Deriving Duality for l_p-norm Optimization Using Conic Optimization

Image Technical Report 9904

François Glineur*
Service de Mathématique et de Recherche Opérationnelle,
Faculté Polytechnique de Mons,
Rue de Houdain, 9, B-7000 Mons, Belgium,
`Francois.Glineur@fpms.ac.be`.

Tamás Terlaky
Faculty of Technical Mathematics and Informatics,
Delft University of Technology,
NL-2600 GA Delft, The Netherlands,
`T.Terlaky@its.tudelft.nl`.

December 99

Abstract

In this paper, we formulate the l_p-norm optimization problem as a conic optimization problem and derive its standard duality properties.

We first define an *ad hoc* closed convex cone $\mathcal{L}$ and derive its dual. We express then the standard l_p-norm optimization primal problem as a conic problem involving $\mathcal{L}$. Using convex conic duality, we derive the dual of this problem and prove the well-known regularity properties of this primal-dual pair, i.e. zero duality gap and dual attainment.

Keywords. l_p-norm optimization, conic optimization, duality

1 Introduction

l_p-norm optimization problems form an important class of convex problems, which includes as special cases linear programming, quadratic programming, quadratically constrained quadratic programming and l_p-norm approximation problems.

*This author is supported by a grant from the F.N.R.S. (Belgian National Fund for Scientific Research). This research has been carried out during a visit to the Delft University of Technology supported by a grant from the Communauté Française de Belgique.

Some duality results are known for l_p-norm optimization. Namely, a pair of feasible primal-dual l_p-norm optimization problems satisfies the weak duality property, has a zero duality gap and has its dual optimum objective attained by a feasible solution.

These results were first obtained by Peterson and Ecker [PE70a, PE67, PE70b] and later simplified by Terlaky [Ter85]. The aim of this paper is to derive these results using the machinery of convex conic duality, which simplifies further the proofs.

In order to use this setting, we define an appropriate convex cone that allows us to express l_p-norm optimization problems as conic programs. In order to analyze duality for this problem, we first study some properties of this cone (closedness, interior) and determine its dual cone, which happens to have nearly the same structure.

We apply then the general duality theory for conic optimization to our programs [Stu97], to prove the above mentioned duality results about l_p-norm optimization.

This leads in our opinion to clearer proofs, the specificity of the problem being confined to the convex cone we use in our formulation. Moreover, the reason why this problem has better duality properties than a general conic problem becomes clear: this is essentially due to the existence of a strictly interior primal solution. However, such an interior solution does not always exist and a reduction procedure involving an equivalent regularized problem has to be introduced to prove the results in the general case.

2 The $\mathcal{L}$ cone

Definition 1 *Let $p \in \mathbb{R}^n$ with $p_i > 1$. We define the following set*

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

using in the case of a zero denominator the following convention:

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

This means that $(x, 0, \kappa) \in \mathcal{L}^p$ implies $x = 0$. We start by proving that $\mathcal{L}^p$ is a convex cone.

Theorem 1 *$\mathcal{L}^p$ is a convex cone.*

Proof. Let us first introduce the following function

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

With the convention mentioned above, its effective domain is $\mathbb{R}^n \times \mathbb{R}_{++} \cup (0, 0)$. It is straightforward to check that f_p is positively homogeneous, i.e. $f_p(\lambda x, \lambda\theta) = \lambda f_p(x, \theta)$ for $\lambda \geq 0$. Moreover, f_p is subadditive, i.e. $f_p(x + x', \theta + \theta') \leq f_p(x, \theta) + f_p(x', \theta')$. In order to show it, we only need to prove the following inequality:

$$\frac{|x|^{p_i}}{\theta^{p_i - 1}} + \frac{|x'|^{p_i}}{\theta'^{p_i - 1}} \geq \frac{|x + x'|^{p_i}}{(\theta + \theta')^{p_i - 1}} .$$

