Semidefinite Programming: Applications, Duality, and Interior-Point Methods

Michael J. Todd,
School of OR & IE, Cornell University

miketodd@cs.cornell.edu
http://www.orie.cornell.edu/
~miketodd/todd.html

INFORMS meeting, Seattle, October 1998.

Resources:

Survey articles:

- L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

- A. S. Lewis and M. L. Overton. Eigenvalue optimization. *Acta Numerica* 1996, pp. 149–160.

- M. X. Goemans. Semidefinite programming in combinatorial optimization. *Mathematical Programming*, 79:143–161, 1997.

- (Forthcoming) Handbook on Semidefinite Programming, eds. R. Saigal, L. Vandenberghe, and H. Wolkowicz.

Web Sites:

- F. Alizadeh:
  karush.rutgers.edu/~alizadeh/
  Sdppage/index.html

- C. Helmberg:
  www.zib.de/helmberg/semidef.html

- Interior-Point Methods Online:
  www-c.mcs.anl.gov/home/otc/
  InteriorPoint/index.html

Bibliography of H. Wolkowicz:
liinwww.ira.uka.de/bibliography/Math/psd.html

Tutorial: MC15, Some Applications
of SDP to Combinatorial Optimization,
David Williamson.

Problems and Notation

Primal form SDP:

$$
\begin{array}{llrcl}
(P) & \min_X & C \bullet X & & \\
 & & A_i \bullet X & = & b_i, \quad i = 1, \ldots, m \\
 & & X & \succeq & 0,
\end{array}
$$

where $C, A_1, \ldots, A_m$ and X are real symmetric $n \times n$ matrices (in $S\mathbb{R}^{n \times n}$) and $b \in \mathbb{R}^m$.

Here $P \bullet Q$ represents the usual inner product $\text{Trace}\,(P^T Q) = \sum_{ij} P_{ij} Q_{ij}$ on $n \times n$ matrices, and $X \succeq 0$ indicates that X is symmetric positive semidefinite (psd). [Similarly $U \succeq V$ or $V \preceq U$ means $U - V$ is psd, and $U \succ 0$ means U is symmetric positive definite (pd).]

Dual form SDP:

$$(D) \qquad \max_y \qquad \begin{array}{rcl} & b^T y & \\ \sum_{i=1}^m y_i A_i & \preceq & C, \end{array}$$

or

$$(D) \qquad \max_{y,S} \qquad \begin{array}{rcrcl} & b^T y & & & \\ \sum_{i=1}^m y_i A_i & + & S & = & C \\ & & S & \succeq & 0, \end{array}$$

or

$$(D) \qquad \max_y \qquad \begin{array}{rcrcl} & & b^T y & & \\ C & - & \sum_{i=1}^m y_i A_i & \succeq & 0. \end{array}$$

I.e., optimize a linear function of several variables subject to the constraint that a symmetric matrix that depends affinely on the variables be psd.

Some simple facts:

The following are equivalent for a symmetric $n \times n$ matrix U:

- $U \succeq 0$ ($\succ 0$);
- $z^T U z \geq 0$ (> 0) for all nonzero $z \in IR^n$;
- all eigenvalues $\lambda_j(U)$ of U are nonnegative (positive) (helpful to think in SDP $\lambda(X) \geq 0 \leftrightarrow x \geq 0$ in LP);
- $U = P^T P$ for some (square nonsingular) P.

Hence, for any $x \in IR^n$, $X := xx^T \succeq 0$.

If $U \in SIR^{n \times n}$, $U = Q \Lambda Q^T$, Q orthogonal, Λ diagonal.

$\text{Trace}\, PQ = \text{Trace}\, QP$.

Hence
$x^T U x = \text{Trace}\, x^T U x = \text{Trace}\, U x x^T = U \bullet X$,
where $X = x x^T$.

If U, $V \succeq 0$, then $U \bullet V \geq 0$. Indeed, if
$K := \{U : U \succeq 0\}$, then

$$K = K^* := \{V : U \bullet V \geq 0 \text{ for all } U \succeq 0\}.$$

If $U \succ 0$, then $\{V : U \bullet V \leq \beta, V \succeq 0\}$ is bounded.

