Conic optimization: an elegant framework for convex optimization

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.

Abstract

The purpose of this survey article is to introduce the reader to a very elegant formulation of convex optimization problems called *conic optimization* and outline its many advantages.

After a brief introduction to convex optimization, the notion of convex cone is introduced, which leads to the conic formulation of convex optimization problems. This formulation features a very symmetric dual problem, and several useful duality theorems pertaining to this conic primal-dual pair are presented.

The usefulness of this approach is then demonstrated with its application to a well-known class of convex problems called l_p-norm optimization. A suitably defined convex cone leads to a conic formulation for this problem, which allows us to derive its dual and the associated weak and strong duality properties in a seamless manner.

Keywords. convex optimization, conic optimization, duality, l_p-norm optimization, separable optimization.

*The author is a postdoctoral researcher supported by the F.N.R.S. (Belgian National Fund for Scientific Research).

1 Introduction

1.1 Optimization

The main goal of *operations research* is to model real-life situations where some decisions have to be taken and help to identify the best one(s). One may for example want to choose between several available alternatives, tune numerical parameters in an engineering design or schedule the use of machines in a factory.

The concept of *best decision* obviously depends on the problem considered and it is therefore far from easy to coin a universal mathematical definition for it. Most of the time, a decision is described as a set of parameters called *decision variables*, and one seeks in fact to minimize (or maximize) a single *objective function* depending on these variables. This function may for example represent the cost associated to the decision. Moreover, we are very frequently faced with a situation where some combinations of parameters are not allowed (e.g. physical dimensions cannot be negative, a system must satisfy some performance requirements, ...), which leads us to consider a set of constraints acting on the decision variables.

Optimization is the field of mathematics whose goal is to minimize or maximize an objective function depending on several decision variables under a set of constraints. The main topic of this article is a special category of optimization problems called *convex optimization*[1].

1.2 Computational complexity

When mathematicians started to consider optimization problems more than a century ago, they were mainly interested in the existence of a solution method and its finiteness, i.e. the were seeking algorithms guaranteed to terminate after a finite number of steps. However, the introduction and generalization of computers in the second half of the 20th century resulted in a true revolution in this domain, which led scientists to become more and more interested in the practical application of these methods to actual problems. Indeed, it became possible to formulate an overwhelming amount of real-world problems as optimization problems.

However, the previously neglected concept of *computational complexity* surfaced at that time [GJ79] and refrained the enthusiasm of practitioners: a solution method, even if proven to be finite, is of little practical use if the time or the number of elementary operations it requires grows too fast with the size of the problem to be solved.

Indeed, many optimization problems are provably very hard to solve, which means it is often not possible to solve these problems in practice. To fix ideas, let us consider as an example the class of Lipschitz-continuous functions of ten variables with a Lipschitz constant equal to 2 and the seemingly simple problem that consists in computing an approximate minimizer of such a function f on the unit hypercube. Without further assumptions, it is possible to prove that, for some choices of f, the computation of this minimizer with 1% accuracy can require more than 10^{20} evaluations of f (see e.g. [Nes96]). This example, built on a quite simple class of objective functions, elementary constraints and a modest accuracy requirement, is sufficient to demonstrate that some optimization problems are unsolvable in practice.

[1] This class of problems is sometimes called *convex programming* in the literature. However, following other authors [RTV97, Ren99], we prefer to use the more natural word "optimization" since the term "programming" is nowadays strongly connected to computer science. The same treatment will be applied to the other classes of problems that will be mentioned in this article, such as linear optimization, semidefinite optimization, etc.

There are basically two fundamentally different ways to react to this distressing fact:

a. Ignore it, i.e. design a method that is supposed to solve all optimization problems. Because of the above-mentioned result, it will be slow (or fail) on some problems, but hopefully will be efficient on most real-world problems we are interested in. This is the approach that generally prevails in the field of *nonlinear optimization.*

b. Restrict the set of problems that the method is supposed to solve and design a method that is provably efficient on this restricted class of problems. This is for example the approach taken in the field of *linear optimization*, where one requires the objective function and the constraints to be linear.

Each of these two approaches has its advantages and drawbacks. The major advantage of the first approach is its potentially very wide applicability, but this is counterbalanced by a less efficient analysis of the behaviour of the corresponding algorithms. In more technical terms, methods arising from the first approach can most of the time only be proven to converge to an optimum solution (in some weak sense), but their computational complexity cannot be evaluated. On the other hand, one can often estimate the efficiency of methods following the second approach, i.e. bound the number of arithmetic operations they need to attain an optimum with a given accuracy. We chose to focus on this second approach in the rest of this article.

1.3 Why convex optimization ?

The next relevant question that has to be answered consists thus in determining which specific classes of optimization problems are going to be studied. It is rather clear that there is a tradeoff between generality and algorithmic efficiency: the more general your problem, the less efficient your methods. Linear optimization is in this respect an extreme case: it is a very particular (yet useful) type of problem for which very efficient algorithms are available (namely the simplex method and, more recently, interior-point methods, see e.g. [Dan63, RTV97]).

However, some problems simply cannot be formulated within the framework of linear programs, which leads us to consider a much broader class of problems called *convex optimization.* Basically, a problem belongs to this category if its objective function is convex and its constraints define a feasible convex set. This class of problems is much larger than linear optimization, but still retains the main advantage of the second approach mentioned above, namely the existence of solution methods whose computational complexity can be estimated. This choice is supported by the following properties:

a. Many problems that cannot be expressed within the framework of linear optimization are convex or can be convexified (which means that one can find an essentially equivalent convex reformulation of the problem).

b. Convex optimization possesses a very rich duality theory (see Section 3).

c. Convex problems can be solved efficiently using interior-point methods. The theory of *self-concordant* barriers developed by Nesterov and Nemirovski [NN94] provides an algorithm with polynomial complexity applicable to all classes of problems that are known to be convex.

Unfortunately, checking that a given optimization problem is convex is far from straightforward (and might even be more difficult than solving the problem itself). The convex problems we deal with are thus convex *by design*, i.e. are formulated in a way that guarantees their convexity. This is done by using specific classes of objective functions and constraints, and is called *structured convex optimization*.

We would also like to mention that although it is not possible to model all optimization problems of interest with a convex formulation, one can nevertheless do it in a surprisingly high number of situations. The reward for the added work of formulating the problem as a structured convex optimization problem is the great efficiency of the methods that can be then applied to it.

1.4 Conic optimization

The purpose of this survey article is to introduce the reader to a very elegant formulation of convex optimization problems called *conic optimization* and outline its many advantages.

After a brief introduction to convex optimization, the notion of convex cone is introduced in Section 2, which leads to the conic formulation of convex optimization problems. This formulation features a very symmetric dual problem, which allows us to state in Section 3 several useful duality theorems pertaining to this conic primal-dual pair. Section 4 presents a classification of conic optimization problems with respect to feasibility, attainability and optimal duality gap and features some illustrative examples.

Section 5 demonstrates then the usefulness of this conic approach by applying it to a well-known class of convex problems called l_p-norm optimization. A suitably defined convex cone leads to a conic formulation for this problem, which allows us to derive its dual and the associated weak and strong duality properties in a seamless manner. The final Section 6 summarizes the advantageous features of this conic formulation. Material surveyed in Sections 2–4 is quite classical, see e.g. [Roc70, SW70, Stu97], while the approach taken in Section 5 is more recent and can be found with greater detail in [Gli01b, GT00].

2 Convex optimization

2.1 Traditional formulation

Let us first recall that a set is convex if and only if it contains the whole segment joining any two of its points. Moreover, a function $f : D \mapsto \mathbb{R}$ is convex if and only if its domain D and its epigraph, defined by

$$\operatorname{epi} f = \{(x,t) \mid x \in D \text{ and } f(x) \leq t\} \, ,$$

are two convex sets.

The standard convex optimization problem deals with the minimization[2] of a convex function on a convex set, and can be written as follows

$$\inf_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad x \in \mathcal{S} \, , \tag{C}$$

[2]Note that contrary to the case of linear optimization, the problem of maximizing a convex function on a convex set cannot be transformed into a convex minimization problem in the form (C), and is very often much more difficult to solve (see [Roc70, §32]).

where $\mathcal{S} \subseteq \mathbb{R}^n$ is a closed convex set and $f_0 : \mathcal{S} \mapsto \mathbb{R}$ is a convex function defined on $\mathcal{S}$. The convexity of both the objective function f_0 and the feasible region $\mathcal{S}$ plays a very important role in this problem, since it is responsible for the following two important properties [Roc70, SW70]:

- Any local optimum for (C) is also a global optimum, which implies that the objective value is equal for all local optima. Moreover, the set of all local optima is always convex (and thus connected).

