(\begin{bmatrix} -Aw^* + b\lambda^* \\ \sigma\lambda^* \end{bmatrix}\right) = \mathcal{N}_{\mathrm{dom}\,\rho}\left(-Aw^* + b\lambda^*\right) \times \{0\}$$

Hence

$$\partial F(w^*, \lambda^*, 0) = \left\{ \begin{bmatrix} v - A^T z \\ \langle b, z \rangle - 1 \\ z \end{bmatrix} \middle| \begin{array}{l} z \in \mathcal{N}_{\mathrm{dom}\,\rho}\left(-Aw^* + b\lambda^*\right), \\ v \in \mathcal{N}_{\mathcal{U}_\kappa}\left(w^*\right) \end{array} \right\}$$

So we have

$$\begin{cases} y^* \in \mathcal{N}_{\mathrm{dom}\,\rho}\left(-Aw^* + b\lambda^*\right) & \Longrightarrow y^* \in (\mathrm{dom}\,\rho)^\circ \\ & \Longrightarrow \rho^\circ(y^*) = 0 \\ & \Longrightarrow (a) \text{ holds} \\ v^* = A^T y^* \in \mathcal{N}_{\mathcal{U}_\kappa}\left(w^*\right) & \Longrightarrow (c) \text{ holds} \\ \langle b, y^* \rangle = 1 & \end{cases}$$

and

$$\langle b - Ax^*, y^* \rangle \overset{(c),(b)}{=} 1 + \sigma\rho^\circ(y^*) - \kappa(x^*)\kappa^\circ(A^T y^*) \overset{s.d.}{=} \sigma\rho^\circ(y^*)$$
$$\Longrightarrow (d) \text{ holds}$$

Recover primal solution from the dual

Corollary 1

Suppose that the primal-dual pair Gp and Gd are each relatively strictly feasible. If y^* is optimal for Gd, then for any primal feasible x the following conditions are equivalent:

(a) x *is optimal for Gp.*

(b) $\langle x, A^T y^* \rangle = \kappa(x)\kappa^\circ(A^T y^*)$ *and* $b - Ax \in \partial(\sigma\rho^\circ)(y^*)$

(c) $A^T y^* \in \kappa^\circ(A^T y^*)\partial\kappa(x)$ *and* $b - Ax \in \partial(\sigma\rho^\circ)(y^*)$

where by convention, $\sigma\rho^\circ = \delta_{\mathrm{cl}\,\mathrm{dom}\,\rho^\circ}$ *when* $\sigma = 0$.

Example

Example 1

Consider basis pursuit denoising:

$$\min_x \{\|x\|_1 \mid \|b - Ax\|_2 \leq \sigma\} \tag{4}$$

$$\min_y \{\|A^T y\|_\infty \mid \langle b, y\rangle - \sigma\|y\|_2 \geq 1\} \tag{5}$$

After solving the dual, we have $z^* = A^T y^*$. Set $I(z) = \{i \mid |z_i| = \|z\|_\infty$. By Corollary 1 part (b),

$$\begin{aligned}\langle x, z^*\rangle &= \|x\|_1\|z^*\|_\infty \\ b - Ax &= \sigma(y^*/\|y^*\|_2)\end{aligned}$$

Hence $x_i = 0$ for all $i \notin I(z^*)$ and $\mathrm{sign}\,(x_i) = \mathrm{sign}\,(z_i^*)$ for all $i \in I(z^*)$. To recover primal solution, we can solve

$$\min_x \{\|b - Ax - \sigma(y^*/\|y^*\|_2)\|_2^2 \mid x_i = 0, \forall i \notin I(z^*)\}$$

The relationship between Lagrange and gauge multipliers

Theorem 5

Suppose that the gauge dual Gd is relatively strictly feasible and the primal Gp is feasible. Let Lp denote the Lagrange dual of Gd, and let ν_L denote its optimal value. Then

$$z^* \text{ is optimal for Lp} \iff z^*/\nu_L \text{ is optimal for Gp}$$

Proof sketch:

- Perturbation function for Lagrange duality:
  $h(w) = \inf_y \{\kappa^\circ(A^T y + w) + \delta_{\langle b,\cdot\rangle - \sigma\rho^\circ(\cdot) \geq 1}\}$
- Relate $\partial h(0)$ and $\partial v_d(0,0)$ by

$$\partial v_d(0,0) = \{(z, -h^*(z)) \mid z \in \partial h(0)\}$$

- Note $0 = \langle 0, z\rangle = h^*(z) + h(0) = h^*(z) + \nu_L$, so

$$\partial v_d(0,0) = \partial h(0) \times \{\nu_L\}$$

Extension to perspective duality

Given $f : \mathbb{R}^n \to \overline{\mathbb{R}}_+$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}_+$ closed, convex.

