DUALITY FOR SEMIDEFINITE AND CONVEX PROGRAMMING

Henry Wolkowicz
University of Waterloo

(work with: Serge Kruk, Levent Tuncel, Motakuri Ramana, Jon Borwein)

DUALITY
FOR
SEMIDEFINITE AND CONVEX
PROGRAMMING

HISTORY

From various sources, e.g.

article by Harold Kuhn in:
NLP SIAM-AMS Proceedings Volume IX, 1976.

The Lagrange multiplier method leads to several transformations which are important both theoretically and practically.

By means of these transformations new problems equivalent to a given problem can be so formulated that stationary conditions occur simultaneously in equivalent problems. In this way we are led to transformations of the problems which are important because of their symmetric character. Moreover, for a given maximum problem with maximum M, we shall often be able to find an equivalent minimum problem with the same value M as minimum; this is a useful tool for bounding M from above and below. (Courant and Hilbert)

Von Neumann in 1940's had a duality for LPs which was based on using game theory (see collected works);

it formed the basis for Dantzig's simplex method

Later: Gale, Kuhn, Tucker developed a duality theory for general problems

Problem 1:

Given three points in the plane, find a fourth point such that the sum of its distances to the three given points is minimum. (17th Century Fermat Problem)

Problem 2:

In the three Sides of an equiangular Field stand three Trees, at the Distances of 10, 12, and 16 Chains from one another: To find the Content of the Field, it being the greatest the Data will admit of? (The Ladies Diary or Woman's Almanack (1755) problem posed by Mr Tho. Moss (pg 47))

Given any triangle, circumscribe the largest possible equilateral triangle about it. (Annales de mathematiques Pures et Appliquees, edited by J.D. Gergonne, Vol I (1810-11), Problem posed on Page 384)

Solution:

Thus the largest equilateral triangle circumscribing a given triangle has sides perpendicular to the lines joining the vertices of the given triangle to the point such that the sum of the distances to these vertices is a minimum.
One can conclude that the altitude of the largest equilateral triangle that can be circumscribed about a given triangle is equal to the sum of distances from the vertices of the given triangle to the point at which the sum of distances is a minimum. (Rochat, Vecten, Fauguier, and Pilatte in vol II (1811-12) page 88-93)

More recent references:

Duffin paper, 1956

Convex Programming:
Rockafellar books: (i)-Convex Analysis, 1969; (ii)-Conjugate Duality and Optimization, 1974

Semi-infinite Programming:
Perfect Duality and various duality states (Ben-Israel, Charnes, Kortanek, 1962,1977)

Monique Guignard, generalized optimality conditions for abstract programming, 1969

BBZ optimality conditions without a CQ, 1977

Borwein and W., Cone programming without CQ, 1979

Standard Convex Programming Problem:

$$\min\{f(x) : f^k(x) \leq 0, k \in \mathcal{P} := \{1, \ldots, m\}\}$$

$\mathcal{P}^=$ implicit equality constraints

$\mathcal{P}^<$ complement of implicit equality constraints

$D^=_{\mathcal{P}^=}(x)$ cone of directions of constancy at x

f is *faithfully convex* if: f is affine on a line segment only if it is affine on the whole line containing that segment. (e.g. analytic convex, strictly convex), In which case

$f(x) = h(Ax + b) + a^t x + \alpha$, h strictly convex

$D_f^{=}$ is null space of A intersect null space of a, independent of x

$$K^{+} = \{\phi : \phi \cdot x \geq 0,\ \forall x \in K\}$$

K, L closed convex cones, then

$$(K \cap L)^{+} = \mathrm{cl}(K^{+} + L^{+})$$

without closure if $\mathrm{int}(K) \cap L \neq \emptyset$

tangent cone of set M at point x is
$T(M,x) = \text{cl}(\text{cone}(M-x))$

linearizing cone of an active set of constraints $\Omega \subset \mathcal{P}$ is $C_\Omega(x)$ and it equals $(-\text{cone(gradients)})^+$, $-B_\Omega(x)^+$ (by Farkas Lemma).

Various ways to derive optimality conditions, e.g.