- There exists a dual convex problem strongly related to (C). Namely, the pair of problems consisting of a convex optimization problem and its dual satisfies a weak duality property (the objective value of any feasible solution for one of these problems provides a bound on the optimum objective value for the dual problem) and, under certain conditions, a strong duality property (equality and attainment of the optimum objective values for the two problems). These properties are discussed with more detail in Section 3.

We note that the objective function f_0 can be assumed with any loss of generality to be linear, so that we can define it as $\hat{f}_0(x) = c^{\mathrm{T}}x$ using a vector $c \in \mathbb{R}^n$. Indeed, it is readily seen that problem (C) is equivalent to the following problem with a linear objective:

$$\inf_{(x,t)\in\mathbb{R}^{n+1}} t \quad \text{s.t.} \quad (x,t) \in \hat{\mathcal{S}} ,$$

where $\hat{\mathcal{S}} \subseteq \mathbb{R}^{n+1}$ is suitably defined as

$$\hat{\mathcal{S}} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \mathcal{S} \text{ and } f_0(x) \leq t\} .$$

Equivalence follows from the fact that the constraint defining $\hat{\mathcal{S}}$ is necessarily satisfied with equality at any optimal solution (x^*, t^*), i.e. we must have $t^* = f_0(x^*)$. Moreover, this equivalent problem is convex, since $\hat{\mathcal{S}}$ is the epigraph of the convex function f_0. We will thus work without any loss of generality in the rest of this article with the problem

$$\inf_{x\in\mathbb{R}^n} c^{\mathrm{T}}x \quad \text{s.t.} \quad x \in \mathcal{S} . \tag{CL}$$

Let us now ask ourselves how one can specify the data of problem (CL), i.e. how one can describe its objective function and feasible set. While specifying the objective function is easily done by providing vector c, describing the feasible set $\mathcal{S}$, which is responsible for the *structure* of problem (CL), can be done in several manners.

The traditional way to proceed in nonlinear optimization is to provide a list of convex constraints defining $\mathcal{S}$, i.e.

$$\mathcal{S} = \left\{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \ \forall i \in I = \{1,2,\ldots,m\}\right\} ,$$

where each of the m functions $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ is convex. This guarantees the convexity of $\mathcal{S}$, as an intersection of convex level sets, and problem (CL) becomes

$$\inf_{x\in\mathbb{R}^n} c^{\mathrm{T}}x \quad \text{s.t.} \quad f_i(x) \leq 0 \ \forall i \in I = \{1,2,\ldots,m\} , \tag{CF}$$

which is the most commonly encountered formulation of a convex optimization problem.

However, a much more elegant way to describe the feasible region consists in defining $\mathcal{S}$ as the intersection of a convex cone and an affine subspace, which leads to *conic optimization*.

2.2 Conic formulation

Convex cones are the main ingredients involved in conic optimization.

Definition 2.1. A set $\mathcal{C}$ is a *cone* if and only if it is closed under nonnegative scalar multiplication, i.e.

$$x \in \mathcal{C} \Rightarrow \lambda x \in \mathcal{C} \text{ for all } \lambda \in \mathbb{R}_+ .$$

Establishing convexity is easier for cones than for general sets, because of the following elementary theorem [Roc70, Theorem 2.6]:

Theorem 2.1. *A cone $\mathcal{C}$ is* convex *if and only if it is closed under addition, i.e.*

$$x \in \mathcal{C} \text{ and } y \in \mathcal{C} \Rightarrow x + y \in \mathcal{C} .$$

In order to avoid some technical nuisances, the convex cones we are going to consider will be required to be closed, pointed and solid, according to the following definitions. A cone is said to be *pointed* if it doesn't contain any straight line passing through the origin, which can be expressed as

Definition 2.2. A cone $\mathcal{C}$ is *pointed* if and only if $\mathcal{C} \cap -\mathcal{C} = \{0\}$, where $-\mathcal{C}$ stands for the set $\{x \mid -x \in \mathcal{C}\}$.

Furthermore, a cone is said to be *solid* if it has a nonempty interior, i.e. it is full-dimensional.

Definition 2.3. A cone $\mathcal{C}$ is *solid* if and only if $\operatorname{int} \mathcal{C} \neq \emptyset$ (where $\operatorname{int} S$ denotes the interior of set S).

For example, the positive orthant $\mathbb{R}^n_+$ is a pointed and solid convex cone. A linear subspace is a convex cone that is neither pointed, nor solid (except $\mathbb{R}^n$ itself, which is solid, and $\{0\}$, which is pointed).

We are now in position to define a conic optimization problem: let $\mathcal{C} \subseteq \mathbb{R}^n$ be a pointed, solid, closed convex cone. The (primal) conic optimization problem is

$$\inf_{x \in \mathbb{R}^n} c^{\mathrm{T}} x \quad \text{s.t.} \quad Ax = b \text{ and } x \in \mathcal{C} , \tag{CP}$$

where $x \in \mathbb{R}^n$ is the column vector we are optimizing and the problem data is given by cone $\mathcal{C}$, a $m \times n$ matrix A and two column vectors b and c belonging respectively to $\mathbb{R}^m$ and $\mathbb{R}^n$. This problem can thus be viewed as the minimization of a linear function over the intersection of a convex cone and an affine subspace. As an illustration, we observe that a linear optimization problem in the standard form can be formulated by choosing cone $\mathcal{C}$ to be the positive orthant $\mathbb{R}^n_+$, which leads to the well-known primal formulation

$$\inf_{x \in \mathbb{R}^n} c^{\mathrm{T}} x \quad \text{s.t.} \quad Ax = b \text{ and } x \geq 0 .$$

When comparing the primal conic problem (CP) with the standard convex problem (CL), one observes that the only difference resides in the special choice of the feasible set $\mathcal{S} = \mathcal{C} \cap \mathcal{L}$, where $\mathcal{L}$ is the affine subspace defined by $\mathcal{L} = \{x \in \mathbb{R}^n \mid Ax = b\}$. Since $\mathcal{C}$ and $\mathcal{L}$ are convex, $\mathcal{C} \cap \mathcal{L}$ is also convex and problem (CP) is clearly a convex optimization problem. However, one can show that the class of problems in the traditional form (CL) is not larger than the class of problems in conic form (CP). Indeed, one can readily check that problem (CP) is equivalent to

$$\inf_{\hat{x} \in \mathbb{R}^{n+1}} \hat{c}^{\mathrm{T}} \hat{x} \quad \text{s.t.} \quad \hat{A}\hat{x} = b \text{ and } \hat{x} \in \mathcal{H} , \tag{CL'}$$

where $\hat{x} = (x_0, x)$, $\hat{c} = (0, c)$, $\hat{A} = (1, 0^{1\times n})$, $b = 1$ and $\mathcal{H}$ is the so-called conical hull of $\mathcal{S}$ defined by

$$\mathcal{H} = \{(x_0, x) \in \mathbb{R}_+ \times \mathbb{R}^n \mid \exists y \in \mathcal{S} \text{ such that } x = x_0 y\} .$$

The set $\mathcal{H}$ is closed under nonnegative scalar multiplication and is thus a cone, and its convexity stems from the convexity of $\mathcal{S}$: if (x_0, x) and (x_0', x') both belong to $\mathcal{H}$, we have $x = x_0 y$ and $x' = x_0' y'$ with $y, y' \in \mathcal{S}$, implying

$$x + x' = (x_0 + x_0') \left(\frac{x_0}{x_0 + x_0'} y + \frac{x_0'}{x_0 + x_0'} y' \right)$$

with

$$\frac{x_0}{x_0 + x_0'} y + \frac{x_0'}{x_0 + x_0'} y' \in \mathcal{S} \text{ (since } \mathcal{S} \text{ is convex)},$$

which is enough to prove that $(x_0 + x_0', x + x') \in \mathcal{H}$ and thus that $\mathcal{H}$ is convex by virtue of Theorem 2.1. Problem (CL') is thus expressed in the standard conic form. Moreover, its linear constraints being simply equivalent to $x_0 = 1$, the definition of $\mathcal{H}$ implies then that $(1, x) \in \mathcal{H}$ if and only if $x \in \mathcal{S}$, proving thus the strict equivalence between problems (CL) and (CL'), besides the additional variable $x_0 = 1$ which doesn't play any role in (CL'). In conclusion, the conic formulation (CP) allows us to express all convex optimization problems without any loss of generality.

The special treatment of the linear constraints in the conic formulation, i.e. their representation as an intersection with an affine subspace, can be justified by the fact that these constraints are easier to handle than general nonlinear constraints, from the following points of view

a. theoretical: linear constraints cannot cause a nonzero duality gap, in the sense that strong duality is valid without a Slater-type assumption for linear optimization (this topic is developed further in Section 3),

b. practical: it is often easy for algorithms to preserve feasibility with respect to these constraints, as opposed to the case of nonlinear constraints.

