A Robust Algorithm for Linear and Semidefinite Programming

M. Gonzales-Lima, X.V. Doan, S. Kruk, H. Wei,
Henry Wolkowicz

Dept. of Combinatorics and Optimization
University of Waterloo

GERAD

ÉCOLE POLYTECHNIQUE MONTRÉAL

Motivation: Linear Models are Pervasive in OR

Linear Programming, LP, Dominant Paradigm

- Dantzig's simplex method 1947
- Karmarkar interior point methods 1984 [13] (large scale problems)
- Currently huge problems with billions of variables and millions of constraints can be solved to optimality (e.g. Gondzio/'06, [9, 10])

(Linear) Semidefinite Programming, SDP

- Around since the 1940's (e.g. Lyapunov, Riccati equations)
- applications and interior point methods in 1990's.
- "world in nonlinear"; SDP provides better/stronger relaxation for hard problems (based on quadratic models).

Motivation cont...: Avoiding ill-conditioning

- Current LP (large scale) and SDP methods are based on primal-dual interior-point approaches; these involve a (i) symmetrization for SDP step to apply Newton's method, (ii) block elimination to reduce size of Newton equation.
- Both these steps create ill-conditioning in the Newton equation and singularity of the Jacobian at optimality
- We avoid the ill-conditioning and singularity using preprocessing and a backwards stable primal-dual interior/exterior-point approach based on an inexact (Gauss)-Newton method with a preconditioned (matrix-free) iterative method for finding the search direction
- improved accuracy and a lower iteration count

Outline:

primal-dual interior point (p-d i-p) for LP:

- path following, linearization/Newton method
- block elimination leading to ill-conditioning
- alternate block elimination to avoid ill-conditioning, benefits/cost

primal-dual interior point (p-d i-p) for SDP:

- path following
- symmetrization for linearization/Newton-method, leading to ill-conditioning
- block elimination leading to ill-conditioning
- alternate symmetrization/Gauss-Newton and alternate block elimination to avoid ill-conditioning, benefits/cost

Central LP Problem

Primal-dual LP pair

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

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

- $A \in \mathbb{R}^{m \times n}$ full rank (onto); LP and DLP strictly feasible
- $p^* \geq d^*$ Weak Duality $p^* = d^*$ Strong Duality
- $z \cdot x = 0$ Complementary Slackness
  (elementwise product is 0)

Central Problem SDP

Primal-dual SDP pair (*looks* similar to LP pair)

$$\text{(PSDP)} \qquad \begin{array}{rl} p^* := \min & C \cdot X \\ \text{s.t.} & \mathcal{A}(X) = b \qquad (\operatorname{trace} A_i X = b_i) \\ & X \succeq 0 \end{array}$$

$$\text{(DSDP)} \qquad \begin{array}{rl} d^* := \max & b^T y \\ \text{s.t.} & \mathcal{A}^*(y) + Z = C \\ & Z \succeq 0 \qquad (Z = C - \sum_i y_i A_i) \end{array}$$

- $C, X, Z \in \mathcal{S}^n$, $n \times n$ real symmetric matrices
- $C \cdot X = \operatorname{trace}(CX)$, trace inner product
- $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ linear transformation; $\mathcal{A}^*$ adjoint
- $X \succeq Y$ denotes the Löwner partial order $X - Y$ is psd

Central Path/Path Following LP

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

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

stationary point of the Lagrangian / optimality conditions

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

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

central path: set of these solutions

$(x_\mu, y_\mu, z_\mu), \mu > 0$

Ill-Conditioning

Due to Z^{-1} (Z singular at optimality)

As $\mu \to 0$, the Jacobian $F'_\mu(x, y, z)$ becomes ill-conditioned near the optimal solution/central path.

Cure/Fix, dates back to Fiacco-McCormick '68, [7]: Make the nonlinear equations *less nonlinear*, i.e. a form of preconditioning for Newton type methods;

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

Linearization; and Damped Newton Method

Special structure of linearization for Newton direction

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

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

Damped Newton steps

$x \leftarrow x + \alpha_p \Delta x, \quad y \leftarrow y + \alpha_d \Delta y, \quad z \leftarrow z + \alpha_d \Delta z,$ *backtrack* to maintain positivity/interiority, $x > 0, z > 0$

On central path, $F_\mu(x, y, z) = 0$, (if exact p-d feasibility holds!)

barrier param. $\mu = \frac{1}{n} z^T x = \frac{1}{n}($ duality gap$)$,

barrier parameter $\mu \cong$ duality gap

$$\begin{aligned} &= c^T x - b^T y \\ &= x^T \left(c - A^T y\right) \\ &= x^T z. \end{aligned}$$

Block Elimination for Normal Equations, NEQ

Linear System

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

Step 1 (Using row 1, Eliminate Δz from row 3):

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

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

RHS

with right-hand side

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

Normal Equations, NEQ

Step 2 (Use row 3, Eliminate Δx from row 2):

NO partial pivoting; well posed $\rightarrow$ ill-posed

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

$$= \begin{pmatrix} 0 & A^T & I_n \\ 0 & \boxed{AZ^{-1}XA^T} & 0 \\ I_n & -Z^{-1}XA^T & 0 \end{pmatrix}; \qquad \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$$

- $AZ^{-1}XA^T$ can have:
  - uniformly bounded condition number, e.g. Güler et al 1993, [11]
  - structured singular values, e.g. 1997: M. Wright [25]; S. Wright [26]
- But $\mathrm{cond}(F_n) \rightarrow \infty$ (backsolve step for for Δx is unstable)

The right-hand side

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

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

Ill-Posed Linear System/Jacobian at Optimality

Proposition

The condition number of NEQ $F_n^T F_n$ 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 original Jacobian $(F_\mu')^T F_\mu'$ is uniformly bounded if there exists a unique primal-dual solution (nondegeneracy).

Note that for NEQ system, as $\mu \downarrow 0$

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

$D_{ii} := (Z^{-1}X)_{ii} \to \infty$;
by interlacing of eigenvalues, ... $\lambda_{\min}(F_n^T F_n) \leq \lambda_{\min}(I_n) = 1 \leq D_{ii} A_{:,i} A_{:,i}^T \leq \lambda_{\max}(AZ^{-1}XA^T)^2 \leq \lambda_{\max}(F_n^T F_n)$ ∎

Condition Number Growth

Remark

- 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 (short/long step methods), then $\min_i(z_i/x_i)$ is $O(\mu)$.

