Large Sparse Semidefinite Programming with Applications to the Nearest Correlation Matrix Problem

Henry Wolkowicz
Department of Combinatorics & Optimization
University of Waterloo

(M. Anjos, N. Higham, M. Takouda; (H. Wei))

OUTLINE

- Background on SDP; Motivation
- Robust, ‘non-interior’, path-following algorithm for SDP (and LP) (Gauss-Newton direction, preconditioned conjugate gradients)
- Nearest Correlation Matrix Problem: duality and algorithm
- Numerics
- Other SDP models

1 What is SDP? And Motivation

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

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

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

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

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

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

1.1 SDP Arises Naturally in Optimization

1.1.1 Quadratic Model for $\min_x f(x)$

A quadratic model at the current estimate x_c:

$$\text{(Quad)} \qquad \begin{array}{rll} q^* = & \min & q(d) := f(x_c) + \nabla f(x_c)^T d + \frac{1}{2} d^T \nabla^2 f(x_c) d \\ & \text{s.t.} & \|d\| \leq s, \qquad (\|d\|^2 \leq s^2) \end{array}$$

where the normalization/constraint avoids unboundedness and steplengths that are *too* long.

Lagrangian dual (with Hessian $\frac{1}{2}A$ and gradient $-2a$):

$$\begin{aligned} q^* = \nu^* \quad &:= \quad \max_{\lambda \le 0} \min_x x^T(A - \lambda I)x - 2a^T x + \lambda s^2 \\ &= \quad \max_{\lambda \le 0} h(\lambda) \end{aligned}$$

hidden constraint: inner minimization bounded below
$\text{dom}(h)$ restricted to $\nabla^2 L(x, \lambda) = A - \lambda I \succeq 0$

(convex) Lagrangian is $L(x, \lambda) := x^T(A - \lambda I)x - 2a^T x + \lambda s^2$;
(concave) dual functional is $h(\lambda) := \min_x \; L(x, \lambda)$

(This can be exploited to solve large sparse trust region subproblems, e.g. classical More-Sorensen (1983) using Cholesky factorization; $n > 10^6$, cf. C. Fortin and W. (2002) using Lanczos/eigenvalues)

SDP has surprisingly many applications arising from

quadratic models

1. Engineering (control theory, design)
2. Hard Combinatorial Problems
3. Robust Optimization
4. Mathematics of Finance (below)
5. ...

2 (Non) Interior Path-Following

2.1 Illustration/Motivation on LP Case

$$\text{(LP)} \qquad \begin{array}{rll} p^* := & \min & c^T x \quad (\text{or } \langle c, x\rangle) \\ & \text{s.t.} & Ax = b \\ & & x \geq 0 \quad (\text{or } x \succeq 0) \end{array} \tag{2.1}$$

$$\text{(DLP)} \qquad \begin{array}{rll} d^* := & \max & b^T y \\ & \text{s.t.} & A^T y + z = c \\ & & z \geq 0 \quad (\text{or } z \succeq 0) \end{array} \tag{2.2}$$

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

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

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

stationary point of the Lagrangian / optimality conditions

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

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

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):

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

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

$$F'_\mu(x,y,z)\Delta s = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & 0 & X \end{pmatrix} \Delta s = -F_\mu(x,y,z). \qquad (2.3)$$

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'

$$\begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \mathcal{S} \end{pmatrix} \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & 0 & X \end{pmatrix} \tag{2.4}$$

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

2.2 Reduction/Block Elimination for the Normal Equations

Step 1 (Eliminate Δz):

$$\begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -X & 0 & I \end{pmatrix} \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & 0 & X \end{pmatrix} = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & -XA^T & 0 \end{pmatrix}. \tag{2.5}$$

We let

$$P_Z = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -X & 0 & I \end{pmatrix}, \quad K = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & -XA^T & 0 \end{pmatrix}. \tag{2.6}$$

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):

$$
\begin{aligned}
F_n := P_n K \quad &:= \begin{pmatrix} I & 0 & 0 \\ 0 & I & -AZ^{-1} \\ 0 & 0 & Z^{-1} \end{pmatrix} \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & -XA^T & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & A^T & I_n \\ 0 & AZ^{-1}XA^T & 0 \\ I_n & -Z^{-1}XA^T & 0 \end{pmatrix}
\end{aligned} \tag{2.7}
$$

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

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

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

The right-hand side becomes

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

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

Proposition 2.1 *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, ... ■

EXAMPLE: *getting too close to boundary* (worse for SDP)

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

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

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:

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

5 decimals rounding before/after arithmetic

centering with $\sigma = .1$

search directions found using full matrix F'_μ and 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.