But the main advantage of the conic formulation resides in the very symmetrical formulation of the dual problem, which we present now.

2.3 Dual problem

As mentioned earlier, to each (primal) convex optimization problem corresponds a strongly related dual problem that can be found using the theory of Lagrange duality. However, the expression of this dual problem in the general case of problem (CF) is far from symmetric and involves some kind of bilevel optimization of the objective function (i.e. optimizing an objective function that is itself the result of an inner optimization process).

However, the Lagrangean dual of a conic problem such as (CP) can be expressed very nicely in a conic form, using the notion of dual cone.

Definition 2.4. The *dual* of a cone $\mathcal{C} \subseteq \mathbb{R}^n$ is defined by

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

For example, the dual of $\mathbb{R}^n_+$ is $\mathbb{R}^n_+$ itself (we say it is *self-dual*). Another example is the dual of the linear subspace $\mathcal{L}$, which is $\mathcal{L}^* = \mathcal{L}^\perp$, the linear subspace orthogonal to $\mathcal{L}$ (note that in that case the inequality of Definition 2.4 is always satisfied with equality).

The following theorem stipulates that the dual of a closed convex cone is always a closed convex cone [Roc70, Theorem 14.1].

Theorem 2.2. *If $\mathcal{C}$ is a closed convex cone, its dual $\mathcal{C}^*$ is another closed convex cone. Moreover, the dual $(\mathcal{C}^*)^*$ of $\mathcal{C}^*$ is equal to $\mathcal{C}$.*

Closedness is essential for $(\mathcal{C}^*)^* = \mathcal{C}$ to hold (without the closedness assumption on $\mathcal{C}$, we only have $(\mathcal{C}^*)^* = \operatorname{cl}\mathcal{C}$ where $\operatorname{cl} S$ denotes the closure of set S [Roc70, Theorem 14.1]). The additional notions of solidness and pointedness also behave well when taking the dual of a convex cone: indeed, these two properties are dual to each other [Stu97, Corollary 2.1], which allows us to state the following theorem:

Theorem 2.3. *If $\mathcal{C}$ is a solid, pointed, closed convex cone, its dual $\mathcal{C}^*$ is another solid, pointed, closed convex cone and $(\mathcal{C}^*)^* = \mathcal{C}$.*

The (Lagrangean) dual of our primal conic problem (CP) is defined by

$$\sup_{y\in\mathbb{R}^m, s\in\mathbb{R}^n} b^{\mathrm{T}}y \quad \text{s.t.} \quad A^{\mathrm{T}}y + s = c \text{ and } s \in \mathcal{C}^* , \tag{CD}$$

where $y \in \mathbb{R}^m$ and $s \in \mathbb{R}^n$ are the column vectors we are optimizing, the other quantities A, b and c being the same as in (CP). It is immediate to notice that this dual problem has the same kind of structure as the primal problem, i.e. it also involves optimizing a linear function over the intersection of a convex cone and an affine subspace. The only differences are the direction of the optimization (maximization instead of minimization) and the way the affine subspace is described (it is here a translation of the range space of A^{T}, while the primal involved a translation of the null space of A). It is also easy to show that the dual of this dual problem is equivalent to the primal problem, using the fact that $(\mathcal{C}^*)^* = \mathcal{C}$.

As a simple illustration, the dual of a linear optimization problem in standard form involves the dual of the positive orthant $\mathbb{R}^n_+$, which is self-dual and gives thus the well-known

$$\sup_{y\in\mathbb{R}^m, s\in\mathbb{R}^n} b^{\mathrm{T}}y \quad \text{s.t.} \quad A^{\mathrm{T}}y + s = c \text{ and } s \geq 0 .$$

The reason why the conic dual (CD) is much simpler than the usual Lagrangean dual of problem (CF), i.e. involves a single-level optimization, is simple: this is due to the fact that the inner optimization that must be performed in the dual of (CF) is here implicitly carried out when the dual cone is introduced. Indeed, when considering a primal conic problem built with a new convex cone, formulating the dual problem requires to compute first the corresponding dual cone. This frequently necessitates a few analytic calculations, which have to be done only once and can then be used to express the dual of any conic optimization problem using the primal cone.

An interesting interpretation of the conic formulation (CP) is based on the fact that we may view the constraint $x \in \mathcal{C}$ as a generalization of the traditional nonnegativity constraint

$x \geq 0$ of linear optimization. Indeed, let us define the relation $\succeq$ on $\mathbb{R}^n \times \mathbb{R}^n$ according to $x \succeq y \Leftrightarrow x - y \in \mathcal{C}$. This relation is reflexive, since $x \succeq x \Leftrightarrow 0 \in \mathcal{C}$ is always true. It is also transitive, since we have

$$x \succeq y \text{ and } y \succeq z \Leftrightarrow x - y \in \mathcal{C} \text{ and } y - z \in \mathcal{C} \Rightarrow (x-y) + (y-z) = x - z \in \mathcal{C} \Leftrightarrow x \succeq z$$

(where we used the fact that a convex cone is closed under addition, see Theorem 2.1). Finally, using the fact that $\mathcal{C}$ is pointed, we can write

$$x \succeq y \text{ and } y \succeq x \Leftrightarrow x - y \in \mathcal{C} \text{ and } -(x-y) \in \mathcal{C} \Rightarrow x - y = 0 \Rightarrow x = y ,$$

which shows that relation $\succeq$ is antisymmetric and is thus a partial order on $\mathbb{R}^n \times \mathbb{R}^n$. Defining $\succeq^*$ to be the relation induced by the dual cone $\mathcal{C}^*$, we can rewrite our primal-dual pair (CP)–(CD) as

$$\begin{aligned} &\inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax = b \text{ and } x \succeq 0 \\ &\sup_{y \in \mathbb{R}^m} b^T y \quad \text{s.t.} \quad c \succeq^* A^T y , \end{aligned}$$

which looks very much like a generalization of the well-known primal-dual pair of linear optimization problems.

To conclude this section, we state another example of primal-dual pair of conic problems, using this time one of the most versatile cones one can encounter in convex optimization, the *positive semidefinite* cone $\mathbb{S}^n_+$.

Definition 2.5. The positive semidefinite cone $\mathbb{S}^n_+$ is a subset of $\mathbb{S}^n$, the set of symmetric $n \times n$ matrices. It consists of all positive semidefinite matrices, i.e.

$$M \in \mathbb{S}^n_+ \Leftrightarrow z^T M z \geq 0 \; \forall z \in \mathbb{R}^n \Leftrightarrow \lambda(M) \geq 0$$

where $\lambda(M)$ denotes the vector of eigenvalues of M.

It is straightforward to check that $\mathbb{S}^n_+$ is a closed, solid, pointed convex cone. A conic optimization problem of the form (CP) or (CD) that uses a cone of the type $\mathbb{S}^n_+$ is called a *semidefinite problem*[3]. This cone provides us with the ability to model many more types of constraints than a linear problem (see Section 4 and e.g. [VB96] for a list of applications in various domains such as engineering, control, combinatorial optimization, finance, etc.).

The examples of conic optimization problems we have given so far, namely linear and semidefinite optimization, were both self-dual. This is however not always the case, and Section 5 will present an example of nonsymmetric conic duality.

[3]The fact that our feasible points are in this case matrices instead of vectors calls for some explanation. Since our convex cones are supposed to belong to a real vector space, we have to consider that $\mathbf{S}^n$, the space of symmetric matrices, is isomorphous to $\mathbf{R}^{n(n+1)/2}$. In that setting, an expression such as the objective function $c^T x$, where c and x belong to $\mathbf{R}^{n(n+1)/2}$, is to be understood as the inner product of the corresponding symmetric matrices C and X in the space $\mathbf{S}^n$, which is defined by $\langle C, X \rangle = \operatorname{trace} CX$. Moreover, A can be seen in this case as an application (more precisely a tensor) that maps $\mathbf{S}^n$ to $\mathbf{R}^m$, while A^T is the adjoint of A which maps $\mathbf{R}^m$ to $\mathbf{S}^n$.

3 Duality theory

The two conic problems of a primal-dual pair are strongly related to each other, as demonstrated by the duality theorems stated in this section. Conic optimization enjoys the same kind of rich duality theory as linear optimization, albeit with some complications regarding the strong duality property.

The results we present in this Section are well-known and can be found for example in the Ph.D. thesis of Sturm [Stu97, Stu99] with similar notations, more classical references presenting equivalent results are [SW70] and [ET76, §III, Section 5]).

