From Linear Optimization to Conic Optimization

Kees Roos
Technische Universiteit Delft
Faculteit Electrotechniek. Wiskunde en Informatica
e-mail: C.Roos@ewi.tudelft.nl
URL: http://www.isa.ewi.tudelft.nl/~roos

NMC 2004
Tilburg
April 17, A.D. 2004
Outline

• Milestones in the history of Linear Optimization (LO)
• Duality via Farkas’ Lemma
• Nonlinear problems
• Cone underlying the inequalities in LO
• Three ’good’ cones
• Conic optimization problems
• Examples
• Dual of a conic optimization problem
• Conic duality theorem
• Example: nonnegativity of polynomials
• Remembering Jos Sturmi
• Concluding remarks
• Some references
Milestones in the history of linear optimization (LO)

**LO problem**

\[(P) \quad \min \left\{ c^T x : Ax \geq b \right\}, \quad A \text{ is } m \times n \]

- **1947**: Simplex Method (Dantzig)
- **1955**: Logarithmic Barrier Method (Frisch)
- **1967**: Affine Scaling Method (Dikin)
- **1967**: Center Method (Huard)
- **1968**: Barrier Methods (Fiacco, McCormick)
- **1972**: Exponential example (Klee and Minty)
- **1979**: Ellipsoid Method (Khachiyan)
- **1984**: Projective Method (Karmarkar) \(\Rightarrow\) Interior Point Methods for
  - Convex Optimization (1989)
  - Semidefinite Optimization (1994)
  - Second Order Cone Optimization (1994)
  - Discrete Optimization (1994)
Some keyplayers

[Images of historical figures and modern group members]
Farkas’ lemma (1904)

Let \( a, c \in \mathbb{R}^n \) and \( b, d \in \mathbb{R} \). Then it can easily be seen that the implication

\[
a^T x \geq b \implies c^T x \geq d
\]

holds if and only if

\[
\exists \lambda \in \mathbb{R} : \lambda a = c, \quad \lambda b \geq d, \quad \lambda \geq 0.
\]

In other words, the inequality \( c^T x \geq d \) is a consequence of the inequality \( a^T x \geq b \) if and only if (2) holds.

Farkas’ lemma generalizes this property. If \( A \) is an \( n \times m \) matrix, \( b \in \mathbb{R}^m \), \( c \in \mathbb{R}^n \) and \( d \in \mathbb{R} \) then it says that the implication

\[
A^T x \geq b \implies c^T x \geq d
\]

holds if and only if

\[
\exists \lambda \in \mathbb{R}^m : A\lambda = c, \quad \lambda^T b \geq d, \quad \lambda \geq 0.
\]

Remark: Note that if \( A^T x \geq b \) then for every \( \lambda \in \mathbb{R}^m_+ \) one has the inequality

\[
\lambda^T A^T x = (A\lambda)^T x \geq \lambda^T b.
\]

is a consequence of \( A^T x \geq b \). Farkas’ lemma essentially says that the converse is also true: every implied inequality can be obtained in this way.
**Question:** When a real \( d \) is a lower bound for the optimal value of the LO problem

\[
(P) \quad \min \{ c^T x : A x \geq b \}?
\]

**Answer:** This is the case if and only if \( A x \geq b \) implies \( c^T x \geq d \), i.e.,

\[
A x \geq b \quad \Rightarrow \quad c^T x \geq d
\]

By Farkas’ lemma, this holds if and only if there exists a \( y \) such that

\[
\begin{align*}
 b^T y & \geq d \\
 A^T y & = c \\
 y & \geq 0
\end{align*}
\]

Hence, whenever \( d \) is such that the latter system is solvable, \( d \) is a lower bound on the optimal value in \( (P) \), and vice versa. The largest lower bound on the optimal value in \( (P) \) clearly is the optimal value of the LO problem

\[
(D) \quad \max \{ b^T y : A^T y = c, \ y \geq 0 \}.
\]

called the problem dual to \( (P) \).
Duality theorem for linear optimization

\[(P) \quad \min \left\{ c^T x : Ax \geq b \right\}\]
\[(D) \quad \max \left\{ b^T y : A^T y = c, \quad y \geq 0 \right\}.\]

1. The value of the dual objective at every dual feasible solution is \(\leq\) the value of the primal objective at every primal feasible solution (weak duality).

