Foundations of Gauge and Perspective Duality

James V Burke

Joint work with
Aleksandr Aravkin (UW), Michael Friedlander (UBC),
Dmitriy Drusvyatskiy (UW) and Kellie MacPhee (UW)

2018 SIAM Annual Meeting, Portland
July 11, 2018

This work partially supported by the NSF under grant number DMS-1514559.

Gauge Optimization and Duality

Suppose κ and ρ are gauges.

$$\min_x \quad \kappa(x) \qquad \text{s.t.} \quad \rho(b - Ax) \leq \tau, \tag{G_p}$$

$$\max_y \quad \langle b, y\rangle - \tau\rho^\circ(y) \qquad \text{s.t.} \quad \kappa^\circ(A^T y) \leq 1, \tag{L_d}$$

$$\min_y \quad \kappa^\circ(A^T y) \qquad \text{s.t.} \quad \langle b, y\rangle - \tau\rho^\circ(y) \geq 1. \tag{G_d}$$

When $\tau = 0$, we define $\tau\rho^\circ := \delta_{\operatorname{cl}\operatorname{dom}\rho^\circ}$.

Minkowski (gauge) functionals and polarity

Let $0 \in C \subset \mathbb{R}^n$ be nonempty, closed, and convex.
The *gauge function* for C is given by

$$\gamma_C\,(x) := \inf\,\{t \mid 0 \leq t,\; x \in tC\},$$

where the infimum over the empty set is $+\infty$.

Minkowski (gauge) functionals and polarity

Let $0 \in C \subset \mathbb{R}^n$ be nonempty, closed, and convex.
The *gauge function* for C is given by

$$\gamma_C(x) := \inf\{t \mid 0 \le t,\ x \in tC\},$$

where the infimum over the empty set is $+\infty$.

Example: $\|x\| = \gamma_{\mathbb{B}}(x)$ for any norm with unit ball $\mathbb{B}$.

Minkowski (gauge) functionals and polarity

Let $0 \in C \subset \mathbb{R}^n$ be nonempty, closed, and convex.
The *gauge function* for C is given by

$$\gamma_C(x) := \inf\{t \mid 0 \leq t,\ x \in tC\},$$

where the infimum over the empty set is $+\infty$.

Example: $\|x\| = \gamma_{\mathbb{B}}(x)$ for any norm with unit ball $\mathbb{B}$.

Gauge functions are sublinear, and so by Hörmander,

$$\gamma_C(x) = \sigma_D(x) := \sup\{\langle x, y\rangle \mid y \in D\},$$

where

$$D = \{z \mid \langle z, x\rangle \leq 1\ \forall\ x \in C\} =: C^\circ$$

and σ_D is the *support function* for the set D.

Polar Gauges

Set $\mathcal{U}_\kappa := \{x \mid \kappa(x) \le 1\}$ and define the *polar* gauge by

$$\kappa^\circ(y) = \sup\{\, \langle y, x\rangle \mid \kappa(x) \le 1 \,\} = \sigma_{\mathcal{U}_\kappa}(y)\,.$$

If κ is a norm then κ° is the corresponding dual norm.

$$\operatorname{epi}\kappa^\circ = \{(y, -\lambda) : (y, \lambda) \in (\operatorname{epi}\kappa)^\circ\}.$$

The generalized Hölder inequality

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

is known as the *polar-gauge inequality.*

In addition, for $\mathcal{H}_\kappa := \{\, u \mid \kappa(u) = 0 \,\}$, we have

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

Approaches to Duality

Additive using the conjugate: If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is closed proper and convex, the Fenchel-Young inequality is

$$\langle x, y\rangle \;\le\; f(x) \;+\; f^*(y) \qquad \forall\, x, y \in \mathbb{R}^n,$$

with

$$\langle x, y\rangle = f(x) + f^*(y) \quad \iff \quad y \in \partial f(x) \quad \iff \quad x \in \partial f^*(y).$$

Multiplicative using the polar: If $\kappa : \mathbb{R}^n \to \bar{\mathbb{R}}$ is a closed gauge, the polar-gauge inequality is

$$\langle x, y\rangle \;\le\; \kappa(x) \cdot \kappa^\circ(y) \qquad \forall\, x, y \in \mathbb{R}^n$$

with

$$\langle x, y\rangle = \kappa(x) \cdot \kappa^\circ(y) \iff y \in N\left(x \,|\, \kappa(x)\mathcal{U}_\kappa\right) \iff x \in N\left(y \,|\, \kappa^\circ(y)\mathcal{U}_{\kappa^\circ}\right)$$

with the convention that $\kappa(x)\mathcal{U}_\kappa = \mathcal{H}_\kappa$ when $\kappa(x) = 0$ (similarly for κ°).

