



HELLENIC REPUBLIC

**National and Kapodistrian
University of Athens**

— EST. 1837 —



DEPARTMENT OF
INFORMATICS +
TELECOMMUNICATIONS

Algorithmic Operation Research From Linear to Conic Duality

Theodora Panagea - 1115201400135

Anna-Aikaterini Kavvada - 1115201500050

14-1-2020

Professor: Anna Karasoulou

Contents

1	Introduction	3
1.1	Definitions	3
2	What is Conic Duality?	8
2.1	The Dual Cone	9
2.2	The Dual Problem of a Conic Problem	10
3	Conic Duality Theorem	10
3.1	Symmetry	11
3.2	Weak Conic Duality Theorem	12
3.3	Regular Conic Duality Theorem	12
3.4	Strong Conic Duality Theorem	15
4	Geometry of Primal and Dual Problems	16
4.1	From Algebra to Geometry	16
4.2	Relation between (P) and (D) Problems	17
5	Applications of Conic Duality	18
5.1	Self-Dual Cones	18
5.2	Second Order Cone Programming (SOCP)	19
5.3	Semidefinite Programming (SDP)	19
5.4	Distributionally Robust Learning (DRL)	20
6	Is Something Wrong?	20
7	Consequences	20
7.1	Sufficient Condition for Infeasibility	20
7.2	The Linear Inequality between a Scalar and a Linear Vector .	21
7.3	Robustness	22
8	Bibliography	25

1 Introduction

Linear Programming (LP) models cover numerous applications. Whenever applicable, LP allows to obtain useful quantitative and qualitative information on the problem at hand. The specific analytic structure of LP programs gives rise to a number of general results, which provide us in many cases with valuable insight and understanding. This analytic structure underlies some specific computational techniques for LP, which by now are perfectly well developed, allowing us to solve routinely quite large LP programs. However, there are situations in reality which cannot be covered by the LP models. To handle these "essentially nonlinear" cases, one needs to extend the basic theoretical result and computational techniques known for LP beyond the bounds of Linear Programming.

The widest class of optimization problems to which the basic results of LP were extended, is the class of convex optimization problems. Below, follows a definition of a general convex optimization problem. This definition is not the traditional, but it suits well to the applications we intend to cover.

When passing from a generic LP problem

$$\max_x \{c^T x | Ax \geq b\} \quad [A : m \times n] \quad (LP)$$

to its nonlinear extensions, we should expect to encounter some nonlinear components in the problem. The traditional way here is to say that in (LP), there are a linear objective function $f(x) = c^T x$ and inequality constraints $f_i(x) \geq b_i$ with linear functions $f_i(x) = a_i^T x$, $i = 1, \dots, m$. Let us allow some/all of these functions f, f_1, \dots, f_m to be nonlinear. In contrast to this traditional way, we intend to keep the objective and the constraints linear, but introduce "non-linearity" in the inequality sign \geq .

1.1 Definitions

The constraint inequality $Ax \geq b$ in (LP) is an inequality between vectors; as such, it requires a definition: given two vectors $a, b \in \mathbb{R}^m$, we write $a \geq b$, if the coordinates of a majorate the corresponding coordinates of b :

$$a \geq b \Leftrightarrow \{a_i \geq b_i, \quad i = 1, \dots, m\} \quad (\geq)$$

The above “coordinate-wise” partial ordering of vectors in \mathbb{R}^m satisfies a number of basic properties of the standard ordering of reals; namely, for all vectors $a, b, c, d, \dots \in \mathbb{R}^m$ one has:

1. *Reflexivity*: $a \geq a$
2. *Anti-Symmetry*: if both $a \geq b$ and $b \geq a$, then $a = b$
3. *Transitivity*: if both $a \geq b$ and $b \geq c$, then $a \geq c$
4. *Compatibility with linear operations*:
 - *Homogeneity*: if $a \geq b$ and λ is a nonnegative real, then $\lambda a \geq \lambda b$ (“One can multiply both sides of an inequality by a nonnegative real”)
 - *Additivity*: if both $a \geq b$ and $c \geq d$, then $a + c \geq b + d$ (“One can add two inequalities of the same sign”)

Thus, we observe that

- A significant part of the nice features of LP programs comes from the fact that the vector inequality \geq in the constraint of (LP) satisfies the properties 1. – 4.;
- The standard inequality “ \geq ” is neither the only possible, nor the only interesting way to define the notion of a vector inequality fitting the axioms 1. – 4.

Therefore, we can conclude to the following:

A generic optimization problem which looks exactly the same as (LP), up to the fact that the inequality \geq in (LP) is now replaced with and ordering which differs from the component-wise one, inherits a significant part of the properties of (LP) problems. Specifying properly the ordering of vectors, one can obtain from (LP) generic optimization problems covering many important applications which cannot be treated by the standard (LP).