For $\min_{x \in S} f(x)$, $x \in S$ is optimal iff

$$\partial f(x) \cap T^+(S, x) \neq \emptyset.$$

For S feasible set from constraints f_k
translate the tangent cone to the linearizing cone
get GEOMETRIC *weakest constraint qualification*

$$T(S, x) = C_{active}(x)$$

$$T(S,x) = cl(conv D^{=}_{\overline{\mathcal{P}^{=}}}(x)) \cap C_{active}(x)$$

$$T^{+}(S,x) = (D^{=}_{\overline{\mathcal{P}^{=}}}(x))^{+} - B_{active}(x)$$

missing set was added on in order to complete the optimality conditions

$$\nabla f(x) + \sum_k \lambda_k \nabla f_k(x) \in K^{+}$$

Primal approach: split constraints into two parts

$$\min\{f(x) : f_k(x) \leq 0,\ k \in \mathcal{P}^{<}, x \in T\}$$

$$T = \{x : f_k(x) = 0,\ k \in \mathcal{P}^{=}\}$$

apply generalized Slater condition with definition of cone of directions of constancy

---

equivalently $F(x) \preceq_{\mathcal{R}^f} 0$, i.e. change nonnegative orthant to *minimal face* of nonnegative orthant. AND add set constraint

$x \in F^{-1}(span\ (\mathcal{R}^f)) = F^{-1}(\mathcal{R}^f - \mathcal{R}^f) = F^{-1}(\mathcal{R}^f - \Re^m_+)$

Cone programming: S a closed convex cone induces a linear partial order (Borwein & W 1980)

$$\min\{f(x) : g(x) \preceq_S 0,\ x \in \Omega\}$$

feasible set A

---

rewrite using minimal face $g(A) \subset -S^f$

$$\min\{f(x) : g(x) \preceq_{S^f} 0,\ x \in \Omega^f\}$$

where $\Omega^f = \Omega \cap g^{-1}(S^f - S^f) = \Omega \cap g^{-1}(S^f - S)$

---

optimality conditions for some $\Lambda \in (S^f)^+$:

$$f(x) + \Lambda g(x) \geq \mu^*,\ \forall x \in \Omega^f$$

Special case of linear cone programming, K, L closed convex cones (W81):

primal:

$$\mu^* = \text{min}\{cx : Ax \succeq_K 0,\ x \succeq_L 0\}$$

minimal cones K^f, L^f

dual:

$$\mu^* = \nu^* = \text{max}\{by : A^* y \preceq_{(L^f)^+} b,\ y \succeq_{(K^f)^+} 0\}$$

Linear Programming ($A : \Re^m \to \Re^n$)

$\Re^n_+$ is a closed convex cone

$$
\begin{array}{llll}
 & p^* = & \sup & c^t x \\
(\mathbf{P}) & & \text{s.t.} & Ax \preceq b \quad (b - Ax \in \Re^n_+) \\
 & & & x \in \Re^m
\end{array}
$$

Lagrangian (payoff function):
$L(x, U) = c^t x + \langle U, b - Ax \rangle$

$$p^* = \max_x \min_{U \succeq 0} L(x, U)$$

(the constraint $U \succeq 0$ is needed to recover the hidden constraint $Ax \preceq b$.)

The dual is obtained from the optimal strategy of the competing player

$$p^* \leq d^* = \min_{U \succeq 0} \max_x L(x,U) = \langle U, b \rangle + x^t(c - A^*U)$$

The hidden constraint $c - A^*U = 0$ yields the dual

$$\textbf{(D)} \quad \begin{array}{rll} d^* = & \inf & \operatorname{trace} bU \\ & \text{s.t.} & A^*U = c \\ & & U \succeq 0. \end{array}$$

for the primal

$$\textbf{(P)} \quad \begin{array}{rll} p^* = & \sup & c^t x \\ & \text{s.t.} & Ax \preceq b \\ & & x \in \Re^m \end{array}$$

If Slater's condition fails for the primal LP, then there are an infinite number of different dual programs.

