Deriving duality for l_p-norm optimization using conic optimization

François Glineur

Francois.Glineur@fpms.ac.be

Faculté Polytechnique de Mons

Delft University of Technology

LP 99

Copenhagen, August 28, 1999

Outline

Introduction

- $\diamond$ l_p-norm optimization

Conic optimization

- $\diamond$ Primal-dual pair
- $\diamond$ Duality theorems

The $\mathcal{L}^p$ cone

- $\diamond$ Definition
- $\diamond$ Properties

l_p-norm optimization

- $\diamond$ Formulation with $\mathcal{L}^p$ cone
- $\diamond$ Weak duality
- $\diamond$ Strong duality

Concluding remarks

Introduction

Convex optimization

Let $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ convex functions, $i = 0, \ldots, m$

$$\inf_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0 \text{ for } i = 1, \ldots, m$$

Properties

- ◇ All local optima are global
- ◇ Lagrange duality $\Rightarrow$ nice dual problem

Conic optimization

Let $K \subset \mathbb{R}^n$ be a convex cone

$$\inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax = b \text{ and } x \in K$$

Equivalent to convex optimization.

Nice form for the dual:

$$\sup_{(y,s) \in \mathbb{R}^{m+n}} b^T y \quad \text{s.t.} \quad A^T y + s = c \text{ and } s \in K^*$$

where K^* is the *dual cone*.

Example : for LO, we have $K = \mathbb{R}^n_+ = K^*$.

l_p-NORM OPTIMIZATION

Let $K = \{1, 2, \ldots, r\}$ and $I = \{1, 2, \ldots, n\}$.

Let $\{I_k\}_{k \in K}$ be a partition of I into r classes.

Data: $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times r}$, $\eta \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^r$. Variables $x \in \mathbb{R}^n$, $z \in \mathbb{R}^r$, $y \in \mathbb{R}^m$.

Primal:

$$\inf \ \psi(x, z) \quad = \quad c^T x + d^T z + \sum_{k=1}^{r} z_k \sum_{i \in I_k} \frac{1}{p_i} \left| \frac{x_i}{z_k} \right|^{p_i}$$
$$\text{s.t.} \quad Ax + Bz = \eta \text{ and } z \geq 0$$

Dual: (with $\frac{1}{p_i} + \frac{1}{q_i} = 1$)

$$\begin{array}{ll} \sup & \eta^T y \\ \text{s.t.} & \displaystyle\sum_{i \in I_k} \frac{1}{q_i} \left| c_i - a_i^T y \right|^{q_i} \leq d_k - b_k^T y \quad \forall k \in K \end{array}$$

Includes linear optimization, convex quadratic optimization, l_p-norm approximation.

Properties (Peterson and Ecker, 1967)

- ◇ Convex program ⇒ weak duality
- ◇ No duality gap and dual attainment

Conic optimization

Let $K \subset \mathbb{R}^n$ be a *solid, pointed, closed* convex cone. Its dual cone is defined by

$$K^* = \left\{ x^* \in \mathbb{R}^n \mid x^T x^* \geq 0 \text{ for all } x \in K \right\}$$

and is also convex, solid, pointed and closed.

Primal-dual pair

$$\inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax = b \text{ and } x \in K$$

$$\sup_{(y,s) \in \mathbb{R}^{m+n}} b^T y \quad \text{s.t.} \quad A^T y + s = c \text{ and } s \in K^*$$

Let p^* and d^* be their optimum objective values (not necessarily attained).

Duality theorems

Weak duality. We have $b^T y \leq c^T x$ for every feasible x and y. Equality occurs iff $x^T s = 0$ (*orthogonality* condition).

Hence we have in general $d^* \leq p^* \Rightarrow p^* - d^*$ is called the *duality gap* (always nonnegative).

Definition: x is *strictly feasible* $\Leftrightarrow$ x is feasible and $x \in \operatorname{int} K$

Slater condition

If the primal has a strictly feasible point, the duality gap is zero, i.e. $p^* = d^*$.

Moreover, if the dual is feasible, its maximum objective value d^* is attained, i.e. $d^* = \max b^T y$

(the same result holds for the dual)

Corollary

If both the primal and the dual have a strictly feasible point, the duality gap is zero and both optimal values are attained.