We have to define the "geometry" of a "vector inequality", bearing in mind the axioms 1. – 4. Consequently, we pick vectors from a finite-dimensional Euclidean space E with an inner product $\langle \cdot, \cdot \rangle$ and reach the assumption that E has a partial ordering (called also vector inequality), let's refer to it as \succeq . In other words, we define the pairs of vectors a, b from E linked by the inequality $a \succeq b$. We call the ordering "good", if it obeys the axioms 1. – 4., and will examine what these good orderings are.

At first, we observe that:

A good vector inequality \succeq is completely identified by the set K of \succeq -nonnegative vectors:

$$K = \{a \in E : a \succeq 0\}$$

Furthermore,

$$a \succeq b \Leftrightarrow a - b \succeq 0 \quad [\Leftrightarrow a - b \in K]$$

Let's prove the above statement. Let $a \succeq b$. By 1. we have $-b \succeq -b$, and by 4.(b) we may add the two inequalities with the same sign, to get $a - b \succeq 0$. If $a - b \succeq 0$, then, adding to this inequality the $b \succeq b$, we get $a \succeq b$.

It is clear that the set K cannot be arbitrary. It must satisfy the following conditions:

1. K is nonempty and closed under addition:

$$a, a' \in K \Rightarrow a + a' \in K$$

2. K is a conic set:

$$a \in K, \lambda \geq 0 \Rightarrow \lambda a \in K$$

3. K is pointed:

$$a \in K \text{ and } -a \in K \Rightarrow a = 0$$

Therefore, it is easy to verify that K is a pointed cone and geometrically, K does not contain any straight lines passing through the origin.

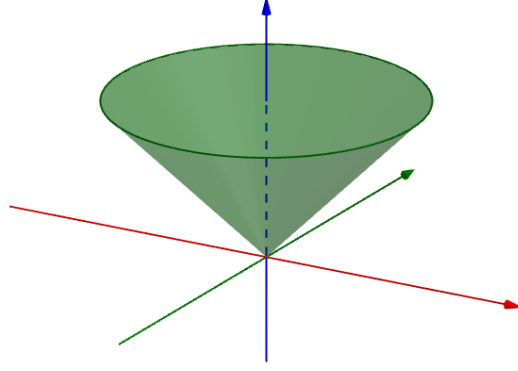


Figure 1: A pointed cone

Hence, every pointed cone K in E leads to a partial ordering on E , satisfying the axioms 1.-4. We symbolize this ordering with \geq_κ :

$$a \geq_\kappa b \Leftrightarrow a - b \geq_\kappa 0 \Leftrightarrow a - b \in K$$

The cone comprised of vectors with nonnegative entries (the nonnegative orthant) is the one responsible for the standard coordinate-wise ordering \geq on $E = \mathbb{R}^m$:

$$\mathbb{R}_+^m = \{x = (x_1, \dots, x_m)^T \in \mathbb{R}^m : x_i \geq 0, i = 1, \dots, m\}$$

We are going to use the standard shorthand notation $a \geq b$, instead of $a \geq_{\mathbb{R}_+^m} b$

The nonnegative orthant \mathbb{R}_+^m is not just a pointed cone, but it also possesses the following properties:

- I. The cone is closed: if a sequence of vectors a_i from the cone has a limit, the latter also belongs to the cone.
- II. The cone possesses a nonempty interior: there exists a vector such that a ball of positive radius centered at the vector is contained in the cone.

From now on, speaking about vector inequalities \geq_κ , we always assume that the underlying set K is a pointed and closed cone with a nonempty interior.

Note that the closedness of K makes it possible to pass to limits in \geq_κ -inequalities:

$$a_i \geq_\kappa b_i, \quad a^i \rightarrow a, b^i \rightarrow b \text{ as } i \rightarrow \infty \Rightarrow a \geq_\kappa b$$

. The nonemptiness of the interior of K allows to define, along with the “non-strict” inequality $a \geq_\kappa b$, also the strict inequality according to the rule

$$a >_\kappa b \Leftrightarrow a - b \in \text{int}K$$

where $\text{int}K$ is the interior of the cone K .

Examples: Some interesting cones with partial orderings, are the following:

- The nonnegative orthant \mathbb{R}_+^m in \mathbb{R}^n
- The Lorentz (or the second-order, or the less scientific name the ice-cream) cone

$$L^m = \left\{ x = (x_1, \dots, x_{m-1}, x_m)^T \in \mathbb{R}^m : x_m \geq \sqrt{\sum_{i=1}^{m-1} x_i^2} \right\}$$

- The semidefinite cone S_+^m . This cone is in the space $E = S^m$ of $m \times m$ symmetric matrices (equipped with the Frobenius inner product $\langle A, B \rangle = \text{Tr}(A, B) = \sum_{ij} A_{ij} B_{ij}$) and consists of all $m \times m$ matrices A which are positive semidefinite, i.e.,

$$A = A^T; \quad x^T A x \geq 0 \quad \forall x \in \mathbb{R}^m$$

2 What is Conic Duality?

Apart from the algorithmic issues, the most important result in LP is the LP Duality Theorem. This theorem, can be extended further to cover convex and conic problems, the latter being a subcategory of the former.