First observe that this inequality is obviously true if θ or θ' is equal to 0. When θ and θ' are both different from 0, we use the well known fact that x^{p_i} is a convex function on $\mathbb{R}_+$, implying that $\lambda a^{p_i} + \lambda' a'^{p_i} \geq (\lambda a + \lambda' a')^{p_i}$ for any nonnegative a, a', λ and λ' satisfying $\lambda + \lambda' = 1$. Choosing $a = \frac{1}{\theta}|x|$, $a' = \frac{1}{\theta'}|x'|$, $\lambda = \frac{\theta}{\theta+\theta'}$ and $\lambda' = \frac{\theta'}{\theta+\theta'}$, we find that

$$\begin{aligned}
\frac{\theta}{\theta+\theta'}\left(\frac{|x|}{\theta}\right)^{p_i} + \frac{\theta'}{\theta+\theta'}\left(\frac{|x'|}{\theta'}\right)^{p_i} &\geq \left(\frac{\theta}{\theta+\theta'}\frac{|x|}{\theta} + \frac{\theta'}{\theta+\theta'}\frac{|x'|}{\theta'}\right)^{p_i} \\
\frac{1}{\theta+\theta'}\left(\frac{|x|^{p_i}}{\theta^{p_i-1}} + \frac{|x'|^{p_i}}{\theta'^{p_i-1}}\right) &\geq \left(\frac{|x|+|x'|}{\theta+\theta'}\right)^{p_i} \\
\frac{|x|^{p_i}}{\theta^{p_i-1}} + \frac{|x'|^{p_i}}{\theta'^{p_i-1}} &\geq \frac{(|x|+|x'|)^{p_i}}{(\theta+\theta')^{p_i-1}} \geq \frac{|x+x'|^{p_i}}{(\theta+\theta')^{p_i-1}} .
\end{aligned}$$

Positive homogeneity and subadditivity imply that f_p is a convex function. Since $f_p(x,\theta) \geq 0$ for all x and θ, we notice that $\mathcal{L}^p$ is the epigraph of f_p, i.e.

$$\operatorname{epi} f = \left\{(x,\theta,\kappa) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R} \mid f_p(x,\theta) \leq \kappa\right\} = \mathcal{L}^p .$$

$\mathcal{L}^p$ is thus the epigraph of a convex positively homogeneous function, hence a convex cone. □

We would like to compute the interior of this cone. According to Lemma 7.3 in [Roc70] we have

$$\operatorname{int} \mathcal{L}^p = \operatorname{int} \operatorname{epi} f_p = \{(x,\theta,\kappa) \mid (x,\theta) \in \operatorname{int} \operatorname{dom} f_p \text{ and } f_p(x,\theta) < \kappa\} .$$

Using the fact that $\operatorname{int} \operatorname{dom} f_p = \mathbb{R}^n \times \mathbb{R}_{++}$, we find that

$$\operatorname{int} \mathcal{L}^p = \left\{(x,\theta,\kappa) \in \mathbb{R}^n \times \mathbb{R}_{++} \times \mathbb{R}_{++} \mid \sum_{i=1}^n \frac{|x_i|^{p_i}}{p_i \theta^{p_i-1}} < \kappa\right\} .$$

We also notice that in the special case where $p_i = 2$ for all i, our cone $\mathcal{L}^p$ becomes

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

which is the *hyperbolic* or *rotated* second-order cone. We are now going to determine the dual cone of $\mathcal{L}^p$. We first introduce a useful inequality, which lies at the heart of duality for the $\mathcal{L}$ cones.

Lemma 1 *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 $a^\alpha = b^\beta$.

Proof. This is an application of the well-known weighted arithmetic-geometric inequality on a^α and b^β with weights $\frac{1}{\alpha}$ and $\frac{1}{\beta}$ (whose sum is equal to one). This gives

$$\frac{a^\alpha}{\alpha} + \frac{b^\beta}{\beta} \geq (a^\alpha)^{1/\alpha}(b^\beta)^{1/\beta} = ab .$$

$\square$

For ease of notation, we also introduce the *switched* cone $\mathcal{L}_s$ as the $\mathcal{L}$ cone with its last two components exchanged, i.e.

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

We are now ready to describe the dual of $\mathcal{L}^p$.