Gauge Optimization and Duality

Suppose κ and ρ are gauges.

$$\min_{x} \quad \kappa(x) \qquad \text{s.t.} \quad \rho(b - Ax) \le \tau, \tag{G_p}$$

$$\max_{y} \quad \langle b, y\rangle - \tau\rho^\circ(y) \qquad \text{s.t.} \quad \kappa^\circ(A^T y) \le 1, \tag{L_d}$$

$$\min_{y} \quad \kappa^\circ(A^T y) \qquad \text{s.t.} \quad \langle b, y\rangle - \tau\rho^\circ(y) \ge 1. \tag{G_d}$$

When $\tau = 0$, we define $\tau\rho^\circ := \delta_{\operatorname{cl}\operatorname{dom}\rho^\circ}$.

Feasibility

Primal, Dual Domains:

$$\mathcal{F}_p := \{\, x \mid \rho(b - Ax) \le \tau \,\} \qquad \text{and} \qquad \mathcal{F}_d := \{\, y \mid \langle b, y\rangle - \tau\rho^\circ(y) \ge 1 \,\}\,.$$

$$\left.\begin{array}{rll} \textbf{Feasibilty}: & \text{Primal} & \mathcal{F}_p \cap (\operatorname{dom}\kappa) \\ & \text{Dual} & A^T\,\mathcal{F}_d \cap (\operatorname{dom}\kappa^\circ) \\[2em] \textbf{Relative Strict Feasibilty}: & \text{Primal} & \operatorname{ri}\mathcal{F}_p \cap (\operatorname{ri}\operatorname{dom}\kappa) \\ & \text{Dual} & A^T\,\operatorname{ri}\mathcal{F}_d \cap (\operatorname{ri}\operatorname{dom}\kappa^\circ) \\[2em] \textbf{Strict Feasibilty}: & \text{Primal} & \operatorname{int}(\mathcal{F})_p \cap (\operatorname{ri}\operatorname{dom}\kappa) \\ & \text{Dual} & A^T\operatorname{int}(\mathcal{F})_d \cap (\operatorname{ri}\operatorname{dom}\kappa^\circ) \end{array}\right\} \ne \emptyset$$

Freund (1987), Friedlander-Macedo-Pong (2014)

$$v_p \;=\; \min_{\rho(b-Ax)\leq\tau} \kappa(x) \qquad\qquad v_d = \min_{\langle b,y\rangle-\tau\rho^\circ(y)\geq 1} \kappa^\circ(A^T y)$$

Theorem: (2014)

1. (Weak duality)
   If x and y are P-D feasible, then

$$1 \leq v_p v_d \leq \kappa(x)\cdot\kappa^\circ(A^T y).$$

2. (Strong duality)
   If the dual is feasible and the primal is relatively strictly feasible, then $\nu_p\nu_d = 1$ and the gauge dual attains its optimal value.
   One can interchange "primal" and "dual" in the above.

Infimal Projection Duality Theory

Let $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be closed proper convex, and define the following optimal value functions by inf-projection:

$$p(y) := \inf_x \ F(x,y) \quad \text{and} \quad q(w) := \inf_z \ F^*(w,z).$$

Infimal Projection Duality Theory

Let $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be closed proper convex, and define the following optimal value functions by inf-projection:

$$p(y) := \inf_x \, F(x, y) \quad \text{and} \quad q(w) := \inf_z \, F^*(w, z).$$

This set-up yields the primal-dual pair

$$p(0) = \inf_x \, F(x, 0) \quad \text{and} \quad p^{**}(0) = \sup_z \, -F^*(0, z) \; (= -q(0)).$$

Infimal Projection Duality Theory

Let $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be closed proper convex, and define the following optimal value functions by inf-projection:

$$p(y) := \inf_x F(x,y) \quad \text{and} \quad q(w) := \inf_z F^*(w,z).$$

This set-up yields the primal-dual pair

$$p(0) = \inf_x F(x,0) \quad \text{and} \quad p^{**}(0) = \sup_z -F^*(0,z) \;(= -q(0)).$$

$$p(0) \;\geq\; p^{**}(0) \;=\; -q(0) \text{ always holds}$$

Duality Theory

$$p(y) := \inf_x \; F(x,y) \quad \text{and} \quad q(w) := \inf_z \; F^*(w,z).$$

Duality Theory

$$p(y) := \inf_x \, F(x, y) \quad \text{and} \quad q(w) := \inf_z \, F^*(w, z).$$

1. If $0 \in \text{ri}\,(\text{dom}\, p)$, then $p(0) = -q(0)$ and the infimum $q(0)$ is attained, if finite, in which case $\partial p(0) = \text{argmax}_z \, -F^*(0, z)$.