  THEN:

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

- 1997: M. Wright [25]; S. Wright [26]

Example: Catastrophic Roundoff Error

Data

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

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

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:

$\|r_p\| = 0.081699, \|R_d\| = 0.36537, \mu = \frac{x^T z}{n} = 2.2528e-008$

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.

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

Use Alternate Second Step; Get Stable Reduction

Assuming! $A = [I_m \; E]$ (for simplicity)

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

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

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

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; $\mathrm{range}(N) = \mathrm{nullspace}(A)$

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

e.g. symmetric form

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

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

More generally (find, S sparse, well conditioned)

$$A = \begin{bmatrix} S & V \end{bmatrix}, \quad N = \begin{bmatrix} -S^{-1}E \\ I_{n-m} \end{bmatrix}; Nv = \begin{pmatrix} -S\backslash(Ev) \\ v \end{pmatrix}$$

(S triangular and sparse easy to find for Netlib data set)

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
$F(v, y) := \mathrm{Diag}\,(c - A^T y)\mathrm{Diag}\,(\hat{x} + Nv)e = 0$ ■
Get a single (perturbed) optimality condition for 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, $\Delta s^T := \begin{pmatrix} \Delta v & \Delta y \end{pmatrix}$

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

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

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

Well Conditioned Jacobian at Optimality

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

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

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

Primal-Dual Algorithm

follow usual primal-dual interior-point framework

Newton's method is applied to perturbed system of optimality conditions with damped step lengths for maintaining nonnegativity constraints.

Differences:

- eliminate, primal-dual linear feasibility (exact feas. once obtained is maintained)
- search direction found using PCG (LSQR)
- no backtracking from boundary for sufficient positivity
- 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

$$Z := Z(y) = \mathrm{Diag}\,(c - A^T y), \quad X := X(v) = \mathrm{Diag}\,(\hat{x} + Nv)$$
$J := F'_\mu(v, y) = \begin{pmatrix} ZN & -XA^T \end{pmatrix}$ Jacobian

find a preconditioner; *simple, nonsingular*

M such that JM^{-1} is well conditioned and then solve the better conditioned systems
$JM^{-1}\Delta q = -F_\mu$ and $M\Delta s = \Delta q$.

We look for: $M^T M \cong J^T J$

We us: LSQR (Paige-Saunders); implicitly solves normal equations

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

simplest of preconditioners; given square matrix K

Optimal diagonal preconditioner

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

If $M = \arg\min \omega((JD)^T(JD))$ over all positive diagonal matrices D then (Dennis-W. 1990, [5])

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

Crossover Criteria/Quadratic Convergence

Assumption: nonsingular Jacobian at optimality

(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 $J \in \mathrm{Lip}_\gamma(\mathcal{N}(s_0, r))$ with $J(s_0)$ nonsingular, and that there exist constants β, $\eta \geq 0$ such that

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

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

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

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

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

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

■

Lipschitz Constant for Region of Convergence

Lemma The Jacobian

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

is Lipschitz continuous with constant

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

with respect to (v, y) ■

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

Stable Jacobian/asymptotic quadratic convergence allows:

- detect zero variables/active constraints at optimality
- use e.g. the *T*apia indicators 1995 [6, 23].

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

- perform a pivot step to eliminate the variables converging to 0.

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 of alternate block eliminations

  no loss of sparsity

  high accuracy solutions available if desired

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

  warm starts for small perturbations of data

  fast convergence (no backtracking from boundary)

Optimality Paradigm for SDP - Interesting Differences

Perturbed overdetermined optimality conditions:
(i) dual/(ii) primal feas., (iii) pert. compl. slack. $:\to \mathbb{R}^{n\times n}$

$$F_\mu(X,y,Z) := \begin{pmatrix} (i) & \mathcal{A}^*(y)+Z-C \\ (ii) & \mathcal{A}(X)-b \\ (iii) & ZX-\mu I \end{pmatrix} = 0, \quad \mu > 0$$

Characterization of optimality

Under constraint qualification, the primal-dual solution (X,y,Z) with $X,Z \succeq 0$ is optimal iff $F_0(X,y,Z)=0$

P-d i-p methods: maintain $X,Z \succ 0$; with $\mu \downarrow 0$,

These methods are based on path following; the perturbed system is overdetermined; under nondegeneracy, Jacobian is full rank at optimality, e.g., AHO97[1].

Overdetermined: Symmetrize for Newton's method

Two most popular symmetrizations

(i) HKM: HRVW/KSM/M96 [12, 14, 16], and
(ii) NT97 [17, 18]

But: algorithms are not backwards stable! precond? sparsity?

- The linearizations (Jacobian) are singular at optimum e.g., T99[21]
- unlike the LP case there is no structured singularity; the block elimination schemes result in increased ill-conditioning e.g., W01[24]
- finding reasonable preconditioners is difficult (impossible?)
- block eliminations make it difficult to exploit sparsity

Notation

$A_i \in \mathcal{S}^n, \forall i$

- $\mathcal{A}(X) = b \Leftrightarrow A_i \cdot X = b_i, \quad \forall i = 1, \ldots, m,$ ($\mathcal{A}$ full rank m)
- $\mathcal{A}^*(y) = \sum_{i=1}^{m} y_i A_i, \forall y \in \mathbb{R}^m$
- $\mathrm{vec} : \mathcal{M}^n \to \mathbb{R}^{n^2}$ (columnwise) $\quad \mathrm{Mat} := \mathrm{vec}^{-1}$
- triangular number, $t(n) = n(n+1)/2$
- $\mathrm{svec} : \mathcal{S}^n \to \mathbb{R}^{t(n)}$ (isometry)
  $\mathrm{sMat} = \mathrm{svec}^{-1}$ (note $\mathrm{sMat}^* = \mathrm{svec}$)

Set $A \in \mathbb{R}^{m \times t(n)}$ with rows $A_{i,:} = \mathrm{svec}(A_i), \forall i = 1, \ldots, m$

$$\mathcal{A}(X) = b \Leftrightarrow A\mathrm{svec}(X) = b.$$

Single bilinear optimality conditions

Null space representation of primal feas.; dual feas.

