

Robust Algorithms for Large Sparse Semidefinite Programming (SDP)

with Applications to the Nearest Euclidean Distance Matrix Problem

Henry Wolkowicz

hwolkowicz@uwaterloo.ca

Department of Combinatorics and Optimization
University of Waterloo



Toulouse 2004

First Joint Canada-France meeting of the Mathematical
Sciences



Toulouse 2004

Robust Algorithms
for
Large Sparse Semidefinite Programming (SDP)

Friday July 15, 2004.

OUTLINE

- Background on SDP; Notation and Motivation
- Robust, (‘non-interior’) path-following algorithm for SDP
(outline of GN PCG method using LP)
- Application to Nearest Euclidean Distance Matrix Problem
- Numerics (Comparisons with a dual algorithm)

Notation and Motivation

$$\begin{array}{ll}
 \min & f(X) \\
 \text{(SDP)} & \text{subject to } \mathcal{A}X = b \\
 & X \succeq 0,
 \end{array}$$

where:

$f : \mathcal{S}^n \rightarrow \mathbb{R}$ convex function
 $\mathcal{S}^n$ $n \times n$ real symmetric matrices
 $X(\succeq) \prec 0$ denotes positive (semi)definite
 $\mathcal{A} : \mathcal{S}^n \rightarrow \mathbb{R}^m$ linear transformation

$$(\mathcal{A}X)_i = \langle A_i, X \rangle = \text{trace } A_i X, \quad A_i = A_i^T, i = 1 \dots n$$

Linear Primal-Dual Pair of SDPs

(looks/behaves like Linear Program, LP)

$$\begin{array}{ll} \min & \langle C, X \rangle = \text{trace } CX \\ \text{(PSDP)} & \text{subject to } AX = b \\ & X \succeq 0 \end{array}$$

$$\begin{array}{ll} \max & b^T y \\ \text{(SDP)} & \text{subject to } A^*y + Z = C \\ & Z \succeq 0 \end{array}$$

adjoint operator: $A^*y = \sum_{i=1}^m y_i A_i$

(some of the) APPLICATIONS

- Relaxations of hard combinatorial problems: e.g. max-cut; graph partitioning; quadratic assignment problem; max-clique.
- NLP 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.

SIMILARITIES to LP: (i) Duality

payoff function, player Y to player X (Lagrangian)

$$L(X, y) := \text{trace}(CX) + y^t(b - AX)$$

Optimal (worst case) strategy for player X :

$$p^* = \max_{X \succeq 0} \min_y L(X, y)$$

Using the *hidden constraint* $b - AX = 0$,
recovers primal problem.

apply adjoint

$$\begin{aligned} L(X, y) &= \text{trace}(CX) + y^t(b - AX) \\ &= b^t y + \text{trace}(C - A^* y) X \end{aligned}$$

adjoint operator, $A^* y = \sum_i y_i A_i$

$$\langle A^* y, X \rangle = \langle y, AX \rangle, \quad \forall X, y$$

Hidden Constraint: $C - A^* y \preceq 0$

exploit Hidden Constraint

$$p^* = \max_{X \succeq 0} \min_y L(X, y) \leq d^* := \min_y \max_{X \succeq 0} L(X, y)$$

dual obtained from optimal strategy of competing player, Y.

Hidden Constraint: $C - \mathcal{A}^*y \preceq 0$ yields the dual

$$\begin{aligned} (\text{DSDP}) \quad d^* &= \min && b^t y \\ &\text{s.t.} && \mathcal{A}^* y \succeq C \end{aligned}$$

for the primal

$$\begin{aligned} (\text{PSDP}) \quad p^* &= \max && \text{trace } CX \\ &\text{s.t.} && \mathcal{A}X = b \\ &&& X \succeq 0 \end{aligned}$$

Characterization of Optimality

primal-dual pair X, y (slack $Z \succeq 0$)

$$A^*y - Z = C \quad \text{dual feasibility}$$

$$AX = b \quad \text{primal feasibility}$$

$$ZX = 0 \quad \text{complementary slackness}$$

$$ZX = \mu I \quad \text{perturbed C.S., } \mu > 0$$

Basis for methods:

- primal simplex (maintain: primal feas. & compl. slack.)
- dual simplex (maintain: dual feas. & compl. slack.)
- interior point (maintain: primal feas. & dual feas.)

SDP Application: (Direct) Max-Cut Relaxation

Graph $G = (E, V)$; $|V| = n$ (nodes); w_{ij} weights on edges;

$$\max \frac{1}{2} \sum_{i < j} w_{ij} (1 - x_i x_j), \quad x \in \{\pm 1\}^n.$$

Equate $x_i = 1$ with i in set $\mathcal{I}$ and $x_i = -1$ otherwise.