Our approach

l_p-norm optimization has nice duality properties, somewhere between the general convex case and linear optimization.

We are going to prove these duality properties using these general theorems $\Rightarrow$ we have to find a suitable convex cone.

The $\mathcal{L}^p$ cone

Definition

Let $p \in \mathbb{R}^n$ with $p_i > 1$. We define the function

$$\begin{aligned} f_p \quad &: \quad \mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbb{R}_+ \cup \{+\infty\} \\ &: \quad (x, \theta) \mapsto \sum_{i=1}^{n} \frac{|x_i|^{p_i}}{p_i \theta^{p_i - 1}} \end{aligned}$$

with the convention

$$\frac{|x|}{0} = \begin{cases} +\infty & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Properties

- $\operatorname{dom} f_p = \mathbb{R}^n \times \mathbb{R}_{++} \cup \ (0, 0)$
- $f_p(\lambda x, \lambda \theta) = \lambda f_p(x, \theta)$ for $\lambda \geq 0$, i.e. f_p is positively homogeneous
- $f_p(x + x', \theta + \theta') \leq f_p(x, \theta) + f_p(x', \theta')$, i.e. f_p is subadditive

$\Rightarrow f_p$ is convex

We define the cone $\mathcal{L}^p$ by

$$\begin{aligned}\mathcal{L}^p &= \left\{(x,\theta,\kappa) \in \mathbb{R}^n \times \mathbb{R}^2_+ \mid \sum_{i=1}^n \frac{|x_i|^{p_i}}{p_i \theta^{p_i-1}} \leq \kappa\right\} \\ &= \left\{(x,\theta,\kappa) \in \mathbb{R}^n \times \mathbb{R}^2_+ \mid f_p(x,\theta) \leq \kappa\right\}\end{aligned}$$

i.e. $\mathcal{L}^p$ is the *epigraph* of f_p.

This cone will enable us to express l_p-norm optimization in a conic form.

Properties

- $\mathcal{L}^p$ is a convex cone
- The interior of $\mathcal{L}^p$ is $\operatorname{int} \mathcal{L}^p = \operatorname{int} \operatorname{epi} f_p = \{(x,\theta,\kappa) \mid (x,\theta) \in \operatorname{int} \operatorname{dom} f_p \text{ and } f_p(x,\theta) < \kappa\}$
- Special case: when $p_i = 2$ for all i we have

  $$\mathcal{L}^p = \left\{(x,\theta,\kappa) \in \mathbb{R}^n \times \mathbb{R}^2_+ \mid \sum_{i=1}^n x_i^2 \leq 2\theta\kappa\right\},$$

  i.e. the *hyperbolic* or *rotated* second-order cone.

Dual cone

Define the *switched* cone $\mathcal{L}^p_s$ as $\mathcal{L}^p$ with its last two components exchanged, i.e.

$$(x, \theta, \kappa) \in \mathcal{L}^p_s \Leftrightarrow (x, \kappa, \theta) \in \mathcal{L}^p$$

Theorem. Let $p, q \in \mathbb{R}^n$ such that $p_i, q_i > 1$ and $\frac{1}{p_i} + \frac{1}{q_i} = 1$ for each i. The dual of $\mathcal{L}^p$ is $\mathcal{L}^q_s$.

The proof relies on the following

Lemma. Let $a, b \in \mathbb{R}_+$, $\alpha, \beta > 1$ and $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. We have the inequality

$$\frac{a^\alpha}{\alpha} + \frac{b^\beta}{\beta} \geq ab \ ,$$

with equality if and only if $a^\alpha = b^\beta$.

Corollaries

- $(\mathcal{L}^p_s)^* = \mathcal{L}^q$, $(\mathcal{L}^q)^* = \mathcal{L}^p_s$ and $(\mathcal{L}^q_s)^* = \mathcal{L}^p$.
- $\mathcal{L}^p$ and $\mathcal{L}^q_s$ are closed.