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

2.3 Alternate Second Step; Stable Reduction

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

Partition diagonal matrix Z, X using vectors

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

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

The right-hand side becomes

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

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

2.4 Solving the Last Two Rows

We can change to a symmetric indefinite system by changing the rows and scaling by the diagonal matrix $\begin{pmatrix} X_v & 0 \\ 0 & Z_m \end{pmatrix}$,

$$\begin{pmatrix} Z_v X_v & -X_v E^T Z_m \\ -Z_m E X_v & -X_m Z_m \end{pmatrix}. \tag{2.8}$$

If E is sparse, then this system stays sparse after the scaling.

But – we use a conjugate gradient type method below.

3 Finding the Nearest Correlation Matrix using SDP

Given symmetric matrix $A \in \mathcal{S}^n$:

$$\mu^* = \min \frac{1}{2}\|A - X\|_F^2 \quad \text{s. t.} \qquad \operatorname{diag} X = e,\ X \succeq 0,\ X \in \mathcal{S}^n\,.$$

Applications e.g. Finance:
approximate correlation matrices are constructed from vectors of stock returns; sample correlations are taken only from days on which both stocks have data available (from inconsistent data).
Ref: N. Higham (2002) uses a projection technique to take advantage of low rank.

3.1 Mixed-Cone Formulation

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

$$\begin{array}{rll} \sqrt{2\mu^*} = & \min & \alpha \\ & \text{s.t.} & \operatorname{diag} X = e \\ & & Y + X = A, \ \|Y\|_F \leq \alpha \\ & & X, Y \in \mathcal{S}^n, X \succeq 0. \end{array} \tag{3.1}$$

(Public domain software packages are available - numeric comparison below.)

3.1.1 Operator Notation: us2vec , us2Mat

$$x = \mathrm{us2vec}\,(X) \in \mathbb{R}^{\binom{n}{2}} \qquad (a = \mathrm{us2vec}\,A, s = \mathrm{us2vec}\,S)$$

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

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

$\mathrm{us2Mat} := \mathrm{us2vec}^{\dagger}$ mapping into $\mathcal{S}^n$

adjoint transformation $\mathrm{us2Mat}^{*} = \mathrm{us2vec}$ since:

$$\begin{aligned}\langle \mathrm{us2Mat}\,(v), S\rangle &= \mathrm{trace}\,\mathrm{us2Mat}\,(v)S = \mathrm{trace}\,\mathrm{us2Mat}\,(v)\mathrm{offDiag}\,(S)\\ &= v^T\mathrm{us2vec}\,(S) = \langle \mathrm{us2vec}\,(S), v\rangle\end{aligned}$$

orthogonal projection: $\mathrm{offDiag}\,(S) = \mathrm{us2Mat}\,\mathrm{us2vec}\,(S)$

3.2 Duality and Optimality Conditions

(using $X = \mathrm{us2Mat}\,(x) + I$) an equivalent problem is:

$$\mu^* := \min \frac{1}{2}\|x - a\|_2^2 \quad \text{subject to} \quad \mathrm{us2Mat}\,(x) + I \succeq 0,\ x \in \mathbb{R}^{\binom{n}{2}}.$$

strong (Lagrangian) duality:

$$\mu^* = \nu^* := \max_{S \succeq 0} \min_x \frac{1}{2}||x - a||^2 - \mathrm{trace}\, S(\mathrm{us2Mat}\,(x) + I).$$

change to **Wolfe dual** using stationarity of inner minimization:

$$0 = (x-a) - \mathrm{us2Mat}^*(S) = (x-a) - \mathrm{us2vec}\,(S),$$

and since $\operatorname{trace} S(\mathrm{us2Mat}\,(x) + I) = x^T \mathrm{us2vec}\,(S) + \operatorname{trace} S$.