If $U \succeq 0$, then U has a *square root*
$U^{1/2} \succeq 0$ with $U^{1/2} U^{1/2} = U$.

If U, $V \succeq 0$, then $U \bullet V = 0 \Leftrightarrow UV = 0$.

Convention: all upper-case Roman letters denote symmetric matrices, except K, P, and Q.

Applications

- Eigenvalue Optimization (also see TC29):

  Suppose we have a symmetric matrix

  $$A(y) := A_0 + \sum_{i=1}^{m} y_i A_i$$

  depending affinely on $y \in I\!R^m$. We wish to choose y to minimize the maximum eigenvalue of $A(y)$.

  Note: $\lambda_{\max}(A(y)) \leq \eta \Leftrightarrow$
  all eigenvalues of $\eta I - A(y)$ are nonnegative
  $\Leftrightarrow A(y) \preceq \eta I$.

  So solve

  $$\begin{array}{lllll} \max_{\eta,y} & -\eta & & & \\ & -\eta I & + \ \sum_{i=1}^{m} y_i A_i & \preceq & -A_0, \end{array}$$

  a problem of form (D).

- Quadratically-Constrained QP (SE27, MD27.1):

  Proposition (Schur complements)
  Suppose $B \succ 0$. Then

$$\begin{pmatrix} B & P \\ P^T & C \end{pmatrix} \succeq 0 \Leftrightarrow C - P^T B^{-1} P \succeq 0.$$

  Hence quadratic constraint
  $(Ax+b)^T(Ax+b) - c^T x - d \leq 0$ holds iff

$$\begin{pmatrix} I & Ax+b \\ (Ax+b)^T & c^T x + d \end{pmatrix} \succeq 0.$$

  QCQP: $\min\{f_0(x) : f_i(x) \leq 0,\ i = 1, \ldots, m\}$
  where $f_i(x) := (A_i x + b_i)^T (A_i x + b_i) - c_i^T x - d_i$ is equivalent to

$$\begin{array}{lcl} \max_{\eta,x} & -\eta & \\ & \begin{pmatrix} I & A_0 x + b_0 \\ (A_0 x + b_0)^T & c_0^T x + d_0 + \eta \end{pmatrix} & \succeq \ 0, \\ & \begin{pmatrix} I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{pmatrix} & \succeq \ 0, \\ & i = 1, \ldots, m. & \end{array}$$

- Systems and Control Theory (MD27.4):

  Suppose the state of a system is defined by a differential inclusion:

$$\dot{x} \in \mathrm{conv}\{P_1, P_2, \ldots, P_m\}\, x.$$

  We want to know whether trajectories remain bounded for all time, whatever convex combination is chosen at each time. Sufficient condition: there is $Y \succ 0$ so that $V(x) := \frac{1}{2}x^T Y x$ is nonincreasing with time, i.e.,

$$\dot{V}(x) = \frac{1}{2}x^T(YP + P^TY)x \leq 0$$

  for all $P \in \mathrm{conv}\{P_1, P_2, \ldots, P_m\}$. This leads to

$$\begin{array}{lrcrcl} \max_{\eta,Y} & -\eta & & & & \\ & -\eta I & + & Y & \preceq & 0, \\ & & & -Y & \preceq & -I, \\ & YP_i & + & P_i^T Y & \preceq & 0,\ i = 1, \ldots, m. \end{array}$$

- Maximum Cut (for more, go to MC15!):

  Given an undirected (wlog complete) graph on $V = \{1, \ldots, n\}$ with nonnegative edge weights $W = (w_{ij})$, find a cut $\delta(S) := \{\{i, j\} : i \in S, j \notin S\}$ with maximum weight.

  (IP): $\max\{\frac{1}{4}\sum_i \sum_j w_{ij}(1 - x_i x_j) : x_i \in \{-1, +1\}, i = 1, \ldots, n\}$.

  Constraint same as $x_i^2 = 1$ all i. Now $\{X : x_{ii} = 1, i = 1, \ldots, n, X \succeq 0, \mathrm{rank}(X) = 1\} = \{xx^T : x_i^2 = 1, i = 1, \ldots, n\}$. So a relaxation is:

$$\begin{array}{rcl} \frac{1}{4}\sum\sum w_{ij} - \frac{1}{4}\min_X W \bullet X & & \\ e_i e_i^T \bullet X & = & 1,\ i = 1, \ldots, n, \\ X & \succeq & 0. \end{array}$$

  This gives a good bound (within 14%) and yields a feasible solution within 13% by randomization.