2. The following 5 properties are equivalent to each other:

   (i) The primal is feasible and below bounded.
   (ii) The dual is feasible and above bounded.
   (iii) The primal is solvable.
   (iv) The dual is solvable.
   (v) Both primal and dual are feasible.

   Strong duality: whenever \((i) \equiv (ii) \equiv (iii) \equiv (iv) \equiv (v)\) is the case, the optimal values in the primal and the dual problems are equal to each other (strong duality):
   \[\text{Opt}(P) = \text{Opt}(D).\]

3. The duality is symmetric: the problem dual to the dual is equivalent to the primal.
Optimality conditions

\((P)\) \quad \min \left\{ c^T x : A x \geq b \right\}

\((D)\) \quad \max \left\{ b^T y : A^T y = c, \quad y \geq 0 \right\}.

Let \((x, y)\) be a pair of primal and dual feasible solutions. The pair consists of optimal solutions of the two problems if and only if

\[ c^T x - b^T y = 0 \quad \text{[vanishing duality gap]} \]

as well as if and only if

\[ y_i [Ax - b]_i = 0, \quad i = 1, \ldots, m, \quad \text{[complementary slackness]} \]

**Proof:** Since \((P)\) and \((D)\) are feasible, they are solvable with equal optimal values, hence for primal-dual feasible \((x, y)\)

\[ \text{Duality gap}(x, y) = \underbrace{c^T x - \text{Opt}(P)}_{\geq 0} + \underbrace{\text{Opt}(D) - b^T y}_{\geq 0} \]

Moreover, since \(\text{Opt}(P) = \text{Opt}(D)\), we have

\[ \text{Duality gap}(x, y) = c^T x - b^T y = (A^T y)^T x - b^T y = [Ax - b]^T y \]

\[ \Downarrow \]

\[ c^T x - b^T y = 0 \quad \Leftrightarrow \quad \underbrace{[Ax - b]^T y}_{\geq 0} = 0 \quad \Leftrightarrow \quad y_i [Ax - b]_i = 0 \quad \forall i. \]
From linear to nonlinear optimization

When passing from a generic LP problem

$$(P) \quad \min \left\{ c^T x : Ax \geq b \right\}$$

to its nonlinear extensions, some components of the problem become nonlinear.

The traditional way is to allow nonlinearity of the objective and the constraints:

$c^T x \mapsto c(x), \quad a_i^T x - b_i \mapsto a_i(x)$

and to preserve the “coordinate-wise” interpretation of the vector inequality $A(x) \geq 0$:

$$A(x) \equiv \begin{bmatrix} a_1(x) \\ a_2(x) \\ \vdots \\ a_m(x) \end{bmatrix} \geq 0 \iff a_i(x) \geq 0, \quad i = 1, \ldots, m.$$

An (equivalent) alternative is to preserve the linearity of the objective and the constraint functions and to modify the interpretation of the vector inequality "$\geq$"

We prefer the second option, due to its strong “unifying abilities”: it turns out that a lot of quite diverse nonlinear optimization problems can be covered by just 3 “standard” types of vector inequalities.
### Examples of nonlinear problems

\[
\sum_{\ell=1}^{n} x_\ell^2 \rightarrow \min
\]

\[(a)\]
\[
\|Px - p\|_2 \leq c^T x + d;
\]

\[(b)\]
\[
\frac{x_{\ell+1}}{x_\ell} \leq e_\ell^T x + f_\ell, \quad \ell = 1, \ldots, n;
\]

\[(c)\]
\[
\frac{x_{\ell+3}}{x_{\ell+1}} \geq g_\ell^T x + h_\ell, \quad i = 1, \ldots, n - 1;
\]

\[(d)\]
\[
\begin{array}{cccccc}
  x_1 & x_2 & x_3 & \cdots & x_n \\
  x_2 & x_1 & x_2 & \cdots & x_{n-1} \\
  x_3 & x_2 & x_1 & \cdots & x_{n-2} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  x_n & x_{n-1} & x_{n-2} & \cdots & x_1
\end{array}
\geq 1;
\]

\[(e)\]
\[
1 \leq \sum_{\ell=1}^{n} x_\ell \cos(\ell \omega) \leq 1 + \sin^2(5\omega), \quad \forall \omega \in \left[-\frac{\pi}{7}, 1.3\right]
\]