Orthogonality conditions

If $v = (x, \theta, \kappa) \in \mathcal{L}^p$ and $v^* = (x^*, \theta^*, \kappa^*) \in \mathcal{L}^q_s$, we have

- $v^T v^* \geq 0$
- $v^T v^* = 0$ if and only if

$$\left\{ \begin{array}{r} \kappa^* (f_p(x, \theta) - \kappa) = 0 \\ \theta (f_q(x^*, \kappa^*) - \theta^*) = 0 \\ \kappa^* \frac{|x_i|^{p_i}}{\theta^{p_i - 1}} = \theta \frac{|x_i^*|^{q_i}}{\kappa^{* q_i - 1}} \\ x_i x_i^* \leq 0 \text{ for all } i \,. \end{array} \right.$$

We are now ready to apply the general duality theory for conic primal-dual pairs, using our dual cones $\mathcal{L}^p$ and $\mathcal{L}^q_s$, to derive the duality properties of the l_p-norm optimization primal-dual pairs.

Notation: v_I (resp. M_I) $\equiv$ restriction of vector v (resp. matrix M) to indices belonging to I.

l_p-norm optimization

Formulation with $\mathcal{L}^p$ cone

Primal

$$\begin{aligned}
\inf\ \psi(x,z) \quad &= \quad c^T x + d^T z + \sum_{k=1}^{r} z_k \sum_{i\in I_k} \frac{1}{p_i}\left|\frac{x_i}{z_k}\right|^{p_i} \\
&\text{s.t.} \quad Ax + Bz = \eta \text{ and } z \geq 0 \\
&\Updownarrow \\
\inf \bar{\psi}(x,z,v) \quad &= \quad c^T x + d^T z + e^T v \\
&\text{s.t.} \quad Ax + Bz = \eta,\ z \geq 0 \text{ and} \\
&\qquad z_k \sum_{i\in I_k} \frac{1}{p_i}\left|\frac{x_i}{z_k}\right|^{p_i} \leq v_k\ \forall k \in K \\
&\Updownarrow \\
\inf \bar{\psi}(x,z,v) \quad &= \quad c^T x + d^T z + e^T v \\
&\text{s.t.} \quad Ax + Bz = \eta \text{ and} \\
&\qquad (x_{I_k}, z_k, v_k) \in \mathcal{L}^{p_{I_k}}\ \forall k \in K
\end{aligned}$$

where e is the all-one vector.

$\Rightarrow$ mechanically derive the dual, using that

$$\begin{pmatrix} A & B \end{pmatrix}^T = \begin{pmatrix} A^T \\ B^T \end{pmatrix}$$

and

$$(\mathcal{L}^p)^* = \mathcal{L}_s^q$$

Dual

$$\begin{array}{ll} \sup & \eta^T y \\ \text{s.t.} & A^T y + x^* = c,\ B^T y + z^* = d,\ v^* = e \\ & \quad \text{and } (x^*_{I_k}, z^*_k, v^*_k) \in \mathcal{L}_s^{q_{I_k}}\ \forall k \in K \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \sup & \eta^T y \\ \text{s.t.} & (c_{I_k} - A_{I_k}^T y, 1, d_k - b_k^T y) \in \mathcal{L}^{q_{I_k}}\ \forall k \in K \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \sup & \eta^T y \\ \text{s.t.} & \displaystyle\sum_{i \in I_k} \frac{1}{q_i} \left| c_i - a_i^T y \right|^{q_i} \le d_k - b_k^T y \quad \forall k \in K \end{array}$$

where a_i is a shorthand notation for $A_{\{i\}}^T$

Weak duality

If (x, z) is feasible for the primal and y is feasible for the dual, we have $\eta^T y \leq \psi(x, z)$. Equality occurs if and only if we have, for all $k \in K$ and $i \in I_k$,

$$z_k\left(\sum_{i \in I_k} \frac{1}{q_i} \left|c_i - a_i^T y\right|^{q_i} + b_k^T y - d_k\right) = 0,$$

$$x_i(c_i - a_i^T y) \leq 0, \quad \frac{\left|x_i\right|^{p_i}}{z_k^{p_i - 1}} = z_k \left|c_i - a_i^T y\right|^{q_i}$$

Proof. Use weak duality theorem on conic optimization primal-dual pair and extend objective values to the l_p-norm optimization primal-dual pair. This proof and the corresponding classical proof are not too difficult.

The classical proof of the strong duality properties is much more involved.

