Department of Math. & Stats.

A Stable Iterative Method for Linear Programming

Thursday, Oct. 14 2004

A Stable Iterative Method for Linear Programming

Henry Wolkowicz

COMBINATORICS & OPTIMIZATION

University of Waterloo

(with: Maria Gonzalez-Lima, Hua Wei)

OUTLINE

- Background on LP and SDP; Notation and Motivation
- Robust, ('non-interior') path-following algorithm for LP (and SDP)
- Details of algorithm on LP with numerics
- Application to Nearest Euclidean Distance Matrix Problem, Numerics, Comparisons with a dual algorithm (time permitting)

Notation and Motivation

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

where: $f : \mathcal{S}^n \longrightarrow \mathbb{R}$ convex function

$\mathcal{S}^n$ $n \times n$ real symmetric matrices

$\mathcal{A} : \mathcal{S}^n \longrightarrow \mathbb{R}^m$ linear operator,

$X(\succeq) \succ 0$ denotes positive (semi)definite

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

Linear Primal-Dual Pair of SDPs

(looks/behaves like Linear Program, LP)

$$
\text{(PSDP)} \qquad \begin{array}{cl} \min & \langle C, X\rangle = \operatorname{trace} CX \\ \text{subject to} & \mathcal{A}X = b \\ & X \succeq 0, \end{array}
$$

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

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

(Perturbed) Optimality Conditions

For *barrier parameter* $\mu > 0$:

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

For SDP:

$$F_\mu : \mathcal{S}^n \times \Re^m \times \mathcal{S}^n \to \mathcal{S}^n \times \Re^m \times \mathcal{M}^n$$

i.e. overdetermined nonlinear system

(Non) Interior Path-Following on LP

$$
\text{(LP)} \qquad \begin{array}{rll} p^* := & \min & c^T x \quad (\text{or } \langle c, x \rangle) \\ & \text{s.t.} & Ax = b \\ & & x \geq 0 \quad (\text{or } x \succeq 0) \end{array}
$$

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

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

Central Path/Path Following

dual log-barrier problem with parameter $\mu > 0$ is

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

stationary point of the Lagrangian / optimality conditions

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

$X = \operatorname{Diag}(x), Z = \operatorname{Diag}(z)$

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

Ill-Conditioning

As $\mu \to 0$, Jacobian $F'_\mu(x,y,z)$ grows 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):

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

$$=: \begin{pmatrix} R_d \\ r_p \\ R_{ZX} \end{pmatrix}$$

Linearization

Special structure of linearized system can be exploited;

linearization for the Newton direction $\Delta s = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$ is

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

Damped Newton Method

Damped Newton steps

$$x \leftarrow x + \alpha_p \Delta x, \quad y \leftarrow y + \alpha_d \Delta y, \quad z \leftarrow z + \alpha_d \Delta z,$$

are taken that *backtrack* from the nonnegativity boundary to maintain the positivity/interiority, $x > 0, z > 0$.
On central path, $F_\mu(x, y, z) = 0$,

$$\mu = \frac{1}{n}\mu e^T e = \frac{1}{n} e^T Z X e = \frac{1}{n} z^T x = \frac{1}{n}(\text{ duality gap}),$$

$$\begin{aligned} \text{barrier parameter } \mu \cong \text{duality gap} &= c^T x - b^T y \\ &= x^T \left(c - A^T y\right) \\ &= x^T z. \end{aligned}$$

if exact feasibility holds!

SDP Case

overdetermined system in SDP case:

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

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 is linearization of:
$ZX + XZ - 2\mu I = 0$ (AHO search direction)

Reduction/Block Elimination for the Normal Equations, NEQ

Step 1 (Eliminate Δz):

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

We let

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

with right-hand side

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

Normal Equations, NEQ

Step 2 (Eliminate Δx):

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

$$= \begin{pmatrix} 0 & A^T & I_n \\ 0 & AZ^{-1}XA^T & 0 \\ I_n & -Z^{-1}XA^T & 0 \end{pmatrix}$$

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

- uniformly bounded condition number, e.g. Güler et al 1993
- structured singular values, e.g. M. Wright 1997

But $\mathrm{cond}(F_n) \to \infty$.

The right-hand side

$$-P_n P_Z \begin{pmatrix} R_d \\ r_p \\ R_{ZX} \end{pmatrix} =$$

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

Ill-Posed System

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

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

By interlacing of eigenvalues, ... ■

Condition Number Growth

observe

• condition number of $F_n^T F_n$ is greater than the largest eigenvalue of the block $AZ^{-2}X^2A^T$;

• equivalently, $\frac{1}{\mathrm{cond}(F_n^T F_n)}$ is smaller than the reciprocal of this largest eigenvalue.

• If x, z stay in neighbourhood of central path, then $\min_i(z_i/x_i)$ is $O(\mu)$.

THEN:

reciprocal of the condition number of F_n is $O(\mu)$.

Ex. Catastrophic Roundoff Error

$A = (\,1 \quad 1\,)\,, c = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, b = 1.$

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

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:

$\|r_b\| = 0.081699, \quad \|R_d\| = 0.36537, \quad \mu = x^T z/n = 2.2528e-008$

5 decimals rounding before/after arithmetic

centering with $\sigma = .1$

BUT: residuals are NOT order μ.