Theorem 2 (Dual of $\mathcal{L}$) *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 $\mathcal{L}_s^q$.*

Proof. By definition of the dual cone, we have

$$(\mathcal{L}^p)^* = \left\{ v^* \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid v^T v^* \geq 0 \text{ for all } v \in \mathcal{L}^p \right\} \ .$$

We start by showing that $\mathcal{L}_s^q \subseteq (\mathcal{L}^p)^*$. Let $v^* = (x^*, \theta^*, \kappa^*) \in \mathcal{L}_s^q$ and $v = (x, \theta, \kappa) \in \mathcal{L}^p$. We are going to prove that $v^T v^* \geq 0$, which will imply the desired inclusion. The case when $\theta = 0$ is easily handled: we have then $x = 0$ implying $v^T v^* = \kappa\kappa^* \geq 0$. Similarly we can eliminate the case where $\kappa^* = 0$. In the remaining cases, we use the definitions of $\mathcal{L}$ and $\mathcal{L}_s$ to get

$$f_p(x, \theta) = \sum_{i=1}^n \frac{|x_i|^{p_i}}{p_i \theta^{p_i - 1}} \leq \kappa \text{ and } f_q(x^*, \kappa^*) = \sum_{i=1}^n \frac{|x_i^*|^{q_i}}{q_i {\kappa^*}^{q_i - 1}} \leq \theta^* \ .$$

Dividing respectively by θ and κ^* and adding we find

$$\sum_{i=1}^n \Big(\frac{|x_i|^{p_i}}{p_i \theta^{p_i}} + \frac{|x_i^*|^{q_i}}{q_i {\kappa^*}^{q_i}} \Big) \leq \frac{\kappa}{\theta} + \frac{\theta^*}{\kappa^*} \ . \tag{2.1}$$

Applying Lemma 1 to each pair $\dfrac{|x_i|}{\theta}, \dfrac{|x_i^*|}{\kappa^*}$ we get

$$\sum_{i=1}^n \frac{|x_i|}{\theta} \frac{|x_i^*|}{\kappa^*} \leq \frac{\kappa}{\theta} + \frac{\theta^*}{\kappa^*} \ , \tag{2.2}$$

which is equivalent to

$$\sum_{i=1}^n |x_i| \, |x_i^*| \leq \kappa\kappa^* + \theta\theta^* \ .$$

Finally, noting that $x_i x_i^* \geq -|x_i| \, |x_i^*|$ we conclude that

$$v^T v^* = x_i^T x_i^* + \kappa\kappa^* + \theta\theta^* = \sum_{i=1}^n x_i x_i^* + \kappa\kappa^* + \theta\theta^* \geq \sum_{i=1}^n -|x_i| \, |x_i^*| + \kappa\kappa^* + \theta\theta^* \geq 0 \ , \tag{2.3}$$

showing that $\mathcal{L}_s^q \subseteq (\mathcal{L}^p)^*$. Let us prove now the reverse inclusion, i.e. $(\mathcal{L}^p)^* \subseteq \mathcal{L}_s^q$. Let $v^* = (x^*, \theta^*, \kappa^*) \in (\mathcal{L}^p)^*$. We have to show that $v^* \in \mathcal{L}_s^q$, using that $v^T v^* \geq 0$ for every $v = (x, \theta, \kappa) \in \mathcal{L}^p$. Choosing $v = (0, 0, 1)$, we have $v^T v^* = \kappa^* \geq 0$. We distinguish the cases $\kappa^* = 0$ and $\kappa^* > 0$. If $\kappa^* = 0$, we have that $v^T v^* = x^T x^* + \theta\theta^* \geq 0$ for every $v = (x, \theta, \kappa) \in \mathcal{L}^p$. Choosing $\theta = 1$ and $\kappa \geq f_p(x, 1)$ for any $x \in \mathbb{R}^n$, we find that $x^T x^* + \theta^* \geq 0$ for all $x \in \mathbb{R}^n$, which implies $x^* = 0$ and $\theta^* \geq 0$ and thus $v^* \in \mathcal{L}_s^q$. When $\kappa^* > 0$, we can always choose a $v \in \mathcal{L}^p$ such that

$$\frac{|x_i|^{p_i}}{\theta^{p_i}} = \frac{|x_i^*|^{q_i}}{\kappa^{*q_i}} \quad , \quad x_i x_i^* \leq 0 \quad \text{and} \quad f_p(x,\theta) = \sum_{i=1}^n \frac{|x_i|^{p_i}}{p_i \theta^{p_i - 1}} = \kappa \; . \tag{2.4}$$