Equivalent problem: homogeneous (± 1) -QQP

$$\mu^* := \max q(x) := x^t Q x = \text{trace } Q x x^T, \quad x \in \{\pm 1\}^n.$$

REPLACE $x \in \{\pm 1\}^n$ WITH CONSTRAINTS $x_i^2 = 1$??!

LIFTING :

$$X = x x^t$$

Relax the rank-1 condition on X to get linear SDP.

$$\mu^* \leq \max \{ \text{trace } QX : \text{diag}(X) = e, X \succeq 0 \}$$

SDP from general quadratic approx? (Lagr. Relax!)

$$q_i(y) := \frac{1}{2}y^t Q_i y + y^t b_i + c_i, \quad y \in \mathbb{R}^n$$

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

$$\text{Lagrangian :} \quad L(y, x) = q_0(y) + \sum_{i=1}^m x_i q_i(y)$$

or equivalently

$$\begin{aligned} L(y, x) &= \frac{1}{2}y^t (Q_0 + \sum_{i=1}^m x_i Q_i) y && \text{(quadratic in } y) \\ &\quad + y^t (b_0 + \sum_{i=1}^m x_i b_i) && \text{(linear in } y) \\ &\quad + (c_0 + \sum_{i=1}^m x_i c_i) && \text{(constant in } y) \end{aligned}$$

Weak Duality

follows from definition of dual program and hidden constraints:

$$d^* = \max_{x \geq 0} \min_y L(y, x) \leq q^* = \min_y \max_{x \geq 0} L(y, x).$$

Now **homogenize**; multiply linear term by new variable y_0

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

and add new constraint to Lagrangian (Lagrange multiplier t)

$$t(y_0^2 - 1)$$

Homogenization

$$\begin{aligned}
 d^* &= \max_{x \geq 0} \min_y L(y, x) \\
 &= \max_{x \geq 0} \min_{y_0^2=1} \frac{1}{2} y^t (Q_0 + \sum_{i=1}^m x_i Q_i) y + t y_0^2 \\
 &\quad + y_0 y^t (b_0 + \sum_{i=1}^m x_i b_i) \\
 &\quad + (c_0 + \sum_{i=1}^m x_i c_i) - t \\
 &= \max_{x \geq 0, t} \min_y \frac{1}{2} y^t (Q_0 + \sum_{i=1}^m x_i Q_i) y + t y_0^2 \\
 &\quad + y_0 y^t (b_0 + \sum_{i=1}^m x_i b_i) \\
 &\quad + (c_0 + \sum_{i=1}^m x_i c_i) - t
 \end{aligned}$$

hidden semidefinite constraint yields SDP

Apply Hidden SDP Constraint (Hessian psd)

$$B := \begin{pmatrix} 0 & b_0^t \\ b_0 & Q_0 \end{pmatrix} \text{ and } A : \Re^{m+1} \rightarrow \mathcal{S}_{n+1}$$

$$A \begin{pmatrix} t \\ x \end{pmatrix} := - \begin{bmatrix} t & \sum_{i=1}^m x_i b_i^t \\ \sum_{i=1}^m x_i b_i & \sum_{i=1}^m x_i Q_i \end{bmatrix}$$

$$\text{Lagrangian psd : } B - A \begin{pmatrix} t \\ x \end{pmatrix} \succeq 0.$$

NOTE There is **NO** hidden constraint on the Q_i if all q_i are convex. Better algorithms exist for the convex case, e.g. proximal methods, using quadratic cones, ...

Dual of Dual $\rightarrow$ SDP Relaxation

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

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

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

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

Tractable Relaxations

In some sense, Lagrangian relaxation is **best tractable relaxation**.

There are *higher order* relaxations:

e.g. from $X = xx^T$ from max-cut relaxation (from $x_j^2 = 1$)

$$\text{2nd LIFTING : } x_i x_j^2 x_k = x_i x_k, \quad Y = \begin{pmatrix} 1 \\ \text{svec } X \end{pmatrix} (1 \quad \text{svec } X)$$

Public domain software: e.g. **NEOS**

URL: www-neos.mcs.anl.gov

(Perturbed) Optimality Conditions

For barrier parameter $\mu > 0$:

$$F_\mu(X, y, Z) := \begin{pmatrix} A^*y + Z - C \\ AX - b \\ ZX - \mu I \end{pmatrix} = 0 \quad \begin{pmatrix} \text{dual feasibility} \\ \text{primal feasibility} \\ \text{pert. compl. slack.} \end{pmatrix}$$

For SDP:

$$F_\mu : \mathcal{S}^n \times \mathfrak{R}^m \times \mathcal{S}^n \rightarrow \mathcal{S}^n \times \mathfrak{R}^m \times \mathcal{M}^n$$

i.e. overdetermined nonlinear system

(Non) Interior Path-Following

Illustration/Motivation on LP Case

$$\begin{aligned} p^* &:= \min c^T x \quad (\text{or } \langle c, x \rangle) \\ (\text{LP}) \quad &\text{s.t. } Ax = b \end{aligned}$$

$$x \geq 0 \quad (\text{or } x \succeq 0)$$

$$\begin{aligned} d^* &:= \max b^T y \\ (\text{DLP}) \quad &\text{s.t. } A^T y + z = c \\ &z \geq 0 \quad (\text{or } z \succeq 0) \end{aligned}$$