- columns of $Q \in \mathbb{R}^{t(n)\times(t(n)-m)}$ are taken from $\{q_1,\ldots,q_{t(n)-m}\}$ orthonormal basis of $\mathcal{N}(A)$
- $\hat{X} \succ 0$ primal feasible solution; $\hat{x} = \operatorname{svec}(\hat{X})$, $v \in \mathbb{R}^{t(n)-m}$
  $\mathcal{A}(X) = b \Leftrightarrow \operatorname{svec}(X) = \hat{x} + Qv$
- $c = \operatorname{svec}(C)$; dual feasibility:
  $\mathcal{A}^*(y) + Z = C \quad \Leftrightarrow \quad \operatorname{svec}(Z) = c - A^T y$

Equivalent (overdetermined) optimality conditions:

$$\boxed{G_\mu(v,y) := \overbrace{\operatorname{sMat}(c - A^T y)}^{Z}\ \overbrace{\operatorname{sMat}(\hat{x} + Qv)}^{X}\ \overbrace{-\mu I}^{-\mu I} = 0}$$

Apply Gauss-Newton

$G_\mu(v,y) = \operatorname{sMat}(c - A^T y)\operatorname{sMat}(\hat{x} + Qv) - \mu I = 0$

- single bilinear overdetermined system;
  $t(n)$ variables; n^2 equations
- apply Newton's method to $\min \frac{1}{2}\|G_\mu(v,y)\|_2^2$; discard the second order terms; i.e. use GN method as residual $\to 0$
- For each $\mu > 0$, there exists unique p-d solution
  (X_μ, y_μ, Z_μ) with $X_\mu \succ 0, Z_\mu \succ 0$:
  lies on (defines) central path
  corresponding $\operatorname{svec}(Z_\mu) = c - A^T y_\mu$ and
  $\operatorname{svec}(X_\mu) = \hat{x} + Qv_\mu$, for appropriate v_μ

$G_\mu(v,y) = \mathrm{sMat}\,(c - A^T y)\mathrm{sMat}\,(\hat{x} + Qv) - \mu I = 0$

Gauss-Newton search direction; least-squares solution

$$-G_\mu(v,y) = G'_\mu(v,y)\begin{pmatrix}\Delta v\\ \Delta y\end{pmatrix},$$

where

- Jacobian $J = G'_\mu(v,y) : \mathbb{R}^{t(n)-m} \times \mathbb{R}^m \to \mathcal{M}^n$
- $\begin{cases} Z := C - \mathrm{sMat}\,(A^T y), & X := \hat{X} + \mathrm{sMat}\,(Qv), \\ \mathcal{Z}(v) := Z\,\mathrm{sMat}\,(Qv), & \mathcal{X}(y) := -\mathrm{sMat}\,(A^T y)X \\ J = [\mathcal{Z} \mid \mathcal{X}] \end{cases}$

(overdetermined) Gauss-Newton equation

$$\begin{aligned} -G_\mu(v,y) &= \mathcal{Z}(\Delta v) + \mathcal{X}(\Delta y) \\ &= Z\,\mathrm{sMat}\,(Q\Delta v) - \mathrm{sMat}\,(A^T\Delta y)X \end{aligned}$$

Adjoint Jacobian, J^*; normal equations

$$J = [\mathcal{Z} \quad \mathcal{X}]; \quad J^* = \begin{bmatrix} \mathcal{Z}^* \\ \mathcal{X}^* \end{bmatrix}$$

- $J^* \circ J = \begin{bmatrix} \mathcal{Z}^* \circ \mathcal{Z} & \mathcal{Z}^* \circ \mathcal{X} \\ \mathcal{X}^* \circ \mathcal{Z} & \mathcal{X}^* \circ \mathcal{X} \end{bmatrix},$
- normal equations $\qquad J^* \circ J \begin{pmatrix} \Delta v \\ \Delta y \end{pmatrix} = -J^* \circ G_\mu(v, y)$
- $\mathcal{X}^*(M) = -\frac{1}{2} A \operatorname{svec}(XM^T + MX).$
  $\mathcal{Z}^*(M) = \frac{1}{2} Q^T \operatorname{svec}(M^T Z + ZM).$

Infeasible start

Residuals: r_d, r_p, R_c

$$F_\mu(v,y,x,z) := \begin{pmatrix} z - c + A^T y \\ x - \hat{x} - Qv \\ \mathrm{sMat}(z)\mathrm{sMat}(x) - \mu I \end{pmatrix} =: \begin{pmatrix} r_d \\ r_x \\ R_c \end{pmatrix} = 0.$$

- $\Delta x = Q\Delta v - r_x$, $\Delta z = -A^T \Delta y - r_d$
- $G'_\mu(v,y)\begin{pmatrix} \Delta v \\ \Delta y \end{pmatrix} := Z\,\mathrm{sMat}(Q\Delta v - r_x) - \mathrm{sMat}(A^T\Delta y + r_d)X = -R_c$
  equivalently:
  $Z\,\mathrm{sMat}(Q\Delta v) - \mathrm{sMat}(A^T\Delta y)X = Z\,\mathrm{sMat}(r_x) + \mathrm{sMat}(r_d)X - R_c.$
- with step sizes α_p, α_d, new residuals;
  ($\alpha = 1$, exact feasibility maintained)!:
  $r_d \leftarrow (z - \alpha_d(A^T\Delta y + r_d)) - c + A^T(y + \alpha_d\Delta y) = (1 - \alpha_d)r_d,$
  $r_x \leftarrow x + \alpha_p(Q\Delta v - r_x) - \hat{x} - Q(v + \alpha_p\Delta v) = (1 - \alpha_p)r_p.$

Local quadratic convergence

Consider the following general problem

$$F(x,y) = \begin{pmatrix} L(x,y) \\ R(x) \end{pmatrix} = 0, \quad L(x,y) = x - \hat{x} - Ty, T \text{ full rank}$$

linearization

$$\left\{ F'(x,y) \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -F(x,y) \right\} \Leftrightarrow \left\{ \begin{array}{l} \Delta x - T\Delta y = -L(x,y) \\ R'(x)\Delta x = -R(x) \end{array} \right\}.$$

with $\Delta x = T\Delta y - L(x,y)$ get:

$R'(x)T\Delta y = -R(x) + R'(x)L(x,y)$ assumed overdetermined

Local convergence (Kantorovitch type) theorem

$J_x(x) := R'(x)$; $J_y(x) := J_x(x)T$; $R(x) \in \mathcal{C}^2$ in open convex set D; $J_x(x) \in \mathrm{Lip}_\gamma(D)$; $||J_x(x)|| \leq \alpha, \forall x \in D$;
$\lambda = \lambda_{\min}(J_y(x_*)^T J_y(x_*))$; $\exists x_* \in D, y_*$ with $\boxed{L(x_*, y_*) = 0, J_x(x_*)^T R(x_*) = 0}$;
$\left|\left|(J_x(x) - J_x(x_*))^T R(x_*)\right|\right|_2 \leq \sigma \, ||x - x_*||_2 \, , \quad \forall x \in D.$

Define iteration:

$$\Delta x_k = T\Delta y_k - L(x_k, y_k), \ \boxed{x_{k+1} = x_k + \Delta x_k};$$
$$\Delta y_k = (J_y(x_k)^T J_y(x_k))^{-1} J_y(x_k)^T(-R(x_k) + J_x(x_k) L(x_k, y_k)),$$
$$\boxed{y_{k+1} = y_k + \Delta y_k}.$$

Let $\tau := ||T||_2$, $c \in (1, \lambda/(\tau^2\sigma))$. If $\tau^2\sigma < \lambda$, then $\exists \epsilon > 0$ with $\forall x_0 \in N(x_*, \epsilon)$, y_0 arbitrary, we have that:

the sequences $\{x_k\}, \{y_k\}$ converge to (x_*, y_*) with $L(x_k, y_k) = 0, \forall k \geq 1$; and

$$\boxed{||x_{k+1} - x_*|| \leq \frac{\tau^2 c}{\lambda}\left(\sigma \, ||x_k - x_*|| + \frac{\alpha\gamma}{2} \, ||x_k - x_*||^2\right),}$$

$$\boxed{||x_{k+1} - x_*|| \leq \frac{\lambda + \tau^2 c\sigma}{2\lambda} \, ||x_k - x_*|| < ||x_k - x_*|| \, .}$$

Applying Theorem to Gauss-Newton equation

Key: lower bound to smallest singular value of Jacobian

$$L(z,x,y,v) = \begin{pmatrix} z - c + A^T y \\ x - \hat{x} - Qv \end{pmatrix}, \text{ with } T = \begin{pmatrix} -A^T & 0 \\ 0 & Q \end{pmatrix};$$

$R_\mu(z,x) = \mathrm{sMat}(z)\mathrm{sMat}(x) - \mu I$; $R_\mu(z_*, x_*) = 0$. $R'(z,x)\begin{pmatrix} \Delta z \\ \Delta x \end{pmatrix} = Z\mathrm{sMat}(\Delta x) + \mathrm{sMat}(\Delta z)X$ implies $||R'(z,x)|| \leq \sqrt{||Z||^2 + ||X||^2} = ||(z,x)||$. Therefore the constants are:

$$\boxed{\sigma = 0, \tau = ||T|| = \sqrt{||A||^2 + ||Q||^2}, \gamma = 1, \alpha = \sup_{(z,x)\in D} ||(z,x)||}$$

Poly. time global convergence

But it is difficult to estimate $\sigma_{\min}(J_y(x_*))$; which is needed for global poly. time convergence.

(a global polynomial convergence result for a scaled Gauss-Newton method is given in KPRT99[4].)

Crossover technique

Set (Gauss) Newton free: affine scaling $\mu = 0$, step length 1

(under nondegeneracy) Theorem indicates that there is a neighborhood around optimal solution where free Gauss-Newton method converges quadratically.

- affine scaling: set barrier parameter $\mu = 0$
- set G-N free: take full step lengths $= 1$; no nonnegativity checks/no backtracking
- region of quadratic convergence: build heuristic rule based on the infeasibility of the current solution (z, x)

Well-conditioned, sparse bases for range and nullspace of A

- motivation: apply a matrix-free Krylov type method for search direction and exploit sparsity.
- Given A sparse, find permutations $P_r A P_c = [S \mid E]$, with $S \in \mathbb{R}^{m \times m}$ well-conditioned, (approx) triangular
- columns of $Q = \begin{bmatrix} -S^{-1}E \\ I \end{bmatrix}$ gives basis of nullspace of A

ω measure for $X \succ 0$: $\omega(X) := \frac{\operatorname{trace}(X)/n}{\det(X)^{\frac{1}{n}}}$

Use ω-optimal diagonal and ω-optimal block diagonal preconditioners.

Lovász Theta Function Problem

Primal TN

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ undirected graph; $n = |\mathcal{V}|$ nodes; $m = |\mathcal{E}|$ edges
E matrix all ones, $E_{ij} = (e_i e_j^T + e_j e_i^T)/\sqrt{2}$ ij-th unit matrix; e_i i-th unit vector

$$
\text{(TN)} \qquad
\begin{array}{rll}
\vartheta(\mathcal{G}) := p^* := & \max & E \cdot X \\
 & \text{s.t.} & I \cdot X = 1 \\
 & & E_{ij} \cdot X = 0, \quad \forall (i,j) \in \mathcal{E} \\
 & & X \succeq 0
\end{array}
$$

Dual DTN

$$
\text{(DTN)} \qquad
\begin{array}{rll}
d^* := & \min & z \\
 & \text{s.t.} & zI + \sum_{(i,j)\in\mathcal{E}} y_{ij} E_{ij} - Z = E, \\
 & & Z \succeq 0,
\end{array}
$$

Gauss-Newton for Lovász theta function problem

Let $V := \begin{pmatrix} -e^T \\ I \end{pmatrix} \in \mathbb{R}^{n\times(n-1)}$, (nullspace V is $e^\perp$)

Primal-Dual feasibility: X, Z

$$\begin{array}{llcl} X \succeq 0 & \text{primal feas.} & \iff & X = \frac{1}{n}I + \mathrm{Diag}\,(Vw) + \mathrm{sMat}_{\mathcal{E}^c}(v) \\ Z \succeq 0 & \text{dual feasible} & \iff & Z = -E + zI + \mathrm{sMat}_{\mathcal{E}}(y) \end{array}$$

solve the perturbed optimality condition $ZX - \mu I = 0$

Jacobian

$J := G'_\mu = [\mathcal{Z} \mid \mathcal{X}]$

where: $\mathcal{Z}(w,v) = Z\left(\mathrm{Diag}\,(Vw) + \mathrm{sMat}_{\mathcal{E}^c}(v)\right)$