Search Direction

search direction is found using:

(i) full matrix F'_μ; (ii) backsolve matrix F_n

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} 8.17000e-02 \\ -1.35440e-08 \\ 1.63400e-01 \\ -2.14340e-08 \\ 1.63400e-01 \end{pmatrix}; \quad = \begin{pmatrix} -6.06210e+01 \\ -1.35440e-08 \\ 1.63400e-01 \\ 0.00000e+00 \\ 1.63400e-01 \end{pmatrix}$$

error in Δy is small; 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

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

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

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

The right-hand side

$$
-P_s P_Z \begin{pmatrix} A^T y + z - c \\ Ax - b \\ ZXe - \mu e \end{pmatrix} = -P_s \begin{pmatrix} R_d \\ r_p \\ -XR_d + ZXe - \mu e \end{pmatrix}
$$

$$
= \begin{pmatrix} -R_d \\ -r_p \\ -Z_m r_p - X_m (R_d)_m + Z_m X_m e - \mu e \\ -X_v (R_d)_v + Z_v X_v e - \mu e \end{pmatrix}
$$

Equivalent View of Stable Linearization

find (hopefully sparse) representation
range of N is nullspace of A

$$Ax = b \quad \text{if and only if} \quad x = \hat{x} + Nv, \;\text{for some } v \in \Re^{n-m}.$$

e.g. symmetric form

$$Ex_v \leq b, x_v \geq 0, \quad E \in \Re^{m\times(n-m)} \quad \Leftrightarrow \quad x_m + Ex_v = b, x \geq 0$$

Therefore assume E sparse and

$$A = \begin{pmatrix} I_m & E \end{pmatrix}, \quad N = \begin{pmatrix} -E \\ I_{n-m} \end{pmatrix}.$$

Substitute for z, x; Eliminate

Theorem The primal-dual variables x, y, z, with $x = \hat{x} + Nv \geq 0,\ z = c - A^T y \geq 0$, are optimal for (LP),(DLP) if and only if they satisfy the single bilinear optimality equation

$$F(v, y) := \mathrm{Diag}\,(c - A^T y)\mathrm{Diag}\,(\hat{x} + Nv)e = 0.$$

■

A single (perturbed) optimality conditions to use for the primal-dual method,

$$F_\mu(v, y) := \mathrm{Diag}\,(c - A^T y)\mathrm{Diag}\,(\hat{x} + Nv)e - \mu e = 0.$$

Linearization for Search Direction, Δs

$$-F_\mu(v,y) = F'_\mu(v,y)\Delta s \quad \Delta s := \begin{pmatrix} \Delta v \\ \Delta y \end{pmatrix}$$

Jacobian matrix is

$$F'_\mu(v,y) = \begin{pmatrix} \mathrm{Diag}\,(c - A^T y)N & -\mathrm{Diag}\,(\hat{x} + Nv)A^T \end{pmatrix}$$

system to solve for search direction is

$$-F_\mu(v,y) = \mathrm{Diag}\,(c - A^T y)N\Delta v - \mathrm{Diag}\,(\hat{x} + Nv)A^T\Delta y.$$

first part usually large, $n-m$ variables
second part usually small, only m variables.

Well Conditioned

Theorem Consider the primal-dual pair (LP),(DLP). Suppose that A is onto (full rank), the range of N is the null space of A, N is full column rank, and (x, y, z) is the *unique* primal-dual optimal solution. Then the matrix of the linear system

$$
\begin{aligned}
-F_\mu &= F'_\mu \Delta s \\
&= ZN\Delta v - XA^T \Delta y
\end{aligned}
\tag{0.1}
$$

(F'_μ is Jacobian of F_μ) is nonsingular.

Proof of Theorem

PROOF: Suppose that $F'_\mu(v,y)\Delta s = 0$. We need to show that $\Delta s = (\Delta v, \Delta y) = 0$.

Let $\mathcal{B}$ and $\mathcal{N}$ denote the set of indices j such that $x_j = \hat{x}_j + (Nv)_j > 0$ and set of indices i such that $z_i = c_i - (A^T y)_i > 0$, respectively. Under the nondegeneracy (uniqueness) and full rank assumptions, we get $\mathcal{B} \bigcup \mathcal{N} = \{1, ...n\}$, $\mathcal{B} \bigcap \mathcal{N} = \emptyset$, and the cardinalities $|\mathcal{B}| = m$, $|\mathcal{N}| = n - m$. Moreover, the submatrix $A_{\mathcal{B}}$, formed from the columns of A with indices in $\mathcal{B}$, is nonsingular.

By our assumption and the linearization definition, we get that

$$\left(F'_\mu(v,y)\Delta s\right)_k = (c - A^T y)_k (N\Delta v)_k - (\hat{x} + Nv)_k (A^T \Delta y)_k = 0, \quad \forall k.$$

Proof of Theorem cont...

From the definitions of $\mathcal{B}, \mathcal{N}$, this implies that

$$(0.2) \qquad (A^T\Delta y)_j = 0, \forall j \in \mathcal{B}, \qquad (N\Delta v)_i = 0, \forall i \in \mathcal{N}.$$

The left part of (0.2) implies $A_{\mathcal{B}}^T\Delta y = 0$, i.e. we obtain $\Delta y = 0$. It remains to show that $\Delta v = 0$. From the definition of N we have $AN = 0$. Therefore, using the right part of (0.2) implies

$$\begin{aligned} 0 &= \begin{pmatrix} A_{\mathcal{B}} & A_{\mathcal{N}} \end{pmatrix} \begin{pmatrix} (N\Delta v)_{\mathcal{B}} \\ (N\Delta v)_{\mathcal{N}} \end{pmatrix} \\ &= A_{\mathcal{B}}(N\Delta v)_{\mathcal{B}} + A_{\mathcal{N}}(N\Delta v)_{\mathcal{N}} \\ &= A_{\mathcal{B}}(N\Delta v)_{\mathcal{B}}. \end{aligned}$$

Proof of Theorem cont...