Duality Theory

$$p(y) := \inf_x \; F(x,y) \quad \text{and} \quad q(w) := \inf_z \; F^*(w,z).$$

1. If $0 \in \operatorname{ri}(\operatorname{dom} p)$, then $p(0) = -q(0)$ and the infimum $q(0)$ is attained, if finite, in which case $\partial p(0) = \operatorname{argmax}_z -F^*(0,z)$.

2. If $0 \in \operatorname{ri}(\operatorname{dom} q)$, then $p(0) = -q(0)$ and the infimum $p(0)$ is attained, if finite, in which case $\partial q(0) = \operatorname{argmin}_x F(x,0)$.

Duality Theory

$$p(y) := \inf_x\ F(x,y) \quad \text{and} \quad q(w) := \inf_z\ F^*(w,z).$$

1. If $0 \in \operatorname{ri}(\operatorname{dom} p)$, then $p(0) = -q(0)$ and the infimum $q(0)$ is attained, if finite, in which case $\partial p(0) = \operatorname{argmax}_z -F^*(0,z)$.

2. If $0 \in \operatorname{ri}(\operatorname{dom} q)$, then $p(0) = -q(0)$ and the infimum $p(0)$ is attained, if finite, in which case $\partial q(0) = \operatorname{argmin}_x F(x,0)$.

3. Optimal solutions are characterized by

$$\left.\begin{array}{l} \bar{x} \in \operatorname{argmin}_x\ F(x,0) \\ \bar{y} \in \operatorname{argmax}_z\ -F^*(0,z) \\ F(\bar{x},0) = -F^*(0,\bar{z}) \end{array}\right\} \iff (0,\bar{z}) \in \partial F(\bar{x},0) \iff (\bar{x},0) \in \partial F^*(0,\bar{z}).$$

Fenchel-Rockafellar Duality

$$F(x, y) = h(Ax + y) + g(x)$$

$$p(0) = \inf_x \{ h(Ax) + g(x) \} \quad \text{and} \quad p^{\star\star}(0) = \sup_z \{ -h^*(z) - g^*(-A^*z) \}$$

A prototype problem:

$$\mathcal{P} \qquad \begin{array}{l} \min \ \|x\|_1 \\ \text{s.t. } \|Ax - b\|_2 \le \tau \end{array}$$

$$g(x) = \|x\|_1 = \delta^*(x \mid \mathbb{B}_\infty) \qquad g^*(w) = \delta(w \mid \mathbb{B}_\infty)$$

$$h(y) = \delta(y - b \mid \tau\mathbb{B}_2) \qquad h^*(z) = -\langle z, b\rangle + \delta^*(z \mid \tau\mathbb{B}_2) = -\langle z, b\rangle + \tau\|z\|_2$$

$$\mathcal{D}_{\mathrm{L}} \qquad \begin{array}{l} \sup \ \langle b, z\rangle - \tau\|z\|_2 \\ \text{s.t. } \left\|A^T z\right\|_\infty \le 1. \end{array}$$

Gauge Duality and Sensitivity

$$v_p(y) := \inf_{\mu > 0,\, x} \{\mu \mid \rho\,(b - Ax + \mu y) \leq \tau,\ \kappa(x) \leq \mu\}$$

Gauge Duality and Sensitivity

$$v_p(y) := \inf_{\mu>0,\, x} \{\mu \mid \rho\,(b - Ax + \mu y) \le \tau,\ \kappa(x) \le \mu\}$$

$\lambda := 1/\mu$ and $w := x/\mu$

Gauge Duality and Sensitivity

$$v_p(y) := \inf_{\mu > 0,\, x} \{\mu \mid \rho\,(b - Ax + \mu y) \leq \tau,\ \kappa(x) \leq \mu\}$$

$\lambda := 1/\mu$ and $w := x/\mu$

$$= \inf_{\lambda > 0,\, w} \{1/\lambda \mid \rho(\lambda b - Aw + y) \leq \tau\lambda,\ w \in \mathcal{U}_\kappa\},$$

or

$$\inf_{\lambda > 0,\, w} \{-\lambda \mid \rho(\lambda b - Aw + y) \leq \tau\lambda,\ w \in \mathcal{U}_\kappa\}.$$