Writing

$$\begin{aligned} 0 \leq \frac{v^T v^*}{\theta \kappa^*} &= \left(\frac{x}{\theta}\right)^T \left(\frac{x^*}{\kappa^*}\right)^T + \frac{\theta^*}{\kappa^*} + \frac{\kappa}{\theta} \\ &= \sum_{i=1}^n \frac{x_i}{\theta} \frac{x_i^*}{\kappa^*} + \frac{\theta^*}{\kappa^*} + \frac{\kappa}{\theta} \\ &= \sum_{i=1}^n -\frac{|x_i|}{\theta} \frac{|x_i^*|}{\kappa^*} + \frac{\theta^*}{\kappa^*} + \frac{\kappa}{\theta} \; , \end{aligned}$$

using the case of equality of Lemma 1 on the pairs $\frac{|x_i|}{\theta}, \frac{|x_i^*|}{\kappa^*}$ and the choice of v in (2.4),

$$\begin{aligned} &= -\sum_{i=1}^n \left(\frac{|x_i|^{p_i}}{p_i \theta^{p_i}} + \frac{|x_i^*|^{q_i}}{q_i \kappa^{*q_i}} \right) + \frac{\theta^*}{\kappa^*} + \frac{\kappa}{\theta} \\ &= \frac{\theta^*}{\kappa^*} - \sum_{i=1}^n \frac{|x_i^*|^{q_i}}{q_i \kappa^{*q_i}} \; , \end{aligned}$$

and finally multiplying by κ^* leads to

$$\sum_{i=1}^n \frac{|x_i^*|^{q_i}}{q_i \kappa^{*q_i - 1}} \leq \theta^* \; ,$$

i.e. $v^* \in \mathcal{L}_s^q$, showing that $(\mathcal{L}^p)^* \subseteq \mathcal{L}_s^q$ and thus $(\mathcal{L}^p)^* = \mathcal{L}_s^q$. $\square$

The dual of a $\mathcal{L}$ cone is thus equal, up to a permutation of two variables, to another $\mathcal{L}$ cone with a dual set of exponents.

Corollary 1 *We also have* $(\mathcal{L}_s^p)^* = \mathcal{L}^q$, $(\mathcal{L}^q)^* = \mathcal{L}_s^p$ *and* $(\mathcal{L}_s^q)^* = \mathcal{L}^p$.

Proof. Obvious considering both the symmetry between $\mathcal{L}$ and $\mathcal{L}_s$ and the symmetry between p and q. $\square$

Corollary 2 $\mathcal{L}^p$ *and* $\mathcal{L}_s^q$ *are closed.*

Proof. Starting with $(\mathcal{L}^p)^* = \mathcal{L}_s^q$ and taking the dual of both sides, we find $((\mathcal{L}^p)^*)^* = (\mathcal{L}_s^q)^*$. Since $(\mathcal{L}_s^q)^* = \mathcal{L}^p$ by Corollary 1 and $((\mathcal{L}^p)^*)^* = \operatorname{cl} \mathcal{L}^p$ [Roc70, page 121], we have $\operatorname{cl} \mathcal{L}^p = \mathcal{L}^p$, hence $\mathcal{L}^p$ is closed. The switched cone $\mathcal{L}_s^q$ is obviously also closed. $\square$

We can also provide a direct proof of the closedness of $\mathcal{L}^p$: using the fact that it is the epigraph of f_p, it is enough to show that f_p is a lower semicontinuous function [Roc70, Theorem 7.1]. Being convex, f_p is continuous on the interior of its effective domain, i.e. when $\theta > 0$. When $\theta = 0$, we have to prove that

$$\lim_{(x,\theta)\to(x^0,0)} f_p(x,\theta) \geq f_p(x^0, 0) \; .$$

On the one hand, if $x_i^0 \neq 0$ for some index i, we have that $f_p(x^0, 0) = +\infty$ but also that $\lim_{(x,\theta)\to(x^0,0)} f_p(x,\theta) = +\infty$, hence the inequality is true. On the other hand, if $x^0 = 0$, we have to check that $\lim_{(x,\theta)\to(0,0)} f_p(x,\theta) \geq f_p(0,0) = 0$, which is obviously also true. From this we can conclude that f_p is lower semicontinuous and $\mathcal{L}^p$ is closed. Please note however that f_p is not continuous in $(0,0)$. Choosing an arbitrary positive constant K and taking for example $x_i = (Kp_i)^{1/p_i}\theta^{1/q_i}$, so that $x \to 0$ when $\theta \to 0$, we have that $\lim_{\theta\to 0} f(x(\theta),\theta) = nK \neq f(0,0) = 0$. The limit of f_p at $(0,0)$ can indeed take any positive value[1].