By the right part of (0.2) and the nonsingularity of $A_{\mathcal{B}}$, we get

$$N\Delta v = 0.$$

Now, full rank of N implies $\Delta v = 0$. ■

Primal-Dual Algorithm

follow usual primal-dual interior-point framework: Newton's method applied to perturbed system of optimality conditions with damped step lengths for maintaining nonnegativity constraints.

Differences:

- eliminate, primal-dual linear feasibility (exact feas. maintained)
- search direction found using PCG (LSQR)
- no backtracking to preserve sufficient positivity of z, x
- crossover step (affine scaling, perturbation parameter $\mu = 0$ and full Newton step close to optimum; exterior method; NLF - Newton Liberation Front)
- identify zero values for x, *purification step*.

Preconditioning Techniques

$$Z := Z(y) = \mathrm{Diag}\,(c - A^T y), \qquad X := X(v) = \mathrm{Diag}\,(\hat{x} + Nv)$$

$$J := F'_\mu(v, y) = \begin{pmatrix} ZN & -XA^T \end{pmatrix} \quad \text{Jacobian}$$

find a preconditioner *simple nonsingular* M such that JM^{-1} is well conditioned and solve better conditioned systems $JM^{-1}\Delta q = -F_\mu$ and $M\Delta s = \Delta q$.

$$\text{We look for:} \quad M^T M \cong J^T J$$

LSQR (Paige-Saunders) implicitly solves normal equations

$$J^T J ds = -J^T F'_\mu$$

Optimal Diagonal Column Preconditioning

simplest of preconditioners; given square matrix K

$$\omega(K) = \frac{\mathrm{trace}\,(K)/n}{\det(K)^{1/n}} \quad \text{condition number}$$

If $M = \arg\min \omega((JD)^T(JD))$ over all positive diagonal matrices D then (Dennis-W. 1990)

$$M_{ii} = 1/\|J_{:i}\| \quad i\text{-th column norm}$$

Partial (Block) Cholesky Preconditioner

$$J^TJ = \begin{pmatrix} N^TZ^2N & -N^TZXA^T \\ -AXZN & AX^2A^T \end{pmatrix}.$$

For z, x near central path, i.e. $ZX \cong \mu I$, off diagonal terms $\cong 0$
block (partial) Cholesky preconditioning is good preconditioner
Q-less QR factorization $Q_ZR_Z = ZN,\ Q_XR_X = XA^T$

$$R_Z^TR_Z = N^TZ^2N, \quad R_X^TR_X = AX^2A^T.$$

$$J^TJ \cong M^TM, \quad M = \begin{pmatrix} R_Z & 0 \\ 0 & R_X \end{pmatrix}$$

expensive!

Crossover Criteria/Quadratic Convergence

assume nonsingular Jacobian at optimality (so unique primal and dual solutions, s^*)

standard theory for Newton's method:

$\exists$ quadratic convergence neighbourhood of s^*

Theorem (Kantorovich) Let $r > 0$, $s_0 \in \mathbb{R}^n$, $F : \mathbb{R}^n \to R^n$, and assume that F is continuously differentiable in $\mathcal{N}(s_0, r)$. Assume for a vector norm and the induced operator norm that $J \in \text{Lip}_\gamma(\mathcal{N}(s_0, r))$ with $J(s_0)$ nonsingular, and that there exist constants $\beta, \eta \geq 0$ such that

$$\|J(s_0)^{-1}\| \leq \beta, \;\; \|J(s_0)^{-1} F(s_0)\| \leq \eta.$$

Define $\alpha = \beta\gamma\eta$. If $\alpha \leq \frac{1}{2}$ and $r \geq r_0 := (1 - \sqrt{1 - 2\alpha}\,)/(\beta\gamma)$.

Quadratic Convergence cont...

Then the sequence $\{s_k\}$ produced by

$$s_{k+1} = s_k - J(s_k)^{-1}F(s_k), \quad k = 0, 1, \ldots,$$

is well defined and converges to s_*, a unique zero of F in the closure of $\mathcal{N}(s_0, r_0)$. If $\alpha < \frac{1}{2}$, then s_* is the unique zero of F in $\mathcal{N}(s_0, r_1)$, where $r_1 := \min[r, (1 + \sqrt{1 - 2\alpha})/(\beta\gamma)]$ and

$$\|s_k - s_*\| \le (2\alpha)^{2^k} \frac{\eta}{\alpha}, \quad k = 0, 1, \ldots,$$

■

Lipschitz Constant for Region of Convergence

Lemma The Jacobian

$$F'(v,y)\begin{pmatrix} \cdot \\ \cdot \end{pmatrix} := \begin{pmatrix} \mathrm{Diag}\,(c - A^T y) N \cdot & \mathrm{Diag}\,(\hat{x} + Nv) A^T \cdot \end{pmatrix}$$

is Lipschitz continuous with constant

$$\gamma = \sqrt{2}\|A\|\|N\|$$

with respect to (v, y) ■

Proof

PROOF: We let $\Delta s = \begin{pmatrix} \Delta v \\ \Delta y \end{pmatrix}$. Since

$$\begin{aligned} \|F'(s) - F'(\bar{s})\| &= \max \frac{\|(F'(s) - F'(\bar{s}))\Delta s\|}{\|\Delta s\|} \\ &= \max \frac{\|\mathrm{Diag}\,(A^T(y - \bar{y}))N\Delta v - \mathrm{Diag}\,(A^T\Delta y)N(v -}{\|\Delta s\|} \\ &\le \max \frac{\|A^T(y - \bar{y})\|\|N\Delta v\| + \|A^T\Delta y\|\|N(v - \bar{v})\|}{\|\Delta s\|} \\ &\le \|A\|\|N\|\|y - \bar{y}\| + \|A\|\|N\|\|v - \bar{v}\| \\ &\le \sqrt{2}\|A\|\|N\|\|s - \bar{s}\|. \end{aligned}$$

