An Introduction to Semidefinite Programming and its Applications

Henry Wolkowicz
Department of Combinatorics & Optimization
University of Waterloo

• CRM • 2002

A Tutorial on Semidefinite Programming and its Applications

Henry Wolkowicz

Department of Combinatorics & Optimization
University of Waterloo

CORS • SCRO • 2002

OUTLINE

- What is SDP?

  Basic Properties

  Duality and Optimality Conditions

  (Comparisons with Linear Programming)

- Why use SDP?

  Quadratic Models better than Linear Models

  (Examples/Applications/Efficient Solutions)

- How to solve SDPs?

  Interior-Point Algorithms

  (State of the Art, Sparsity Issues, Open Problems)

HISTORY

- Lyapunov (stability) 1890
- SDP (cone optimization/duality) studied back in 60's
- Engineering applications in 60's
- matrix completion problems in 80's
- polynomial time algorithms Nesterov-Nemirovski 80's
- combinatorial appl., primal-dual interior-point algorithms (explosion) 90's

WHAT?

primal and dual SDPs (look like LPs)

$$\text{(PSDP)} \qquad \begin{array}{rll} p^* := & \max & \operatorname{trace} CX \\ & \text{s.t.} & \mathcal{A}X = b \\ & & X \succeq 0. \end{array}$$

$$\text{(DSDP)} \qquad \begin{array}{rll} d^* := & \min & b^T y \\ & \text{s.t.} & \mathcal{A}^* y - Z = C \\ & & Z \succeq 0, \end{array}$$

$\mathcal{S}^n$ space of $n \times n$ real symmetric matrices, $X, Z, C \in \mathcal{S}^n$

$\succ 0$ ($\succeq 0$) denotes positive (semi)definiteness

\- (i.e. Loewner partial order)

$\mathcal{A} : \mathcal{S}^n \to \Re^m$ linear operator

$\mathcal{A}^*$ adjoint operator (compare transpose)

$$
(PSDP) \qquad \begin{array}{rcc} p^* := \min & C \bullet X & (= \operatorname{trace} CX) \\ \text{s.t.} & \mathcal{A}X = b & \\ & X \succeq 0, & \end{array}
$$

---

weak duality (using hidden constraints)

$$
\begin{array}{rcl} p^* & = & \min\limits_{X \succeq 0} \max\limits_{y} C \bullet X + y^T(b - \mathcal{A}X) \\ & \geq & \max\limits_{y} \min\limits_{X \succeq 0} y^T b + (C - \mathcal{A}^* y) \bullet X \\ & = & \max\limits_{C - \mathcal{A}^* y \succeq 0} y^T b =: d^* \quad (DSDP) \end{array}
$$

$\mathcal{A}^*$ adjoint operator of $\mathcal{A}$

$$
\mathcal{A}(X) \bullet y = X \bullet \mathcal{A}^*(y), \quad \forall X, \forall y
$$

The duality theory gives rise to the characterization of optimality

$$\begin{array}{rcll} \mathcal{A}^*y + Z - C & = & 0 & \text{dual feasibility} \\ b - \mathcal{A}(X) & = & 0 & \text{primal feasibility} \\ ZX & = & 0 & \text{complementary slackness} \\ Z, X \succeq 0. & & & \end{array}$$

$X, (y, Z)$ a primal-dual optimal pair. Z is the (dual) slack variable.

$$ZX = \mu I, \quad \mu > 0$$

perturbed compl. slack. used for interior-point methods

$Z, X \in \mathcal{S}^n$ but ZX is not necessarily symmetric!

More Comparisons with LP

- *strict complementarity* for SDP: $Z + X \succ 0$
- *zero duality gap*: $p^* = d^*$ iff $Z \circ X = 0$ iff $ZX = 0$.
- *strong duality* for SDP: zero duality gap and d^* is attained. Strong duality can fail, though regularization is possible (ref. Borwein-W/80, Ramana/97)
- *Nondegeneracy and strict complementarity* (Theorem of Goldman and Tucker) do **not** directly follow through from LP to SDP (though conditions hold generically - ref. Shapiro/99, Pataki-Tuncel/98, Alizadeh-Haeberly-Overton/98)
- *Complexity/algorithms*: SDP are convex programs - can be approximately solved in polynomial time by interior point algorithms (ref. Nesterov-Nemirovski/88)

WHY?