The source of the LP Duality Theorem was the desire to get in a systematic way a lower bound on the optimal value c^* in an LP program

$$c^* = \min_x \{c^T x | Ax \geq b\}. \quad (LP)$$

The bound was obtained by looking at the inequalities of the type

$$\langle \lambda, Ax \rangle \equiv \lambda^T Ax \geq \lambda^T b \quad (Cons(\lambda))$$

with weight vectors $\lambda \geq 0$. This type of inequality is, by its origins, a consequence of the system of constraints $Ax \geq b$ of LP, i.e., it is satisfied at every solution to the system. As a consequence, whenever we are lucky to get, as the left hand side of $(Cons(\lambda))$, the expression $c^T x$, i.e., whenever a nonnegative weight vector λ satisfies the relation

$$A^T \lambda = c,$$

then, the inequality $(Cons(\lambda))$ produces a lower bound $b^T \lambda$ on the optimal value in LP and the dual problem

$$\max \{b^T \lambda : \lambda \geq 0, A^T \lambda = c\}$$

was nothing but the problem of finding the best lower bound one can get in this way.

Coming to think of a conic problem, the same strategy can be used to develop its dual

$$\min \{c^T x | Ax \geq_{\kappa} b\}, \quad K \subset E \quad (CP)$$

At this point, we need to clarify the following step:

What are the “admissible” weight vectors λ , i.e., the vectors such that the scalar inequality

$$\langle \lambda, Ax \rangle \geq \langle \lambda, b \rangle$$

is a consequence of the vector inequality $Ax \geq_{\kappa} b$?

In the particular case of coordinate-wise partial ordering, i.e., in the case of $E = R^m, K = R_+^m$, the admissible vectors were those with nonnegative coordinates. These vectors, however, not necessarily are admissible for an ordering \geq_{κ} when K is different from the nonnegative orthant.

2.1 The Dual Cone

In order to define Conic Duality, we need to introduce the dual cone (K_*) and its properties.

The set K_* is comprised of vectors whose inner products with all vectors from K are nonnegative. K_* is called the cone dual to K .

Theorem (Properties of the dual cone). *Let E be a finite-dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and let $K \subset E$ be a nonempty set. Then*

(i) *The set*

$$K_* = \{\lambda \in E^m \mid \langle \lambda, a \rangle \geq 0 \quad \forall a \in K\}$$

is a closed cone

(ii) *If $\text{int}K \neq \emptyset$, then K_* is pointed.*

(iii) *If K is a closed convex pointed cone, then $\text{int} K_* \neq \emptyset$.*

(iv) *If K is a closed cone, then so is K_* , and the cone dual to K_* is K itself:*

$$(K_*)_* = K$$

An immediate corollary of the Theorem is as follows:

Corollary. *A set $K \subset E$ is a regular (i.e., closed, convex, pointed and with a nonempty interior) cone if and only if the set K_* is so.*

2.2 The Dual Problem of a Conic Problem

Following the above, as in the case of (LP), we observe, as an immediate consequence of the definition of K_* , that whenever x is a feasible solution to conic primal problem (P) and λ is an admissible weight vector, i.e., $\lambda \in K_*$, then x satisfies the scalar inequality:

$$(A^*\lambda)^T x \equiv \langle \lambda, Ax \rangle \geq \langle \lambda, b \rangle$$

Thus, whenever λ_* is an admissible weight vector satisfying the relation

$$A^*\lambda = c$$

we have

$$c^T x = (A^*\lambda)^T x = \langle \lambda, Ax \rangle \geq \langle b, \lambda \rangle$$

for all x feasible for (P), so that $\langle b, \lambda \rangle$ is a lower bound on the optimal value of (P). The best bound is the optimal value in the problem:

$$\max\{\langle b, \lambda \rangle : A^*\lambda = c, \lambda \succeq_{K_*} 0\}$$

and this program is called the dual program (D) to (P). To conclude, we have reached to the following proposition, which we will analyse further, at the next section.

Proposition (Weak Conic Duality Theorem). *The optimal value of (D) is a lower bound on the optimal value of (P).*

3 Conic Duality Theorem

For this section, let us call the following program as the conic primal, and name it (CP):

$$\begin{aligned} & \text{Maximize} && \langle c, x \rangle \\ & \text{subject to} && b - Ax \in L \\ & && x \in K \end{aligned} \tag{P}$$

Then, we define its dual (CD) as the conic program below:

$$\begin{aligned} & \text{Minimize} && \langle b, y \rangle \\ & \text{subject to} && A^T y - c \in K^* \\ & && y \in L^* \end{aligned} \tag{D}$$

In linear programming, we assume that one of the two programs, let's say (LD) is feasible. Firstly we prove the *weak duality*: If the primal program (LP) is subfeasible, the subvalue of (LP) is upper-bounded by the value of (LD). The weak duality has the important consequence that (LD) - a minimization problem - has finite value if (LP) is subfeasible.