$\mathcal{X}(z,y) = (zI + \mathrm{sMat}_{\mathcal{E}}(y))\,X$

Numerics

Three versions of the Gauss-Newton algorithm

- general version for full matrices
- sparse version for sparse matrices
- specialized version for the Lovász theta function problem

Three different preconditioners

- diagonal
- two block diagonal
- multiple block diagonal

Random Well-Conditioned Instances

Sparse version

Data inputs are $C \in \mathcal{S}^n$, $A_i \in \mathcal{S}^n$, $i = 1, \ldots, m$, $b \in \mathbb{R}^m$; coded in MATLAB; G-N directions computed using LSMR by Fong-Saunders10[8] and LSQR Paige-Saunders82[20].
$n = 10:100$, $m = \lceil n(n+1)/4 \rceil$; sparseness density $1/(n+m)$; tolerance $\epsilon = 10^{-12}$; crossover is used.

different versions of the code with different preconditioners

without preconditioner (**SRSD**$_{\mathrm{o}}$);
diagonal preconditioner (**SRSD**$_{\mathrm{Diag}}$);
block preconditioner (**SRSD**$_{\mathrm{BDiag}}$).

Random well conditioned instances

Reduction in iterations; tradeoff – diagonal precond.

n	10	20	30	40	50	60	70	80	90	100
($\mathbf{SRSD}_{\mathrm{Diag}}$)	0.67	0.55	0.55	0.39	0.42	0.42	0.47	0.36	0.34	0.32
($\mathbf{SRSD}_{\mathrm{BDiag}}$)	0.30	0.16	0.13	0.09	0.09	0.08	0.08	0.07	0.07	0.06

Table: Ratios of numbers of LSMR iterations of ($\mathbf{SRSD}_{\mathrm{Diag}}$) and ($\mathbf{SRSD}_{\mathrm{BDiag}}$) to ($\mathbf{SRSD}_{\mathrm{o}}$)

n	10	20	30	40	50	60	70	80	90	100
($\mathbf{SRSD}_{\mathrm{Diag}}$)	0.63	0.58	0.59	0.45	0.49	0.51	0.62	0.45	0.48	0.44
($\mathbf{SRSD}_{\mathrm{BDiag}}$)	0.33	0.27	0.46	0.57	1.05	1.44	2.20	2.26	3.50	3.26

Table: Ratios of total computational time of ($\mathbf{SRSD}_{\mathrm{Diag}}$) and ($\mathbf{SRSD}_{\mathrm{BDiag}}$) to ($\mathbf{SRSD}_{\mathrm{o}}$)

Accuracy/Random sparse instances

SRSD$_{\mathrm{Diag}}$ with crossover

$n = 100, m = 100$, ($N = 1000$ instances, log-average) tolerance 10^{-12}

$\mathrm{RelZXnorm} = \frac{||ZX||_F}{|C \cdot X|+1}$, $\quad \mathrm{Relmineig} = \frac{\min\{\lambda_{\min}(X), \lambda_{\min}(Z)\}}{|C \cdot X|+1}$

Accuracy measures for random sparse instances

	Iteration	RelZXnorm	Relmineig
Average	18.66	2.62×10^{-15}	-8.79×10^{-16}
Best	14.00	2.78×10^{-16}	-3.86×10^{-17}
Worst	26.00	8.36×10^{-13}	-1.97×10^{-13}

Accuracy comparisons

20 random instances; four SDP solvers

SeDuMi 1.3, CSDP 6.1.0, SDPA 7.3.1, SDPT3 4.0.

tolerances set small; six DIMACS error measures [15]

$$
\begin{aligned}
\epsilon_1(X,y,Z) &= \frac{\|\mathcal{A}(X)-b\|_2}{1+\|b\|_\infty} \\
\epsilon_2(X,y,Z) &= \max\left\{0, \frac{-\lambda_{\min}(X)}{1+\|b\|_\infty}\right\} \\
\epsilon_3(X,y,Z) &= \frac{\|\mathcal{A}^{\mathrm{adj}}(y)+Z-C\|_F}{1+\|C\|_{\max}} \\
\epsilon_4(X,y,Z) &= \max\left\{0, \frac{-\lambda_{\min}(Z)}{1+\|C\|_{\max}}\right\} \\
\epsilon_5(X,y,Z) &= \frac{C\bullet X-b^Ty}{1+|C\bullet X|+|b^Ty|} \\
\epsilon_6(X,y,Z) &= \frac{Z\bullet X}{1+|C\bullet X|+|b^Ty|}
\end{aligned}
$$

High accuracy in comparison with 4 solvers

Performance measures for random sparse instances

	($\mathbf{SRSD}_{\mathrm{Diag}}$)	SeDuMi	CSDP	SDPA	SDPT3
Iteration	23.45	19.15	16.50	16.30	24.50
RelZXnorm	7.18×10^{-15}	1.50×10^{-07}	8.92×10^{-08}	5.86×10^{-08}	1.01×10^{-10}
Relmineig	-1.12×10^{-15}	-8.72×10^{-15}	3.28×10^{-13}	1.46×10^{-13}	5.99×10^{-17}
DIMACS1	1.28×10^{-15}	7.93×10^{-11}	1.24×10^{-13}	2.48×10^{-13}	3.19×10^{-13}
DIMACS2	2.28×10^{-14}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}
DIMACS3	9.39×10^{-16}	3.53×10^{-16}	1.64×10^{-08}	1.44×10^{-15}	8.72×10^{-14}
DIMACS4	3.22×10^{-15}	3.42×10^{-13}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}
DIMACS5	1.19×10^{-15}	9.72×10^{-13}	1.54×10^{-09}	4.96×10^{-10}	1.10×10^{-13}
DIMACS6	4.65×10^{-16}	2.84×10^{-14}	1.16×10^{-09}	4.96×10^{-10}	2.01×10^{-13}
Time	61.89	1.23	0.60	0.74	1.13