Many applications: engineering, combinatorial optimization, statistics, matrix completions, approximation theory, nonlinear programming, ...

Many computationally hard problems can be modelled as quadratic programs (rather than linear programs).

These quadratic programs are themselves hard to solve numerically.

But, the Lagrangian relaxation can be solved efficiently using SDP.

MORE APPLICATIONS

Finding bounds and good feasible solutions for **hard combinatorial problems** such as: max-cut; graph partitioning; quadratic assignment problem; max-clique.

Unconstrained and constrained **NLP** techniques, e.g.: quasi-Newton updates that preserve positive definiteness; Trust region algorithms for large scale minimization; Extended SQP techniques for constrained minimization.

Partial Hermitian **matrix completion problems** and Euclidean distance matrix completion problems.

Engineering problems such as: Ricatti equations; min-max eigenvalue problems; matrix norm minimization; eigenvalue localization.

Tractable **Relaxations of Max-Cut**

(quadratic boolean programming)

$$(MCQ) \qquad \mu^* := \max_{x \in \mathcal{F}} q_0(x) \quad (:= x^T Q x - 2c^T x).$$

where $\mathcal{F} = \{\pm 1\}^n$

---

perturbing diagonal of Q on $\mathcal{F}$ yields an equivalent problem:

$$\begin{aligned} q_u(x) \quad &:= \quad x^T(Q + \mathrm{Diag}\,(u))x - 2c^T x - u^T e \\ &= \quad q_0(x), \quad \forall x \in \mathcal{F}, \end{aligned}$$

(e - vector of ones; $\mathrm{Diag}\,(u)$ diagonal matrix formed from u)

Bound 0: trivial bound from diagonal perturbations:

$$\mu^* \le f_0(u) := \max_x q_u(x).$$

$$S := \left\{u : u^T e = 0, Q + \mathrm{Diag}\,(u) \preceq 0\right\}.$$

$$\boxed{\begin{array}{c} \mu^* \le B_0 := \min_u f_0(u) \\ \left(= \min_{u^T e = 0} f_0(u), \text{ if } S \neq \emptyset \right). \end{array}}$$

Use hidden semidefinite constraint:

$$\mu^* \le B_0 = \min_{Q + \mathrm{Diag}\,(u) \preceq 0} f_0(u).$$

Bound 1: relax feasible set to **SPHERE** of radius $\sqrt{n}$ (tractable trust region subproblem, TRS):

$$\mu^* \leq f_1(u) := \max_{||x||^2=n} \; q_u(x)$$

and

$$\boxed{\mu^* \leq B_1 := \min_u f_1(u).}$$

The inner maximization problem is called a trust region subproblem and is tractable.

Bound 2: relax to **BOX CONSTRAINT**:

$$\mu^* \leq f_2(u) := \max_{|x_i| \leq 1} \; q_u(x).$$

add the semidefinite constraint to make bound tractable.

$$\mu^* \leq \min_u f_2(u)$$

and

$$\boxed{\mu^* \leq B_2 := \min_{Q + \mathrm{Diag}\,(u) \preceq 0} f_2(u).}$$

Bound B_1^c: lift to **EIGENVALUE BOUND**:

$$Q^c := \begin{bmatrix} 0 & -c^T \\ -c & Q \end{bmatrix}$$

$$q_u^c(y) := y^T(Q^c + \mathrm{diag}\,(u))y - u^T e$$

$$\mu^* \le f_1^c(u) := \max_{||y||^2=n+1} q_u^c(y)$$

where

$$\max_{||y||^2=n+1} q_u^c(y) = (n+1)\lambda_{\max}(Q^c + \mathrm{diag}\,(u)) - u^T e$$

$$\boxed{\mu^* \le B_1^c := \min_u f_1^c(u).}$$

Similarly: equivalent bounds B_0^c and other homogenized bounds

Bound B_3: SDP BOUND:

After homogenization ($c = 0$), use

$$x^T Q x = \operatorname{trace} x^T Q x = \operatorname{trace} Q x x^T$$

and, for $x \in \mathcal{F}$, $y_{ij} = x_i x_j$ defines a symmetric, rank one, positive semidefinite matrix Y with diagonal elements 1. Relax the rank one condition.