Then we prove *regular duality*: there is no gap between the subvalue of (LP) and the.

Similarly, we have to prove the following, in order to establish the conic duality:

1. Symmetry
2. Weak Conic Duality Theorem
3. Strong Conic Duality Theorem

3.1 Symmetry

Let's change the (D), in order to have the same format as the (CP).

$$\begin{aligned} \text{Maximize} \quad & -\langle b, y \rangle \\ \text{subject to} \quad & -c + A^T y \in K_* \\ & y \in L^* \end{aligned} \tag{D}$$

Having done this, we can easily compute the dual of (D') which leads us back to (P). Now, we can continue with the Weak and Strong Duality.

First, we prove *weak duality*: If the primal program (P) is subfeasible, the subvalue of (P) is upper-bounded by the value of (D). Weak duality has the important consequence that (D)—a minimization problem—has finite value if (P) is subfeasible.

Then we prove *regular duality*: there is in fact no gap between the subvalue of (P) and the value of (D). There was no need for that in LP, but here the following scenario is possible: both (P) and (D) are feasible, but there is a gap between their values γ and β .



Figure 2: Gap between the values of (P) and (D) cone program

In order to get *strong duality*, we apply Slater's constraint qualification: If one of the programs has an interior point, then the other program is feasible as well, and there is no gap between their values.

3.2 Weak Conic Duality Theorem

Theorem. *If the dual program (D) is feasible, and if the primal program (P) is subfeasible, then the subvalue of (P) is upper-bounded by the value of (D).*

If (P) is feasible as well, this implies that the value of (P) is upper-bounded by the value of (D), and that both values are finite. This is weak duality as we know it from linear programming.

Proof. We pick any feasible solution y of (D) and any feasible sequences $(x_k)_{k \in \mathbb{N}}$, $(x'_k)_{k \in \mathbb{N}}$ of (P). Then we know that

$$0 \leq \underbrace{\langle A^T y - c, x_k \rangle}_{\in K_*} + \underbrace{\langle y, x'_k \rangle}_{\in L_*} = \langle y, Ax_k + x'_k \rangle - \langle c, x_k \rangle, k \in \mathbb{N}$$

Hence,

$$\limsup_{k \rightarrow \infty} \langle c, x_k \rangle \leq \limsup_{k \rightarrow \infty} \langle y, Ax_k + x'_k \rangle = \langle c, b \rangle$$

Since the feasible sequences were arbitrary, this means that the subvalue of (P) is upper-bounded by $\langle y, b \rangle$, and since y was an arbitrary feasible solution of (D), the lemma follows. \square

3.3 Regular Conic Duality Theorem

Here is the Regular duality theorem. The proof essentially consists of applications of the Farkas Lemma, to carefully crafted systems.

Lemma (Farkas Lemma). *Let $K \subseteq V$ be a closed convex cone, and $b \in W$. The system $Ax = b, x \in K$ is subfeasible if and only if every $y \in W$ with $A^T y \in K_*$ also satisfies $\langle y, b \rangle \geq 0$*

Theorem. *The dual program (D) is feasible and has finite value β if and only if the primal program (P) is subfeasible and has finite subvalue γ . Moreover, $\beta = \gamma$.*

Proof. If (D) is feasible and has value β , we know that

$$A^T y - c \in K_*, y \in L_* \quad \Rightarrow \quad \langle b, y \rangle \geq \beta \quad (\text{I})$$

We also know that

$$A^T y \in K_*, y \in L_* \quad \Rightarrow \quad \langle b, y \rangle \geq 0 \quad (\text{II})$$

Indeed, if we had some y that fails to satisfy the latter implication, we could add a large positive multiple of it to any feasible solution of (D) and in this way obtain a feasible solution of value smaller than β . Merging the (I) and (II), we get:

$$A^T y - zc \in K_*, y \in L_*, z \geq 0 \rightarrow \langle b, y \rangle \geq z\beta. \quad (\text{III})$$

For $z > 0$, we obtain this from (I) by multiplication of all terms with z and then renaming $zy \in L_*$ back to y . For $z = 0$, it is simply (II). In matrix form, we can rewrite (III) as follows:

$$\left(\begin{array}{c|c} A^T & -c \\ \hline \text{id} & 0 \\ \hline 0 & 1 \end{array} \right) (y, z) \in K_* \oplus L_* \oplus R_+ \quad \Rightarrow \quad (b^T | \beta)(y, z) \geq 0 \quad (\text{IV})$$