Gauge Duality and Sensitivity

$$v_p(y) := \inf_{\mu > 0,\, x} \{\mu \mid \rho\,(b - Ax + \mu y) \le \tau,\ \kappa(x) \le \mu\}$$

$$\lambda := 1/\mu \text{ and } w := x/\mu$$

$$= \inf_{\lambda > 0,\, w} \{1/\lambda \mid \rho(\lambda b - Aw + y) \le \tau\lambda,\ w \in \mathcal{U}_\kappa\},$$

or

$$\inf_{\lambda > 0,\, w} \{-\lambda \mid \rho(\lambda b - Aw + y) \le \tau\lambda,\ w \in \mathcal{U}_\kappa\}.$$

Variational framework:

$$F(w, \lambda, y) := -\lambda + \delta_{(\mathrm{epi}\,\rho) \times \mathcal{U}_\kappa}\left(W \begin{pmatrix} w \\ \lambda \\ y \end{pmatrix}\right), \quad W := \begin{pmatrix} -A & b & I \\ 0 & \tau & 0 \\ I & 0 & 0 \end{pmatrix}$$

$$F^*(w, \lambda, y) = \delta_{\mathrm{epi}\,\rho^\circ}\begin{pmatrix} y \\ -\sigma^{-1}(1 + \lambda - \langle b, y\rangle) \end{pmatrix} + \kappa^\circ(w + A^T y)$$

Gauge Duality and Sensitivity

$$p(y) := \inf_{w,\lambda} F(w, \lambda, y)$$

Theorem: The following relationships hold for the gauge primal-dual pair G_p and G_d.

(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 Duality and Optimality Conditions

$$v_p \;=\; \min_{\rho(b-Ax)\le\tau} \kappa(x) \qquad\qquad v_d = \min_{\langle b,y\rangle-\tau\rho^\circ(y)\ge 1} \kappa^\circ(A^T y)$$

Theorem: Suppose both the gauge primal and gauge dual problems 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

$$\rho(b - Ax^*) = \tau \quad \text{or} \quad \rho^\circ(y^*) = 0 \qquad \text{(primal activity)}$$

$$\langle b, y^*\rangle - \tau\rho^\circ(y^*) = 1 \qquad \text{(dual activity)}$$

$$\langle x^*, A^T y^*\rangle = \kappa(x^*)\cdot\kappa^\circ(A^T y^*) \qquad \text{(objective alignment)}$$

$$\langle b - Ax^*, y^*\rangle = \tau\rho^\circ(y^*). \qquad \text{(constraint alignment)}$$

By convention, when $\tau = 0$, $\tau\rho^\circ := \delta_{\mathrm{cl\,dom}\,\rho^\circ}$.

Gauge primal-dual recovery

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

(a) x is optimal for (G_p);

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

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

where, by convention, $\sigma\rho^\circ = \delta_{\mathrm{cl}\,\mathrm{dom}\,\rho^\circ}$ when $\sigma = 0$, in which case

$$\partial(\sigma\rho^\circ)(y^*) = N\left(y^* \,|\, \mathcal{H}_\rho^\circ\right).$$

Gauge primal-dual recovery from the Lagrange dual

Theorem:
Suppose that the gauge dual G_d is relatively strictly feasible and the primal G_p is feasible. Let L_p denote the Fenchel-Rockafellar dual of G_d, and let ν_L denote its optimal value. Then

$$z^* \text{ is optimal for } L_p \iff z^*/\nu_L \text{ is optimal for } G_p.$$

Perspective Duality

The Perspective Transform

$$f^{\pi}(x,\mu) := \begin{cases} \mu f(\mu^{-1}x), & \mu > 0 \\ f^{\infty}(x), & \mu = 0 \\ +\infty, & \mu < 0 \end{cases}$$

where

$$f^{\infty}(x) := \sup_{z \in \mathrm{dom}\,(f)} [f(x+z) - f(x)]$$

is the horizon function of f.

$$h^{\pi}(y,\mu) = \sigma_{\mathrm{epi}\,h^*}((y,-\mu))$$

The Perspective-Polar Transform

$$
\begin{aligned}
f^\sharp(x,\xi) &:= (f^\pi)^\circ(x,\xi) \\
&= \sigma_{\operatorname{epi}(f^*)^\circ}(x,-\xi) \\
&= \gamma_{\operatorname{epi}(f^*)}(x,-\xi) \\
&= \inf\{\mu > 0 \mid \xi + \langle z, x\rangle \le \mu f(z),\ \forall z\}
\end{aligned}
$$

$f^\sharp$ is a gauge.