Therefore a Lipschitz constant is $\gamma = \sqrt{2}\|A\|\|N\|$. ■

Region of Quadratic Convergence

$\|[ZN \; - XA^T]^{-1}\| \leq \beta$, e.g. using smallest singular value

$$\|J^{-1}F_0(v,y)\| = \|[ZN \; - XA^T]^{-1}(-XZe)\| \leq \eta.$$

Theorem Suppose that

$$\alpha = \gamma\beta\eta < \frac{1}{2}.$$

Then the sequence s_k generated by

$$s_{k+1} = s_k - J(s_k)^{-1}F(s_k)$$

converges (quadratically) to s^*, the unique zero of F in the neighbourhood $\mathcal{N}(s_0, r_1)$. ■

Purify Step

- detect zero variables/active constraints at optimality
- use the Tapia indicators 1995,

$$\frac{(x_{k+1})_i}{(x_k)_i} \qquad \text{ratio of } i\text{-th component of iterates}$$

- perform a pivot step to eliminate these variables

Summary; Path-following and NOT Interior-point

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

  no loss of sparsity

  high accuracy solutions available if desired

  exact primal and dual feasibility at each iteration (true duality gap)

  warm starts

  fast convergence (no backtracking from boundary)

Numerical Tests

- randomly generated data with known optimum
- well-conditioned basis matrix
- stopping condition relative gap 10^{-12}
- MATLAB 6.5, Pentium 3 733MHz, 256MB RAM
- iterative approach; LSQR (Paige-Saunders); different preconditioners
- NEQ stalls with relative gap approximately 10^{-11} on many problems

NEQ vs Stable Method-Direct Solver

data	m	n	nnz(E)	cond$(A_{\mathcal{B}})$	cond(J)	NEQ		Stable direct	
						D_time	its	D_Time	its
1	100	200	1233	51295	32584	0.03	$*$	0.06	6
2	200	400	2526	354937	268805	0.09	6	0.49	6
3	200	400	4358	63955	185503	0.10	$*$	0.58	6
4	400	800	5121	14261771	2864905	0.61	$*$	3.66	6
5	400	800	8939	459727	256269	0.64	6	4.43	6
6	800	1600	10332	11311945	5730600	5.02	6	26.43	6
7	800	1600	18135	4751747	1608389	5.11	$*$	33.10	6

nnz(E) - number of nonzeros in E;

cond$(\cdot)$ - condition number; $J = (ZN \ - XA^T)$ at optimum;

D_time - avg. for search direction per iter.;

its - for interior point method

$*$ denotes NEQ stalls at 10^{-11}

Stable Method with LSQR and Two Precond.

data set	LSQR with ILU				LSQR with Diag			
	D_Time	its	L_its	Pre_time	D_Time	its	L_its	Pre_time
1	0.15	6	37	0.06	0.41	6	556	0.01
2	3.42	6	343	0.28	2.24	6	1569	0.00
3	2.11	6	164	0.32	3.18	6	1595	0.00
4	NA	Stalling	NA	NA	13.37	6	4576	0.01
5	NA	Stalling	NA	NA	21.58	6	4207	0.01
6	NA	Stalling	NA	NA	90.24	6	9239	0.02
7	NA	Stalling	NA	NA	128.67	6	8254	0.02

Same data sets as above;

two different preconditioners (diagonal and incomplete Cholesky with drop tolerance 0.001);

D_time - average time for search direction;

its - iteration number of interior point methods;

L_its - average number LSQR iterations per major iteration;

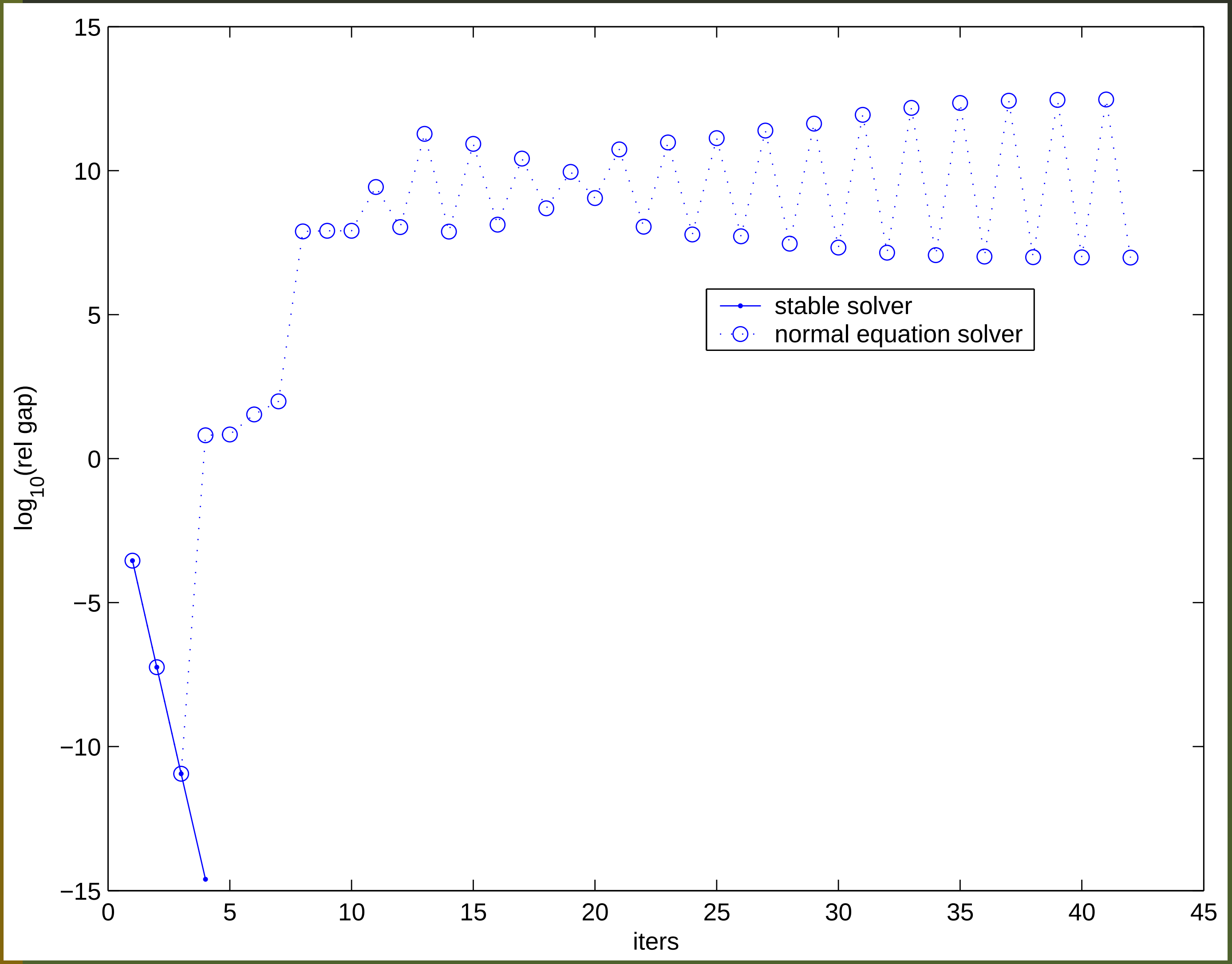
Pre_time - average time for preconditioner;