Increasing number constraints

$n = 100$, $m = 500 : 5000$; sparseness density $1/(4n)$;
block diagonal preconditioner

Random instances; varying number constraints

	($\mathbf{SRSD}_{\mathrm{BDiag}}$)	SeDuMi	CSDP	SDPA	SDPT3
Iteration	14.82	15.73	15.36	16.00	21.09
RelZXnorm	6.93×10^{-15}	1.99×10^{-07}	1.91×10^{-07}	1.77×10^{-08}	2.27×10^{-09}
Relmineig	-3.69×10^{-16}	-4.94×10^{-15}	4.06×10^{-13}	9.06×10^{-14}	2.31×10^{-16}

Time ratios relative to ($\mathbf{SRSD}_{\mathrm{BDiag}}$); varying number constraints

m	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
SeDuMi	168.74	49.99	21.47	4.18	1.92	1.89	1.09	0.66	0.68	0.32
CSDP	144.02	39.62	21.64	7.58	5.89	5.56	3.53	2.40	2.88	1.62
SDPA	488.28	203.06	134.49	50.99	37.09	33.37	20.98	15.55	17.83	9.41
SDPT3	87.49	27.96	16.10	6.00	3.80	4.42	2.11	2.30	2.89	1.81

Varying sparsity

$n = 100, m = 2500$; varying density $1/(4sn), s = 1, \ldots, 10$

Average measures for random instances with different sparsity

	$(\mathbf{SRSD}_{\mathrm{Diag}})/2$	SeDuMi	CSDP	SDPA	SDPT3
Iteration	13.80	16.80	16.20	16.20	24.30
RelZXnorm	4.47×10^{-15}	5.59×10^{-08}	1.30×10^{-07}	3.80×10^{-09}	3.65×10^{-11}
Relmineig	-7.59×10^{-16}	-2.24×10^{-16}	4.96×10^{-13}	8.71×10^{-14}	8.51×10^{-18}

Time ratios relative to $(\mathbf{SRSD}_{\mathrm{Diag}})/2$ for different sparsity

s	1	2	3	4	5	6	7	8	9	10
SeDuMi	2.84	2.74	1.26	1.15	1.49	0.62	0.53	0.49	0.34	0.30
CSDP	7.33	12.17	6.07	6.77	8.06	4.14	3.21	3.39	2.49	2.17
SDPA	43.35	48.63	22.13	20.95	25.47	12.33	9.25	9.82	7.08	6.04
SDPT3	4.60	8.82	5.79	5.74	7.50	3.94	2.86	2.75	2.27	2.09

Random Ill-Conditioned Instances

Strict complementarity fails

code for problem generation WW06 [22]; $n = 50, m = 1000$; block preconditioner without crossover; tolerance 10^{-14}.

GN-method significant accuracy increase $\|ZX\|$, RelZXnorm

	($\textbf{GRSD}_{\text{BDiag}}$)	SeDuMi	CSDP	SDPA	SDPT3
Iteration	29.20	13.60	12.10	12.00	22.70
RelZXnorm	1.32×10^{-12}	4.12×10^{-06}	6.53×10^{-05}	1.91×10^{-07}	2.97×10^{-08}
Relmineig	5.96×10^{-20}	6.20×10^{-14}	1.44×10^{-15}	3.25×10^{-12}	1.21×10^{-18}
DIMACS1	2.16×10^{-12}	2.01×10^{-07}	3.50×10^{-10}	3.79×10^{-07}	4.62×10^{-10}
DIMACS2	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}
DIMACS3	1.81×10^{-11}	6.34×10^{-15}	2.10×10^{-08}	2.09×10^{-14}	1.54×10^{-11}
DIMACS4	0.00×10^{00}	1.48×10^{-14}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}
DIMACS5	2.07×10^{-12}	9.23×10^{-09}	1.23×10^{-08}	1.58×10^{-07}	5.70×10^{-11}
DIMACS6	1.86×10^{-12}	4.99×10^{-09}	4.25×10^{-08}	3.50×10^{-07}	1.06×10^{-10}
Time	281.45	64.02	69.50	27.31	60.25

Table: Performance measures; random instances; strict complementarity fails

Random Ill-Conditioned Instances

dual Slater's CQ (almost) fails

	($\mathbf{GRSD}_{\mathrm{BDiag}}$)	SeDuMi	CSDP	SDPA	SDPT3
Iteration	26.00	17.10	13.80	15.20	20.40
RelZXnorm	9.90×10^{-16}	6.14×10^{-07}	4.61×10^{-07}	2.05×10^{-07}	2.13×10^{-08}
Relmineig	1.05×10^{-17}	1.06×10^{-15}	1.60×10^{-12}	1.06×10^{-10}	1.47×10^{-15}
DIMACS1	2.78×10^{-13}	1.07×10^{-10}	1.10×10^{-12}	1.17×10^{-13}	8.41×10^{-11}
DIMACS2	0.00×10^{00} (*)	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}
DIMACS3	1.42×10^{-12}	3.56×10^{-14}	5.21×10^{-09}	6.83×10^{-08}	2.59×10^{-13}
DIMACS4	0.00×10^{00} (*)	1.48×10^{-14}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}
DIMACS5	1.09×10^{-14}	2.77×10^{-12}	1.38×10^{-09}	2.02×10^{-14}	1.26×10^{-11}
DIMACS6	3.44×10^{-15}	3.16×10^{-12}	8.54×10^{-10}	4.43×10^{-08}	1.82×10^{-12}
Time	58.04	75.02	76.42	32.32	48.98

Random Instances; Lovász Theta Function

random graph with $n = 100$ nodes; $n(n-1)/4$ (approx) edges

	Iteration	RelZXnorm	Relmineig
Average	18.31	4.85×10^{-15}	2.47×10^{-13}
Best	16.00	6.42×10^{-17}	-1.28×10^{-15}
Worst	23.00	9.74×10^{-13}	-3.58×10^{-11}