Here and in the following, we use a column vector $c \in V$ as the linear operator $z \mapsto zc$ from \mathbb{R} to V and the row vector c^T as the (adjoint) linear operator $x \mapsto \langle c, x \rangle$ from V to \mathbb{R} . The form (IV) now allows us to apply the Farkas lemma. The implication (IV) holds if and only if

$$\left(\begin{array}{c|c|c} A & \text{id} & 0 \\ \hline -c^T & 0^T & 1 \end{array} \right) (x, x', z) = (b, -\beta), (x, x', z) \in (K_* \oplus L_* \oplus R_+)_* = K \oplus L \oplus R_+ \quad (\text{V})$$

is subfeasible. The above system is subfeasible if and only if there are sequences $(x_k)_{k \in \mathbb{N}}, (x'_k)_{k \in \mathbb{N}}, (z_k)_{k \in \mathbb{N}}$ with $x_k \in K, x'_k \in L, z_k \geq 0$ for all k , such that

$$\lim_{k \rightarrow \infty} Ax_k + x'_k = b \quad (\text{VI})$$

and

$$\lim_{k \rightarrow \infty} \langle c, x_k \rangle - z_k = \beta \quad (\text{VII})$$

Now (VI) shows that (P) is subfeasible, and (VII) shows that the subvalue of (P) is at least β . Weak duality shows that it is at most β , concluding the “only if” direction.

For the “if” direction, let (P) be subfeasible with finite subvalue γ and assume for the purpose of obtaining a contradiction that (D) is infeasible. This yields the implication

$$A^T y - zc \in K_*, y \in L_*, \rightarrow z \leq 0 \quad (\text{VIII})$$

since for any pair (y, z) that violates it, $\frac{1}{z}y$ would be a feasible solution of (D). We now play the same game as before and write this in Farkas-lemma-compatible matrix form:

$$\left(\begin{array}{c|c} A^T & -c \\ \hline \text{id} & 0 \end{array} \right) (y, z) \in K_* \oplus L_* \Rightarrow (0^T | -1)(y, z) \geq 0 \quad (\text{IX})$$

That means that the system

$$\left(\begin{array}{c|c} A & \text{id} \\ \hline -c^T & 0 \end{array} \right) (x, x') = (0, -1), (x, x') \in K \oplus L \quad (\text{X})$$

is subfeasible, which in turn means that there are sequences $(x_k)_{k \in \mathbb{N}}, (x'_k)_{k \in \mathbb{N}}$ with $x_k \in K, x'_k \in L$ for all k , such that

$$\lim_{k \rightarrow \infty} Ax_k + x'_k = 0 \quad (\text{XI})$$

$$\lim_{k \rightarrow \infty} \langle c, x_k \rangle = 1 \quad (\text{XII})$$

But this is a contradiction: Elementwise addition of $(x_k)_{k \in \mathbb{N}}, (x'_k)_{k \in \mathbb{N}}$ to any feasible sequences of (P) that witness subvalue γ would result in feasible sequences that yield subvalue at least $\gamma + 1$.

Consequently, the dual program (D) must have been feasible. Weak duality yields that (D) has finite value $\beta \geq \gamma$. But then $\beta = \gamma$ follows from the previous “only if” direction. \square

3.4 Strong Conic Duality Theorem

Here is the **strong Duality Theorem of Cone Programming**, under Slater’s constraint qualification.

Theorem. *If the primal program (P) is feasible, has finite value and has an interior point \tilde{x} , then the dual program (D) is feasible and has finite value $\beta = \gamma$.*

Proof. (P) is in particular subfeasible, and since the program has an interior point, that means that the subvalue of (P) is also γ . Using the regular Duality Theorem (“if” direction), the statement follows. \square

The strong Duality Theorem is not applicable if the primal cone program (P) is in equational form ($L = \{0\}$), since the cone $L = \{0\}$ has no interior points. But there is a different variant of Slater’s constraint qualification that we can use in this case, given that V, W are finite-dimensional.

Theorem. *Assume that the Hilbert spaces V and W are finite-dimensional. If the primal program*

$$\begin{aligned} & \text{Maximize} && \langle c, x \rangle \\ & \text{subject to} && Ax = b \\ & && x \in K \end{aligned} \tag{P}$$

is feasible, has finite value γ and has a point $\tilde{x} \in \text{int}(K)$ such that $A\tilde{x} = b$, then the dual problem

$$\begin{aligned} & \text{Minimize} && \langle b, y \rangle \\ & \text{subject to} && A^T y - c \in K_* \end{aligned} \tag{D}$$

is feasible and has finite value $\beta = \gamma$.

We remark that for general L , just requiring a point $\tilde{x} \in \text{int}(K)$ with $b - A\tilde{x} \in L$. is not enough to achieve strong duality.

4 Geometry of Primal and Dual Problems

Let's consider the geometry of primal (P) and dual (D) conic problems:

$$\min\{c^T x \mid Ax \geq_{\kappa} b\} \quad (P)$$

$$\max\{\langle b, \lambda \rangle \mid A^* \lambda = c, \lambda \geq_{\kappa^*} 0\} \quad (D)$$

where we can observe that their in-between structure looks quite different and this difference is due to data representation. However, we will examine them more carefully, and prove that geometrically, the problems (P) and (D) are completely similar.