Assume: $A \in \Re^{m \times n}$ full rank (onto); LP, DLP strictly feasible

dual log-barrier problem; parameter $\mu > 0$

$$\begin{aligned} d_\mu^* := \quad & \max \quad b^T y + \mu \sum_{j=1}^n \log z_j \quad (+\mu \log \det(z)) \\ \text{s.t.} \quad & A^T y + z = c \quad (A^T \cong A^*) \\ & z > 0 \quad (z \succ 0). \end{aligned}$$

stationary point of the Lagrangian / optimality conditions

$$F_\mu(x, y, z) = \begin{pmatrix} A^T y + z - c \\ Ax - b \\ X - \mu Z^{-1} \end{pmatrix} = 0, \quad \begin{aligned} & x, z > 0, \quad (\lambda \succ 0) \\ & X = \text{Diag}(x) \\ & Z = \text{Diag}(z) \end{aligned}$$

central path := set of solutions $(x_\mu, y_\mu, z_\mu), \mu > 0$

Jacobian Ill-conditioning

As $\mu \rightarrow 0$, Jacobian $F'_\mu(x, y, z)$ ill-conditioned near central path
Cure/Fix: Make nonlinear equations *less nonlinear*, i.e. **preconditioning** for Newton type methods;

premultiply by block-diag matrix with blocks (I, I, Z) :

$$\begin{aligned} F_\mu(x, y, z) &\leftarrow \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & Z \end{pmatrix} F_\mu(x, y, z) &= \begin{pmatrix} A^T y + z - c \\ Ax - b \\ ZX - \mu I \end{pmatrix} \\ &=: \begin{pmatrix} R_d \\ r_p \\ R_{ZX} \end{pmatrix} \end{aligned}$$

recovers modern primal-dual optimality paradigm

Exploited Special Structure

linearization for the Newton direction

$$\Delta s = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$$

$$F'_\mu(x, y, z) \Delta s = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & 0 & X \end{pmatrix} \Delta s = -F_\mu(x, y, z).$$

Overdetermined system in SDP case

$$\mathcal{S}^n \times \Re^m \times \mathcal{S}^n \rightarrow \mathcal{S}^n \times \Re^m \times \mathcal{M}^n$$

apply symmetrization; **undoes preconditioning**

$$\begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \mathcal{S} \end{pmatrix} \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & 0 & X \end{pmatrix}$$

e.g. last equation after symmetrization:

$$ZX + XZ - 2\mu I = 0 \text{ (AHO search direction)}$$

Reduction/Block-Elimination → Normal Equations

Step 1 (Eliminate Δz from row 3):

$$\begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -X & 0 & I \end{pmatrix} \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & 0 & X \end{pmatrix} = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & -XA^T & 0 \end{pmatrix}.$$

Define:

$$P_Z := \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -X & 0 & I \end{pmatrix}, \quad K := \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & -XA^T & 0 \end{pmatrix}.$$

with right-hand side

$$-\begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -X & 0 & I \end{pmatrix} \begin{pmatrix} R_d \\ r_p \\ R_{ZX} - \mu e \end{pmatrix} = \begin{pmatrix} -R_d \\ -r_p \\ XR_d - R_{ZX} \end{pmatrix}$$

Step 2: Eliminate Δx from row 2

(and scale row 3)

$$\begin{aligned} F_n := P_n K &:= \begin{pmatrix} I & 0 & 0 \\ 0 & I & -AZ^{-1} \\ 0 & 0 & Z^{-1} \end{pmatrix} \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & -XA^T & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & A^T & I_n \\ 0 & AZ^{-1}XA^T & 0 \\ I_n & -Z^{-1}XA^T & 0 \end{pmatrix} \end{aligned}$$

$AZ^{-1}XA^T$ can have:

- uniformly bounded condition number, e.g. Güler et al 1993
- structured singularity, e.g. S. Wright 95,97/ M. Wright 1999

But $\text{cond}(F_n) \rightarrow \infty$

The right-hand side becomes

$$-P_n P_Z \begin{pmatrix} R_d \\ r_p \\ R_{ZX} \end{pmatrix} = \begin{pmatrix} -R_d \\ -r_p + A(x - Z^{-1} X R_d - \mu Z^{-1} e) \\ Z^{-1} X R_d - x + \mu Z^{-1} e \end{pmatrix}$$

Ill-conditioning

Proposition The condition number of $F_n^T F_n$ diverges to infinity if $x(\mu)_i / z(\mu)_i$ diverges to infinity, for some i , as μ converges to 0. The condition number of $(F'_\mu)^T F'_\mu$ is uniformly bounded if there exists a unique primal-dual solution.