Stalling - LSQR cannot converge due to poor preconditioning.

LSQR with Block Cholesky preconditioner

data set	LSQR with block Chol. Precond.			
	D_Time	its	L_its	Pre_time
1	0.09	6	4	0.07
2	0.57	6	5	0.48
3	0.68	6	5	0.58
4	5.55	6	6	5.16
5	6.87	6	6	6.45
6	43.28	6	5	41.85
7	54.80	6	5	53.35

Iterations/Degeneracy



Sparse; Well conditioned $A_{\mathcal{B}}$

- about 3-4 nonzeros per row in E
- the Jacobian nonsingular at optimum.
- well-conditioned basis matrix, $A_{\mathcal{B}}$
- the same dimensions and two dense columns, while total number of nonzeros increases

- The loss in sparsity has essentially no effect on NEQ, since the ADA^T matrix is dense due to the two dense columns. But we can see the negative effect that the loss of sparsity has on the stable direct solver.

However, we see that for these problem instances, using LSQR with the stable system can be up to twenty times faster then NEQ solver.

Sparse; Well conditioned $A_{\mathcal{B}}$, cont...

data sets				NEQ		Stable Direct		LSQR		
Name	cond($A_{\mathcal{B}}$)	cond(J)	nnz(E)	D_Time	its	D_Time	its	D_Time	its	L_its
nnz2	19	13558	4490	9.87	7	18.78	7	0.55	7	81
nnz4	21	19540	6481	10.09	7	20.74	7	0.86	7	106
nnz8	28	10170	10456	10.05	7	29.48	7	1.51	7	132
nnz16	76	11064	18346	10.08	7	35.48	7	3.65	7	210
nnz32	201	11778	33883	10.04	9	41.49	9	8.96	8	339

$\mathrm{cond}(\cdot)$ - (rounded) condition number;

nnz(E) - number of nonzeros in E;

D_time - average time for search direction;

its - number of iterations;

L_its - average number LSQR iterations per major iteration;

All data sets have the same dimension, 1000×2000, and have 2 dense columns.

Sparse; Well conditioned $A_{\mathcal{B}}$, ... Size

The time for the NEQ solver is proportional to m^3. The stable direct solver is about twice that of NEQ. LSQR is the best among these 3 solvers on these instances. The computational advantage of LSQR becomes more apparent as the dimension grows.

data sets				NEQ		Stable Direct		LSQR	
name	size	cond($A_{\mathcal{B}}$)	cond(J)	D_Time	its	D_Time	its	D_Time	its
sz1	400×800	20	2962	0.63	7	1.37	7	0.15	7
sz2	400×1600	15	2986	0.63	7	1.36	7	0.26	7
sz3	400×3200	13	2358	0.63	7	1.39	7	0.53	7
sz4	800×1600	19	12344	5.08	7	9.60	7	0.32	7
sz5	800×3200	15	15476	5.06	7	9.64	7	0.76	7
sz6	1600×3200	20	53244	39.01	7	72.12	7	1.35	7
sz7	1600×6400	16	56812	38.83	7	72.16	7	2.32	8
sz8	3200×6400	19	218664	346.24	7	549.44	7	2.99	7

$\mathrm{cond}(\cdot)$ - (rounded) condition number;