Each of these problems can be converted, in a systematic way, into an equivalent problem

\[
\min \left\{ c^T x : Ax \succeq b \right\}
\]

"\succeq" being one of our 3 standard vector inequalities, so that seemingly highly diverse constraints of the original problem allow a unified treatment.
General conic optimization problems

A general conic optimization problem is a problem in the conic form

$$\min_{x \in \mathbb{R}^n} \left\{ c^T x : Ax - b \in K \right\},$$

where $K$ is a closed convex pointed cone. Examples of such cones are

- the nonnegative orthant:
  $$\mathbb{R}_+^m = \{ x \in \mathbb{R}^m : x \geq 0 \}.$$

- the Lorentz (or second order, or ice-cream) cone:
  $$L^m = \left\{ x \in \mathbb{R}^m : x_m \geq \sqrt{m-1} \sum_{i=1}^m x_i^2 \right\}.$$

- the semidefinite cone:
  $$S_+^m = \left\{ A \in \mathbb{R}^{m \times m} : A = A^T, x^T A x \geq 0, \forall x \in \mathbb{R}^m \right\}.$$

- a direct product of such cones.

In all these cases conic optimization problems can be solved efficiently by an interior-point method.
Examples of conic optimization problems

Let $A$ be any $m \times n$ matrix and $b \in \mathbb{R}^m$. The least squares problem

$$\min_x ||Ax - b||$$

can be modelled as a second order cone problem:

$$\min_x \{\tau : ||Ax - b|| \leq \tau\} \equiv \min_x \left\{\tau : \begin{bmatrix} Ax - b \\ \tau \end{bmatrix} \succeq_{Lm+1} 0 \right\}$$

Finding the smallest eigenvalue $\lambda_{\min}$ of a symmetric matrix $M$, of size $n \times n$, can be modelled as semidefinite problem:

$$\max \left\{\lambda : M - \lambda I \succeq_{S^n} 0 \right\}.$$  

This can be understood as follows. If $\lambda_{\min} = \lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of $M$ then the eigenvalues of $M - \lambda I$ are $\lambda_1 - \lambda \leq \ldots \leq \lambda_n - \lambda$. Hence $M - \lambda I$ is positive semidefinite if and only if $\lambda_{\min} \geq \lambda$. 

Optimization Group
Every nonlinear optimization problem is a conic optimization problem

Suppose we want to minimize a (nonlinear) function \( f(x) \) over an arbitrary domain \( X \subseteq \mathbb{R}^n \). Then we can equivalently solve

\[
\min \{ t : f(x) \leq t, \, x \in X \},
\]

which has a linear objective function. Thus we may assume that the objective function is linear, i.e., \( f(x) = c^Tx \) for some vector \( c \). The optimal value is the same as for the problem

\[
\min \{ c^Tx : x \in \bar{X} \},
\]

where \( \bar{X} \) is the convex hull of \( X \). Now, since \( \bar{X} \) is convex, the set

\[
\mathcal{K} = \{(t, tx) : t \geq 0, \, x \in \bar{X} \}
\]

is a convex cone. Defining \( \bar{c} = (0, c^T)^T \), the last problem can be reformulated as

\[
\min \{ \bar{c}^T z : z \in \mathcal{K}, \, z_1 = 1 \},
\]

which is a conic optimization problem. Whether this problem can be solved efficiently depends on two questions:

- is the convex hull of the domain \( X \) computationally tractable?
- has the cone \( \mathcal{K} \) a barrier function that is computationally tractable?
Dual cone

Notation: $a \in \mathcal{K}$ is also denoted as $a \geq_{\mathcal{K}} 0$ and $a \geq_{\mathcal{K}} b$ means $a - b \geq_{\mathcal{K}} 0$.