Duality

Rewrite (P) as

$$\begin{array}{lrcl} \min_X & C \bullet X & & \\ & -X & \preceq & 0, \\ & b - \mathcal{A}X & = & 0, \end{array}$$

where $\mathcal{A} : S\mathbb{R}^{n \times n} \to \mathbb{R}^m$ is defined by $\mathcal{A}X :=$ $(A_i \bullet X)$. The Lagrangian is

$$L(X, U, v) := \begin{cases} C \bullet X - U \bullet X + v^T(b - \mathcal{A}X) & U \succeq 0, \\ -\infty & \text{else} \end{cases}$$

and it is not hard to see that (P) is equivalent to

$$\min_X \max_{U,v} L(X, U, v).$$

The Lagrangian dual is then

$$\max_{U,v} \min_X L(X, U, v).$$

Write $L(X, U, v)$ for $U \succeq 0$ as

$$b^T v + (C - \mathcal{A}^* v - U) \bullet X,$$

where $\mathcal{A}^* : I\!R^m \to SI\!R^{n \times n}$ is defined by $\mathcal{A}^* v := \sum v_i A_i$ (so that $(\mathcal{A}^* v) \bullet X = v^T(\mathcal{A}X)$). Then we see that the dual is

$$\begin{array}{lrcl} \max_{U,v} & b^T v & & \\ & C - \mathcal{A}^* v - U & = & 0, \\ & U & \succeq & 0, \end{array}$$

which is exactly (D).

Weak duality is easy: if X and (y, S) are feasible in (P) and (D), then

$$C \bullet X - b^T y = (\mathcal{A}^* y + S) \bullet X - (\mathcal{A}X)^T y = S \bullet X \geq 0.$$

So a sufficient condition for optimality is primal and dual feasibility plus complementary slackness in the form $S \bullet X = 0$. Unfortunately not necessary.

Examples:

$$\max \ -y_1$$
$$y_1 \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \preceq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

y is feasible iff $\begin{pmatrix} y_1 & 1 \\ 1 & y_2 \end{pmatrix} \succeq 0$. This has optimal value zero, but no optimal solution. The corresponding primal problem does have an optimal solution, and the same optimal value.

The next example shows that there may be a positive finite duality gap. For both primal and dual, use the fact that a psd matrix with a zero on the diagonal must have zeroes in the corresponding row and column. (Try perturbing c_{22} or b_1 to small positive numbers and see how the duality gap disappears!)

$$\begin{array}{rcl} \min \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \bullet X & & \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \bullet X & = & 0, \\ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \bullet X & = & 2, \\ X & \succeq & 0, \end{array}$$

has optimal solution $X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and optimal value 1, while its dual

$$\begin{array}{ll} \max & 2y_2 \\ y_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} & \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \end{array}$$

has optimal solution $y = (0, 0)^T$ and optimal value 0.

Define

$$F(P) := \{X : \mathcal{A}X = b, X \succeq 0\},$$

$$F^0(P) := \{X : \mathcal{A}X = b, X \succ 0\},$$

$$F(D) := \{(y, S) : \mathcal{A}^* y + S = C, S \succeq 0\},$$

$$F^0(D) := \{(y, S) : \mathcal{A}^* y + S = C, S \succ 0\}.$$

Theorem Suppose $F(P)$ and $F^0(D)$ are nonempty. Then (P) has an optimal solution; indeed its set of optimal solutions is nonempty and bounded. Also, there is no duality gap.

Proof sketch The set of optimal solutions to (P) is the same as that for (P) with the added constraint $C \bullet X \leq C \bullet \hat{X}$, where $\hat{X}$ is an arbitrary feasible solution. But this constraint is equivalent to $\widehat{S} \bullet X \leq \widehat{S} \bullet \hat{X}$, where $(\hat{y}, \widehat{S}) \in F^0(D)$, and the set of psd X satisfying this inequality is bounded.

Hence we are minimizing a continuous function over a nonempty compact set, so the first part is proved.