If f is a gauge, then $f^\sharp(x,\xi) = f^\circ(x) + \delta_{\mathbb{R}_-}(\xi)$.

Perspective duality

Suppose $f : \mathbb{R}^n \to \bar{\mathbb{R}}_+$ and $g : \mathbb{R}^m \to \bar{\mathbb{R}}_+$ are closed, convex and nonnegative over their domains.

$$N_p \quad \min_{x} \quad f(x) \qquad \text{s.t.} \quad g(b - Ax) \leq \sigma,$$

$$N_d \quad \min_{y,\ \alpha,\ \mu} \quad f^{\sharp}(A^T y, \alpha) \qquad \text{s.t.} \quad \langle b, y \rangle - \sigma \cdot g^{\sharp}(y, \mu) \geq 1 - (\alpha + \mu)$$

The Perspective-Polar of a PLQ Penalty

Piecewise linear-quadratic (PLQ) penalties:

$$g(y) := \sup_{u \in U} \left\{ \langle u, y \rangle - \tfrac{1}{2}\|Lu\|_2^2 \right\}, \quad U := \left\{ u \in \mathbb{R}^l \mid Wu \leq w \right\},$$

$$g^{\sharp}(y, \mu) = \delta_{\mathbb{R}_-}(\mu) + \max \left\{ \gamma_U(y), -(1/2\mu)\|Ly\|^2 \right\}$$

$$= \delta_{\mathbb{R}_-}(\mu) + \max \left\{ -(1/2\mu)\|Ly\|^2, \max_{i=1,\ldots,k} \left\{ W_i{}^T y / w_i \right\} \right\},$$

where $W_1^T, \ldots, W_k^T$ are the rows of W.

The Perspective Duality for PLQ Penalties

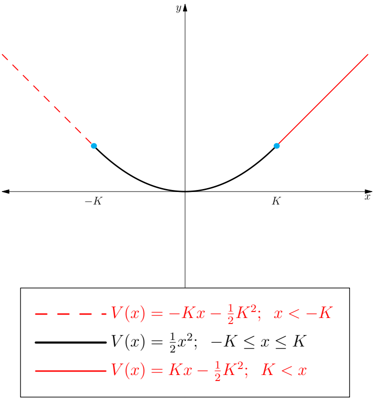
Assume f is a gauge and g is a PLQ penalty, then

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

Perspective Duality Numerics

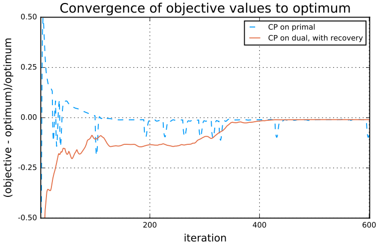
$$\min_{x} \quad \|x\|_1$$
$$\text{s.t.} \quad \sum_{i=1}^{m} V((Ax-b)_i) \le \sigma,$$

where V is the Huber function

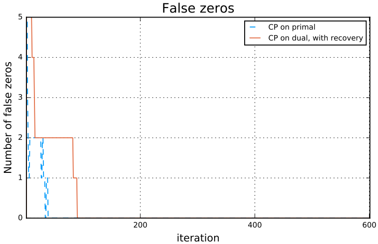


Experiment:
$m = 120$, $n = 512$, $\sigma = 0.2$, $\eta = 1$, and A is a Gaussian matrix. The true solution $x_{\text{true}} \in \{-1, 0, 1\}$ is a spike train which has been constructed to have 20 nonzero entries, and the true noise $b - Ax_{\text{true}}$ has been constructed to have 5 outliers.

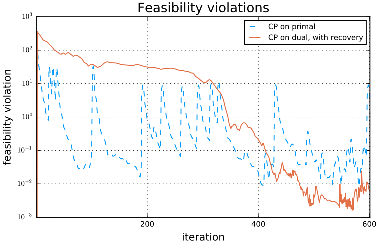
Perspective Duality Numerics



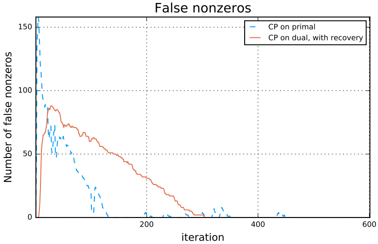
(a)



(c)



(b)



(d)

Chambolle- Pock (CP) algorithm