PROOF: Note that

$$F_n^T F_n = \begin{pmatrix} I_n & -Z^{-1} X A^T \\ -A X Z^{-1} & (A Z^{-1} X A^T)^2 + A Z^{-2} X^2 A^T \\ 0 & A^T \\ & I_n \end{pmatrix}$$

By interlacing of eigenvalues, ...

Corollary The condition number of F_n is at least $O(1/\mu)$.

EXAMPLE

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, c = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, b = 1,$$

$$x^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, y^* = -1, z^* = \begin{pmatrix} 0 \\ 2 \end{pmatrix};$$

initial points:

$$x = \begin{pmatrix} 9.183012e - 001 \\ 1.356397e - 008 \end{pmatrix}, z = \begin{pmatrix} 2.193642e - 008 \\ 1.836603e + 000 \end{pmatrix},$$

$$y = -1.163398e + 000.$$

residuals and duality gap:

$\|r_b\| = 0.081699$, $\|R_d\| = 0.36537$, $\mu = x^T z / n = 2.2528e - 008$
5 decimals rounding before/after arithmetic
centering with $\sigma = .1$

BUT: residuals are NOT order μ .

search directions found

using :

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} 8.17000e - 02 \\ -1.35440e - 08 \\ 1.63400e - 01 \\ -2.14340e - 08 \\ 1.63400e - 01 \end{pmatrix} ;$$

full matrix F'_μ and

$$\text{backsolve matrix : } \begin{pmatrix} -6.06210e + 01 \\ -1.35440e - 08 \\ 1.63400e - 01 \\ 0.00000e + 00 \\ 1.63400e - 01 \end{pmatrix}$$

error in Δy is small;
but error after backsubstitution for $(\Delta x)_1$ is **large**.

$$\begin{pmatrix} AZ^{-1}XA^T \\ -Z^{-1}XA^T \end{pmatrix} = \begin{pmatrix} 4.18630e + 07 \\ -4.18630e + 07 \\ -7.38540e - 09 \end{pmatrix}$$

Alternate Second Step; Stable Reduction

Assuming! $A = [I_m \ E]$.

Partition diagonal matrix Z, X using vectors

$$z = \begin{pmatrix} z_m \\ z_v \end{pmatrix}, x = \begin{pmatrix} x_m \\ x_v \end{pmatrix}, XA^T = \begin{pmatrix} X_m \\ X_v E^T \end{pmatrix}$$

$$F_s : = P_s K = \begin{pmatrix} I_n & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 \\ 0 & -Z_m & I_m & 0 \\ 0 & 0 & 0 & I_v \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ I_m & E & A^T \\ Z_m & 0 & 0 \\ 0 & Z_v & -X_m E^T \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & A^T & I_n \\ I_m & E & 0 & 0 \\ 0 & -Z_m E & -X_m E^T & 0 \\ 0 & Z_v & -X_v E^T & 0 \end{pmatrix}.$$

The right-hand side becomes

$$\begin{aligned}
 -P_s P_Z \begin{pmatrix} A^T y + z - c \\ Ax - b \\ ZXe - \mu e \end{pmatrix} &= -P_s \begin{pmatrix} R_d \\ r_p \\ -XR_d + ZXe - \mu e \end{pmatrix} \\
 &= \begin{pmatrix} -R_d \\ -r_p \\ -Z_m r_p - X_m(R_d)_m + Z_m X_{me} - \mu e \\ -X_v(R_d)_v + Z_v X_{ve} - \mu e \end{pmatrix}
 \end{aligned}$$

Summary: Path-following; **NOT Interior-point**

- staying interior is a heuristic for staying within a neighbourhood of the central path
- staying interior is required for numerical accuracy when solving the *current* ill-conditioned reduced systems

(Nearest) Euclidean Distance Matrix Completion using SDP

Given:

pre-distance matrix $A \in \mathcal{S}^n$ (nonnegative with zero diagonal)

weight matrix $H \in \mathcal{S}^n$:

$$\text{(NEDM)} \quad \mu^* = \min \frac{1}{2} \|H \circ (A - D)\|_F^2 \text{ subject to: } D \in \text{EDM}$$

$\text{EDM} = \{D = (d_{ij}) \in \mathcal{S}^n : d_{ij} = \|x_i - x_j\|^2, \text{ for some } x_i \in \mathfrak{X}^k\}$, k is *embedding dimension*

◦ denotes Hadamard (elementwise) matrix product

Applications

e.g. molecular conformation problems in chemistry;
multidimensional scaling and multivariate analysis problems in
statistics; genetics, geography,