The next theorem describes the cases where two vectors from $\mathcal{L}^p$ and $\mathcal{L}_s^q$ are orthogonal, which will be used in the study of the duality properties.

Theorem 3 *Let $v = (x,\theta,\kappa) \in \mathcal{L}^p$ and $v^* = (x^*,\theta^*,\kappa^*) \in \mathcal{L}_s^q$. We have $v^T v^* = 0$ if and only if*

$$\kappa^*(f_p(x,\theta) - \kappa) = 0 \tag{2.5a}$$

$$\theta(f_q(x^*,\kappa^*) - \theta^*) = 0 \tag{2.5b}$$

$$\kappa^* \frac{|x_i|^{p_i}}{\theta^{p_i-1}} = \theta \frac{|x_i^*|^{q_i}}{\kappa^{*\,q_i-1}} \tag{2.5c}$$

$$x_i x_i^* \leq 0 \text{ for all } i\ . \tag{2.5d}$$

Proof. When $\theta > 0$ and $\kappa^* > 0$, a careful reading of the first part of the proof of Theorem 2 shows that equality occurs if and only if all conditions in (2.5) are fulfilled. Namely, (2.5a) and (2.5b) are responsible for equality in (2.1), (2.5c) ensures that we are in the case of equality of Lemma 1 for inequality (2.2) and the last condition (2.5d) is necessary for equality in (2.3).

When $\theta = 0$ but $\kappa^* > 0$, we have $x = 0$ and thus $v^T v^* = \kappa\kappa^*$. This quantity is zero if and only if $\kappa = 0$, which is equivalent in this case to $f_p(x,\theta) = \kappa$ and occurs if and only if (2.5a) is satisfied (all the other conditions being trivially fulfilled). A similar reasoning takes care of the case $\theta > 0,\ \kappa^* = 0$.

Finally, when $\theta = \kappa^* = 0$, we have $x = x^* = 0$ and $v^T v^* = 0$, while the set of conditions (2.5) is also always satisfied. $\square$

3 Duality for the l_p-norm optimization problem

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. Moreover, let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times r}$, $\eta \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^r$. We will also use the following notation: v_I (resp. M_I) denotes the restriction of column vector v (resp. matrix M) to the components (resp. rows) whose indices belong to set I. The primal l_p-norm optimization problem is the following

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

The variables are $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^r$. Introducing an extra vector $v \in \mathbb{R}^r$, we reformulate (P) as

$$\inf c^T x + d^T z + e^T v \quad \text{s.t.} \quad Ax + Bz = \eta,\ z_k \sum_{i\in I_k} \frac{1}{p_i}\left|\frac{x_i}{z_k}\right|^{p_i} \leq v_k\ \forall k \in K \text{ and } z \geq 0\ , \tag{P'}$$

[1]However, taking x proportional to θ, namely $x_i = K_i\theta$, we have $\lim_{\theta\to 0} f(x(\theta),\theta) = f(0,0) = 0$, i.e. f_p is continuous on its restrictions to lines passing through the origin.

where e stands for the all-one vector. This program is indeed equivalent to (P), since in every optimal solution to (P') v_k must be equal to its lower bound. (P') can be again reformulated using $\mathcal{L}$ cones as

$$\inf c^T x + d^T z + e^T v \quad \text{s.t.} \quad Ax + Bz = \eta, \quad (x_{I_k}, z_k, v_k) \in \mathcal{L}^{p_{I_k}} \ \forall k \in K \ . \tag{P$_\mathrm{c}$}$$

This problem being a conic optimization problem in its standard form, we can easily derive its dual and find

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

Let us simplify this dual problem. Using the equality constraint to express x^*, z^* and v^* in terms of y

$$x_i^* = c_i - a_i^T y, \quad z_k^* = d_k - b_k^T y, \quad v_k^* = 1 \ , \tag{3.1}$$

where a_i is simply shorthand notation for $A^T_{\{i\}}$, i.e. the transpose of the i^{th} column of A, we find

$$\sup \eta^T y \quad \text{s.t.} \quad (c_{I_k} - A^T_{I_k} y, 1, d_k - b_k^T y) \in \mathcal{L}^{q_{I_k}} \ \forall k \in K \tag{D'}$$

and making the conic constraint explicit we finally get the usual l_p-norm optimization dual:

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

We are now able to prove the weak duality property for the l_p-norm optimization problem.