Theorem 3.1 (Weak duality). *Let x a feasible (i.e. satisfying the constraints) solution for* (CP), *and (y, s) a feasible solution for* (CD). *We have*

$$b^T y \leq c^T x \; ,$$

equality occurring if and only if the following orthogonality condition *is satisfied:*

$$x^T s = 0 \; .$$

This theorem shows that any primal (resp. dual) feasible solution provides an upper (resp. lower) bound for the dual (resp. primal) problem. Its proof is quite easy to obtain: elementary manipulations give

$$c^T x - b^T y = x^T c - (Ax)^T y = x^T (A^T y + s) - x^T A^T y = x^T s \; ,$$

this last inner product being always nonnegative because of $x \in C$, $s \in C^*$ and Definition 2.4 of the dual cone C^*. The nonnegative quantity $x^T s = c^T x - b^T y$ is called the *duality gap*.

Obviously, a pair (x, y) with a zero duality gap must be optimal. It is well known that the converse is true in the case of linear optimization, i.e. that all primal-dual pairs of optimal solutions for a linear optimization problem have a zero duality gap (see e.g. [Sch86]), but this is not in general the case for conic optimization.

Let us denote for convenience the optimum objective values of problems (CP) and (CD) by p^* and d^*. We will say that the primal (resp. dual) problem is *unbounded* if $p^* = -\infty$ (resp. $d^* = +\infty$) and that it is *infeasible* if there is no feasible solution, i.e. when $p^* = +\infty$ (resp. $d^* = -\infty$). We emphasize the fact that although our cone C is closed, it may happen that the infimum in (CP) or the supremum in (CD) is not attained, either because the problem is infeasible or unbounded, or because there exists a sequence of feasible points whose objective values tend to p^* (or d^*) but whose limit is not feasible (some examples of these situations will be given in Section 4). We will say that the primal (resp. dual) problem is *solvable* or *attained* if the optimum objective value p^* (resp. d^*) is achieved by at least one feasible primal (resp. dual) solution.

The weak duality theorem implies that $p^* - d^* \geq 0$, a nonnegative quantity which will be called the *duality gap (at optimality)*. Under certain circumstances, it can be proven to be equal to zero, which shows that the optimum values of problems (CP) and (CD) are equal. Before we describe the conditions guaranteeing such a situation, called *strong* duality, we need to introduce the notion of strictly feasible point.

Definition 3.1. A point x (resp. (y, s)) is said to be *strictly feasible* for the primal (resp. dual) problem if and only if it is feasible and belongs to the interior of the cone C (resp. C^*), i.e.

$$Ax = b \text{ and } x \in \operatorname{int} C \quad (\text{resp. } A^T y + s = c \text{ and } s \in \operatorname{int} C^*) \; .$$

Strictly feasible points, sometimes called *Slater points*, are also said to satisfy the *interior-point* or *Slater condition.*

Theorem 3.2 (Strong duality). *If the dual problem* (CD) *admits a strictly feasible solution, we have either*

- *an infeasible primal problem* (CP) *if the dual problem* (CD) *is unbounded, i.e.* $p^* = d^* = +\infty$
- *a feasible primal problem* (CP) *if the dual problem* (CD) *is bounded. Moreover, in this case, the primal optimum is finite and attained with a zero duality gap, i.e. there is at least an optimal feasible solution* x^* *such that* $c^T x^* = p^* = d^*$.

The first case in this theorem (see e.g. [Stu97, Theorem 2.7] for a proof) is a simple consequence of Theorem 3.1, which is also valid in the absence of a Slater point for the dual, as opposed to the second case which relies on the existence of such a point. It is also worth to mention that boundedness of the dual problem (CD), defining the second case, is implied by the existence of a feasible primal solution, because of the weak duality theorem (however, the converse implication is not true in general, since a bounded dual problem can admit an infeasible primal problem ; an example of this situation is provided in Section 4).

This theorem is important, because it provides us with way to identify when both the primal and the dual problems have the same optimal value, and when this optimal value is attained by one of the problems. Obviously, this result can be dualized, meaning that the existence of a strictly feasible primal solution implies a zero duality gap and dual attainment. The combination of these two theorems leads to the following well-known corollary:

Corollary 3.1. *If both the primal and the dual problems admit a strictly feasible point, we have a zero duality gap and attainment for both problems, i.e. the same finite optimum objective value is attained for both problems.*

When the dual problem has no strictly feasible point, nothing can be said about the duality gap (which can happen to be strictly positive) and about attainment of the primal optimum objective value. However, even in this situation, we can prove an alternate version of the strong duality theorem involving the notion of primal problem subvalue. The idea behind this notion is to allow a small constraint violation in the infimum defining the primal problem (CP).

Definition 3.2. The *subvalue* of primal problem (CP) is given by

$$p^- = \lim_{\epsilon \to 0^+} \left[\inf_x c^T x \quad \text{s.t.} \quad \|Ax - b\| < \epsilon \text{ and } x \in \mathcal{C} \right]$$

(a similar definition is holding for the dual subvalue d^-).

It is readily seen that this limit always exists (possibly being $+\infty$), because the feasible region of the infimum shrinks as ϵ tends to zero, which implies that its optimum value is a nonincreasing function of ϵ. Moreover, the inequality $p^- \leq p^*$ holds, because all the feasible regions of the infima defining p^- as ϵ tends to zero are larger than the actual feasible region of problem (CP).

The case $p^- = +\infty$, which implies that primal problem (CP) is infeasible (since we have then $p^* \geq p^- = +\infty$), is called primal *strong infeasibility*, and essentially means that the affine subspace defined by the linear constraints $Ax = b$ is strongly separated from cone $\mathcal{C}$. We are now in position to state the following alternate strong duality theorem:

Theorem 3.3 (Strong duality, alternate version). *We have either*

- $p^- = +\infty$ *and* $d^* = -\infty$ *when primal problem* (CP) *is strongly infeasible and dual problem* (CD) *is infeasible.*
- $p^- = d^*$ *in all other cases.*

This theorem (see e.g. [Stu97, Theorem 2.6] for a proof) states that there is no duality gap between p^- and d^*, except in the rather exceptional case of primal strong infeasibility and dual infeasibility. Note that the second case covers situations where the primal problem is infeasible but not strongly infeasible (i.e. $p^- < p^* = +\infty$).

To conclude this section, we would like to mention the fact that all the properties and theorems described in this section can be easily extended to the case of several conic constraints involving disjoint sets of variables.

Note 3.1. Namely, having to satisfy the constraints $x^i \in \mathcal{C}^i$ for all $i \in \{1, 2, \dots, k\}$, where $\mathcal{C}^i \subseteq \mathbb{R}^{n_i}$, we will simply consider the Cartesian product of these cones $\mathcal{C} = \mathcal{C}^1 \times \mathcal{C}^2 \times \cdots \times \mathcal{C}^k \subseteq \mathbb{R}^{\sum_{i=1}^k n_i}$ and express all these constraints simultaneously as $x \in \mathcal{C}$ with $x = (x^1, x^2, \dots, x^k)$. The dual cone of $\mathcal{C}$ will be given by

$$\mathcal{C}^* = (\mathcal{C}^1)^* \times (\mathcal{C}^2)^* \times \cdots \times (\mathcal{C}^k)^* \subseteq \mathbb{R}^{\sum_{i=1}^k n_i},$$

as implied by the following theorem:

Theorem 3.4. *Let* $\mathcal{C}^1$ *and* $\mathcal{C}^2$ *two closed convex cones, and* $\mathcal{C} = \mathcal{C}^1 \times \mathcal{C}^2$ *their Cartesian product. Cone* $\mathcal{C}$ *is also a closed convex cone, and its dual* $\mathcal{C}^*$ *is given by*

$$\mathcal{C}^* = (\mathcal{C}^1)^* \times (\mathcal{C}^2)^*.$$

4 Classification of conic optimization problems

In this section, we describe all possible types of conic programs with respect to feasibility, attainability of the optimum and optimal duality gap, and provide corresponding examples.

Given our standard primal conic program (CP), we define

$$\mathcal{F}_+ = \{x \in \mathbb{R}^n \mid Ax = b \text{ and } x \in \mathcal{C}\}$$

to be its feasible set and $\delta = \operatorname{dist}(\mathcal{C}, L)$ the minimum distance between cone $\mathcal{C}$ and the affine subspace $\mathcal{L} = \{x \mid Ax = b\}$ defined by the linear constraints. We also call $\mathcal{F}_{++}$ the set of strictly feasible solutions of (CP), i.e.

$$\mathcal{F}_{++} = \{x \in \mathbb{R}^n \mid Ax = b \text{ and } x \in \operatorname{int} \mathcal{C}\}.$$

4.1 Feasibility

First of all, the distinction between feasible and infeasible conic problems is not as clear-cut as for linear optimization. We have the following cases[4]

- A conic program is infeasible. This means the feasible set $\mathcal{F}_+ = \emptyset$, and that $p^* = +\infty$. But we have to distinguish two subcases

[4] In the following, we'll mark with a (‡) the cases which never happen in the case of linear optimization.