Mixed-Cone Formulation

direct approach using a mixed SDP and second-order (or Lorentz) cone problem:

$$\begin{array}{ll} \min & \alpha \\ \text{s.t.} & Y = H \circ (\mathcal{L}(X) - A), \quad \|Y\|_F \leq \alpha \\ & X \in \mathcal{S}^{n-1}, Y \in \mathcal{S}^n, X \in \text{SDP} \end{array}$$

where $X \in \text{SDP} \Rightarrow \mathcal{L}(X) \in \text{EDM}$

(Public domain software packages are available, but problem size becomes large)

Connection between SDP and EDM

$$B = [x_1 \ x_2 \ \dots \ x_n], \quad k \times n$$

$$D_{ij} = \|x_i - x_j\|^2 = -2x_i^T x_j + \|x_i\|^2 + \|x_j\|^2$$

$$D = -2B^T B + e \left(\text{diag}(B^T B) \right)^T + \left(\text{diag}(B^T B) \right)^T e^T$$

With $X = B^T B \succeq \underline{0}$

Operator Notation:

us2vec , us2Mat , svec , sMat

$$x = \text{svec}(X) \in \mathbb{R}^{\binom{n+1}{2}}, \quad X = \text{sMat}(x)$$

$\sqrt{2}$ times vector (columnwise) from upper-triang of X .

$$\binom{n+1}{2} = n(n+1)/2; \sqrt{2} \text{ guarantees isometry.}$$

$\text{sMat} := \text{svec}^{-1}$ mapping into $\mathcal{S}^n$

adjoint transformation $\text{sMat}^* = \text{svec}$:

$$\begin{aligned} \langle \text{sMat}(v), S \rangle &= \text{trace sMat}(v) S \\ &= v^T \text{svec}(S) = \langle v, \text{svec}(S) \rangle \end{aligned}$$

Characterization of EDM using SDP

D is EDM $(\subset \mathcal{S}^n)$

iff

$$D = \mathcal{L}(X) := \begin{pmatrix} 0 & \text{diag}(X)^T \\ \text{diag}(X) & \text{diag}(X)e^T + e\text{diag}(X)^T - 2X \end{pmatrix},$$

for some $X \succeq 0, X \in \mathcal{S}^{n-1}$

(e is vector of ones)

$$\mathcal{L} : \mathcal{S}^{n-1} \rightarrow \mathcal{S}^n, \quad \mathcal{L}(\mathcal{S}_+^{n-1}) = \text{EDM}$$

adjoint/generalized inverse

with partition:

$$D = \begin{bmatrix} \alpha & d^T \\ d & \bar{D} \end{bmatrix},$$

where $\alpha \in \mathbb{R}$

$$\mathcal{L}^*(D) = 2 \left(\text{Diag}(d) + \text{Diag}(\bar{D}e) - \bar{D} \right)$$

$$\mathcal{L}^\dagger(D) = \frac{1}{2} \left(de^T + ed^T - \bar{D} \right)$$

$$\mathcal{L}^*, \mathcal{L}^\dagger : \mathcal{S}^n \rightarrow \mathcal{S}^{n-1}, \quad \mathcal{L}^\dagger(EDM) = \mathcal{S}_+^{n-1}$$

Duality and Optimality Conditions

(using $X = \text{sMat}(x) + I$) an equivalent problem is:

$$\mu^* := \min \frac{1}{2} \|H \circ (A - \mathcal{L}(X))\|_F^2 \quad \text{subject to} \quad X \succeq 0$$

strong (Lagrangian) duality holds (Slater's holds for primal and holds for dual if the graph is complete)

$$\mu^* = \nu^* := \max_{\Lambda \succeq 0} \min_X \frac{1}{2} \|H \circ (A - \mathcal{L}(X))\|_F^2 - \text{trace } \Lambda X$$

Wolfe dual and optimality conditions

With

$$C := \mathcal{L}^*(H^{(2)} \circ A),$$

optimality conditions are:

$$\begin{aligned} X &:= \text{sMat}(x) \succeq 0 && \text{(primal feasibility)} \\ \Lambda &:= \mathcal{L}^* \left\{ H^{(2)} \circ (\mathcal{L}(X)) \right\} - C, \quad \Lambda \succeq 0 && \text{(dual feasibility)} \\ \Lambda X &:= 0 && \text{(compl. slack.)} \end{aligned}$$

equivalent dual problem:

$$\begin{aligned} &\max \quad \frac{1}{2} \|H \circ (A - \mathcal{L}(X))\|_F^2 - \text{trace } \Lambda X \\ &\text{subject to} \quad \Lambda = \mathcal{L}^* \left\{ H^{(2)} \circ (\mathcal{L}(X)) \right\} - C \\ &\quad \Lambda \succeq 0. \end{aligned} \tag{0.1}$$