$$\begin{array}{rll} \mu^* = & \max & \frac{1}{2}\|x-a\|^2 - \operatorname{trace} S(\mathrm{us2Mat}\,(x) + I) \\ & \text{subject to} & x - \mathrm{us2vec}\,(S) = a \\ & & S \succeq 0. \end{array} \tag{3.2}$$

Slater's CQ holds for both primal and dual:

Theorem 3.1 *The optimal values* $\mu^* = \nu^*$ *and the primal-dual pair* $x, (y, s)$ *are optimal if and only if*

$$\begin{array}{rcl} X & := & \text{us2Mat}(x) + I \succeq 0 \quad \textit{(primal feasibility)} \\ x & = & a + s, \quad S^y := \text{us2Mat}(s) + \text{Diag}(y) \succeq 0 \quad \textit{(dual feasibility)} \\ XS^y & = & 0 \quad \textit{(complementary slackness)} \end{array}$$

■

(cf $A = [I \ E]$ for LP)

Substitute feasibility equations; exact primal-dual feasibility maintained during iterations; full rank Jacobian at optimality.

single bilinear (perturbed) equation in s, y: $F_\mu(s, y) : \mathbb{R}^{\binom{n+1}{2}} \rightarrow \mathcal{M}^n$

$$F_\mu(s, y) := [A + \mathrm{us2Mat}\,(s) + I]\,[\mathrm{us2Mat}\,(s) + \mathrm{Diag}\,(y)] - \mu I = 0$$

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.

Linearization for search direction $\Delta v = \begin{pmatrix} \Delta s \\ \Delta y \end{pmatrix}$:

$(S^y = \mathrm{us2Mat}\,(s) + \mathrm{Diag}\,(y))$

$$\begin{aligned} & F'_\mu(s, y)\Delta v = \\ & = [A + \mathrm{us2Mat}\,(s) + I]\,(\mathrm{us2Mat}\,(\Delta s) + \mathrm{Diag}\,(\Delta y)) + \mathrm{us2Mat}\,(\Delta s) S^y \\ & = (\mathcal{X}_u + \mathcal{S})\,(\Delta s) + \mathcal{X}_d(\Delta y). \end{aligned}$$

This is a linear, full rank, overdetermined system.

Our search direction Δv is its least squares solution.

3.2.1 Algorithm: p-d i-e-p framework

- **Initialization:**

  •• **Input data:** a real symmetric $n \times n$ matrix A (set $\operatorname{diag}(A) = 0$)

  •• **Positive tolerances:**

  $$\epsilon_1 \text{ (stopping)}, \ \epsilon_2 \text{ (lss accuracy)}, \ \epsilon_3 \text{ (crossover)},$$

  •• **Find initial strictly feasible points:** both $S^0, X^0 := (\operatorname{offDiag}(S + A) + I) \succ 0$; $\mu > 0$

  •• **Set initial parameters:**

  $$\text{gap} = \operatorname{trace} S^0 X^0; \ \ \mu = \text{gap}/n; \ \ \text{objval} = .5 \|X^0 - A\|_F^2; \ \ k = 0.$$

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

•• **solve lss for search direction** (accuracy $\epsilon_2 \min\{\mu, 1\}$)

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

where σ_k centering, $\mu_k = \frac{1}{n}\mathrm{trace}\, S^k(\mathrm{offDiag}\,(S^{k+1} + A) + I)$

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

so that both $S^{k+1}, \mathrm{offDiag}\,(S^{k+1} + A) + I \succeq 0$

($\alpha_k = 1$ after the crossover.)

•• **update**

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

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

- **end (while)**.
- **Conclusion:** X is approx. $\mathrm{us2Mat}\,(s) + A + I$

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)

Perturbations (remove smaller nonzeros) to given A are done at start to reduce *nnz(A)*. Nonzeros are added back to A once the duality gap is *small*. Tests show that q-quadratic convergence is maintained.