$$\boxed{\begin{array}{lcc} B_3 := & \max & \operatorname{trace} QY \\ & \text{subject to} & \operatorname{diag}(Y) = e \\ & & Y \succeq 0. \end{array}}$$

Summary:

(without restrictions e.g. $u^T e = 0$)

$$
\begin{aligned}
B_0 &= \min_u \max_x q_u(x) \\
B_1 &= \min_u \max_{x^T x = n} q_u(x) \\
B_2 &= \min_u \max_{-1 \leq x_i \leq 1} q_u(x) \\
B_3 &= \max\{\operatorname{trace} Q^c Y : \operatorname{diag}(Y) = e,\ Y \succeq 0.\} \\
B_1^c &= \min_u \max_{y^T y = n+1} q_u^c(y)
\end{aligned}
$$

Now replace ± 1 constraints with $x_i^2 = 1, \forall i$.

$$
\begin{array}{lll}
(P_E) & \max & q_0(x) = x^T Q x - 2c^T x \\
& \text{subject to} & x_i^2 = 1, \quad i = 1, \cdots, n.
\end{array}
$$

B_L denotes Lagrangian relaxation bound.

$$\boxed{\textbf{Theorem} \;\; B_L \text{ equals all above bounds.}}$$

REF: Polyak-W/94, Polyak-Rendl-W/95.

Recipe for (strengthened) SDP relaxations:

1. add redundant constraints
2. take Lagrangian dual
3. homogenize
4. use hidden semidefinite constraint to obtain SDP equivalent
   (check Slater's constraint qualification - strict feasibility)
5. take Lagrangian dual again
6. check Slater's CQ again - project if it fails
7. delete redundant constraints

Quadratic Assignment Problem, QAP

$$(QAP) \quad \mu^* := \min_{X \in \Pi} \operatorname{trace} AXBX^T - 2CX^T,$$

A, B real symmetric $n \times n$ matrices, C real $n \times n$ matrix, Π set of permutation matrices.

model for e.g. allocating set of n facilities to set of n locations while minimizing quadratic objective (distance-flow)

QAP is NP-hard and, in practice, problems of moderate sizes, such as $n = 16$, are still considered very hard.

recent surveys/books: Burkard/91, Rendl-Pardalos-W/93, Çela/98.

Questions:

1. interesting numerical and theoretical difficulties, e.g. loss of constraint qualification and loss of sparsity in the optimality conditions.
2. Can the new bound compete with other bounding techniques in speed and quality?
3. comparison on the Nugent test problems?
4. add new facet inequalities?

Notation:

$\mathcal{E} := \{X : Xe = X^T e = e\}$
set of matrices satisfying *assignment constraints*

$\mathcal{Z} := \{X : X_{ij} \in \{0, 1\}\}$
set of (0,1)-matrices

$\mathcal{N} := \{X : X_{ij} \geq 0\}$
set of *nonnegative matrices*

$\mathcal{O} := \{X : XX^T = X^T X = I\}$
set of orthogonal matrices

$$\Pi = \mathcal{E} \cap \mathcal{Z} = \mathcal{O} \cap \mathcal{Z}$$

rewrite QAP as (add redundant constraints)

$$
\begin{array}{ll}
\min & \operatorname{trace} AXBX^T - 2CX^T \\
\text{s.t.} & XX^T = X^TX = I \quad \text{(orthog.; redundancy important)} \\
& X_{ij}^2 - X_{ij} = 0, \quad \forall i, j. \\
& \left.\begin{array}{l} ||Xe - e||^2 = 0 \\ ||X^Te - e||^2 = 0 \end{array}\right\} \quad \text{(row \& col. sums 1)} \\
& \left.\begin{array}{l} \operatorname{diag}(X_{:i}X_{:j}^T) = 0, \quad \text{if } i \neq j \\ X_{:i}X_{:j}^T - \operatorname{Diag}\left(\operatorname{diag}(X_{:i}X_{:j}^T)\right) = 0, \ \text{if } i = j \end{array}\right\} \quad \textit{(gangster)}
\end{array}
$$

Lagrangian Relaxation

$$\begin{aligned}
\mu_{\mathcal{O}} &\geq \mu_{\mathcal{L}} \\
&:= \max_{W,u_0,v_0\ldots} \min_{XX^T=X^TX=I} \quad \{\operatorname{trace} AXBX^T - 2CX^T \\
&\qquad + \textstyle\sum_{ij} W_{ij}(X_{ij}^2 - X_{ij}) \\
&\qquad + u_0\|Xe - e\|^2 \\
&\qquad + v_0\|X^Te - e\|^2 \\
&\qquad + \ldots\}.
\end{aligned}$$

homogenize the Lagrangian using scalar x_0 and constraint $x_0^2 = 1$.