- $\delta = 0$, which means an infinitesimal perturbation of the problem data may transform the program into a feasible one. We call the program *weakly infeasible*(‡). This corresponds to the case of a finite subvalue, i.e. $p^- < p^* = +\infty$.
- $\delta > 0$, which corresponds to the usual infeasibility as for linear optimization. We call the program *strongly infeasible*, which corresponds to an infinite subvalue $p^- = p^* = +\infty$.

◇ A conic program is feasible, which means $\mathcal{F}_+ \neq \emptyset$ and $p^* < +\infty$ (and thus $\delta = 0$). We also distinguish two subcases

- $\mathcal{F}_{++} = \emptyset$, which implies that all feasible points belong to the boundary of the feasible set $\mathcal{F}_+$ (this corresponds indeed to the case where the affine subspace L is tangent to the cone $\mathcal{C}$). This also means that an infinitesimal perturbation of the problem data can make the program infeasible. We call the program *weakly feasible*.
- $\mathcal{F}_{++} \neq \emptyset$. We call the program *strongly feasible*. This means there exists at least one feasible solution belonging to the interior of $\mathcal{C}$, which is the main hypothesis of the strong duality Theorem 3.2.

It is possible to characterize these situations by looking at the existence of certain types of directions in the dual problem (level direction, improving direction, improving direction sequence, see [Stu97]). Let us now illustrate these four situations with an example.

Example 4.1. Let us choose

$$\mathcal{C} = \mathbb{S}^2_+ \quad \text{and} \quad x = \begin{pmatrix} x_1 & x_3 \\ x_3 & x_2 \end{pmatrix} .$$

We have that $x \in \mathcal{C} \Leftrightarrow x_1 \geq 0, x_2 \geq 0$ and $x_1 x_2 \geq x_3^2$.

If we add the linear constraint $x_3 = 1$, the feasible set becomes the epigraph of the positive branch of the hyperbola $x_1 x_2 = 1$, i.e. $\mathcal{F}_+ = \{(x_1, x_2) \mid x_1 \geq 0 \text{ and } x_1 x_2 \geq 1\}$ as depicted on Figure 1.

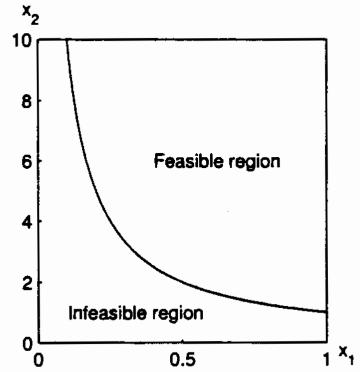

Figure 1: Epigraph of the positive branch of the hyperbola $x_1 x_2 = 1$.

This problem is strongly feasible.

- ◇ If we add another linear constraint $x_1 = -1$, we get a strongly infeasible problem (since x_1 must be positive).
- ◇ If we add $x_1 = 0$, we get a weakly infeasible problem (since the distance between the axis $x_1 = 0$ and the hyperbola is zero but x_1 still must be positive).
- ◇ Finally, adding $x_1 + x_2 = 2$ leads to a weakly feasible problem (because the only feasible point, $x_1 = x_2 = x_3 = 1$, does not belong to the interior of $\mathcal{C}$).

4.2 Attainability

Let us denote by $\mathcal{F}^*$ the set of optimal solutions, i.e. feasible solutions with an objective equal to p^*

$$\mathcal{F}^* = \mathcal{F}_+ \cap \{x \in \mathbb{R}^n \mid c^T x = p^*\} .$$

We have the following distinction regarding attainability of the optimum:

- ◇ A conic program is *solvable* if $\mathcal{F}^* \neq \emptyset$.
- ◇ A conic program is *unsolvable* if $\mathcal{F}^* = \emptyset$, but we have three subcases
  - If $p^* = +\infty$, the program is *infeasible*, as mentioned in the previous subsection.
  - If $p^* = -\infty$, the program is *unbounded.*
  - If p^* is finite, we have a feasible unsolvable bounded program (‡). This situation happens when the infimum defining p^* is not attained, i.e. there exists feasible solutions with objective value arbitrarily close to p^* but no optimal solution.

Let us examine a little further the last situation. In this case, we have a sequence of feasible solutions whose objective value tends to p^*, but no optimal solution. This implies that at least one of the variables in this sequence of feasible solutions tends to infinity. Indeed, if it was not the case, that sequence would be bounded, and since the feasible set $\mathcal{F}$ is closed (it is the intersection of a closed cone and a affine subspace, which is also closed), its limit would also belong to the feasible set, hence would be a feasible solution with objective value p^*, i.e. an optimal solution, which is a contradiction.

Example 4.2. Let us consider the same strongly feasible problem as in Example 4.1 (epigraph of an hyperbola).

- ◇ If we choose a linear objective equal to $x_1 + x_2$,, $\mathcal{F}^*$ is reduced to the unique point $(x_1, x_2, x_3) = (1, 1, 1)$, and the problem is solvable ($p^* = 2$).
- ◇ If we choose another objective equal to $-x_1 - x_2$, $\mathcal{F}^* = \emptyset$ because $p^* = -\infty$, and the problem is unbounded.
- ◇ Finally, choosing x_1 as objective function leads to an unsolvable bounded problem: p^* is easily seen to be equal to zero but $\mathcal{F}^* = \emptyset$ because there is no feasible solution with $x_1 = 0$ since the product $x_1 x_2$ has to be greater than 1.

4.3 Duality gap at optimality

Finally, we state the various possibilities about the *duality gap at optimality*, which is equal to $p^* - d^*$:

- An optimal solution pair (x, y) has a zero duality gap.
- The optimal duality gap is strictly positive (‡).
- The optimal duality gap is zero but there is no optimal solution pair. In this case, there exists pairs (x, y) with an arbitrarily small duality gap (which means that the optimum is not attained for at least one of the two programs (CP) and (CD)) (‡).

Of course, the last two cases can be avoided if we require our problem to satisfy the Slater condition. We can alternatively work with the subvalue p^-, for which there is no duality gap except when both problems are infeasible.

Example 4.3. The first problem described in Example 4.2 has its optimal value equal to $p^* = 2$. Its data can be described as

$$c = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; A : \mathbb{S}^2 \mapsto \mathbb{R} : \begin{pmatrix} x_1 & x_3 \\ x_3 & x_2 \end{pmatrix} \mapsto x_3 \text{ and } b = 1 \; .$$

Using the fact that the adjoint of A can be written as[5]

$$A^{\mathrm{T}} : \mathbb{R} \mapsto \mathbb{S}^2 : y_1 \mapsto \begin{pmatrix} 0 & y_1/2 \\ y_1/2 & 0 \end{pmatrix}$$

and the dual formulation (CD), we can state the dual as

$$\sup y_1 \quad \text{s.t.} \quad \begin{pmatrix} 0 & y_1/2 \\ y_1/2 & 0 \end{pmatrix} + \begin{pmatrix} s_1 & s_3 \\ s_3 & s_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} s_1 & s_3 \\ s_3 & s_2 \end{pmatrix} \in \mathbb{S}^2_+$$

or equivalently, after eliminating the s variables,

$$\sup y_1 \quad \text{s.t.} \quad \begin{pmatrix} 1 & -y_1/2 \\ -y_1/2 & 1 \end{pmatrix} \in \mathbb{S}^2_+ \; .$$

The optimal value d^* of this problem is equal to 2, because the semidefinite constraint is equivalent to $y_1^2 \leq 4$, and the optimal duality gap $p^* - d^*$ is zero as expected.

Changing the primal objective to $c = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, we get an unsolvable bounded problem

$$\inf x_1 \quad \text{s.t.} \quad x_3 = 1 \text{ and } x_1 x_2 \geq 1$$

whose optimal value is $p^* = 0$ but is not attained. The dual becomes

$$\sup y_1 \quad \text{s.t.} \quad \begin{pmatrix} 1 & -y_1/2 \\ -y_1/2 & 0 \end{pmatrix} \in \mathbb{S}^2_+$$

[5] To check this, simply write $\langle Ax, y \rangle = \langle x, A^{\mathrm{T}} y \rangle$, where the first inner product is the usual dot product on $\mathbb{R}^n$ but the second inner product is the trace inner product on $\mathbb{S}^n$.

which admits only one feasible solution, namely $y_1 = 0$, and has thus an optimal value $d^* = 0$. In this case, the optimal duality gap is zero but is not attained (because the primal problem is unsolvable).