The implicit equality constraints are:

$$A_e x = b_e$$

where $A = \begin{bmatrix} A_l \\ A_e \end{bmatrix}$

$$\textbf{(D)} \quad \begin{array}{rll} d^* = & \inf & \text{trace}\, bU \\ & \text{s.t.} & A_l^* U_l + A_e^* U_e = c \\ & & U \in \mathcal{U} \\ & & \{U : U \succeq 0\} \subset \mathcal{U} \\ & & \mathcal{U} \subset \{U : U_l \succeq 0, U_e \text{ free}\}. \end{array}$$

for the equivalent primal program

$$\textbf{(P)} \quad \begin{array}{rll} p^* = & \sup & c^t x \\ & \text{s.t.} & A_l x \preceq b_l \\ & & A_e x = b_e \\ & & x \in \Re^m \end{array}$$

DUALITY THEOREM

1. If one of the problems is inconsistent, then the other is inconsistent or unbounded.

2. **WEAK DUALITY**
   Let the two problems be consistent, and let x^0 be a feasible solution for P and U^0 be a feasible solution for D. Then

$$c^t x^0 \leq \left\langle b, U^0 \right\rangle .$$

3. **STRONG DUALITY**
   If both P and D are consistent, then they have optimal solutions and their optimal values are equal.

4. **COMPLEMENTARY SLACKNESS**
   Let x^0 and U^0 be feasible solutions of P and D, respectively. Then x^0 and U^0 are optimal if and only if

$$\left\langle U^0, (b - Ax^0) \right\rangle = 0.$$

   if and only if

$$U^0 \circ (b - Ax^0) = 0.$$

5. **SADDLE POINT**
   The vectors x^0, U^0 are optimal solutions of P and D, respectively, if and only if (x^0, U^0) is a saddle point of the Lagrangian $L(x, U)$ for all (x,U),

$$L(x, U^0) \leq L(x^0, U^0) \leq L(x^0, U)$$

   and then

$$L(x^0, U^0) = c^t x^0 = \left\langle b, U^0 \right\rangle.$$

Characterization of optimality for the dual pair x, U

$$Ax \preceq b \qquad \text{primal feasibility}$$

$$A^*U = c \qquad \text{dual feasibility}$$

$$U \circ (Ax - b) = 0e \quad \text{complementary slackness}$$

$$U \circ (Ax - b) = \mu e \qquad \text{perturbed}$$

Forms the basis for:

primal simplex method
dual simplex method
interior point methods

What is SEMIDEFINITE PROGRAMMING?

Why use it?

Quadratic approximations are better than linear approximations. And, we can solve relaxations of quadratic approximations efficiently using semidefinite programming.

How does SDP arise from quadratic approximations?

Let

$$q_i(y) = \frac{1}{2} y^t Q_i y + y^t b_i + c_i, \;\; y \in \Re^n$$

$$(\mathbf{QQP}) \quad \begin{array}{rll} q^* = & \min & q_0(y) \\ & \text{s.t.} & q_i(y) = 0 \\ & & i = 1, \ldots m \end{array}$$

Lagrangian:

$$\begin{array}{rcl} L(y,x) & = & \frac{1}{2} y^t (Q_0 - \sum_{i=1}^m x_i Q_i) y \\ & & + y^t (b_0 - \sum_{i=1}^m x_i b_i) \\ & & + (c_0 - \sum_{i=1}^m x_i c_i) \end{array}$$

$$q^* = \min_y \max_x L(y,x) \geq d^* = \max_x \min_y L(y,x).$$

homogenize

$$y_0 y^t (b_0 - \sum_{i=1}^m x_i b_i), \quad y_0^2 = 1.$$

$$
\begin{array}{rcll}
d^* & = & \max_x \min_y & L(y,x) \\
& = & \max_x \min\limits_{y_0^2=1} & \frac{1}{2} y^t (Q_0 - \sum_{i=1}^m x_i Q_i) y \quad (+t y_0^2) \\
& & & + y_0 y^t (b_0 - \sum_{i=1}^m x_i b_i) \\
& & & + (c_0 - \sum_{i=1}^m x_i c_i) \quad (-t)
\end{array}
$$