Performance measures for random Lovász theta function

	SeDuMi	CSDP	SDPA	SDPT3	($\textbf{TRSD}_{\mathrm{MBDiag}}$)
Iteration	21.00	17.10	16.00	22.60	21.00
RelZXnorm	5.00×10^{-08}	1.56×10^{-08}	2.38×10^{-07}	1.73×10^{-11}	2.67×10^{-14}
Relmineig	3.34×10^{-14}	7.09×10^{-14}	3.49×10^{-12}	2.32×10^{-18}	6.74×10^{-13}
DIMACS1	1.84×10^{-11}	6.07×10^{-14}	5.09×10^{-14}	1.27×10^{-13}	1.25×10^{-16}
DIMACS2	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}	3.11×10^{-15}
DIMACS3	9.09×10^{-16}	1.09×10^{-07}	1.52×10^{-14}	8.87×10^{-13}	8.75×10^{-15}
DIMACS4	2.39×10^{-13}	0.00×10^{00}	0.00×10^{00}	0.00×10^{00}	4.82×10^{-12}
DIMACS5	9.09×10^{-12}	4.44×10^{-10}	9.61×10^{-09}	1.47×10^{-13}	5.92×10^{-16}
DIMACS6	9.93×10^{-14}	2.10×10^{-10}	9.61×10^{-09}	7.08×10^{-14}	4.19×10^{-16}
Time	90.46	13.05	6.02	12.91	158.09

Summary

- Current p-d i-p methods for SDP have two basic steps that lead to ill-conditioning: (i) symmetrize to be able to apply Newton's method; (ii) block elimation. (Implies difficulty for: accurate solutions; exploit sparsity.)
- We avoid these two steps using stable block elimination steps and applying Gauss-Newton method
- The numerical tests illustrate that we get improved accuracy and can exploit sparsity.
- Future research: To improve computation time, we need to exploit structure and find further stable eliminations.

Thanks for your attention!

A Robust Algorithm for Linear and Semidefinite Programming

M. Gonzales-Lima, X.V. Doan, S. Kruk, H. Wei, Henry Wolkowicz

Dept. of Combinatorics and Optimization
University of Waterloo

GERAD

ÉCOLE POLYTECHNIQUE MONTRÉAL

Numerical Tests for LP

- 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	$\mathrm{nnz}(E)$	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

$\mathrm{nnz}(E)$ - 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.

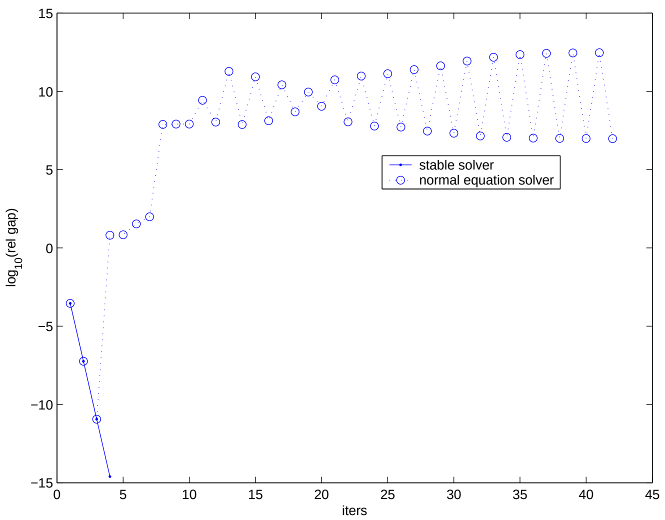
data set	LSQR with ILU				LSQR with Diag			
	D_Time	its	L_its	Pre_time	D_Time	its	L_its	Pre_time
1	0.15	6	37	0.06	0.41	6	556	0.01
2	3.42	6	343	0.28	2.24	6	1569	0.00
3	2.11	6	164	0.32	3.18	6	1595	0.00
4	NA	Stalling	NA	NA	13.37	6	4576	0.01
5	NA	Stalling	NA	NA	21.58	6	4207	0.01
6	NA	Stalling	NA	NA	90.24	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 $\times$ 800	20	2962	0.63	7	1.37	7	0.15	7
sz2	400 $\times$ 1600	15	2986	0.63	7	1.36	7	0.26	7
sz3	400 $\times$ 3200	13	2358	0.63	7	1.39	7	0.53	7
sz4	800 $\times$ 1600	19	12344	5.08	7	9.60	7	0.32	7
sz5	800 $\times$ 3200	15	15476	5.06	7	9.64	7	0.76	7
sz6	1600 $\times$ 3200	20	53244	39.01	7	72.12	7	1.35	7
sz7	1600 $\times$ 6400	16	56812	38.83	7	72.16	7	2.32	8
sz8	3200 $\times$ 6400	19	218664	346.24	7	549.44	7	2.99	7

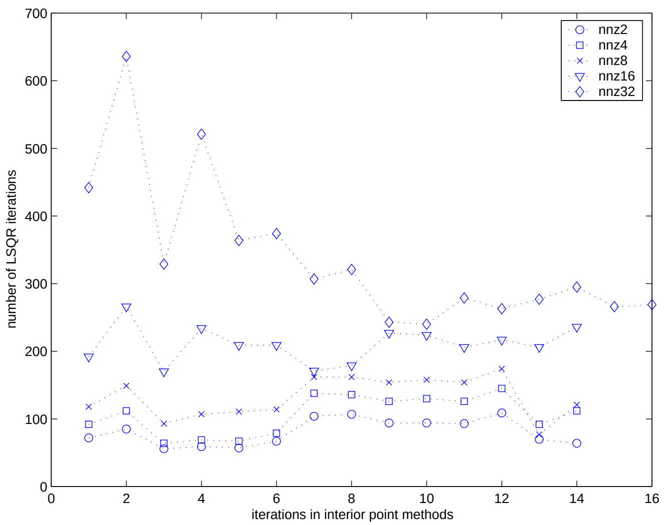
$\mathrm{cond}(\cdot)$ - (rounded) condition number;
D_time - average time for search direction;
its - number of iterations