Bilinear System

eliminate Λ
exact primal-dual feasibility during iterations
full rank Jacobian at optimality.
single bilinear (perturbed) equation in x ;

$$F_{\mu}(x) : \mathbb{R}^{\binom{n}{2}} \rightarrow \mathcal{M}^{n-1}$$

$$F_{\mu}(x) := \left[\mathcal{L}^* \left\{ H^{(2)} \circ (\mathcal{L}(X)) \right\} - C \right] X - \mu I = 0$$

typical SDP - overdetermined system of bilinear equations
current approach is to symmetrize - which results in
ill-conditioning! from rank deficient Jacobian at optimality.
BUT, here, no symmetrization used;
solve using (an inexact) Gauss-Newton method - with PCG

Linearization

Let $\mathcal{W}(x) := \mathcal{L}^* \left\{ H^{(2)} \circ (\mathcal{L}(x)) \right\}$

Linearization for search direction Δx at current $x = \text{svec}(X)$:

$$F'_\mu(x) \Delta x = [\mathcal{W}(x) - C] \Delta x + [\mathcal{W}(\Delta x)] X$$

This is a linear, full rank, overdetermined system.

Our search direction Δx is its (approx.) least squares solution.

Algorithm: p-d i-e-p framework

- Initialization:

- Input data: a pre-distance $n \times n$ matrix A
- Positive tolerances:

ϵ_1 (stopping), ϵ_2 (loss accuracy), ϵ_3 (crossover),

- Find initial strictly feasible points: both

$X^0, \Lambda^0 := (\mathcal{W}(X) - C) \succ 0; \mu > 0$

- Set initial parameters:

$\text{gap} = \text{trace } \Lambda^0 X^0; \mu = \text{gap}/n; \text{objval} = f(X^0); k = 0.$

Algorithm continued 1

- while $\min\{\frac{\text{gap}}{\text{objval}+1}, \text{objval}\} > \epsilon_1$
- solve lss for search direction (accuracy $\epsilon_2 \min\{\mu, 1\}$)

$$F'_{\sigma\mu}(x^k) \left(\Delta x^k \right) = -F_{\sigma\mu}(x^k),$$

where σ_k centering, $\mu_k = \frac{1}{n} \text{trace}(\mathcal{W}(X^k) - C) X^k$

$$X^{k+1} = X^k + \alpha_k \Delta X^k, \quad \alpha_k > 0,$$

so that both X^{k+1} , $(\mathcal{W}(X^{k+1}) - C) \succeq 0$
 ($\alpha_k = 1$ after the crossover.)

- update

$$k \leftarrow k + 1 \quad \text{and then}$$

$$\sigma_k \left(\text{set } \sigma_k = 0 \text{ if } \min\left\{ \frac{\text{gap}}{\text{objval}+1}, \text{objval} \right\} > \epsilon_2 \right)$$

Algorithm continued 2

- while $\min\{\frac{\text{gap}}{\text{objval}+1}, \text{objval}\} > \epsilon_1$

- solve lss for search direction

-

- update

$k \leftarrow k + 1$ and then

$\sigma_k \left(\text{set } \sigma_k = 0 \text{ if } \min\{\frac{\text{gap}}{\text{objval} + 1}, \text{objval}\} < \epsilon_3 \right)$

- end (while).

- **Conclusion:** $D = \mathcal{L}(X) \in \text{EDM}$ is approx. to A

Crossover

After the **crossover**, centering $\sigma = 0$ and steplength $\alpha = 1$, we get q-quadratic convergence; allows for *warm starts*.

Long steps can be taken *beyond* the positivity boundary. (tests show improved convergence rates)

Preconditioning

where

$$(\Lambda + \mathcal{X}\mathcal{W}) P^{-1}(\widehat{\Delta}x) = -F_{\mu}(x),$$

$$\widehat{\Delta}x = P(\Delta x)$$

Diagonal Preconditioning

Optimal scaling Dennis and W. (1993) full rank matrix $A \in \mathbb{R}^{m \times n}$, $m \geq n$, with condition number $\omega(K) := n^{-1} \text{trace}(K) / \det(K)^{1/n}$, the optimal scaling

$\min \omega((AD)^T(AD))$ subject to: D positive and diagonal

solution: $d_{ii} = 1 / \|A_{:i}\|_2$, $i = 1, \dots, n$
explicit expressions for preconditioner
inexpensive