We get the lower bound
(quadratic, linear constant in X)

$$\begin{aligned}\max_{W,S_b,S_o,u_0v_0,,w_0}\ \min_{X,\ x_0}\quad &\{\operatorname{trace}[AXBX^T \\ &+ u_0\|Xe\|^2 + v_0\|X^Te\|^2 \\ &+ W(X\circ X)^T + w_0x_0^2 \\ &+ S_bXX^T + S_oX^TX] \\ &- \operatorname{trace} x_0(2C+W)X^T \\ &- 2x_0u_0e^T(X+X^T)e \\ &+ \dots \\ &- w_0 - \operatorname{trace} S_b - \operatorname{trace} S_o + 2nu_0x_0^2\}.\end{aligned}$$

Apply the hidden semidefinite constraint to get an SDP:

$$\begin{array}{ll} \max & -w_0 - \operatorname{trace} S_b - \operatorname{trace} S_o + \ldots \\ \text{s.t.} & L_Q + \operatorname{Arrow}(w) + \mathrm{B}^0\operatorname{Diag}(S_b) \\ & \quad + \mathrm{O}^0\operatorname{Diag}(S_o) + u_0 D + \ldots \succeq 0. \end{array}$$

check Slater's condition - OK

take Lagrangian dual again

$$
\begin{array}{ll}
\min & \operatorname{trace} L_Q Y \\
\text{s.t.} & \mathrm{b}^0\operatorname{diag}(Y) = I, \quad \mathrm{o}^0\operatorname{diag}(Y) = I \\
& \operatorname{arrow}(Y) = e_0, \quad \operatorname{trace} DY = 0 \\
& \dots \\
& Y \succeq 0,
\end{array}
$$

check Slater's condition again - $D \succeq 0$ so it fails!! But we can project onto the minimal face!!

Now remove redundant constraints to get simplified SDP relaxation simple projected relaxation with $n^3 - 2n^2 + 1$ constraints.

$$\begin{array}{rll} \mu_{R2} := & \min & \operatorname{trace}(\hat{V}^T L_Q \hat{V})R \\ & \text{s.t.} & \mathcal{G}_{\bar{J}}(\hat{V}R\hat{V}^T) = E_{00} \\ & & R \succeq 0. \end{array}$$

The dual problem is

$$\begin{array}{rll} \mu_{R2} = & \max & -Y_{00} \\ & \text{s.t.} & \hat{V}^T(L_Q + \mathcal{G}_{\bar{J}}^*(Y))\hat{V} \succeq 0. \end{array}$$

Remarks:

In each case of a tractable bound for a nonconvex problem, the structure allows for redundant constraints to be added to close the Lagrangian duality gap.

What is the correct question about "best" relaxation? Can we close or reduce the duality gap in general?

	Sol.	GLB	ELI	EVB3	SDP	rel. error SDP
Nug20	2570	2057	2196	2290	2386	0.0771
Nug21	2438	1833	1979	2116	2253	0.0821
Nug22	3596	2483	2966	3174	3396	0.0589
Nug24	3488	2676	2960	3074	3235	0.0782
Nug25	3744	2869	3190	3287	3454	0.0840
Nug30	6124	4539	5266	5448	5695	0.0753

QAPLIB (Nugent) instances

NEOS Server time for n=30; 1400 hours on SUN E6500

Nug 30 **finally** solved to optimality using CONDOR; equivalent computation time is 7 years on a single HP9000 C3000 workstation; ref. Anstreicher-Brixius-Groux-Linderoth/02.

HOW?

Use Newton's method on the characterization of optimality for the primal-dual pair

$$X \succeq 0, y, Z \succeq 0$$

$$\begin{aligned} \mathcal{A}^* y - Z &= C & \text{dual feasibility} \\ AX &= b & \text{primal feasibility} \\ ZX &= 0 & \text{complementary slackness} \end{aligned}$$

$$ZX = \mu I \qquad \text{perturbed}$$

Forms the basis for:
primal simplex; dual simplex; interior point methods

However, there are several subtle
DIFFERENCES with LP.