A basic question is for which $\lambda \in \mathbb{R}^n$ the implication

$$a \geq_{\mathcal{K}} b \Rightarrow \lambda^T a \geq \lambda^T b$$

always holds. Obviously, this holds if and only if $a \geq_{\mathcal{K}} 0$ implies $\lambda^T a \geq 0$, or if

$$\lambda \in \mathcal{K}_* := \left\{ y \in \mathbb{R}^n : y^T x \geq 0, \ \forall x \in \mathcal{K} \right\}.$$

**Theorem 1** If $\mathcal{K}$ is a closed convex pointed cone with nonempty interior, then so is the dual cone $\mathcal{K}_*$, and then the duality is symmetric: $(\mathcal{K}_*)_* = \mathcal{K}$.

**Remark:** The three 'standard' cones $\mathbb{R}_+^n$, $\mathbb{L}_m^m$ and $\mathbb{S}_+^M$ are self-dual!
Conic Farkas’ lemma

Let $\mathcal{K}$ be a closed convex pointed cone with nonempty interior. Then the implication

$$A^T x \geq_{\mathcal{K}} b \Rightarrow c^T x \geq d$$

(5)

holds if

$$\exists \lambda \in \mathcal{K}^* : A\lambda = c, \quad \lambda^T b \geq d.$$  

(6)

Proof: If $A^T x \geq_{\mathcal{K}} b$ then for every $\lambda \in \mathcal{K}^*$ one has the inequality

$$\lambda^T A^T x = (A\lambda)^T x \geq \lambda^T b.$$  

Hence, if $A\lambda = c$ and $\lambda^T b \geq d$ then $c^T x \geq \lambda^T b \geq d$, proving that (6) implies (5).

In general (5) does not imply (6).

The Conic Farkas’ lemma says that (5) does imply (6) if $A^T x \geq_{\mathcal{K}} b$ is strictly feasible, i.e., if $A^T x >_{\mathcal{K}} b$ has a solution.
Dual problem of a conic optimization problem

**Question:** When a real $d$ is a lower bound for the optimal value of the conic problem

$$
(CP) \quad \min \left\{ c^T x : Ax \succeq_K b \right\}?
$$

**Answer:** This is the case if and only if $Ax \succeq b$ implies $c^T x \succeq d$, i.e.,

$$Ax \succeq_K b \implies c^T x \succeq d$$

Assuming that $Ax \succeq_K b$ is strictly feasible, by the conic Farkas’ lemma this holds if and only if there exists a $y$ such that

$$b^T y \succeq d$$

$$A^T y = c$$

$$y \succeq_K 0$$

Hence, whenever $d$ is such that the latter system is solvable, $d$ is a lower bound on the optimal value in $(CP)$, and vice versa. The largest lower bound on the optimal value in $(CP)$ clearly is the optimal value of the conic problem

$$
(CD) \quad \max \left\{ b^T y : A^T y = c, \quad y \succeq_K 0 \right\}.
$$

$(CD)$ is called the dual problem of $(CP)$. 

Optimization Group

TU Delft
Duality theorem for conic optimization

\[(CP) \quad \min \left\{ c^T x : Ax \succeq_K b \right\} \]
\[(CD) \quad \max \left\{ b^T y : A^T y = c, \quad y \succeq_K 0 \right\}. \]

1. The value of the dual objective at every dual feasible solution is \( \leq \) the value of the primal objective at every primal feasible solution (weak duality).

2. The following 2 properties are equivalent to each other:
   (i) The primal is strictly feasible and below bounded.
   (ii) The dual is solvable.

3. The following 2 properties are equivalent to each other:
   (iii) The dual is strictly feasible and below bounded.
   (iv) The primal is solvable.