Finally, we give here an example where the optimal duality gap is nonzero. Choosing a nonnegative parameter λ and

$$\mathcal{C} = \mathbb{S}^3_+,\ c = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & \lambda \end{pmatrix},\ A : \mathbb{S}^3 \mapsto \mathbb{R}^2 : \begin{pmatrix} x_1 & x_4 & x_5 \\ x_4 & x_2 & x_6 \\ x_5 & x_6 & x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_3 + x_4 \\ x_2 \end{pmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

we have for the primal

$$\inf \lambda x_3 - 2x_4 \quad \text{s.t.} \quad x_3 + x_4 = 1,\ x_2 = 0 \text{ and } \begin{pmatrix} x_1 & x_4 & x_5 \\ x_4 & x_2 & x_6 \\ x_5 & x_6 & x_3 \end{pmatrix} \in \mathbb{S}^3_+ .$$

The fact that $x_2 = 0$ implies $x_4 = x_6 = 0$, which in turn implies $x_3 = 1$. We have thus that all solutions have the form

$$\begin{pmatrix} x_1 & 0 & x_5 \\ 0 & 0 & 0 \\ x_5 & 0 & 1 \end{pmatrix}$$

which is feasible as soon as $x_1 \geq x_5^2$. All these feasible solutions have an objective value equal λ, and hence are all optimal: we have $p^* = \lambda$. Using the fact that the adjoint of A is

$$A^{\mathrm{T}} : \mathbb{R}^2 \mapsto \mathbb{S}^3 : \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto \begin{pmatrix} 0 & y_1/2 & 0 \\ y_1/2 & y_2 & 0 \\ 0 & 0 & y_1 \end{pmatrix}$$

we can write the dual (after eliminating the s variables with the linear equality constraints) as

$$\sup y_1 \quad \text{s.t.} \quad \begin{pmatrix} 0 & -1 - y_1/2 & 0 \\ -1 - y_1/2 & -y_2 & 0 \\ 0 & 0 & \lambda - y_1 \end{pmatrix} \in \mathbb{S}^3_+$$

The above matrix can only be positive semidefinite if $y_1 = -2$. In that case, any nonnegative value for y_2 will lead to a feasible solution with an objective equal to -2, i.e. all these solutions are optimal and $d^* = -2$. The optimal duality gap is equal to $p^* - d^* = \lambda + 2$, which is strictly positive fear all values of λ. Note that in this case, as expected from the theory, none of the two problems satisfies the Slater condition since every feasible primal or dual solution has at least a zero on its diagonal, which implies a zero eigenvalue and hence that it does not belong to the interior of $\mathbb{S}^3_+$.

5 Application to l_p-norm optimization

5.1 Problem definition

l_p-norm optimization [PE70a, Ter85] is an important class of convex problems, which includes as special cases linear optimization, quadratically constrained convex quadratic optimization and l_p-norm approximation problems. The purpose of this section is to show how this class of problems can be modelled with a conic formulation and the advantages of such a procedure (see [GT00] for a more detailed exposition).

Let us start by introducing the primal l_p-norm optimization problem, which is basically a slight modification of a linear optimization problem where the use of l_p-norms applied to linear terms is allowed within the constraints. In order to state its formulation in the most general setting, we need to introduce the following sets: let $K = \{1, 2, \ldots, r\}$, $I = \{1, 2, \ldots, n\}$ and let $\{I_k\}_{k \in K}$ be a partition of I into r classes, i.e. satisfying

$$\cup_{k \in K} I_k = I \text{ and } I_k \cap I_l = \emptyset \text{ for all } k \neq l \, .$$

The problem data is given by two matrices $A \in \mathbb{R}^{m \times n}$ and $F \in \mathbb{R}^{m \times r}$ (whose columns will be denoted by a_i, $i \in I$ and f_k, $k \in K$) and four column vectors $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $d \in \mathbb{R}^r$ and $p \in \mathbb{R}^n$ such that $p_i > 1 \ \forall i \in I$. Our primal problem consists in optimizing a linear function of a column vector $y \in \mathbb{R}^m$ under a set of constraints involving l_p-norms of linear forms, and can be written as

$$\sup \ b^T y \quad \text{s.t.} \quad \sum_{i \in I_k} \frac{1}{p_i} \left| c_i - a_i^T y \right|^{p_i} \leq d_k - f_k^T y \quad \forall k \in K \, . \qquad (\mathrm{P}l_p)$$

Defining a vector $q \in \mathbb{R}^n$ such that $\frac{1}{p_i} + \frac{1}{q_i} = 1$ for all $i \in I$, the dual problem for $(\mathrm{P}l_p)$ can be defined as (see e.g. [Ter85])

$$\inf \ \psi(x, z) = c^T x + d^T z + \sum_{k \in K \mid z_k > 0} z_k \sum_{i \in I_k} \frac{1}{q_i} \left| \frac{x_i}{z_k} \right|^{q_i} \quad \text{s.t.} \quad \begin{cases} Ax + Fz = b \text{ and } z \geq 0 \, , \\ z_k = 0 \Rightarrow x_i = 0 \ \forall i \in I_k \, . \end{cases} \qquad (\mathrm{D}l_p)$$

We note that a special convention has been taken to handle the case when one or more components of z are equal to zero: the associated terms are left out of the first sum (to avoid a zero denominator) and the corresponding components of x have to be equal to zero. When compared with the primal problem $(\mathrm{P}l_p)$, this problem has a simpler feasible region (mostly defined by linear equalities and nonnegativity constraints) at the price of a highly nonlinear (but convex) objective.

5.2 Cones for l_p-norm optimization

Let us now introduce the $\mathcal{L}^p$ cone, which will allow us to give a conic formulation to l_p-norm optimization problems.

Definition 5.1. Let $n \in \mathbb{N}$ and $p \in \mathbb{R}^n$ 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:

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

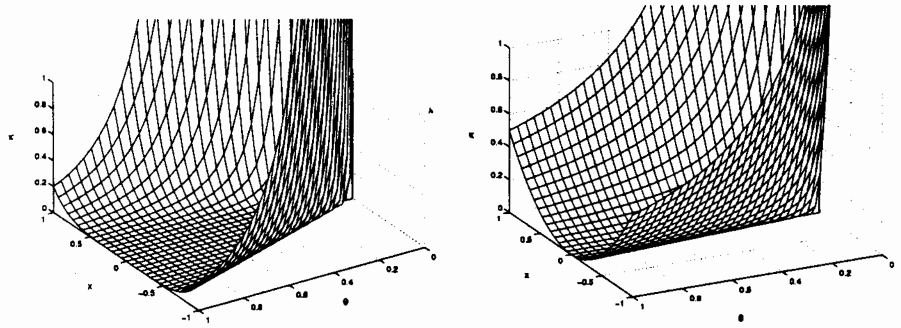

Figure 2: The boundary surfaces of $\mathcal{L}^{(5)}$ and $\mathcal{L}^{(2)}$ (in the case $n = 1$).

This convention means that if $(x, \theta, \kappa) \in \mathcal{L}^p$, $\theta = 0$ implies $x = 0^n$. This set is in fact a solid, pointed closed convex cone (see [GT00]), and is hence suitable for conic optimization.

To illustrate our purpose, we provide in Figure 2 the three-dimensional graphs of the boundary surfaces of $\mathcal{L}^{(5)}$ and $\mathcal{L}^{(2)}$ (corresponding to the case $n = 1$).

The dual cone of $\mathcal{L}^p$ can be explicitly computed. Let us introduce for ease of notation the *switched* cone $\mathcal{L}^p_s$, defined as the $\mathcal{L}^p$ cone with its last two components exchanged, i.e.

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

This allows us to describe the dual of $\mathcal{L}^p$ (which is another solid, pointed closed convex cone by virtue of Theorem 2.3)

Theorem 5.1 (Dual of $\mathcal{L}^p$). *Let $p, q \in \mathbb{R}^n_{++}$ such that $\frac{1}{p_i} + \frac{1}{q_i} = 1$ for each i. The dual of $\mathcal{L}^p$ is $\mathcal{L}^q_s$.*

The dual of a $\mathcal{L}^p$ cone is thus equal, up to a permutation of two variables, to another $\mathcal{L}^p$ cone with a *dual* vector of exponents. We also have $(\mathcal{L}^p_s)^* = \mathcal{L}^q$, $(\mathcal{L}^q)^* = \mathcal{L}^p_s$ and $(\mathcal{L}^q_s)^* = \mathcal{L}^p$, for obvious symmetry reasons.

5.3 Conic formulation for l_p-norm optimization

We now proceed to show how a primal-dual pair of l_p-norm optimization problems can be modelled using the $\mathcal{L}^p$ and $\mathcal{L}^q_s$ cones. Let us start with the primal problem (Pl_p). The following notation will be useful in this context: v_S (resp. M_S) denotes the restriction of column vector v (resp. matrix M) to the components (resp. rows) whose indices belong to set S.