If ζ is the optimal value of (P), then apply a separating hyperplane argument to $\{X : X \succeq 0\}$ and $\{X : \mathcal{A}X = b, C \bullet X \leq \zeta - \epsilon\}$ for an arbitrary positive ϵ to get a feasible dual solution within ϵ of the optimal value of (P).

Corollary Suppose both (P) and (D) are strictly feasible. Then there is no duality gap and there are optimal solutions to both satisfying complementary slackness.

Sufficient conditions for optimality:

$$\begin{array}{rcccll} & \mathcal{A}^*y & + & S & = & C, \quad S \succeq 0, \\ \mathcal{A}X & & & & = & b, \quad X \succeq 0, \\ & XS & & & = & 0. \end{array}$$

The Central Path and Interior-Point Algorithms for SDP

Assume $F^0(P) \times F^0(D) \neq \emptyset$. As in LP, we perturb the optimality conditions to get the central path equations:

$$\begin{array}{rcccl} & \mathcal{A}^* y & + & S & = C, \quad S \succeq 0, \\ \mathcal{A}X & & & & = b, \quad X \succeq 0, \quad (*) \\ & XS & & & = \nu I, \end{array}$$

for $\nu > 0$. We can show

Theorem For each $\nu > 0$ there is a unique solution $(X(\nu), y(\nu), S(\nu))$ to (*) and these form a smooth path. As $\nu \to 0$, this path converges to a pair of optimal solutions to (P) and (D).

The proof uses the fact that (*) forms the optimality conditions for the primal (and dual!) barrier problem:

$$(PBP_\nu) \quad \begin{array}{rrcl} \min_X & C \bullet X + \nu f(X) & & \\ & \mathcal{A}X & = & b \\ & (X & \succ & 0), \end{array}$$

where f is the barrier function for the psd cone defined by

$$f(X) := -\ln \det X = -\sum_j \ln \lambda_j(X).$$

Note that $X(\nu)$ and $(y(\nu), S(\nu))$ are strictly feasible in (P) and (D) with duality gap

$$X \bullet S = \text{Trace}\,(XS) = \text{Trace}\,(\nu I) = n\nu,$$

so finding points on (or near to) the central path with small parameter values ν gives almost optimal solutions.

Neighborhoods of the central path:

Let $\mu := \mu(X,S) := X \bullet S/n$. For $\gamma \in (0,1)$, define

$$\mathcal{N}_F(\gamma) := \{(X,y,S) \in F^0(P) \times F^0(D) : \\ \|X^{1/2}SX^{1/2} - \mu I\|_F = [\sum(\lambda_j(XS) - \mu)^2]^{1/2} \leq \gamma\mu\},$$

$$\mathcal{N}_\infty(\gamma) := \{(X,y,S) \in F^0(P) \times F^0(D) : \\ \|X^{1/2}SX^{1/2} - \mu I\| = \max|\lambda_j(XS) - \mu| \leq \gamma\mu\},$$

$$\mathcal{N}_{-\infty}(\gamma) := \{(X,y,S) \in F^0(P) \times F^0(D) : \\ \|X^{1/2}SX^{1/2} - \mu I\|_{-\infty} = \max(\mu - \lambda_j(XS)) \leq \gamma\mu\}.$$

The first and last are usually called the *narrow* and *wide* neighborhoods of the central path.

For the overview to follow, suppose we

have $(X_0, y_0, S_0) \in F^0(P) \times F^0(D)$, on or near the central path, with duality gap

$$X_0 \bullet S_0 \leq \frac{1}{\epsilon}; \text{ and}$$

want $(X, y, S) \in F(P) \times F(D)$ with duality gap

$$X \bullet S \leq \epsilon.$$

- $O(n\,\ln(1/\epsilon))$-iteration primal potential-reduction method (Nesterov and Nemirovskii) [like Karmarkar for LP];

- $O(\sqrt{n}\,\ln(1/\epsilon))$-iteration asymmetric primal-dual potential-reduction method (Alizadeh, Nesterov and Nemirovskii) [Ye];

- $O(\sqrt{n}\,\ln(1/\epsilon))$-iteration symmetric primal-dual potential-reduction method (Nesterov and Todd, Kojima, Shindoh, and Hara) [Kojima, Mizuno, and Yoshise].