   Strong duality: whenever (i) \( \equiv \) (ii) or (iii) \( \equiv \) (iv) is the case, the optimal values in the primal and the dual problems are equal to each other (strong duality):

   \[ \text{Opt}(CP) = \text{Opt}(CD). \]

4. The duality is symmetric: the problem dual to the dual is equivalent to the primal.
Pathological examples

1. \((P)\) may be strictly feasible, below bounded and at the same unsolvable:

\[
\min \begin{cases} 
  x_1 : \begin{bmatrix} x_1 - x_2 \\ 1 \\ x_1 + x_2 \end{bmatrix} \geq \begin{bmatrix} L \end{bmatrix} 0 
\end{cases} \ \Leftrightarrow \ \begin{cases} 
  x_1 \rightarrow \min \\
  4x_1x_2 \geq 1, \\
  x_1 + x_2 > 0
\end{cases}
\]

2. \((P)\) may be solvable, while \((D)\) may be infeasible:

\[
\min \begin{cases} 
  x_2 : \begin{bmatrix} x_1 \\ x_2 \\ x_1 \end{bmatrix} \geq \begin{bmatrix} L \end{bmatrix} 0 
\end{cases} \ \Leftrightarrow \ \begin{cases} 
  x_2 \rightarrow \min \\
  x_2 = 0, \\
  x_1 \geq 0
\end{cases}
\]

3. \((P)\) and \((D)\) may both be solvable with distinct optimal values:

\[
\min \begin{cases} 
  x_2 : \begin{bmatrix} 1 + x_2 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \geq \begin{bmatrix} S_3 \end{bmatrix} 0 
\end{cases} \ \Leftrightarrow \ \begin{cases} 
  x_2 \rightarrow \min \\
  x_1 \geq 0, \\
  x_2 = 0
\end{cases}
\]

and its dual are solvable, but \(\text{Opt}(P) = 0 > -1 = \text{Opt}(D)\).
Nonnegativity of polynomials

It is well known that
\[ ax^2 + bx + c \geq 0, \quad \forall x \in \mathbb{R} \]
if and only if
\[ a \geq 0, \quad c \geq 0, \quad b^2 - 4ac \leq 0. \]

This holds if and only if the matrix
\[
X = \begin{bmatrix}
    c & \frac{b}{2} \\
    \frac{b}{2} & a
\end{bmatrix}
\]
is positive semidefinite. This result can be generalized to polynomials of arbitrary (even!) degree. One has
\[ p(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + \ldots + p_{2m} x^{2m}, \quad \forall x \in \mathbb{R} \]
if and only if there exists a positive semidefinite matrix \( X \) of size \( (2m + 1) \times (2m + 1) \) such that the sums of the respective opposite diagonals are equal to the coefficients \( p_i \) of \( p(x) \).

We therefore say: nonnegativity of a polynomial in one variable is semidefinite representable.
Suppose we want to find the minimum value $z^*$ of a polynomial $p(x)$:

$$z^* = \min \{ p(x) : x \in \mathbb{R} \}.$$  

Then $z^*$ is the largest real number for which the polynomial $p(x) - z^*$ is nonnegative. Thus we have

$$z^* = \max \{ z : p(x) - z \geq 0, \quad \forall x \in \mathbb{R} \}.$$  

The condition $p(x) - z \geq 0$, $\forall x \in \mathbb{R}$ is semidefinite representable, and hence we can solve this problem efficiently.
Example: hidden convexity

Graph of $p(x) = 48 - 28x - 56x^2 + 35x^3 + 7x^4 - 7x^5 + x^6$.

A famous example is due to Theodore S. Motzkin (1908-1970). He showed that for $n \geq 3$ polynomials of the form
\[
\left( x_{1}^{2} + \ldots + x_{n-1}^{2} - nx_{n} \right) x_{1}^{2} \ldots x_{n-1}^{2} + x_{n}^{2n}
\]
are nonnegative (everywhere), but are not a sum of squares (SOS) of (real) polynomials. For $n = 3$ this gives the celebrated example
\[
M(x, y, z) = z^{6} + x^{4}y^{2} + x^{2}y^{4} - 3x^{2}y^{2}z^{2}.
\]
Nonnegativity follows from the arithmetic-geometric means inequality:
\[
\frac{z^{6} + x^{4}y^{2} + x^{2}y^{4}}{3} \geq \sqrt[3]{x^{6}y^{6}z^{6}} = x^{2}y^{2}z^{2}.
\]
$M(x, y)$ is not a SOS (see next slide).