The hidden semidefinite constraint yields the semidefinite program, i.e. we get
$A : \Re^{m+1} \to \mathcal{S}_{n+1}$

$$
B = \begin{pmatrix} 0 & b_0^t \\ b_0 & Q_0 \end{pmatrix}, A \begin{pmatrix} t \\ x \end{pmatrix} = \begin{bmatrix} -t & \sum_{i=1}^m x_i b_i^t \\ \sum_{i=1}^m x_i b_i & \sum_{i=1}^m x_i Q_i \end{bmatrix}
$$

$$
B - A \begin{pmatrix} t \\ x \end{pmatrix} \succeq 0.
$$

The dual program is equivalent to the SDP (with $c_0 = 0$)

$$\textbf{(D)} \qquad \begin{array}{rll} d^* = & \sup & -\sum_{i=1}^m x_i c_i - t \\ & \text{s.t.} & A \begin{pmatrix} t \\ x \end{pmatrix} \preceq B \\ & & x \in \Re^m, t \in \Re \end{array}$$

As in linear programming, the dual is obtained from the optimal strategy of the competing player:

$$\textbf{(DD)} \qquad \begin{array}{rll} d^* = & \inf & \operatorname{trace} BU \\ & \text{s.t.} & A^*U = \begin{pmatrix} -1 \\ -c \end{pmatrix} \\ & & U \succeq 0. \end{array}$$

Example

If the primal is

$$
\textbf{(P)} \qquad \begin{array}{lll} p^* = & \sup & x_2 \\ & \text{s. t.} & \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}
$$

$$
S^f = cone \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\}
$$

Then the dual is

$$
\textbf{(D)} \qquad \begin{array}{llc} d^* = & \inf & U_{11} \\ & \text{s. t.} & U_{22} = 0 \\ & & U_{11} + 2U_{23} = 1 \\ & & U \succeq 0. \end{array}
$$

Then $p^* = 0 < d^* = 1$.

But regularized dual has $U \succeq_{(S^f)^+} 0$, i.e. only constraint is $U_{11} \geq 0$. So new dual optimal value is 0. (Attained.)

What is a proper duality theory?

Do duality gaps occur in practice?

Are there an infinite number of duals if Slater's condition fails?

the cone $T \subset K$ is a *face* of the cone K, denoted $T \lhd K$, if

$$x, y \in K,\ x + y \in T \Rightarrow x, y \in T.$$

Each face, $K \lhd \mathcal{P}$, is characterized by a subspace, $S \subset \Re^n$.

$$K = \{X \in \mathcal{P} : \mathcal{N}(X) \supset S\}.$$

Moreover,

$$\operatorname{relint} K = \{X \in \mathcal{P} : \mathcal{N}(X) = S\}.$$

The complementary face of K is $K^c = K^\perp \cap \mathcal{P}$

$$K^c = \{X \in \mathcal{P} : \mathcal{N}(X) \supset S^\perp\}.$$

Moreover,

$$\operatorname{relint} K^c = \{X \in \mathcal{P} : \mathcal{N}(X) = S^\perp\}.$$

the face K (respectively, K^c) is determined by the supporting hyperplane corresponding to any $X \in \operatorname{relint} K^c$ (respectively, $\operatorname{relint} K$);

and

$$XY = 0, \forall X \in K, Y \in K^c$$

An SDP:

$$
\textbf{(P)} \quad \begin{array}{rll} p^* = & \max & c^t x \\ & \text{s.t.} & Ax \preceq_{\mathcal{P}} b \\ & & x \in \Re^m. \end{array}
$$

Optimality conditions:

$$
c \in A^*(\mathcal{P}^+) \quad \text{(closed?)}
$$

dual program

$$
\textbf{(D)} \quad \begin{array}{rll} p^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & A^*U = c \\ & & U \succeq_{\mathcal{P}^+} 0. \end{array}
$$

Optimality conditions:

$$
b \in \mathcal{R}(A) + \mathcal{P}^+ \quad \text{(closed?)}
$$

The *minimal cone* of P is defined as

$$\mathcal{P}^f = \cap\{\text{faces of } \mathcal{P} \text{ containing } (b - A(F))\}.$$