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & g(b - Ax) \le \sigma \end{array} \tag{Np}$$

$$\begin{array}{ll} \min\limits_{y,\alpha,\mu} & f^\sharp(A^T y, \alpha) \\ \text{s.t.} & \langle b, y\rangle - \sigma g^\sharp(y, \mu) \ge 1 - (\alpha + \mu) \end{array} \tag{Nd}$$

where $f^\sharp := (f^\pi)^\circ$ and f^π is the perspective function of f, i.e.,

$$f^\pi(x, \lambda) = \begin{cases} \lambda f(x/\lambda) & \text{if } \lambda > 0 \\ f^\infty(x) & \text{if } \lambda = 0 \\ +\infty & \text{if } \lambda < 0 \end{cases}$$

Examples

$$
\begin{aligned}
f^\sharp(z,-\xi) &= \inf\{\mu > 0 \mid \langle z, x\rangle - \xi\lambda \le \mu f^\pi(x,\lambda), \forall x, \lambda\} \\
&= \inf\{\mu > 0 \mid \langle z, x\rangle - \xi\lambda \le \mu\lambda f(x/\lambda), \forall x, \lambda > 0\} \\
&= \inf\{\mu > 0 \mid \langle z, \lambda x\rangle - \xi\lambda \le \mu\lambda f(x), \forall x, \lambda > 0\} \\
&= \inf\{\mu > 0 \mid \langle z, x\rangle \le \xi + \mu f(x), \forall x\}
\end{aligned}
$$

Example 2

- Let f be a closed gauge, then $f^\sharp(z,\xi) = f^\circ(z) + \delta_{\mathbb{R}_-}(\xi)$
- Let $f(x) = \sum_{i=1}^n f_i(x_i)$ where each f_i is convex and nonnegative, then $f^\pi(x,\lambda) = \sum_{i=1}^n f_i^\pi(x_i,\lambda)$ and $f^\sharp(z,\xi) = \max_{i=1,\dots,n} f_i^\sharp(z_i,\xi)$

A useful characterization of perspective-polar transform

Theorem 6

For any closed proper convex function f with $0 \in \operatorname{dom} f$, we have $f^{\pi *}(z, -\xi) = \delta_{\operatorname{epi} f^*}(z, \xi)$. If f is also nonnegative, $f^\sharp(z, -\xi) = \gamma_{\operatorname{epi} f^*}(z, \xi)$

Proof.

- $$\begin{aligned} f^{\pi *}(z, -\xi) &= \sup_{x,\lambda}\{\langle z, x\rangle - \lambda\xi - f^\pi(x, \lambda)\} \\ &= \sup_{x,\lambda>0}\{\langle z, x\rangle - \lambda\xi - \lambda f(x/\lambda)\} \\ &= \sup_{y,\lambda>0} \lambda\{\langle z, y\rangle - \xi - f(y)\} \\ &= \delta_{\{f^*(z)\leq\xi\}}(z, \xi) = \delta_{\operatorname{epi} f^*}(z, \xi) \end{aligned}$$
- If f is nonnegative, then $f^\pi(x, \lambda) = f^{\pi **} = \delta^*_{\operatorname{epi} f^*}(x, -\lambda)$, then $f^\sharp(z, -\xi) = f^{\pi\circ}(z, -\xi) = \delta^{*\circ}_{\operatorname{epi} f^*}(z, \xi) = \gamma_{\operatorname{epi} f^*}(z, \xi)$ since $\operatorname{epi} f^*$ contains the origin.

Derive perspective dual via lifting

Np is equivalent to the gauge problem:

$$\begin{array}{ll} \min_{x,\lambda} & f^{\pi}(x,\lambda) \\ \text{s.t.} & \rho\left(\begin{bmatrix} b \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} A & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ \lambda \end{bmatrix}\right) \le \sigma \end{array} \tag{6}$$

where $\rho(z,\mu,\tau) := g^{\pi}(z,\tau) + \delta_{\{0\}}(\mu)$.
Its gauge dual is

$$\begin{array}{ll} \min_{y,\alpha,\mu} & f^{\pi\circ}\left(\begin{bmatrix} A^T & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} y \\ \alpha \\ \mu \end{bmatrix}\right) \\ \text{s.t.} & \langle (y,\alpha,\mu),(b,1,1)\rangle - \sigma\rho^{\circ}(y,\alpha,\mu) \ge 1 \end{array} \tag{7}$$

where $\rho^{\circ}(y,\alpha,\mu) = \max\{g^{\pi\circ}(y,\mu), \delta^{\circ}_{\{0\}}(\alpha)\} = g^{\sharp}(y,\mu)$