4.1 From Algebra to Geometry

In (D) our main goal is to maximize a linear objective $\langle b, \lambda \rangle$, over the intersection of an affine plane $L_* = \{\lambda : A^* \lambda = c\}$ with the cone K_* .

In (P), if x are "true design variables", we get their images $y = Ax - b \in E$. When x runs through \mathbb{R}_n , y runs through the affine plane $L = \{y = Ax - b : x \in \mathbb{R}_n\}$; $x \in \mathbb{R}_n$ is feasible for (P) if and only if the corresponding $y = Ax - b$ belongs to the cone K .

So, in (P) we have the intersection of the affine plane L and the cone K , same fashion as (D). Having the image of x $y = Ax - b$, we can express the objective $c^T x$ it as follows:

$$c^T x = \langle d, Ax - b \rangle + \text{const} \equiv c \in \text{Im} A^* \quad (1)$$

where (P) can be expressed as:

$$\min_y \{\langle d, y \rangle : y \in L, y \geq_{\kappa} 0\}$$

since $L = \text{Im} A - b$, $A^* d = c$ and d is any vector satisfying the above relation.

The primal problem (P) (1), geometrically, is the problem of minimizing a linear form over the intersection of the affine plane L with the cone K . The dual problem (D), similarly, is to maximize another linear form over the intersection of the affine plane L_* with the dual cone K_* .

However, if the conditions for (P) are not met, the problem is either unbounded below, or infeasible. Thus, we reject (P) from the very beginning. Therefore, we assume that (1) is satisfied.

4.2 Relation between (P) and (D) Problems

To wrap this up, we need to explain the relation between the geometric data of (P) and (D) problems. Now we are aware of the association of cone K_* being the dual cone of K , in (D) and (P) problems respectively. Following that, we also know that the feasible L and L_* are orthogonal to each other ($L_* \perp L$). More specifically, L is the translation by vector $-b$, of the linear subspace

$$\mathcal{L} = \text{Im}A \equiv \{y = Ax : x \in \mathbb{R}^n\}$$

In the same fashion, we get the linear subspace \mathcal{L}_* of the dual, which is derived from the translation of L_* with any solution λ_0 of $A^*\lambda = c$.

$$\mathcal{L}_* = \text{Null}(A^*) \equiv \{\lambda : A^*\lambda = 0\}$$

From Linear Algebra we know that \mathcal{L} and \mathcal{L}_* are also orthogonal complements of each other.

We reach the following geometrical conclusion, as far as the conic problem is concerned:

$$\min_y \{\langle d, y \rangle : y \in \mathcal{L} - b, y \geq_{\kappa} 0\} \quad (\text{P})$$

where (P) is a problem of minimizing a linear objective $\langle d, y \rangle$ over the intersection of a cone K with an affine plane $L = \mathcal{L} - b$ given as a translation, by vector $-b$, of a linear subspace \mathcal{L} .

Its dual problem is:

$$\max_{\lambda} \{\langle b, \lambda \rangle : \lambda \in \mathcal{L}^\perp + d, \lambda \geq_{\kappa_*} 0\} \quad (\text{D})$$

where (D) is a problem of maximizing the linear objective $\langle b, \lambda \rangle$ over the intersection of the dual cone K_* with an affine plane $L_* = \mathcal{L}^\perp + d$ given as a translation, by the vector d , of the orthogonal complement \mathcal{L}^\perp of \mathcal{L} .

From Weak Conic Duality, we know that $(K_*)_* = K$ and $(\mathcal{L}^\perp)^\perp = \mathcal{L}$. The geometry and the above symmetry of (P) and (D) is illustrated below:

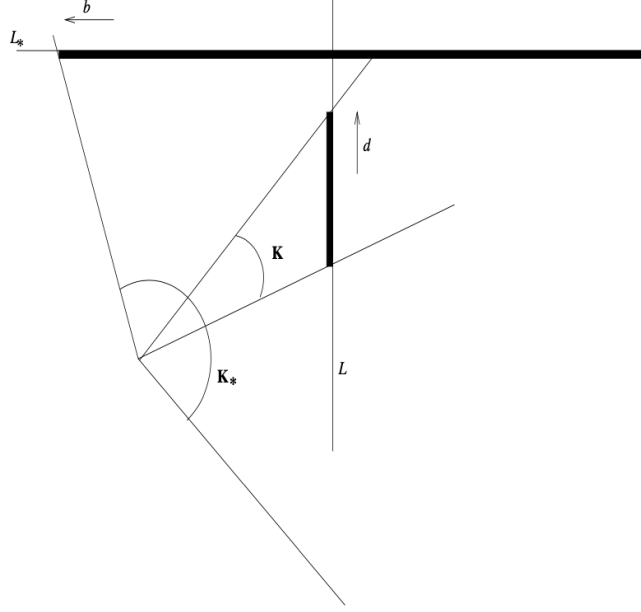


Figure 3: Primal-dual pair of conic problems
[bold: primal (vertical segment) and dual (horizontal ray) feasible sets]