1. Duality gaps can exist for SDP in the absence of strictly feasible solutions (Slater's constraint qualification, CQ).

2. Strict complementarity can fail at the optimum.

$\Rightarrow$ 3. The perturbed optimality conditions involve nondiagonal matrices and map between different spaces
(overdetermined -does it require symmetrization?).

1. Duality gaps:

$$
(\mathbf{P}) \quad
\begin{array}{rll}
p^* = & \max & x_2 \\
& \text{s. t.} & \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\end{array}
$$

$$
(\mathbf{D}) \quad
\begin{array}{rlc}
d^* = & \min & \operatorname{trace} U_{11} \\
& \text{s. t.} & U_{22} = 0 \\
& & U_{11} + 2U_{23} = 1 \\
& & U \succeq 0.
\end{array}
$$

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

2. Strict complementarity:

$$
(\mathbf{P}) \qquad \begin{array}{rll} p^* = & \max & x_1 \\ & \text{s. t.} & \begin{bmatrix} x_1 & x_3 & x_2 \\ x_3 & x_2 & 0 \\ x_2 & 0 & x_3 \end{bmatrix} \preceq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}
$$

$$
(\mathbf{D}) \qquad \begin{array}{rll} d^* = & \min & \operatorname{trace} U_{33} \\ & \text{s. t.} & U_{11} = 1 \\ & & U_{22} + 2U_{13} = 0 \\ & & U_{33} + 2U_{12} = 0 \\ & & U \succeq 0. \end{array}
$$

Then the (unique) optimum pair is:
$U = \operatorname{Diag}(1, 0, 0)$
and $x = [0, 0, 0]$ with slack $Z = \operatorname{Diag}(0, 0, 1)$

INTERIOR-POINT METHODS for LP and SDP

history ....

primal-dual interior-point methods:

1. potential reduction
2. path following
3. infeasible-interior-point
4. predictor-corrector

(see e.g. recent books: Steve Wright/96; Bob Vanderbei/98) log-barrier problems

$$\begin{array}{rrl} d^* = & \min & b^t y \\ (\mathbf{D}) & \text{s.t.} & \mathcal{A}^* y - Z = C \\ & & Z \succeq 0 \end{array}$$

log-barrier problem for dual SDP, with fixed $\mu > 0$ (and $\mu \downarrow 0$)

$$\begin{array}{ll} \min & b^t y - \mu \log\det Z \\ \text{s.t.} & \mathcal{A}^* y - Z = C \end{array}$$

Lagrangian for log-barrier

$$\min b^t y + \operatorname{trace} X\,(C - \mathcal{A}^* y + Z) - \mu \log\det Z$$

Optimality conditions for the dual log-barrier
solve $F_\mu(X, y, Z) = 0$, with $X, Z \succ 0$

$$\begin{array}{rlcl} \frac{\partial}{\partial X}: & C - \mathcal{A}^* y + Z = 0 & & \text{dual feasibility} \\ \frac{\partial}{\partial y}: & b - \mathcal{A} X = 0 & & \text{primal feasibility} \\ \frac{\partial}{\partial Z}: & X - \mu Z^{-1} = 0 & & \text{pert. compl. slack.} \end{array}$$

or, to avoid ill-conditioning, use:

$$ZX - \mu I = 0 \quad \text{perturbed complementary slackness}$$

nonlinear system $F_\mu(X, y, Z) = 0$ is overdetermined if (less nonlinear) $ZX - \mu I = 0$ is used.
(this is **not** the case for LP)

CENTRAL PATH

For each $\mu > 0$, there exists a unique (X_μ, y_μ, Z_μ), with $X_\mu, Z_\mu \succ 0$, that solves the nonlinear system $F_\mu(X, y, Z) = 0$.

$$\mathcal{C} = \{(X_\mu, y_\mu, Z_\mu) : \mu > 0\}$$

well defined,
converges to optimum as $\mu \downarrow 0$,
long steps possible in neighbourhoods near $\mathcal{C}$

In numerical analysis, how do we solve a nonlinear system of equations?

The standard approach is to solve the equivalent nonlinear least squares problem using "Gauss-Newton" method.

$$\min \frac{1}{2}||F_\mu(X, y, Z)||_2^2$$

subject to $X, Z > 0$.