Perspective duality

Theorem 7

Let ν_p and ν_d be the optimal value of Np and Nd, we have

(a) $1/\nu_p \leq \nu_d$ *and* $1/\nu_d \leq \nu_p$. *If* $\nu_p = 0$(*resp.* $\nu_d = 0$), *then* Nd(*resp.* Np) *is infeasible.*

(b) *(Weak duality) If x and (y, α, μ) are primal and dual feasible, then*

$$1 \leq \nu_p \nu_d \leq f(x) f^\sharp(A^T y, \alpha)$$

(c) *If the dual(resp. primal) is feasible and the primal(resp. dual) is relatively strictly feasible, then $\nu_p \nu_d = 1$ and Nd(resp. Np) attains its optimal value.*

- Proof is based on Theorem 2.
- Note Np(resp. Nd) is relatively strictly feasible iff the lifted formulation (6)(resp. (7)) is relatively strictly feasible.

Perspective optimality conditions

Theorem 8

Suppose both problems of the perspective dual pair Np and Nd are relatively strictly feasible, and the pair $(x^*, y^*, \alpha^*, \mu^*)$ is primal-dual feasible. Then $(x^*, y^*, \alpha^*, \mu^*)$ is primal-dual optimal if and only if it satisfies the conditions:

(a)	$g(b - Ax^*) = \sigma$ or $g^\sharp(y^*, \mu^*) = 0$	*(primal activity)*
(b)	$\langle b, y^* \rangle - \sigma g^\sharp(y^*, \mu^*) = 1 - (\alpha^* + \mu^*)$	*(dual activity)*
(c)	$\langle x^*, A^T y^* \rangle + \alpha^* = f(x^*) f^\sharp(A^T y^*, \alpha^*)$	*(objective alignment)*
(d)	$\langle b - Ax^*, y^* \rangle + \mu^* = \sigma g^\sharp(y^*, \mu^*)$	*(constraint alignment)*

Proof directly from Theorem 4.

Examples

Piecewise linear-quadratic constraints

$$g(y) = \sup_{u \in \mathcal{U}} \{\langle u, y \rangle - \frac{1}{2}\|Lu\|_2^2\}, \quad \mathcal{U} := \{u \in \mathbb{R}^\ell \mid Wu \leq w\} \ni 0$$

with conjugate

$$g^*(y) = \delta_{\mathcal{U}}(y) + \frac{1}{2}\|Ly\|^2$$

Hence

$$\begin{aligned} g^\sharp(y, \mu) \quad &= \gamma_{\operatorname{epi} g^*}(y, -\mu) \\ &= \inf\{\lambda > 0 \mid (y, -\mu) \in \lambda \operatorname{epi} g^*\} \\ &= \inf\{\lambda > 0 \mid y \in \lambda\mathcal{U}, \tfrac{1}{2}\|Ly\|^2 \leq -\mu\lambda\} \\ &= \delta_{\mathbb{R}_-}(\mu) + \max\{\gamma_{\mathcal{U}}(y), -\tfrac{1}{2\mu}\|Ly\|^2\} \\ &= \delta_{\mathbb{R}_-}(\mu) + \max\{\max_{i=1,\dots,k}\{W_i^T y / w_i\}, -\tfrac{1}{2\mu}\|Ly\|^2\} \end{aligned}$$

Examples cont.

If f is a closed gauge, then the perspective dual of $\min_x\{f(x) \mid g(b - Ax) \le \sigma\}$ is

$$\begin{array}{ll} \min_{y,\mu,\xi} & f^\circ(A^T y) \\ \text{s.t.} & \langle b, y\rangle + \mu - \sigma\xi = 1 \\ & Wy \le \xi w \\ & -\frac{1}{2\mu}\|Ly\|^2 \le \xi, \mu \le 0, \xi \ge 0 \left(\Leftrightarrow \left\| \begin{bmatrix} 2Ly \\ \xi + 2\mu \end{bmatrix} \right\|_2 \le \xi - 2\mu\right) \end{array}$$

- Sparse robust regression with Huber misfit: $f = \|\cdot\|_1$, $\mathcal{U} = \{u \mid \|u\|_\infty \le 1\}$ and $L = \sqrt{\eta} I$, the dual becomes

$$\begin{array}{ll} \min_{y,\mu,\xi} & f^\circ(A^T y) \\ \text{s.t.} & \langle b, y\rangle + \mu - \sigma\xi = 1 \\ & \|y\|_\infty \le \xi \\ & -\frac{\eta}{2\mu}\|Ly\|^2 \le \xi, \mu \le 0, \xi \ge 0 \end{array}$$

Conclusion

- Gauge duality under perturbation framework
- Extension to perspective duality
- Problems
  - Identify problem classes for which gauge duality is computationally advantageous
  - Primal-dual algorithm based on gauge duality
  - Comparison with Lagrange duality

Thank You! Questions?