Theorem 4 (Weak duality) *If (x, z) is feasible for* (P) *and y is feasible for* (D)*, we have $\eta^T y \leq \psi(x, z)$. Equality occurs if and only if for all $k \in K$ and $i \in I_k$*

$$z_k\Big(\sum_{i \in I_k} \frac{1}{q_i} \left| c_i - a_i^T y \right|^{q_i} + b_k^T y - d_k\Big) = 0, \quad x_i(c_i - a_i^T y) \leq 0, \quad \frac{|x_i|^{p_i}}{z_k^{p_i - 1}} = z_k \left| c_i - a_i^T y \right|^{q_i} \ . \tag{3.2}$$

Proof. Let (x, z) and y be feasible for (P) and (D). Choosing $v_k = f_{p_{I_k}}(x_{I_k}, z_k)$ for all $k \in K$, we have that (x, z, v) is feasible for (P$_\mathrm{c}$) and that $c^T x + d^T z + c^T v = \psi(x, z)$. Moreover, choosing (x^*, z^*, v^*) according to (3.1) we have that (x^*, z^*, v^*, y) is feasible for (D$_\mathrm{c}$). The standard weak duality property for the conic pair { (P$_\mathrm{c}$), (D$_\mathrm{c}$) } then states that $c^T + d^T z + c^T v \geq \eta^T y$, which in turn implies $\eta^T y \geq \psi(x, z)$.

Since it is well known that the duality gap is equal to $(x, z, v)^T (x^*, z^*, v^*)$, equality can only occur if (x, z, v) and (x^*, z^*, v^*) are orthogonal, i.e. $(x_{I_k}, z_k, v_k)^T (x^*_{I_k}, z^*_k, v^*_k) = 0$ for all $k \in K$. Having $(x_{I_k}, z_k, v_k)^T \in \mathcal{L}^{p_{I_k}}$ and $(x^*_{I_k}, z^*_k, v^*_k) \in \mathcal{L}_s^{q_{I_k}}$, Theorem 3 gives us the necessary and sufficient conditions for equality to happen

$$v_k^*(f_p(x_{I_k}, z_k) - v_k) = 0, \ z_k(f_q(x^*, 1) - z_k^*) = 0, \ v_k^* \frac{|x_i|^{p_i}}{z_k^{p_i - 1}} = z_k \frac{|x_i^*|^{q_i}}{v_k^{*\,q_i - 1}}, \ x_i x_i^* \leq 0 \tag{3.3}$$

for all $i \in I_k$ and $k \in K$. The first condition is always satisfied because of our choice of v_k, while the other three conditions can be readily simplified using (3.1) to give the announced inequalities (3.2). □

In the case of a general pair of primal and dual conic problems, the duality gap is not always equal to zero, neither are the primal or dual optimum objective values always attained by feasible solutions. However, in the special case of linear optimization, we always have a zero duality gap and attainment of both optimum objective values. The status of l_p-norm optimization lies somewhere between these two situations: the duality gap is always zero but attainment of the optimum objective value can only be guaranteed for the dual problem, as described by the following theorem:

Theorem 5 *If* (P) *and* (D) *are feasible, the dual optimum objective is attained with a zero duality gap, i.e.*

$$\begin{aligned} d^* &= \max \eta^T y \quad s.t. \quad \sum_{i \in I_k} \frac{1}{q_i} \left| c_i - a_i^T y \right|^{q_i} \leq d_k - b_k^T y \quad \forall k \in K \\ &= \inf \psi(x,z) \quad s.t. \quad Ax + Bz = \eta \text{ and } z \geq 0 \ . \end{aligned}$$