Therefore, an equivalent program is the *regularized P program*

$$
(\mathbf{RP}) \quad
\begin{array}{rll}
p^* = & \max & c^t x \\
& \text{s.t.} & Ax \preceq_{\mathcal{P}^f} b \\
& & x \in \Re^m.
\end{array}
$$

there exists x such that $b - Ax \in \operatorname{relint} \mathcal{P}^f$. (generalized Slater's constraint qualification) strong duality pair is RP and

$$
(\mathbf{DRP}) \quad
\begin{array}{rll}
p^* = & \min & \operatorname{trace} bU \\
& \text{s.t.} & A^*U = c \\
& & U \succeq_{(\mathcal{P}^f)^+} 0.
\end{array}
$$

Find $\mathcal{P}^f$? Properties?

LEMMA 1

Suppose $\mathcal{P}^f \lhd K \lhd \mathcal{P}$. Then the system

$$A^*U = 0, U \succeq_{K+} 0, \text{trace}\, Ub = 0$$

is consistent only if

$$\text{the minimal cone } \mathcal{P}^f \subset \{U\}^\perp \cap K.$$

PROOF

Since $\text{trace}\, U(Ax - b) = 0$, for all x, we get $A(F) - b \subset U^\perp$, i.e. $\mathcal{P}^f \subset \{U\}^\perp$.

$\square$

LEMMA 2 (surprising)
Suppose that $0 \neq K \lhd \mathcal{P}$ (proper face). Then

$\mathcal{P}^{+} + K^{\perp} = cl(\mathcal{P}^{+} + span\ K^{c})$ is always closed.

But

$$\mathcal{P}^{+} + span\ K \text{ is never closed.}$$

$\square$

Define: $\mathcal{P}_0 := \mathcal{P}$ and

$$\mathcal{U}_1 := \{U \succeq_{(\mathcal{P}_0)^+} 0 : A^*U = 0, \operatorname{trace} Ub = 0\}$$

Choose $U_1 \in \mathcal{U}_1 \cap \operatorname{relint} \mathcal{U}_1^f$ (if 0 - **STOP**)

$$\mathcal{P}_1 := \mathcal{U}_1^c = \{U_1\}^\perp \cap \mathcal{P}_0 \lhd \mathcal{P}_0$$

$$(\mathbf{RP_1}) \qquad \begin{array}{rll} p^* = & \max & c^t x \\ & \text{s.t.} & Ax \preceq_{\mathcal{P}_1} b \\ & & x \in \Re^m. \end{array}$$

————————— $p^* \le d_1^* \le d^*$ —————————-

$$(\mathbf{DRP_1}) \qquad \begin{array}{rll} d_1^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & A^*U = c \\ & & U \succeq_{(\mathcal{P}_1)^+} 0. \end{array}$$

———- $(\mathcal{P}_1)^+ = (\mathcal{P} \cap \mathcal{P}_1)^+ = \mathcal{P} + (\mathcal{P}_1)^\perp$ ———-

$$
(\mathbf{ELSD_1}) \qquad \begin{array}{rll} d_1^* = & \min & \operatorname{trace} b(U + (W + W^t)) \\ & \text{s.t.} & A^*(U + (W + W^t)) = c \\ & & A^* U_1 = 0, \operatorname{trace} U_1 b = 0 \\ & & U \succeq 0, \ \left[\begin{array}{cc} I & W^t \\ W & U_1 \end{array} \right] \succeq 0. \end{array}
$$

We used

$$\mathcal{P}_1^{\perp} = \left\{ (W + W^t) : A^* U_1 = 0, \operatorname{trace} U_1 b = 0, \right.$$
$$\left. \begin{bmatrix} I & W^t \\ W & U_1 \end{bmatrix} \succeq 0 \right\}$$

and

$$U_1 \succeq W W^t \qquad \text{iff} \qquad \begin{bmatrix} I & W^t \\ W & U_1 \end{bmatrix} \succeq 0$$
$$\text{implies} \quad W = U_1 H, \text{for some matrix H}$$

$$
\begin{aligned}
\mathcal{U}_2 &:= \{U \succeq_{(\mathcal{P}_1)^+} 0 : A^*U = 0, \operatorname{trace} Ub = 0\} \\
&= \{U + Z : A^*(U+Z) = 0, \operatorname{trace} Ub = 0, \\
&\qquad U \succeq_{(\mathcal{P}_0)^+}, Z \in (\mathcal{P}_1)^\perp\}
\end{aligned}
$$