In the LP case, when we differentiate we get a square system and Gauss-Newton reduces to Newton's method.

i.e. solving

$$\left(F'_\mu\right)^t F'_\mu(\Delta s) = -\left(F'_\mu\right)^t F_\mu$$

(where F'_μ is the Jacobian)
is equivalent to solving

$$F'_\mu(\Delta s) = -F_\mu$$

framework p-d i-p:

Given $(X^0, y^0, Z^0) \in \mathcal{F}^0$

for $k = 0, 1, 2 \ldots$

solve for the search direction

$$F'_\mu(X^k, y^k, Z^k) \begin{pmatrix} \Delta X^k \\ \Delta y^k \\ \Delta Z^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k Z^k + \sigma_k \mu_k I \end{pmatrix}$$

where σ_k centering, $\mu_k = \operatorname{trace} X^k Z^k / n$

$$(X^{k+1}, y^{k+1}, Z^{k+1}) = \\ (X^k, y^k, Z^k) + \alpha_k (\Delta X^k, \Delta y^k, \Delta Z^k)$$

so that $(X^{k+1}, Z^{k+1}) \succ 0$

end (for).

SYMMETRIZATIONS

$$X - \mu Z^{-1} = 0 \quad \text{from log-barrier}$$

$$X + \Delta X - \mu \left(Z^{-1} - Z^{-1} \Delta Z Z^{-1}\right) = 0 \quad \text{after linearization}$$

$$ZX - \mu I = 0 \quad \text{less nonlinear}$$

$$ZX - \mu I + Z\Delta X + \Delta Z X = 0 \quad \text{after linearization}$$

$$ZX + XZ - 2\mu I = 0 \quad \text{symmetrize}$$

$$ZX + ZX - 2\mu I + Z\Delta X + \Delta X Z + \Delta Z X + X \Delta Z = 0 \quad \text{after linearization}$$

Others: $H_P(M) := \frac{1}{2}\left(PMP^{-1} + (PMP^{-1})^t\right)$
$M = XZ$ and P any nonsingular matrix

- $XZ + ZX$, AHO direction:

advantage: fast convergence and stability;
disadvantage: solve Lyapunov equation when finding Newton direction

- Heuristic; symmetrize ΔX after calculation; HKM direction.

advantages: fast convergence and inexpensive Newton direction, especially for max-cut problem!
disadvantage: stability

- Still many open questions, e.g. why symmetrize?

Can apply Gauss-Newton method to solving overdetermined system of nonlinear equations.
Large Sparse Case

Recall: **Max-Cut problem** and **SDP relaxation**:

$$(\text{MC0}) \qquad \begin{array}{rll} \mu^* := & \max & v^T Q v \\ & \text{s.t.} & v_i^2 = 1, \quad i = 1, \ldots, n. \end{array}$$

$$(\text{PSDP}) \qquad \begin{array}{rll} \mu^* \le \nu^* := & \max & \operatorname{trace} QX \\ & \text{s.t.} & \operatorname{diag}(X) = e \\ & & X \succeq 0, X \in \mathcal{S}^n, \end{array}$$

diag : vector from diagonal elements; e vector of ones; Diag diagonal matrix

$$(\text{DSDP}) \qquad \begin{array}{rll} \mu^* \le \nu^* := & \min & e^T y \\ & \text{s.t.} & \operatorname{Diag}(y) - Z = Q \\ & & Z \succeq 0, Z \in \mathcal{S}^n, \end{array}$$

Newton equation (from feasible X, y, Z)

$$\begin{aligned} \operatorname{Diag}(\Delta y) - \Delta Z &= 0 \\ \operatorname{diag}(\Delta X) &= 0 \\ Z\Delta X + \Delta Z X &= \mu I - ZX \end{aligned}$$

$$\begin{aligned} \Delta Z &= \operatorname{Diag}(\Delta y) \quad \text{BACKSUBSTIT.} \\ \Delta X &= -Z^{-1}(\Delta Z)X + \mu Z^{-1} - X \quad \text{BACKSUBSTIT.} \\ \operatorname{diag}(\Delta X) &= 0 \end{aligned}$$

$$\operatorname{diag}(Z^{-1}(\operatorname{Diag}(\Delta y))X = \operatorname{diag}(\mu Z^{-1} - X)$$