Proof. Conic duality tells us that zero duality gap and dual attainment are guaranteed by the existence of a strictly interior primal feasible solution. Let (x, z) be a feasible solution for (P). It is always possible to choose $v_k > f_{p_{I_k}}(x_{I_k}, z_k)$ for all $k \in K$ such that (x, z, k) is feasible for ($\mathrm{P_c}$). However, for (x_{I_k}, z_k, v_k) to belong to the interior of $\mathcal{L}^{p_{I_k}}$, we also need $z_k > 0$. It is unfortunately not always possible to find such a z_k, since it may happen that $z_k = 0$ for all primal feasible solutions. Let us identify such z_k's: since v_k can be chosen as large as needed, the only real constraint on z_k is $z_k \geq 0$. If z_k is zero for all primal feasible solutions, it will thus also be zero for all feasible (and optimal) solutions to the following linear problem

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

Let us call Z the set of all indices k such that z_k is 0 for all optimal solutions in (LP), and N the set of all the other indices. (Z, N) is called the optimal partition of (LP).

Let us outline the procedure we are going to follow to prove the theorem: first define a *restricted* version of ($\mathrm{P_c}$) where variables z_k and x_{I_k} have been removed for all $k \in Z$. This problem has a strictly interior solution, hence its dual problem has zero duality gap and its objective is attained by some optimum solution. The last step to prove our theorem is then convert this solution for the restricted dual problem to an optimal solution for ($\mathrm{D_c}$).

Our primal problem ($\mathrm{P_c}$) is equivalent to

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

where variables z_k and x_{I_k} have been removed for all $k \in Z$. We first show that this problem has a strictly feasible solution. For each $k \in N$, there exists at least one solution to (LP) for which $z_k > 0$. The Goldman-Tucker theorem [GT56] guarantees the existence of a solution $(\bar{x}, \bar{z})$ to (LP) such that $\bar{z}_N > 0$, i.e. $\bar{z}_k > 0$ simultaneously for every $k \in N$. Completing this vector with a sufficiently large v_N, we find a strictly feasible solution $(\bar{x}_N, \bar{z}_N, \bar{v}_N)$ to ($\mathrm{P_r}$). Let us now derive the dual of this problem:

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

Due to the existence of a strictly interior feasible solution to ($\mathrm{P_r}$), the optimum objective in ($\mathrm{D_r}$) is attained and the duality gap is zero (see e.g. [Stu97]). Since we have $p^* = \inf(\mathrm{P}) = \inf(\mathrm{P_c}) = \inf(\mathrm{P_r}) = \max(\mathrm{D_r})$, all we have to do to prove the theorem is to find a way to extend the optimal solution of $(\mathrm{D_r})$ to a solution of (D) with the same objective value. Let $\tilde{y}$ be an optimum solution of $(\mathrm{D_r})$, with $(\tilde{x}^*_{I_N}, \tilde{z}^*_N, \tilde{v}^*_N)$ given by the equality constraints (3.1). This solution is not always feasible for $(\mathrm{D_c})$, because the constraints $(\tilde{x}^*_{I_k}, \tilde{z}^*_k, \tilde{v}^*_k) \in \mathcal{L}_s^{q_{I_k}}$ are not necessarily satisfied for $k \in Z$. We are thus going to perturb $\tilde{y}$ to make it feasible for $(\mathrm{D_c})$. In order to do this, we need to study the dual of (LP), which is

$$\max\ \eta^T y \quad \text{s.t.} \quad A^T y = 0, \quad B^T y + z^* = 0 \quad \text{and} \quad z^* \geq 0\,. \tag{DLP}$$

The Goldman-Tucker theorem also implies the existence of a strictly complementary solution to the pair of problems $\{(\mathrm{LP}), (\mathrm{DLP})\}$, i.e. optimal $(\bar{x}, \bar{z})$ and $(\bar{y}, \bar{z}^*)$ such that $\bar{z} + \bar{z}^* > 0$. Since $\bar{z}_Z = 0$ by definition of Z, we have $\bar{z}^*_Z > 0$. The optimal value of (LP) being 0, the optimum objective $\eta^T \bar{y}$ of (DLP) is also 0. Summarizing the properties of $\bar{y}$ obtained so far, we have

$$\eta^T \bar{y} = 0, \quad A^T \bar{y} = 0, \quad B_N^T \bar{y} = -\bar{z}^*_N \leq 0 \text{ and } B_Z^T \bar{y} = -\bar{z}^*_Z < 0\,.$$