D_time - average time for search direction;

its - number of iterations

Sparse; Well conditioned $A_{\mathcal{B}}$, ...
Dense Cols

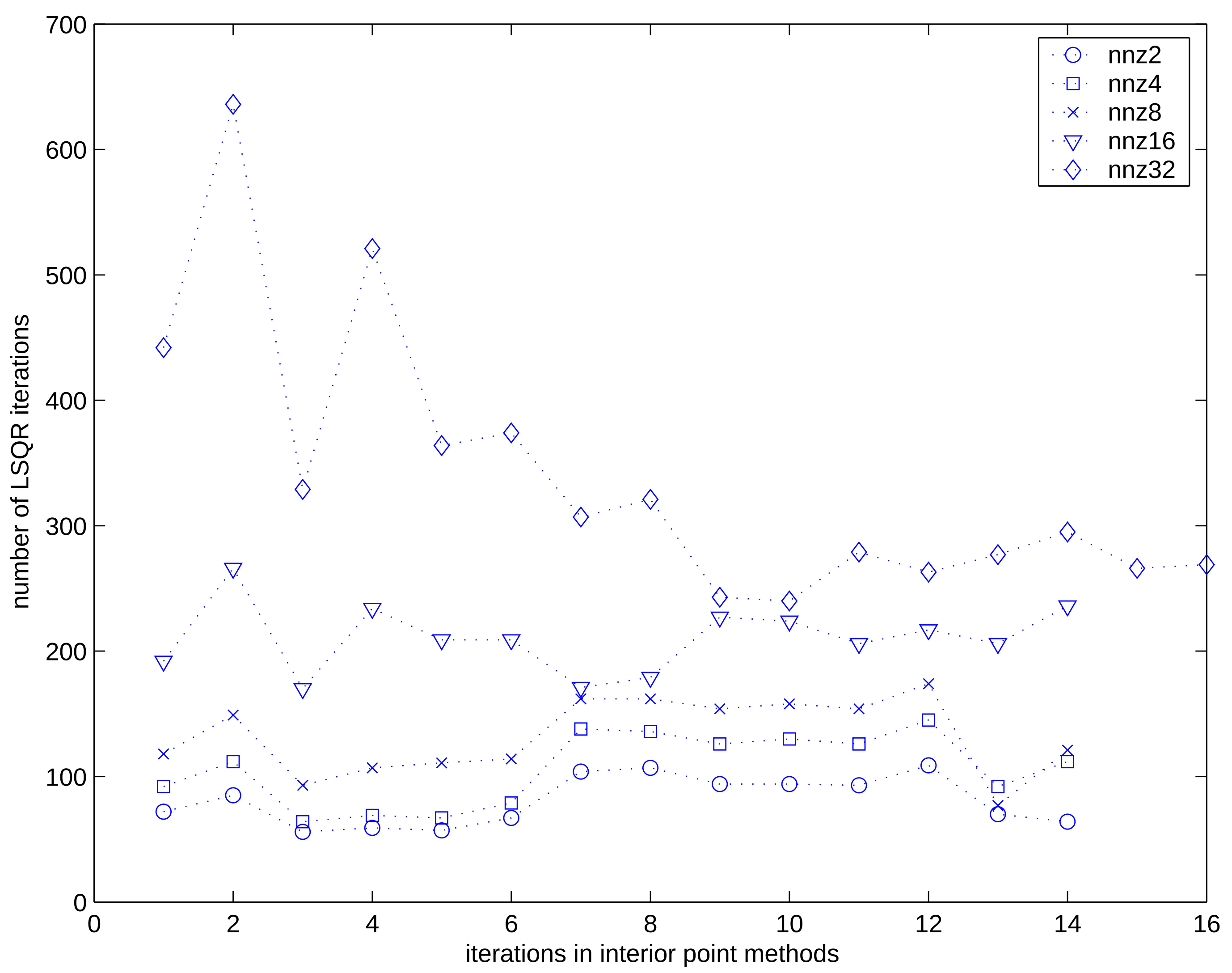
data sets				NEQ		Stable Direct		LSQR	
name	dense cols.	cond($A_{\mathcal{B}}$)	cond(J)	D_Time	its	D_Time	its	D_Time	its
den0	0	18	45	1.21	6	2.96	6	0.41	6
den1	1	19	13341	10.11	7	18.29	7	0.47	7
den2	2	19	18417	9.98	7	19.35	7	0.60	7
den3	3	19	19178	9.92	7	18.64	7	0.72	7
den4	4	18	18513	9.89	7	18.72	7	0.97	7

$\text{cond}(\cdot)$ - (rounded) condition number;