$$(Z^{-1} \circ X)\Delta y = \mu \operatorname{diag}(Z^{-1}) - e \quad \text{SOLVE}$$

```
function [phi, X, y] = psd_ip( L);
% solves: max trace(LX) s.t. X psd, diag(X) = b;  b = ones(n,1)/4
%         min b'y       s.t. Diag(y) - L psd, y unconstrained,
% input:  L ... symmetric matrix
% output: phi ... optimal value of primal, phi =trace(LX)
%         X   ... optimal primal matrix
%         y   ... optimal dual vector
% call:   [phi, X, y] = psd_ip( L);
%%%%%%%%%%%%%%%%%%%
         %%Initialization
digits = 6;                     % 6 significant digits of phi
[n, n1] = size( L);             % problem size
b = ones( n,1 ) / 4;            % any b>0 works just as well
X = diag( b);                   % initial primal matrix is pos. def.
y = sum( abs( L))' * 1.1;       % initial y is chosen so that
Z = diag( y) - L;               % initial dual slack Z is pos. def.
phi = b'*y;                     % initial dual
psi = L(:)' * X( :);            % and primal costs
mu = Z( :)' * X( :)/( 2*n);     % initial complementarity
iter=0;                         % iteration count
```

```
disp(['      iter    alphap    alphad      gap     lower      upper']);

while phi-psi > max([1,abs(phi)]) * 10^(-digits)

      iter = iter + 1;          % start a new iteration
      Zi = inv( Z);             % inv(Z) is needed explicitly
      Zi = (Zi + Zi')/2;
      dy =  (Zi.*X) \ (mu * diag(Zi) - b);      % solve for dy
      dX = - Zi * diag( dy) * X + mu * Zi - X;  % back substitute for dX
      dX = ( dX + dX')/2;       % symmetrize
```

```
% line search on primal
     alphap = 1;               % initial steplength
     [dummy,posdef] = chol( X + alphap * dX );  % test if pos.def
     while posdef > 0,
            alphap = alphap * .8;
            [dummy,posdef] = chol( X + alphap * dX );
            end;
     if alphap < 1, alphap = alphap * .95; end; % stay away from boundary
% line search on dual; dZ is handled implicitly: dZ = diag( dy);
     alphad = 1;
     [dummy,posdef] = chol( Z + alphad * diag(dy) );
     while posdef > 0;
            alphad = alphad * .8;
            [dummy,posdef] = chol( Z + alphad * diag(dy) );
            end;
     if alphad < 1, alphad = alphad * .95; end;
% update
     X = X + alphap * dX;
     y = y + alphad * dy;
     Z = Z + alphad * diag(dy);
     mu = X( :)' * Z( :) / (2*n);
     if alphap + alphad > 1.8, mu = mu/2; end; % speed up for long steps
     phi = b' * y; psi = L( :)' * X( :);
% display current iteration
       disp([ iter alphap alphad (phi-psi) psi  phi ]);

       end;            % end of main loop
```

representation

$$X = \mathrm{u2sMat}\,(x) + I, \quad x = \mathrm{u2svec}\,(X)$$

In the general SDP notation:
$\mathcal{A}X = b$
iff
$X = \mathcal{N}(x) + \hat{X}$
where $\mathcal{N}$ range equals null space of $\mathcal{A}$, $\hat{X}$ particular solution

substitute for both Z, X; eliminate first two blocks of optimality conditions.

THEOREM The primal-dual variables X, y, Z, with $X = \mathrm{u2sMat}\,(x) + I \succeq 0,\ Z = (\mathrm{Diag}\,(y) - Q) \succeq 0$, are optimal for (PSDP),(DSDP) if and only if they satisfy the single bilinear optimality equation

$$F(x,y) := (\mathrm{Diag}\,(y) - Q)(\mathrm{u2sMat}\,(x) + I) = 0$$

$\blacksquare$

$$F(x,y) : \Re^{\binom{n}{2}} \times \Re^n \to \mathcal{M}^n.$$

linearization (or Gauss-Newton equation)

$$\begin{aligned} -F_\mu(x,y) &= F'_\mu(x,y)\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} \\ &= Z\mathrm{u2sMat}\,(\Delta x) + \mathrm{Diag}\,(\Delta y)X \end{aligned}$$

a linear, **full rank**, overdetermined system and we find the least squares solution.