Let us consider $y = \tilde{y} + \lambda \bar{y}$ with $\lambda \geq 0$ as a solution of $(\mathrm{D_c})$ and compute the value of x^* and z^* given by (3.1), distinguishing the N and Z parts (we know that $v^* = e$) :

$$\begin{array}{lllll}
x^*_{I_N} & = c_{I_N} - A^T_{I_N} y & = c_{I_N} - A^T_{I_N} \tilde{y} = \tilde{x}^*_{I_N} & & \text{using } A^T_{I_N} \bar{y} = 0 \\
z^*_N & = d_N - B^T_N y & \geq d_N - B^T_N \tilde{y} = \tilde{z}^*_N & & \text{using } B^T_N \bar{y} \leq 0 \\
x^*_{I_Z} & = c_{I_Z} - A^T_{I_Z} y & = c_{I_Z} - A^T_{I_Z} \tilde{y} & & \text{using } A^T_{I_Z} \bar{y} = 0 \\
z^*_Z & = d_Z - B^T_Z y & = d_Z - B^T_Z \tilde{y} + \lambda \bar{z}^*_Z & & \text{using } -B^T_Z \bar{y} = \bar{z}^*_Z > 0\,.
\end{array}$$

Let us check the validity of the conic constraints for $k \in N$: knowing that $\tilde{y}$ is feasible for $(\mathrm{D_r})$, we have $(\tilde{x}^*_{I_k}, 1, \tilde{z}^*_k) \in \mathcal{L}_s^{q_{I_k}}$ and thus $(x^*_{I_k}, 1, z^*_k) \in \mathcal{L}_s^{q_{I_k}}$ for $k \in N$, since $z^*_N \geq \tilde{z}^*_N$. Considering now the constraints for $k \in Z$, we see that $x^*_{I_Z}$ does not depend on λ, while z^*_Z can be made arbitrarily large by increasing λ, due to the fact that $\bar{z}^*_Z > 0$. Choosing a sufficiently large λ, we can force $(x^*_{I_k}, 1, z^*_k) \in \mathcal{L}_s^{q_{I_k}}$ for $k \in Z$ and thus make (x^*, v^*, z^*, y) feasible for $(\mathrm{D_c})$.

Looking at the objective value, we find that $\eta^T y = \eta^T \tilde{y} + \lambda \eta^T \bar{y} = \eta^T \tilde{y} = \max(\mathrm{D_r})$, i.e. the feasible solution y we constructed has the same objective value for $(\mathrm{D_c})$ as $\tilde{y}$ for $(\mathrm{D_r})$. This proves that $\max(\mathrm{D_r}) \leq \sup(\mathrm{D_c})$, which combined with our previous results gives $p^* = \inf(\mathrm{P}) = \max(\mathrm{D_r}) \geq \sup(\mathrm{D}) = d^*$. Finally, using the weak duality of Theorem 4, i.e. $p^* \geq d^*$, we obtain the desired result $p^* = \inf(\mathrm{P}) = \max(\mathrm{D}) = d^*$. □

References

[GT56] A. J. Goldman and A. W. Tucker, *Theory of linear programming*, Linear Equalities and Related Systems (H. W. Kuhn and A. W. Tucker, eds.), Annals of Mathematical Studies, vol. 38, Princeton University Press, Princeton, New Jersey, 1956, pp. 53–97.

[PE67] E. L. Peterson and J. G. Ecker, *Geometric programming: Duality in quadratic programming and l_p approximation II*, SIAM Journal on Applied Mathematics **13** (1967), 317–340.

[PE70a] E. L. Peterson and J. G. Ecker, *Geometric programming: Duality in quadratic programming and l_p approximation I*, Proceedings of the International Symposium of Mathematical Programming (Princeton, New Jersey) (H. W. Kuhn and A. W. Tucker, eds.), Princeton University Press, 1970.

[PE70b] E. L. Peterson and J. G. Ecker, *Geometric programming: Duality in quadratic programming and l_p approximation III*, Journal on Mathematical Analysis and Applications **29** (1970), 365–383.

[Roc70] R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, N. J., 1970.

[Stu97] J. F. Sturm, *Primal-dual interior-point approach to semidefinite programming*, Ph.D. thesis, Erasmus Universiteit Rotterdam, The Netherlands, 1997.

[Ter85] T. Terlaky, *On l_p programming*, European Journal of Operations Research **22** (1985), 70–100.