We start by introducing an auxiliary vector of variables $x^* \in \mathbb{R}^n$ to represent the argument of the power functions, namely we let

$$x^*_i = c_i - a_i^{\mathrm{T}} y \text{ for all } i \in I \text{ or, in matrix form, } x^* = c - A^{\mathrm{T}} y ,$$

and we also need additional variables $z^* \in \mathbb{R}^r$ for the linear term forming the right-hand side of the inequalities

$$z^*_k = d_k - f_k^{\mathrm{T}} y \text{ for all } k \in K \text{ or, in matrix form, } z^* = d - F^{\mathrm{T}} y .$$

Our problem is now equivalent to

$$\sup\ b^T y \quad \text{s.t.} \quad A^T y + x^* = c,\ F^T y + z^* = d \text{ and } \sum_{i \in I_k} \frac{1}{p_i} |x_i^*|^{p_i} \le z_k^* \quad \forall k \in K\ ,$$

where we can easily plug our definition of the $\mathcal{L}^p$ cone, provided we fix variables θ to 1

$$\sup\ b^T y \quad \text{s.t.} \quad A^T y + x^* = c,\ F^T y + z^* = d \text{ and } (x_{I_k}^*, 1, z_k^*) \in \mathcal{L}^{p^k}\ \forall k \in K$$

(where for convenience we defined vectors $p^k = (p_i \mid i \in I_k)$ for $k \in K$). We finally introduce an additional vector of fictitious variables $v^* \in \mathbb{R}^r$ whose components are fixed to 1 by linear constraints to find

$$\sup\ b^T y \quad \text{s.t.} \quad A^T y + x^* = c,\ F^T y + z^* = d,\ v^* = e \text{ and } (x_{I_k}^*, v_k^*, z_k^*) \in \mathcal{L}^{p^k}\ \forall k \in K$$

(where e stands again for the all-one vector). Rewriting the linear constraints with a single matrix equality, we end up with

$$\sup\ b^T y \quad \text{s.t.} \quad \begin{pmatrix} A^T \\ F^T \\ 0 \end{pmatrix} y + \begin{pmatrix} x^* \\ z^* \\ v^* \end{pmatrix} = \begin{pmatrix} c \\ d \\ e \end{pmatrix} \text{ and } (x_{I_k}^*, v_k^*, z_k^*) \in \mathcal{L}^{p^k}\ \forall k \in K\ , \qquad (\mathrm{CP}l_p)$$

which is exactly a conic optimization problem in the dual[6] form (CD), using variables $(\tilde{y}, \tilde{s})$, data $(\tilde{A}, \tilde{b}, \tilde{c})$ and a cone $\mathcal{C}^*$ such that

$$\tilde{y} = y,\ \tilde{s} = \begin{pmatrix} x^* \\ z^* \\ v^* \end{pmatrix},\ \tilde{A} = \begin{pmatrix} A & F & 0 \end{pmatrix},\ \tilde{b} = b,\ \tilde{c} = \begin{pmatrix} c \\ d \\ e \end{pmatrix} \text{ and } \mathcal{C}^* = \mathcal{L}^{p^1} \times \mathcal{L}^{p^2} \times \cdots \times \mathcal{L}^{p^r}\ ,$$

where $\mathcal{C}^*$ has been defined according to Note 3.1, since we have to deal with multiple conic constraints involving disjoint sets of variables.

Using the properties of $\mathcal{L}^p$, it is straightforward to show that $\mathcal{C}^*$ is a solid, pointed, closed convex cone whose dual is

$$(\mathcal{C}^*)^* = \mathcal{C} = \mathcal{L}_s^{q^1} \times \mathcal{L}_s^{q^2} \times \cdots \times \mathcal{L}_s^{q^r}\ ,$$

another solid, pointed, closed convex cone (where we have defined a vector $q \in \mathbb{R}^n$ such that $\frac{1}{p_i} + \frac{1}{q_i} = 1$ for all $i \in I$ and vectors q^k such that $q^k = (q_i \mid i \in I_k)$ for $k \in K$). This allows us to derive a dual problem to $(\mathrm{CP}l_p)$ in a completely mechanical way and find the following conic optimization problem, expressed in the primal form (CP) (since the dual of a problem in dual form is a problem in primal form):

$$\inf \begin{pmatrix} c^T & d^T & e^T \end{pmatrix} \begin{pmatrix} x \\ z \\ v \end{pmatrix} \quad \text{s.t.} \quad \begin{pmatrix} A & F & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ v \end{pmatrix} = b \text{ and } (x_{I_k}, v_k, z_k) \in \mathcal{L}_s^{q^k} \text{ for all } k \in K\ ,$$

which is equivalent to

$$\inf c^T x + d^T z + e^T v \quad \text{s.t.} \quad Ax + Fz = b \text{ and } (x_{I_k}, v_k, z_k) \in \mathcal{L}_s^{q^k} \text{ for all } k \in K\ , \qquad (\mathrm{CD}l_p)$$

[6]This is the reason why we added a * superscript to the notation of our additional variables, in order to emphasize the fact that the primal l_p-norm optimization problem $(\mathrm{P}l_p)$ is in fact in the dual conic form (CD).

where $x \in \mathbb{R}^n$, $z \in \mathbb{R}^r$ and $v \in \mathbb{R}^r$ are the dual variables we optimize. This problem can be simplified: making the conic constraints explicit, we find

$$\inf c^{\mathrm{T}}x + d^{\mathrm{T}}z + e^{\mathrm{T}}v \quad \text{s.t.} \quad Ax + Fz = b,\ \sum_{i\in I_k} \frac{|x_i|^{q_i}}{q_i z_k^{q_i-1}} \leq v_k\ \forall k \in K \text{ and } z \geq 0\ ,$$

keeping in mind the convention on zero denominators that in effect implies $z_k = 0 \Rightarrow x_{I_k} = 0$. Finally, we can remove the v variables from the formulation since they are only constrained by the sum inequalities, which have to be tight at any optimal solution. We can thus directly incorporate these sums into the objective function, which leads to

$$\inf\ \psi(x,z) = c^{\mathrm{T}}x + d^{\mathrm{T}}z + \sum_{k\in K|z_k>0} z_k \sum_{i\in I_k} \frac{1}{q_i}\left|\frac{x_i}{z_k}\right|^{q_i} \text{ s.t. } \begin{cases} Ax + Fz = b \text{ and } z \geq 0\ , \\ z_k = 0 \Rightarrow x_i = 0\ \forall i \in I_k\ . \end{cases} \quad (\mathrm{D}l_p)$$

Unsurprisingly, the dual formulation ($\mathrm{D}l_p$) we have just found without much effort is exactly the standard form for a dual l_p-norm optimization problem [Ter85].

5.4 Duality properties

Until now, we have shown how a suitably defined convex cone allows us to model a well-known class of problems such as l_p-norm optimization with a conic formulation. The approach we took consisted of the following steps:

a. find the definition of a convex cone that allows the formulation of primal l_p-norm optimization,

b. compute the dual cone,

c. derive from it the expression of dual l_p-norm optimization.

It is worthwhile to note that only the first of these steps requires some creative thinking, while the second step is merely a question of accurate calculations and the third step is so straightforward it can be performed in a quasi-mechanical way. This contrasts with the traditional way of handling a convex problem in the form of (CF), where computing the dual is the most difficult part.

This conic approach can of course be seen as an interesting method to compute a dual problem, but otherwise has not yet brought us any new insight about the problem. However, as stated earlier, the true power of this conic formulation only comes into play when dealing with duality issues. Indeed, the application of the general theorems presented in Section 3 to the conic formulation of a convex problem will allow us to derive its properties in a seamless way.

A few interesting 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, which is a mere consequence of convexity, but can also be shown to satisfy two additional properties that cannot be guaranteed in the general convex case: the optimum duality gap is equal to zero and at least one feasible solution attains the optimum primal objective. These results were first presented by Peterson and Ecker [PE70a, PE67, PE70b] and later greatly simplified by Terlaky [Ter85], using standard convex duality theory (e.g. the convex Farkas theorem).

Let us now state these theorems and show how they can be proved in the conic framework.