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

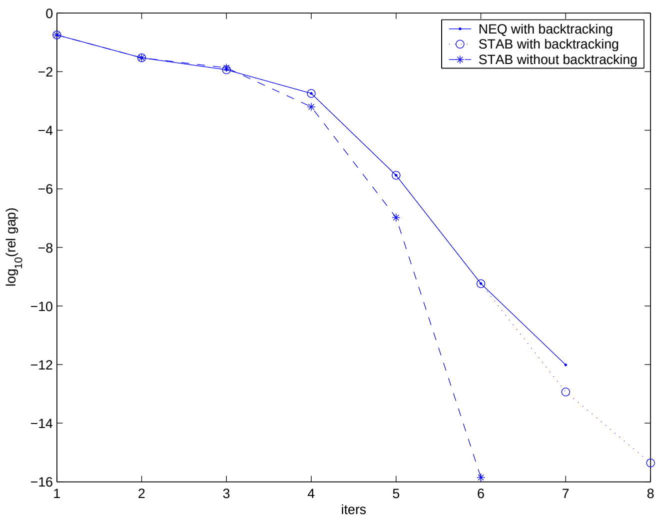
data sets				NEQ		Stable Direct		LSQR	
name	dense cols.	cond($A_{\mathcal{B}}$)	cond(J)	D_Time	its	D_Time	its	D_Time	its
den0	0	18	45	1.21	6	2.96	6	0.41	6
den1	1	19	13341	10.11	7	18.29	7	0.47	7
den2	2	19	18417	9.98	7	19.35	7	0.60	7
den3	3	19	19178	9.92	7	18.64	7	0.72	7
den4	4	18	18513	9.89	7	18.72	7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

$\mathrm{cond}(\cdot)$ - (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



- F. Alizadeh, J-P.A. Haeberly, and M.L. Overton, *Complementarity and nondegeneracy in semidefinite programming*, Math. Programming **77** (1997), 111–128.

- M. Benzi, G.H. Golub, and J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer. **14** (2005), 1–137. MR MR2168342 (2006m:65059)

- M. D'Apuzzo, V. De Simone, and D. di Serafino, *On mutual impact of numerical linear algebra and large-scale optimization with focus on interior point methods*, Comput. Optim. Appl. **45** (2010), no. 2, 283–310. MR 2594606 (2011c:90094)

- E. de Klerk, J. Peng, C. Roos, and T. Terlaky, *A scaled Gauss-Newton primal-dual search direction for semidefinite optimization*, SIAM J. Optim. **11** (2001), no. 4, 870–888 (electronic). MR 2002g:90070

- J.E. Dennis Jr. and H. Wolkowicz, *Sizing and least-change secant methods*, SIAM J. Numer. Anal. **30** (1993), no. 5, 1291–1314. MR 94g:90107

- A.S. El-Bakry, R.A. Tapia, and Y. Zhang, *On the convergence rate of Newton interior-point methods in the absence of strict complementarity*, Comput. Optim. Appl. **6** (1996), no. 2, 157–167. MR 1398265 (97c:90103)

- A.V. Fiacco and G.P. McCormick, *Nonlinear programming*, second (first 1968) ed., Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990, Sequential unconstrained minimization techniques. MR 91d:90089

- D. C.-L. Fong and M. A. Saunders, *LSMR: An iterative algorithm for sparse least-squares problems*, Report SOL 2010-2, Systems Optimization Laboratory, Stanford University, 2010, to appear, SIAM J. Sci. Comp.

- J. Gondzio and A. Grothey, *Direct solution of linear systems of size* 10^9 *arising in optimization with interior point methods*, Parallel Processing and Applied Mathematics (R. Wyrzykowski, J. Dongarra, N. Meyer, and J. Wasniewski, eds.), Lecture Notes in Computer Science, vol. 3911, Springer-Verlag, Berlin, 2006, pp. 513–525.

______, *Exploiting structure in parallel implementation of interior point methods for optimization*, Computational Management Science **6** (2009), 135–160.

O. GÜLER, D. DEN HERTOG, C. ROOS, T. Terlaky, and T. TSUCHIYA, *Degeneracy in interior point methods for linear programming: a survey*, Ann. Oper. Res. **46/47** (1993), no. 1-4, 107–138, Degeneracy in optimization problems. MR 94j:90021

C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz, *An interior-point method for semidefinite programming*, SIAM J. Optim. **6** (1996), no. 2, 342–361. MR 97f:90086

N.K. KARMARKAR, *A new polynomial-time algorithm for linear programming*, Combinatorica **4** (1984), 373–395.

M. Kojima, S. Shindoh, and S. HARA, *Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices*, SIAM J. Optim. **7** (1997), no. 1, 86–125. MR 97m:90088

- H.D. Mittelmann, *An independent benchmarking of SDP and SOCP solvers*, Math. Program. **95** (2003), no. 2, Ser. B, 407–430, Computational semidefinite and second order cone programming: the state of the art. MR MR1976487 (2004c:90033)

- R.D.C. Monteiro, *Primal-dual path-following algorithms for semidefinite programming*, SIAM J. Optim. **7** (1997), no. 3, 663–678.

- Y.E. Nesterov and M.J. Todd, *Self-scaled barriers and interior-point methods for convex programming*, Math. Oper. Res. **22** (1997), no. 1, 1–42.

- ——— , *Primal-dual interior-point methods for self-scaled cones*, SIAM J. Optim. **8** (1998), 324–364.

- D.P. O'Leary, *Symbiosis between linear algebra and optimization*, J. Comput. Appl. Math. **123** (2000), 447–465.

- C.C. Paige and M.A. Saunders, *LSQR: an algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software **8** (1982), no. 1, 43–71. MR 83f:65048

- M.J. Todd, *A study of search directions in primal-dual interior-point methods for semidefinite programming*, Optim. Methods Softw. **11&12** (1999), 1–46.

- H. Wei and H. Wolkowicz, *Generating and solving hard instances in semidefinite programming*, Math. Programming **125** (2010), no. 1, 31–45.

- P.J. Williams, A.S. El-Bakry, and R.A. Tapia, *Using indicators in finite termination procedures*, Optim. Methods Softw. **18** (2003), no. 1, 39–52. MR 1977792 (2004c:90099)

- H. Wolkowicz, *Solving semidefinite programs using preconditioned conjugate gradients*, Optim. Methods Softw. **19** (2004), no. 6, 653–672. MR MR2102220 (2005h:90091)

M.H. Wright, *Ill-conditioning and computational error in interior methods for nonlinear programming*, SIAM J. Optim. **9** (1999), no. 1, 84–111 (electronic). MR 99i:90093

S.J. Wright, *Stability of augmented system factorizations in interior-point methods*, SIAM J. Matrix Anal. Appl. **18** (1997), no. 1, 191–222. MR 97i:90039