Explicit Preconditioning

diagonal operator P ; evaluate using columns of $F'_\mu(v)$.
 $k \cong (i, j)$, $1 \leq i < j \leq n$, strictly upper triangular part

$$\begin{aligned} \|(\Lambda + \mathcal{X}\mathcal{W})(e_k)\|_F^2 &= \|\Lambda(e_k)\|_F^2 + \|(\mathcal{W}(e_k))X\|_F^2 \\ &\quad + \langle \Lambda(E_{ij}), (\mathcal{W}(E_{ij}))X \rangle, \end{aligned}$$

where

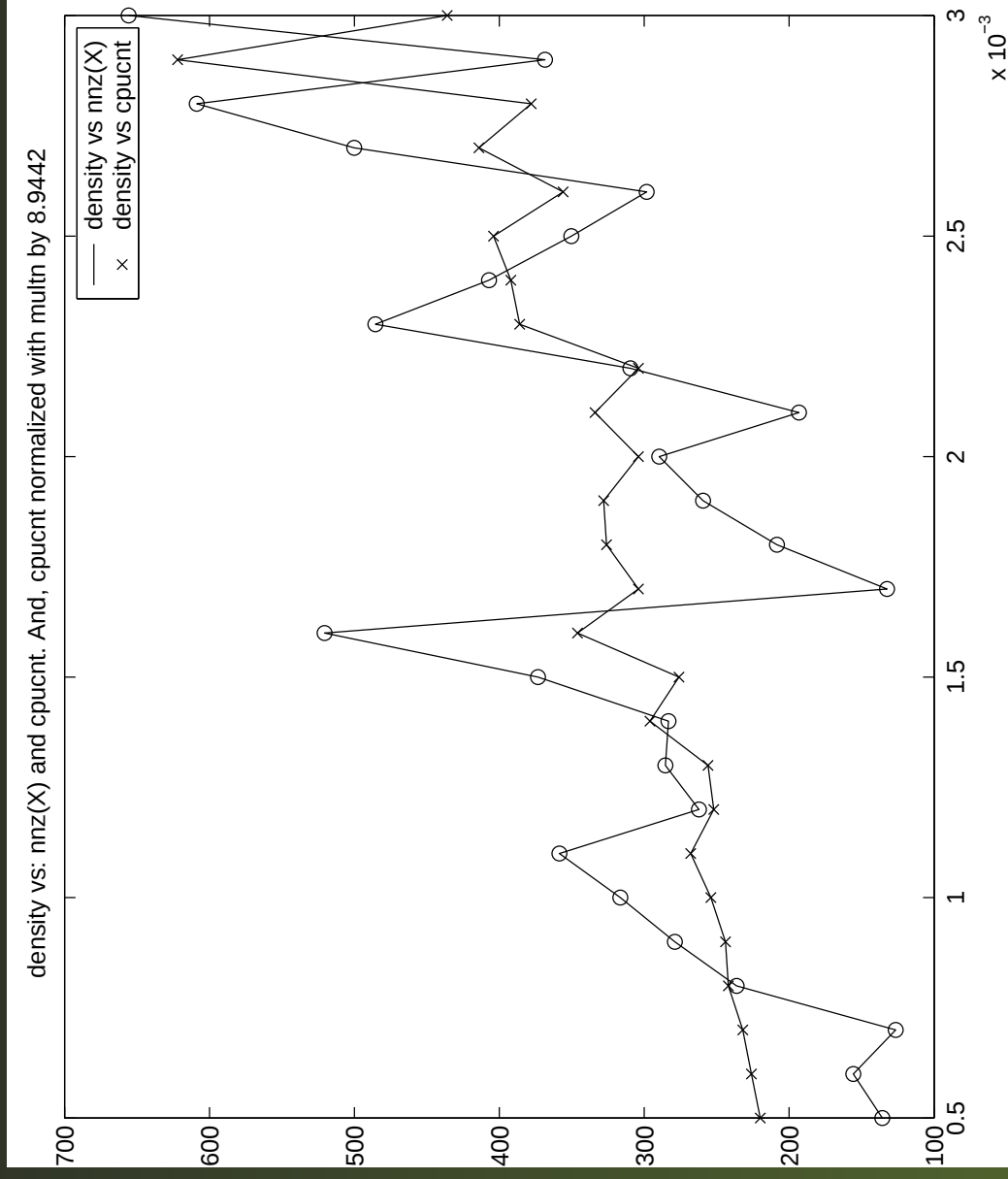
$$\Lambda(e_k) = \begin{cases} \frac{1}{\sqrt{2}} \left(\Lambda_{:i} e_j^T + \Lambda_{:j} e_i^T \right), & \text{if } i < j \\ \Lambda_{:i} e_i^T, & \text{if } i = j. \end{cases}$$

and $\mathcal{X}\mathcal{W}$ inexpensive - 50% reduction in LSQR iterations

Numerical Tests

Pentium 4; MATLAB 6.5; 1 GIG RAM.
crossover heuristic: relative duality gap $< .1$.
Stopping criteria (relative duality gap) $< \epsilon_1 = 1e - 10$.
(But - average accuracy attained $1e - 13$, q-quadratic convergence.)

density .0005:.001:.003, CPU times
and nnz(Λ), n=200



Conclusion

Gauss-Newton direction:

Advantages/Disadvantages:

- Robust, warm starts are simple, longer steps
- exact primal and dual feasibility at each iteration
- Can apply CG-type approaches
- q-quadratic convergence
- scale-invariant on the right