Theorem 5.2 (Weak duality for l_p-norm optimization). *If y is feasible for* $(\mathrm{P}l_p)$ *and* (x, z) *is feasible for* $(\mathrm{D}l_p)$, *we have* $\psi(x, z) \geq b^{\mathrm{T}} y$. *Equality occurs if and only if the following three conditions are satisfied for all $k \in K$ and $i \in I_k$*

$$z_k\Big(\sum_{i\in I_k} \frac{1}{p_i} \left|c_i - a_i^{\mathrm{T}} y\right|^{p_i} + f_k^{\mathrm{T}} y - d_k\Big) = 0, \quad x_i(c_i - a_i^{\mathrm{T}} y) \leq 0, \quad z_k \left|c_i - a_i^{\mathrm{T}} y\right|^{p_i} = \frac{|x_i|^{q_i}}{z_k^{q_i - 1}} . \quad (5.1)$$

Sketch of the proof. The idea of the proof simply consists in showing that the primal-dual pair $(\mathrm{P}l_p)$–$(\mathrm{D}l_p)$ is essentially equivalent to the conic pair $(\mathrm{CP}l_p)$–$(\mathrm{CD}l_p)$, which is known to satisfy a similar weak duality property because of Theorem 3.1 (see [GT00] for the details). □

The weak duality property is a rather straightforward consequence of the convexity of the problems, and in fact can be proved without too many difficulties without sophisticated tools from conic duality theory. However, this is not the case for the next theorem, which deals with a strong duality property.

In the case of a general pair of primal and dual conic problems, the duality gap at the optimum is not always equal to zero, neither are the primal or dual optimum objective values always attained by feasible solutions (see the examples in Section 4). However, it is well-known that 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 equal zero but attainment of the optimum objective value can only be guaranteed for the primal problem.

Theorem 5.3 (Strong duality for l_p-norm optimization). *If both problems* $(\mathrm{P}l_p)$ *and* $(\mathrm{D}l_p)$ *are feasible, the primal optimal objective value is attained with a zero duality gap, i.e.*

$$\begin{aligned} p^* &= \max b^{\mathrm{T}} y \quad s.t. \quad \sum_{i\in I_k} \frac{1}{p_i} \left|c_i - a_i^{\mathrm{T}} y\right|^{p_i} \leq d_k - f_k^{\mathrm{T}} y \quad \forall k \in K \\ &= \inf \psi(x, z) \quad s.t. \quad \left\{ \begin{array}{l} Ax + Fz = b \text{ and } z \geq 0 \\ z_k = 0 \Rightarrow x_i = 0 \ \forall i \in I_k \end{array} \right. \qquad = d^* . \end{aligned}$$

Sketch of the proof. The strong duality Theorem 3.2 tells us that zero duality gap and primal attainment are guaranteed by the existence of a dual strictly feasible solution (excluding the case of an unbounded dual, which cannot happen since the primal is feasible). The idea behind this proof will thus consist in pointing out a strictly feasible solution for $(\mathrm{CD}l_p)$. Unfortunately, it may happen that such a point does not exist, which leads to the application of the following three-step strategy:

a. Since the linear constraints of the problem may prevent the existence of a strictly feasible solution to $(\mathrm{CD}l_p)$, we define a *restricted* version of $(\mathrm{CD}l_p)$ where the problematic components of the dual variables have been removed. Hopefully, this restricted problem $(\mathrm{RD}l_p)$ does not behave too differently from the original problem $(\mathrm{CD}l_p)$ because the removed components did not play a crucial role in it.

b. Since this restricted problem now admits a strictly feasible solution, its dual problem $(\mathrm{RP}l_p)$ (which is a problem in primal form) is solvable with a zero duality gap.

c. The last step of the proof consists in converting this optimal solution with a zero duality gap for the restricted primal problem $(\mathrm{RP}l_p)$ into an optimal solution for the original primal problem (CPl_p), which is enough to show that there is a zero duality gap with primal attainment between our original problems (Pl_p) and (Dl_p).

The whole procedure can be summarized with the following diagram:

(Pl_p)	$\equiv$	(CPl_p)	Weak $\longleftrightarrow$	(CDl_p)	$\equiv$	(Dl_p)
		c. $\updownarrow$		$\updownarrow$ a.		
		(RPl_p)	Strong (zero gap) $\longleftrightarrow$	(RDl_p)		
		$\downarrow$ (Attainment)	b.	$\uparrow$ (Strictly feasible)		

We refer the reader to [GT00] for the technical details of the proof. □

It is worthwhile to note that this proof is simpler and shorter than the original proofs one can find in the literature [PE70a, PE67, PE70b, Ter85], the specificity of the class of problems under study being confined to the convex cone used in the formulation. Moreover, the fundamental reason why this class of optimization problems has better duality properties than a general convex problem becomes clear: this is essentially due to the existence of a strictly interior dual solution (even if a reduction procedure involving an equivalent regularized problem has to be introduced when the original dual lacks a strictly feasible point).

6 Conclusions

In this article, we have presented the conic formulation for convex optimization problems. We hope to have convinced the reader of its various advantages over the traditional formulation, namely

- ◇ The standard primal-dual pair of conic problems does feature a great deal of symmetry.
- ◇ The dual problem can be easily derived once the dual cone has been computed.
- ◇ Duality theorems can be revisited and proved by means of the general conic duality theory, which often leads to shortened and simplified proofs and provides additional insight about the problems (see also [Gli01a] where a work similar to what was outlined in Section 5 was done for geometric optimization).
- ◇ The conic formulation also proves helpful when designing and analysing algorithms: the theory of self-concordant barriers for convex optimization is derived by Nesterov and Nemirovski using a conic formulation [NN94], while Nesterov and Todd generalize a very efficient class of primal-dual methods originally developed for linear optimization to a class of conic problems built with so-called *self-scaled* cones [NT97].
- ◇ Finally, the investigation of the similarities between geometric optimization and l_p-norm optimization led the author to the definition of a common generalization of the corresponding convex cones [Gli01b, Chapter 7]. This resulted in a large class of separable convex problems defined as

  $$\sup\ b^T y \quad \text{s.t.} \quad \sum_{i \in I_k} g_i(c_i - a_i^T y) \le d_k - f_k^T y \quad \forall k \in K ,$$

  where the g_i's are arbitrary scalar convex functions. Duality properties for this seemingly new class of problems appear to be rich and are currently under investigation.

References

[Dan63] G. B. Dantzig, *Linear programming and extensions*, Princeton University Press, Princeton, N.J., 1963.

[ET76] I. Ekeland and R. Temam, *Convex analysis and variational problems*, Studies in mathematics and its applications, vol. 1, North-Holland publishing company, Amsterdam, Oxford, 1976.

[GJ79] M. R. Garey and D. S. Johnson, *Computers and intractability, a guide to the theory of NP-completeness*, Freeman and Co, San Francisco, 1979.

[Gli01a] F. Glineur, *Proving strong duality for geometric optimization using a conic formulation*, Annals of Operations Research **105** (2001), 155–184.

[Gli01b] ———, *Topics in convex optimization: Interior-point methods, conic duality and approximations*, Ph.D. thesis, Faculté Polytechnique de Mons, Mons, Belgium, January 2001.

[GT00] F. Glineur and T. Terlaky, *A conic formulation for l_p-norm optimization*, IMAGE Technical Report 0005, Faculté Polytechnique de Mons, Mons, Belgium, May 2000, submitted.

[Nes96] Y. Nesterov, *Nonlinear optimization*, Notes from a lecture given at CORE, UCL, Belgium, 1996.

[NN94] Y. E. Nesterov and A. S. Nemirovski, *Interior-point polynomial methods in convex programming*, SIAM Studies in Applied Mathematics, SIAM Publications, Philadelphia, 1994.

[NT97] Y. Nesterov and M. J. Todd, *Self-scaled barriers and interior-point methods for convex programming*, Mathematics of Operations Research **22** (1997), no. 1, 1–42.

[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] ———, *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] ———, *Geometric programming: Duality in quadratic programming and l_p approximation III*, Journal on Mathematical Analysis and Applications **29** (1970), 365–383.

[Ren99] J. Renegar, *A mathematical view of interior-point methods in convex optimization*, CORE Lecture series (also to be published by SIAM in the MPS/SIAM Series on Optimization), UCL, Louvain-la-Neuve, Belgium, 1999.

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

[RTV97] C. Roos, T. Terlaky, and J.-Ph. Vial, *Theory and algorithms for linear optimization. an interior point approach*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Chichester, UK, 1997.

[Sch86] A. Schrijver, *Theory of linear and integer programming*, Wiley-Interscience series in discrete mathematics, John Wiley & sons, 1986.

[Stu97] J. F. Sturm, *Primal-dual interior-point approach to semidefinite programming*, Ph.D. thesis, Erasmus Universiteit Rotterdam, The Netherlands, 1997, published in [Stu99].

[Stu99] ______, *Duality results*, High Performance Optimization (H. Frenk, C. Roos, T. Terlaky, and S. Zhang, eds.), Applied optimization, vol. 33, Kluwer Academic Publishers, 1999, pp. 21–60.

[SW70] J. Stoer and Ch. Witzgall, *Convexity and optimization in finite dimensions I*, Springer Verlag, Berlin, 1970.

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

[VB96] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Review **38** (1996), 49–95.