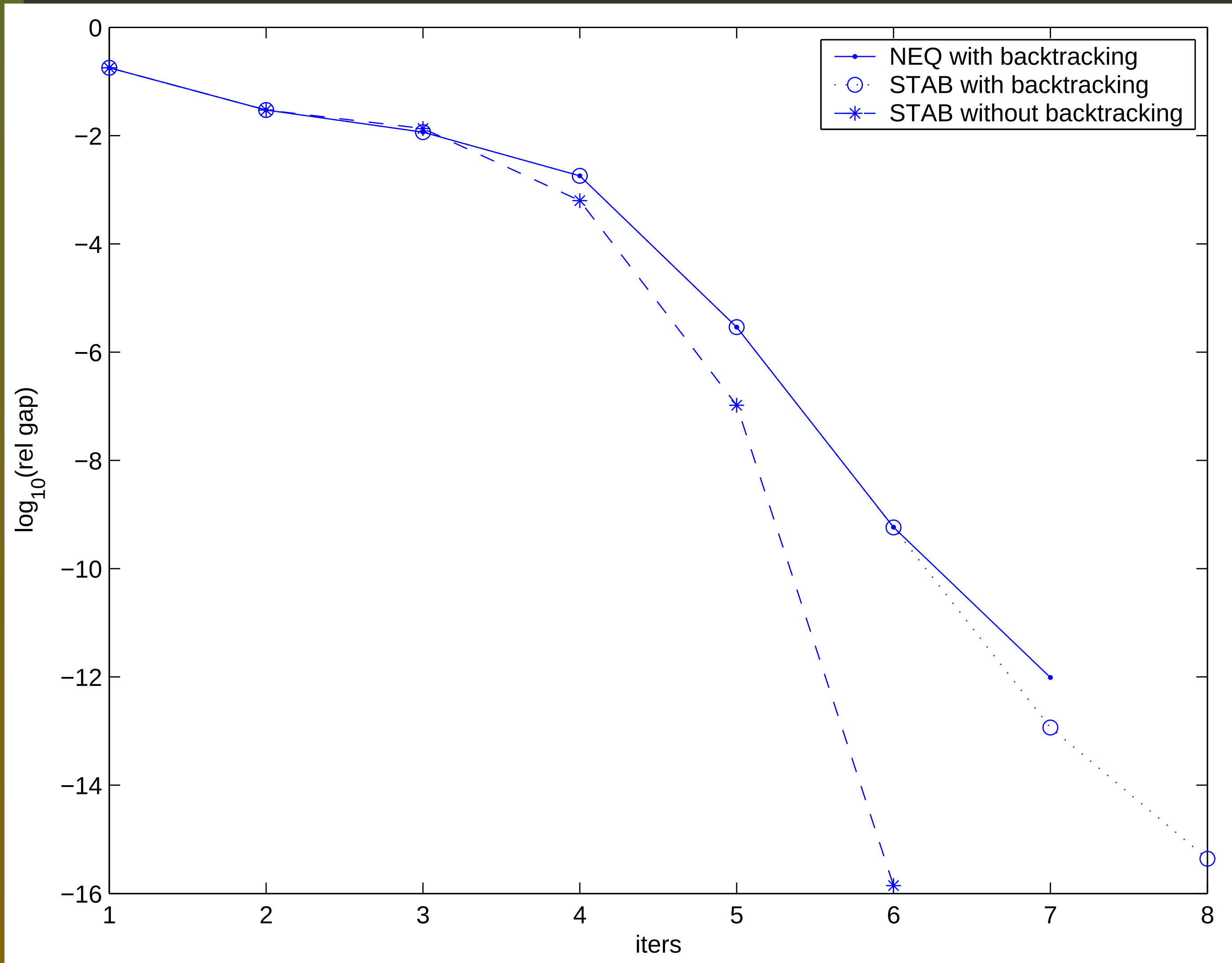
D_time - average time for search direction;

its - number of iterations.

LSQR iterations at different stages



No Backtracking - Complete Step to Boundary



Warm Starts

goal: start from optimal solution as an initial starting point to a perturbed problem.

stable method particularly successful at performing warm starts for small perturbations, since the Jacobian (under nondegeneracy assumptions) is nonsingular

optimality conditions become

$$F(x,y,z) := \begin{pmatrix} (A+\Delta A)^T y + z - c \\ (A+\Delta A)x - b \\ ZXe \end{pmatrix} = \begin{pmatrix} \Delta c \\ \Delta b \\ 0 \end{pmatrix}.$$

Theorem Suppose that the linear programming problem is nondegenerate and $x^*, (y^*, z^*)$ is a strictly complementary primal-dual optimal solution. Let $\mathcal{B}$ and $\mathcal{N}$ be the partition of the index set of variables:

$$\mathcal{B} = \{i : x_i^* > 0\} \text{ and } \mathcal{N} = \{i : 1 \leq i \leq n, i \notin \mathcal{B}\}.$$

Suppose we perturb A to $A + \Delta A$, b to $b + \Delta b$, and c to $c + \Delta c$. Assume $(A + \Delta A)_{\mathcal{B}}$ is nonsingular.

1. Then, a full step in the affine direction from the starting point x^*, y^*, z^* converges in one step to a solution satisfying the optimality conditions.
2. Furthermore, if $\Delta c, \Delta b, \Delta A$ are sufficiently small, then we get both $x \geq 0, z \geq 0$.

PROOF: The affine search direction $\Delta x, \Delta y, \Delta z$ with the starting point x^*, y^*, z^* is the solution to the following system

$$\begin{aligned} (A+\Delta A)\Delta x &= \Delta b - \Delta A x^* \\ (A+\Delta A)^T \Delta y + \Delta z &= \Delta c - \Delta A^T y^* \\ x^* \circ \Delta z + z^* \circ \Delta x &= 0. \end{aligned}$$

By noting that $x^*_{\mathcal{N}} = 0$ and $z^*_{\mathcal{B}} = 0$, solving the above system yields

$$\Delta x_{\mathcal{B}} = (A+\Delta A)^{-1}_{\mathcal{B}}(\Delta b - \Delta A x^*), \;\; \Delta x_{\mathcal{N}} = 0,$$

$$\Delta y = (A+\Delta A)^{-T}_{\mathcal{B}}(\Delta c_{\mathcal{B}} - \Delta A^T_{\mathcal{B}} y^*),$$

$$\Delta z_{\mathcal{B}} = 0, \;\; \Delta z_{\mathcal{N}} = \Delta c_{\mathcal{N}} - \Delta A^T_{\mathcal{N}} y^* - (A+\Delta A)^T_{\mathcal{N}}(A+\Delta A)^{-T}_{\mathcal{B}}(\Delta c_{\mathcal{B}} - \ldots$$

One can now verify that $x^* + \Delta x$, $y^* + \Delta y$, and $z^* + \Delta z$ is a solution to system equation optper. When ΔA, Δb, and Δc are sufficiently small, then $\Delta x_{\mathcal{B}}$ and $\Delta z_{\mathcal{N}}$ are small too, and thus $x^* + \Delta x \geq 0$ and $z^* + \Delta z \geq 0$. ■

Warm Start Convergence Radii for $A + r\Delta A, b + r\Delta b, c + r\Delta c$

Problem 1 with $\|A\| = 10.5$, $\|b\| = 465.6$, $\|c\| = 155.6$	Problem 2 with $\|A\| = 15.1$, $\|b\| = 582.0$, $\|c\| = 217.7$	Problem 3 with $\|A\| = 15.5$, $\|b\| = 755.8$, $\|c\| = 215.9$
0.05	0.01	0.01
0.05	0.03	0.01
0.09	0.03	0.01
0.05	0.03	0.01
0.13	0.01	0.11
0.09	0.01	0.07
0.09	0.01	0.01
0.07	0.03	0.01
0.03	0.02	0.01
0.11	0.07	0.01

(Nearest) Euclidean Distance Matrix Completion using SDP

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

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

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

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:

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

where $X \in SDP \Rightarrow \mathcal{L}(X) \in \text{EDM}$
(Public domain software packages are available)

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

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

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

With $X = B^T B \succeq 0$
connection between SDP and EDM .