These use the potential functions

$$\phi(X,\zeta) := (n+1)\ln(C \bullet X - \zeta) + f(X),$$

$$\Phi(X,S) := (n+\sqrt{n})\ln X \bullet S + f(X) + f(S).$$

- $O(\sqrt{n}\,\ln(1/\epsilon))$-iteration short-step primal path-following method (Nesterov and Nemirovskii) [Renegar, Gonzaga]

- $O(\sqrt{n}\,\ln(1/\epsilon))$-iteration short-step and predictor-corrector primal-dual path-following methods (Kojima-Shindoh-Hara, Nesterov-Todd, Zhang, *Monteiro*, based on algorithms by Helmberg-Rendl-Vanderbei-Wolkowicz, Kojima-Shindoh-Hara, Monteiro, Alizadeh-Haeberly-Overton, Nesterov-Todd) [Kojima-Mizuno-Yoshise, Monteiro-Adler, Ye-Todd-Mizuno]

- $O(n\ln(1/\epsilon))$ and $O(n^{3/2}\ln(1/\epsilon))$-iteration long-step primal-dual path-following method (Zhang, *Monteiro-Zhang*, based on algorithms by HRVW, KSH, M, NT) [Kojima-Mizuno-Yoshise]

There are also "polynomial" infeasible-interior-point methods (KSH, Zhang, Tseng, Potra-Sheng, de Klerk-Roos-Terlaky, Luo-Sturm-Zhang, Nesterov-Todd-Ye)

Why so many variants of methods? What is the search direction?

Natural to take a Newton step for the nonlinear function

$$(X, y, S) \to (\mathcal{A}^* y + S - C,\ \mathcal{A}X - b,\ XS - \sigma\mu I),$$

where $\mu = X \bullet S/n$, $\sigma \in [0, 1]$.

But function takes

$$SIR^{n\times n} \times IR^m \times SIR^{n\times n} \to SIR^{n\times n} \times IR^m \times IR^{n\times n}.$$

So the Newton step is not defined!

Need to symmetrize the last equation.

Change $XS = \sigma\mu I$ to

$$H_P(XS) = \sigma\mu I,$$

where

$$H_P(Q) := [PQP^{-1} + (PQP^{-1})^T]/2.$$

Monteiro-Zhang approach. Gives Newton step solving

$$\begin{array}{rcl} \mathcal{A}^*\Delta y + \Delta S & = & R_D := C - \mathcal{A}^* y - S, \\ \mathcal{A}\Delta X & = & r_P := b - \mathcal{A}X, \\ H_P(\Delta X S) + H_P(X \Delta S) & = & \sigma\mu I - H_P(XS). \end{array}$$

Last equation can be written (or rewritten)

$$\mathcal{E}\Delta X + \mathcal{F}\Delta S = R_C.$$

Solve via

$$M\Delta y := (\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*)\Delta y = \ldots,$$

hence Δy, ΔS, and then ΔX.

Examples (wlog, $P \succ 0$):

- Alizadeh-Haeberly-Overton: $P = I$.

$$\begin{aligned} \mathcal{E}U &= (SU + US)/2, \\ \mathcal{F}V &= (XV + VX)/2. \end{aligned}$$

M not symmetric, maybe singular!
But works very well in practice.

- HRVW-KSH-M: $P = S^{1/2}$.

$$\begin{aligned} \mathcal{E}U &= U, \\ \mathcal{F}V &= (XVS^{-1} + S^{-1}VX)/2. \end{aligned}$$

- Nesterov-Todd: $P = W^{-1/2}$, where $W \succ 0$ solves $X = WSW$.

$$\begin{aligned} \mathcal{E}U &= U, \\ \mathcal{F}V &= WVW. \end{aligned}$$

For latter two, $M \succ 0$.

Plus many more! Families, twenty specific directions, ...

Practical methods: use long-step method, either staying in wide neighborhood or going 99% of the way to the boundary; use one of the directions above and a Mehrotra-type predictor-corrector approach.

Must be able to deal with infeasible starts: use method as above or homogeneous self-dual approach.

Software (see Helmberg's web site):

CSDP	LMITOOL	SDPA	SDPHA
SDPpack	SDPT3	SeDuMi	SP

Can solve medium-sized general problems (m and n of order of 200) without difficulty.