5 Applications of Conic Duality

5.1 Self-Dual Cones

The cone K_* is the dual of the cone K , and if $K_* = K$, then the cone is called **self-dual**. Some self-dual cones are the following:

- The nonnegative orthant \mathbb{R}_+^m in \mathbb{R}^n
- The Lorentz (or the second-order, or the less scientific name the ice-cream) cone

$$L^m = \left\{ x = (x_1, \dots, x_{m-1}, x_m)^T \in \mathbb{R}^m : x_m \geq \sqrt{\sum_{i=1}^{m-1} x_i^2} \right\}$$

- The semidefinite cone S_+^m . This cone is in the space $E = S^m$ of $m \times m$ symmetric matrices (equipped with the Frobenius inner product $\langle A, B \rangle = \text{Tr}(A, B) = \sum_{ij} A_{ij}B_{ij}$) and consists of all $m \times m$ matrices A which are positive semidefinite, i.e.,

$$A = A^T; \quad x^T A x \geq 0 \quad \forall x \in \mathbb{R}^m$$

5.2 Second Order Cone Programming (SOCP)

In a second-order cone program (SOCP) a linear function is minimized over the intersection of an affine set and the product of second-order (quadratic) cones. SOCPs are nonlinear convex problems that include linear and (convex) quadratic programs as special cases, and arise in many engineering problems, such as filter design, antenna array weight design, truss design, robust estimation, and problems involving friction (e.g., robot grasp).

5.3 Semidefinite Programming (SDP)

Semidefinite programming (SDP) is a subfield of convex optimization concerned with the optimization of a linear objective function over the intersection of the cone of positive semidefinite matrices with an affine space, i.e., a spectrahedron.

Semidefinite programming is a relatively new field of optimization which is of growing interest for several reasons. Many practical problems in operations research and combinatorial optimization can be modeled or approximated as semidefinite programming problems. In automatic control theory, SDPs are used in the context of linear matrix inequalities. SDPs are in fact a special case of cone programming and can be efficiently solved by interior point methods. All linear programs can be expressed as SDPs, and via hierarchies of SDPs the solutions of polynomial optimization problems can be approximated. Semidefinite programming has been used in the optimization of complex systems. In recent years, some quantum query complexity problems have been formulated in terms of semidefinite programs.

5.4 Distributionally Robust Learning (DRL)

Distributionally Robust Learning (DRL) is necessary for building reliable machine learning systems. When machine learning is deployed in the real world, its performance can be significantly degraded because test data may follow a different distribution from training data. DRL explicitly considers the worst-case distribution shift by minimizing the adversarially reweighted training loss.

6 Is Something Wrong?

The issue at hand, is that the Conic Duality Theorem is much weaker than the Linear Duality Theorem. In LP, feasibility (even non-strict) and boundedness of either primal or dual problem, implies solvability of both the primal and the dual and equality between their optimal values. In CP, if one of (P) and (D) is essentially strictly feasible and bounded, then the other is solvable with their optimal values are equal. If both are essentially strictly feasible, then both are solvable and their optimal values are equal. As a result, it is clear that, a “word-by-word” extension of the LP Duality Theorem to the conic case is false.

7 Consequences

7.1 Sufficient Condition for Infeasibility

Recall that a necessary and sufficient condition for infeasibility of a (finite) system of scalar linear inequalities (i.e., for a vector inequality with respect to the partial ordering \geq) is the possibility to combine these inequalities in a linear fashion in such a way that the resulting scalar linear inequality is contradictory. In the case of cone-generated vector inequalities a slightly weaker result can be obtained:

Proposition (Conic Theorem on Alternative). *Consider a linear vector inequality:*

$$Ax - b \geq_{\kappa} 0 \quad (I)$$

i. *If there exists λ satisfying*

$$\lambda \geq_{\kappa^*} 0, A^* \lambda = 0, \langle \lambda, b \rangle > 0 \quad (II)$$

then (I) has no solutions.

ii. *If (II) has no solutions, then (I) is "almost solvable"- for every positive ϵ there exists b' such that $\|b' - b\|_2 < \epsilon$ and the perturbed system*

$$Ax - b' \geq_{\kappa} 0$$

is solvable.

iii. *(II) is solvable iff (I) is not "almost solvable".*

7.2 The Linear Inequality between a Scalar and a Linear Vector

We have the following linear vector inequality:

$$Ax \geq_{\kappa} b \quad (V)$$

and a scalar inequality:

$$c^T x \geq d \quad (S)$$

we want to check whether (S) is a consequence of (V). In the case of K being a nonnegative orthant, the answer is trivial, by the inhomogeneous Farkas Lemma:

The linear inequality of scalar (S) is a consequence of a feasible system of linear inequalities $Ax \geq b$ iff (S) can be obtained from linear vector (V) and the trivial $1 \geq 0$

Proposition. (i) If (S) can be obtained from (V) and from the trivial inequality $1 \geq 0$ by admissible aggregation, i.e., there exist weight vector $\lambda \geq_{\kappa_*} 0$ such that

$$A^* \lambda = c, \langle \lambda, b \rangle \geq d$$

then (S) is a consequence of (V) .