Operator Notation:

$\text{us2vec}\,,\ \text{us2Mat}\,,\ \text{svec}\,,\ \text{sMat}$

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

$\sqrt{2}$ times vector (columnwise) from upper-triang of X.
$\binom{n+1}{2} = n(n+1)/2$; $\sqrt{2}$ guarantees isometry.
$\text{sMat} := \text{svec}^{-1}$ mapping into $\mathcal{S}^n$
adjoint transformation $\text{sMat}^* = \text{svec}$:

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

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

iff

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

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

(e is vector of ones)

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

with partition:

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

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

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

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

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

Duality and Optimality Conditions

(using $X = \mathrm{sMat}\,(x) + I$) 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)

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

change to **Wolfe dual** and obtain optimality conditions:

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

eliminate Λ
exact primal-dual feasibility during iterations
full rank Jacobian at optimality.
single bilinear (perturbed) equation in x;
$F_\mu(x) : \mathbb{R}^{\binom{n}{2}} \longrightarrow \mathcal{M}^{n-1}$

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

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

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

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

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

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 (lss 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:**

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

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

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

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

where

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

Diagonal Preconditioning

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

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

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

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

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

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.)

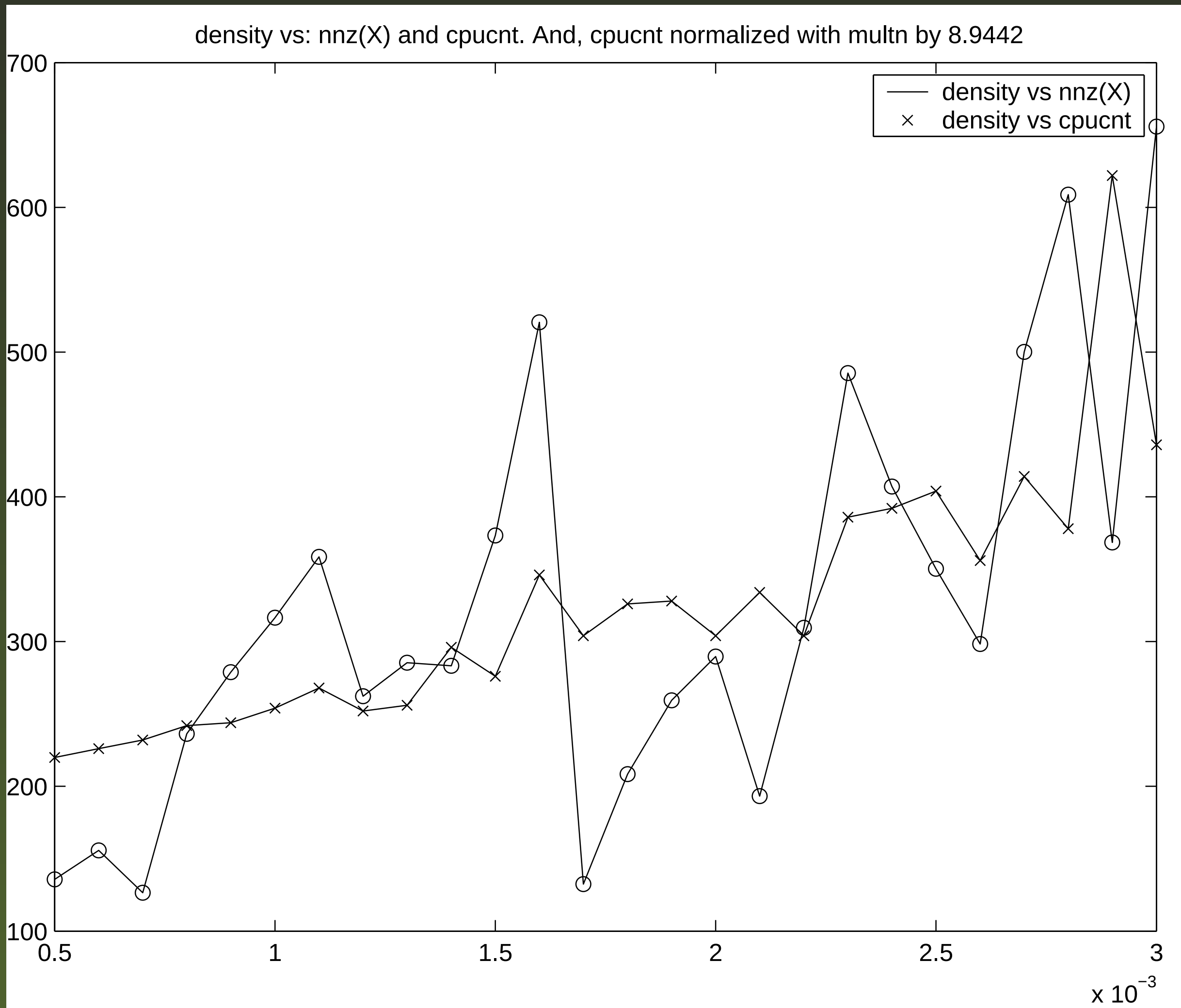


Figure 0.1: density .0005:.001:.003, CPU times and

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