Choose $U_2 \in \mathcal{U}_2 \cap \operatorname{relint} \mathcal{U}_2^f$ (if 0 - **STOP**)

$$\mathcal{P}_2 := \mathcal{U}_2^c = \{U_2\}^\perp \cap \mathcal{P}_1 \lhd \mathcal{P}_1$$

$$
\textbf{(RP}_\mathbf{2}\textbf{)} \qquad \begin{array}{rll} p^* = & \max & c^t x \\ & \text{s.t.} & Ax \preceq_{\mathcal{P}_2} b \\ & & x \in \Re^m. \end{array}
$$

——————$p^* \leq d_2^* \leq d_1^* \leq d^*$——————-

$$
\textbf{(DRP}_\mathbf{2}\textbf{)} \qquad \begin{array}{rll} d_2^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & A^*U = c \\ & & U \succeq_{(\mathcal{P}_2)^+} 0. \end{array}
$$

———$(\mathcal{P}_2)^+ = (\mathcal{P} \cap \mathcal{P}_2)^+ = \mathcal{P} + (\mathcal{P}_2)^\perp$———-

$$
\textbf{(ELSD}_\mathbf{2}\textbf{)} \qquad \begin{array}{rll} d_1^* = & \min & \operatorname{trace} b(U + (W + W^t)) \\ & \text{s.t.} & A^*(U + (W + W^t)) = c \\ & & A^*U_1 = 0, \operatorname{trace} U_1 b = 0 \\ & & A^*(U_2 + (W_1 + W_1^t)) = 0, \\ & & \operatorname{trace}(U_2 + (W_1 + W_1^t))b = 0 \\ & & U \succeq 0, \ \begin{bmatrix} I & W_1^t \\ W_1 & U_1 \end{bmatrix} \succeq 0 \\ & & \begin{bmatrix} I & W^t \\ W & U_2 \end{bmatrix} \succeq 0. \end{array}
$$

HOMOGENIZATION

Alternate view of optimality conditions

$$
\begin{array}{llll}
 & 0 = & \max & c^t x + t(-p^*) \\
(\mathbf{HP}) & & \text{subject to} & Ax + t(-b) + Z = 0 \\
 & & & w \in K = \Re^m \otimes \Re_+ \otimes \mathcal{P}
\end{array}
$$

This defines the objective, constraints, and variables:

$$
\begin{array}{c}
(= \langle a, w \rangle) \\
(Bw = 0) \\
\left(w = \begin{pmatrix} x \\ t \\ Z \end{pmatrix} \right)
\end{array}
$$

the feasible set is

$$F_H = \mathcal{N}(B) \cap K,$$

$$Bw = 0, w \in K \text{ implies } \langle a, w \rangle \leq 0.$$

Optimality conditions:

$$a = \begin{pmatrix} c \\ -p \\ 0 \end{pmatrix} \in -(\mathcal{N}(B) \cap K)^+.$$

$$\begin{pmatrix} -c \\ p \\ 0 \end{pmatrix} \in \overline{\mathcal{R}(B^*) + K^+},$$

WCQ - Weakest Constraint Qualification:

CLOSURE HOLDS

Conditions for closure:

If C, D are closed convex sets and the intersection of their recession cones is empty, then $D - C$ is closed.

$cone(F_H - K)$ is the whole space

(Slater's)

$$\exists \hat{x} \in F \text{ such that } A\hat{x} \prec b.$$

FIX: Find sets, T, to add to attain the closure. Equivalently, find sets, C, $C^+ = T$, to intersect with K to attain the closure since

$$(\mathcal{N}(A) \cap (K \cap C))^+ = \overline{\mathcal{R}(A^*) + K^+ + C^+}.$$