Strong duality

If the primal and the dual are feasible, the dual optimum objective is attained with a zero duality gap, i.e.

$$d^* = \max \eta^T y \ \text{ s.t. } \sum_{i \in I_k} \frac{1}{q_i} \left| c_i - a_i^T y \right|^{q_i} \leq d_k - b_k^T y$$

$$= \inf \psi(x,z) \ \text{ s.t. } Ax + Bz = \eta \text{ and } z \geq 0 = p^*$$

Proof.

Existence of a strictly feasible point for the primal conic program $\Leftrightarrow v_k > f_{p_{I_k}}(x_{I_k}, z_k)$ **and** $z_k > 0$ for all $k \in K$.

But the linear constraints may force $z_k = 0$ for some k for every feasible solution !

$\Rightarrow$ detect these zero z_k components and form a restricted primal-dual pair without these variables.

Detection. Use the following linear problem

$$\min\ 0 \quad \text{s.t.} \quad Ax + Bz = \eta \quad \text{and} \quad z \geq 0$$

Define Z = set of indices k such that z_k is identically zero on the feasible region and N the set of the other indices.

(Z, N) is the optimal partition of this linear problem.

Strategy. Remove variables z_k and x_{I_k} for all $k \in Z \Rightarrow$ restricted primal-dual conic pair $\Rightarrow$ the primal has a strictly feasible solution $\Rightarrow$ zero duality gap and dual attainment.

It remains to prove that

- ◇ Optimal objective values are equal for restricted primal and original primal (easy)
- ◇ Optimal values are equal for restricted dual and original dual and attainment is preserved (more difficult)

Restricted problems

Primal

$$\begin{aligned} \inf \quad & c_{I_N}^T x_{I_N} + d_N^T z_N + e^T v_N \\ \text{s.t.} \quad & A_{I_N} x_{I_N} + B_N z_N = \eta \text{ and} \\ & (x_{I_k}, z_k, v_k) \in \mathcal{L}^{p_{I_k}} \ \forall k \in N \end{aligned}$$

Convert solution to original primal (which has additional variables) : let $z_k = 0$ and $x_{I_k} = 0$ for all $k \in Z \Rightarrow$ same objective.

Dual

$$\begin{aligned} \sup \quad & \eta^T y \\ \text{s.t.} \quad & A_{I_N}^T y + x_{I_N}^* = c_{I_N},\ B_N^T y + z_N^* = d_N, \\ & v_N^* = e \ \text{ and } (x_{I_k}^*, z_k^*, v_k^*) \in \mathcal{L}^{q_{I_k}} \ \forall k \in N \end{aligned}$$

Convert solution to original dual (which has additional constraints): *not straightforward* because feasibility cannot be guaranteed ! We need to *modify* the optimal solution of the restricted dual.

Strategy

Perturb the optimal solution $\tilde{y}$ of the restricted dual with a well-chosen vector.

The *Goldman-Tucker theorem* implies the existence of a strictly complementary solution to any primal-dual pair of linear problems.

$\Rightarrow$ applied to our detection program and its dual, this guarantees the existence of a vector $\bar{y}$ such that

$$\eta^T \bar{y} = 0,\ A^T \bar{y} = 0,\ B_N^T \bar{y} \leq 0 \ \text{ and } B_Z^T \bar{y} < 0$$

These properties imply that $y = \tilde{y} + \lambda \bar{y}$ is feasible for the original dual for a sufficiently large value of λ, while keeping the same objective value as $\tilde{y}$.

$\Rightarrow$ the original dual optimum objective value is attained and equal to the original primal optimum objective value.

Concluding remarks

- ◇ These proofs are shorter than the classical ones. All the specificity of the problems is confined to the pair of dual cones
- ◇ This approach couldn't be applied with the simpler regular p-cone

$$\mathbb{L}_p^n = \left\{ (x, \kappa) \in \mathbb{R}^n \times \mathbb{R}_+ \mid \sum_{i=1}^n x_i^p \leq \kappa^p \right\}$$

- ◇ This strategy can be applied to other types of convex problems, i.e. entropy optimization and geometric optimization
- ◇ Finding a computable *self-concordant* barrier for the $\mathcal{L}^p$ cone provides an algorithm with polynomial complexity (short-step path-following interior-point method, Nesterov & Nemirovsky)
- ◇ The *approximate* self-duality might prove helpful in designing efficient algorithms (cf. *self-scaled cones*)