(ii) If (S) is a consequence of a strictly feasible linear vector inequality (V) , then (S) can be obtained from (V) by an admissible aggregation.

Note that the difference between the case of the partial ordering \geq and a general partial ordering \geq_{κ} is in the word “strictly” in (ii).

7.3 Robustness

In the general conic case we may meet “pathologies” which do not occur in LP. E.g., a feasible and bounded problem may be unsolvable, the dual to a solvable conic problem may be infeasible, etc. Where the pathologies come from? Looking at our “pathological examples”, we arrive at the following guess: the source of the pathologies is that in these examples, the “solvability status” of the primal problem is non-robust – it can be changed by small perturbations of the data. This issue of robustness is very important in modelling.

A question of primary importance is whether the properties of the program (CP) (feasibility, solvability, etc.) are stable with respect to perturbations of the data. The reasons which make this question important are the following:

- In actual applications, especially those arising in Engineering, the data are normally inexact: their true values, even when they “exist in the nature”, are not known exactly when the problem is processed. Consequently, the results of the processing say something definite about the “true” problem only if these results are robust with respect to small data perturbations i.e., the properties of (CP) we have discovered are shared not only by the particular (“nominal”) problem we were processing, but also by all problems with nearby data.
- Even when the exact data are available, we should take into account that processing them computationally we unavoidably add “noise” like rounding errors. As a result, a real-life computational routine can recognize only those properties of the input problem which are stable with respect to small perturbations of the data.

Due to the above reasons, we should study not only whether a given problem (CP) is feasible/bounded/solvable, etc., but also whether these properties are robust – remain unchanged under small data perturbations. As it turns out, the Conic Duality Theorem allows to recognize “robust feasibility/boundedness/solvability...”.

In order to define Robust Feasibility, we have to introduce the following relevant concepts of (CP):

- robust feasible, if all “sufficiently close” problems (i.e., those of the same structure (n, m, K) and with data close enough to those of (CP)) are feasible
- robust infeasible, if all sufficiently close problems are infeasible
- robust bounded below, if all sufficiently close problems are bounded below (i.e., their objectives are bounded below on their feasible sets)
- robust unbounded, if all sufficiently close problems are not bounded
- robust solvable, if all sufficiently close problems are solvable

Proposition (Robust Feasibility). *(CP) is robust feasible if and only if it is strictly feasible, in which case the dual problem (D) is robust bounded above.*

Proposition (Robust Infeasibility). *(CP) is robust infeasible if and only if the system*

$$\langle b, \lambda \rangle = 1, A^* \lambda = 0, \lambda \geq_{\kappa_*} 0$$

is robust feasible, or, which is the same (by Robust feasibility), if and only if the system

$$\langle b, \lambda \rangle = 1, A^* \lambda = 0, \lambda \geq_{\kappa_*} 0$$

has a solution.

Proposition (Robust Solvability). *For a conic problem (CP) the following conditions are equivalent to each other:*

- (i) (CP) is robust feasible and robust bounded (below)*
- (ii) (CP) is robust solvable*
- (iii) (D) is robust solvable*
- (iv) (D) is robust feasible and robust bounded (above)*
- (v) Both (CP) and (D) are strictly feasible*

In particular, under every one of these equivalent assumptions, both (P) and (D) are solvable with equal optimal values.

8 Bibliography

- [1] Ben-Tal, A. and Nemirovskii, A. Lectures on Modern Convex Optimization. pp.52-71. (2013).
- [2] O.Guler. Foundations of Optimization. Graduate Texts in Mathematics. Springer New York, (2010).
- [3] David Williamson. Orie6300: Mathematical programming I, lecture 26. <http://people.orie.cornell.edu/dpw/orie6300/Lectures/lec26.pdf>.
- [4] B. Gartner and Matousek J. Approximation algorithms and semidefinite programming: Cone programming.

<http://www.ti.inf.ethz.ch/ew/lehre/ApproxSDP09/notes/conelp.pdf>
- [5] Weihua Hu et. al, Does Distributionally Robust Supervised Learning Give Robust Classifiers? (2018) arXiv:1611.02041v6
- [6] Yinyu Ye. Conic Duality Theorems and Applications. Department of Management Science and Engineering, Stanford University. Chapters 4.1-4.2 and 6.1-6.4 .
- [7] Semidefinite Programming. Wikipedia. shorturl.at/sBMY5
- [8] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, Applications of Second-Order Cone Programming. shorturl.at/brH01