Future:

- Need large sparse QR efficient as Cholesky predictor-corrector

EDM Completion Problem, EDMC

- given certain fixed elements of a EDM matrix A
- the other elements are unknown (free)
- complete this matrix to an EDM

$$\mathcal{S} = \{(i, j) : A_{i,j} = \frac{1}{\sqrt{2}}b_k \text{ is known, fixed, } i < j\}, \quad |S| = m,$$

$$\begin{array}{ll} \mu^* := & \min \quad f(X) := \frac{1}{2}\|X\|_F^2 \\ \text{(EDMC)} & \text{subject to} \quad \mathcal{A}(X) = b \\ & X \succeq 0, \end{array}$$

constraint $\mathcal{A} = \mathcal{I} \cdot \mathcal{L} : \mathcal{S}^{n-1} \rightarrow \mathbb{R}^{|S|}$ yields interpolation conditions

$$\mathcal{A}(X)_{ij} = \text{trace } E_{ij} \mathcal{L}(X) = b_k, \quad \forall k \cong (ij) \in S,$$

Duality/Optimality for EDMC

- strict convexity, coercivity **implies** compact level sets
 - EDMC attained and no duality gap (actually primal and dual attainment)
- Lagrangian dual

$$\mu^* = \nu^* := \max_{\Lambda \succeq 0, y \in \mathbb{R}^{|s|}} \min_X \frac{1}{2} \|X\|_F^2 + y^T (b - \mathcal{A}(X)) - \text{trace } \Lambda X$$

characterization of optimality

THEOREM Suppose that the feasible set of EDMC is not the empty set. Then the optimal solution of EDMC is $D = \mathcal{L}([\mathcal{A}^*(y)]_+)$, where y is the unique solution of the single equation

$$\mathcal{A}([\mathcal{A}^*(y)]_+) = b,$$

and B_+ denotes the projection of the symmetric matrix $B \in \mathcal{S}^{n-1}$ onto the cone $\mathcal{P}_{n-1}$.

Proof

optimality conditions after differentiation

$$\begin{array}{lll} X = \mathcal{A}^*(y) + \Lambda \succeq 0, & \Lambda \succeq 0, & \text{dual feasibility} \\ \mathcal{A}(X) = b & & \text{primal feasibility} \\ \Lambda X = 0 & & \text{complementary slackness} \end{array}$$

This means that $\mathcal{A}^*(y) = X - \Lambda$, where both $X \succeq 0$, $\Lambda \succeq 0$, and $\Lambda X = 0$. Therefore the three symmetric matrices $W = \mathcal{A}^*(y)$, X , Λ are mutually diagonalizable. We write $X = P D_X P^T$, $\Lambda = P D_\Lambda P^T$, i.e. we conclude that $W = \mathcal{A}^*(y) = P (D_X - D_\Lambda) P^T$, $D_X D_\Lambda = 0$. Therefore $[\mathcal{A}^*(y)]_+ = P D_X P^T = X$. ■

Efficient/Explicit Solution if $y \succeq 0$

large class (**generic?**) can be solved in polytime.

COROLLARY The linear operator $\mathcal{A}$ is onto and $\mathcal{A}\mathcal{A}^*$ is nonsingular. Suppose that $y = (\mathcal{A}\mathcal{A}^*)^{-1}b \in \mathbb{R}_+^m$. Then

$$D = \mathcal{L}(\mathcal{A}^*(y))$$

is the unique solution of EDMC.

PROOF: That $\mathcal{A}$ is onto follows from the definitions.

If $y \succeq 0$, then the matrix $\mathcal{I}(y) \succeq 0$ with 0 diagonal. Therefore, $X = \mathcal{L}^*(\mathcal{I}(y))$ is diagonally dominant with nonnegative diagonal, i.e. $X \succeq 0$ by Gersgorin's disk theorem. This implies that D is a distance matrix and it satisfies the interpolation conditions, i.e. it satisfies the optimality conditions in the Theorem. ■

Numerics: dim vs dens with # of failures in 100 tests

though $y = A^\dagger b \geq 0$ does **not** hold in general, we still get
a distance matrix D , i.e. $\mathcal{A}^*(y) \succeq 0$.
 $n = 10 : 10 : 100$; density $.1 : .1 : .8$.

$n \backslash$ density	.1	.2	.3	.4	.5	.6	.7	.8
10	19	27	29	25	32	27	20	38
20	6	20	23	22	27	21	28	28
30	8	8	9	9	11	16	17	24
40	2	2	6	5	14	17	20	17
50	2	0	2	8	7	8	15	12
60	1	1	1	1	3	8	15	11
70	2	0	3	1	5	7	6	15
80	1	0	0	4	2	4	9	9
90	1	0	0	1	3	2	5	6
100	0	0	0	0	1	6	5	5