However, one may verify that $M(x, y)$ can be decomposed as a SOS of homogeneous forms:
\[
\left( \frac{(x^{2} - y^{2})z^{3}}{x^{2} + y^{2}} \right)^{2} + \left( \frac{x^{2}y(x^{2} + y^{2} - 2z^{2})}{x^{2} + y^{2}} \right)^{2} + \left( \frac{xy^{2}(x^{2} + y^{2} - 2z^{2})}{x^{2} + y^{2}} \right)^{2} + \left( \frac{xyz(x^{2} + y^{2} - 2z^{2})}{x^{2} + y^{2}} \right)^{2},
\]
yielding another proof of nonnegativity. The four terms are squares of homogeneous rational functions.
Proof that Motzkin’s example is not a SOS

Suppose \( M(x, y, z) = \sum_k h_k(x, y, z)^2 \). Writing out \( h_k(x, y, z) \), utilizing the symmetric scheme

\[ A_k x^3 + B_k x^2 y + C_k xy^2 + D_k y^3 + E_k x^2 z + F_k xyz + G_k y^2 z + H_k x z^2 + I_k y z^2 + J_k z^3, \]

and equating coefficients, it follows that

\[ A_k = E_k = H_k = D_k = G_k = I_k = 0. \]

Using also \( M(1 \pm 1, \pm 1) = 0 \) we obtain

\[ B_k + C_k + F_k + J_k = B_k + C_k - F_k - J_k = -B_k + C_k - F_k + J_k = -B_k + C_k + F_k - J_k = 0, \]

whence

\[ B_k = C_k = F_k = J_k = 0. \]

Therefore \( h_k(x, y, z) = 0 \), a contradiction.
The 17th problem of Hilbert reads:

“ob nicht jede definite Form als Quotient von Summen von Formenquadraten dargestellt werden kann”.

Here ’Form’ means: (homogeneous) quotient of polynomials (in one or more variables). Hilbert proved the conjecture for forms of degree $\leq 3$. Emil Artin proved the conjecture in full, in 1927; his proof was non-constructive.

De Klerk en Pasechnik (TUD, 2002) gave a constructive proof for the case of three variables. Using semidefinite optimization they found an efficient method to find the desired composition as a sum of squares.

For those who are interested:
HPOPT 2004 - 8th International Workshop on High Performance Optimization Techniques: Optimization and Polynomials (including a tutorial on Sums of Squares in Optimization)

Semidefinite optimization has a lot of other new and important applications, among these so-called robust engineering design.
‘It was extraordinary to see how Jos had developed his style and personality, starting from being a highly clever but rather clumsy young boy, all the way to a full-rounded, matured, amiable, yet authoritative figure. Given time, we should have no doubt that he would become a superstar. It is beyond words to describe the sadness that this is no longer possible. (. . .) Losing Jos is an incredible pain: a pain that has just started to feel and will last for a long long time.’

Shuzhong Zhang, Hong Kong
Concluding remarks

- Symmetric cones are closely related to Jordan algebras: a symmetric cone is just the set of squares in some Euclidean Jordan algebra (Faraut-Koranyi, 1994).
- The last decade gave rise to a revolution in algorithms and software for linear, convex and semidefinite optimization. For many users of such software, SeDuMi is the favorite package.
- Cone optimization unifies a wide variety of optimization problems. Moreover, they can be solved efficiently. This opens the way to many new applications, including applications which could not be handled some years ago.
- Since 1995, semidefinite relaxations of combinatorial problems have led to numerous improved approximation algorithms for combinatorial optimization problems, like, e.g., MAX CUT, MAX SAT, MAX 2SAT, MAX 3SAT, MAX 4SAT, MAX $k$-CUT, $k$-coloring, scheduling, etc.
Some references