3.3 Preconditioning

$$(\mathcal{X}_u + \mathcal{S})\, P_s^{-1}(\widehat{\Delta s}) + \mathcal{X}_d P_y^{-1}(\widehat{\Delta y}) = -F_\mu(s, y)$$

where

$$\widehat{\Delta s} = P_s(\Delta s), \quad \widehat{\Delta y} = P_y(\Delta y)$$

3.3.1 Diagonal Preconditioning

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

$$\min \omega((AD)^T(AD)) \quad \text{subject to: } D \text{ positive and diagonal} \tag{3.3}$$

solution: $d_{ii} = 1/\|A_{:i}\|_2, i = 1, \ldots, n$

explicit expressions for preconditioner
inexpensive

two *diagonal* operators P_s, P_y
evaluate using columns of $F'_\mu(s,y)$.
$k \cong (i,j),\ 1 \leq i < j \leq n$, strictly upper triangular part of S
$i = 1, \ldots, n$ for y
$X = A + \mathrm{us2Mat}\,(s) + I$ and $S = \mathrm{us2Mat}\,(s) + \mathrm{Diag}\,(y)$

$$\begin{aligned} \mathcal{X}_u(e_k) &= X \mathrm{us2Mat}\,(e_k) \\ &= \tfrac{1}{\sqrt{2}} X \left(e_i e_j^T + e_j e_i^T\right) \\ &= \tfrac{1}{\sqrt{2}} \left\{ \left(X_{:i} e_j^T + X_{:j} e_i^T\right) \right\}. \end{aligned}$$

$$\begin{aligned} \mathcal{S}(e_k) &= \mathrm{us2Mat}\,(e_k)(S + \mathrm{Diag}\,(y)) \\ &= \tfrac{1}{\sqrt{2}} \left(e_i e_j^T + e_j e_i^T\right)(S + \mathrm{Diag}\,(y)) \\ &= \tfrac{1}{\sqrt{2}} \{ (e_i (S + \mathrm{Diag}\,(y))_{j:} + e_j (S + \mathrm{Diag}\,(y))_{i:}) \}. \end{aligned}$$

Therefore

$$\begin{aligned}&\|(\mathcal{X}_u + \mathcal{S})(e_k)\|_F^2 = \\ &\quad = \tfrac{1}{2}\left\{\|(S + \operatorname{Diag}(y))_{:i}\|^2 + \|(S + \operatorname{Diag}(y))_{:j}\|^2 + \right.\\ &\qquad \|X_{:i}\|^2 + \|X_{:j}\|^2 + 2(S + \operatorname{Diag}(y))_{jj} X_{ii} \\ &\qquad \left. +4(S + \operatorname{Diag}(y))_{ji} X_{ij} + 2(S + \operatorname{Diag}(y))_{ii} X_{jj}\right\}.\end{aligned}$$

need three Hadamard products
$X \circ X, (S + \operatorname{Diag}(y)) \circ (S + \operatorname{Diag}(y)), (S + \operatorname{Diag}(y)) \circ X$
and *vector* Kronecker product $\operatorname{diag}((S + \operatorname{Diag}(y))) \otimes \operatorname{diag}(X)$.

$$\mathcal{X}_d(e_i) = X\operatorname{Diag}(e_i)$$

Therefore

$$\|\mathcal{X}_d(e_i)\|_F^2 = ||X_{i,:}||^2.$$

3.3.2 Block-Diagonal Incomplete Cholesky Preconditioner

natural block structure:

$$\left[(\mathcal{X}_u + \mathcal{S}) \mid \mathcal{X}_d \right] \begin{pmatrix} \Delta s \\ \Delta y \end{pmatrix} = -F_\mu,$$

normal equations have the block structure

$$\left[\begin{array}{c|c} ({\mathcal{X}_u}^* + \mathcal{S}^*)(\mathcal{X}_u + \mathcal{S}) & ({\mathcal{X}_u}^* + \mathcal{S}^*)\mathcal{X}_d \\ \hline {\mathcal{X}_d}^*(\mathcal{X}_u + \mathcal{S}) & {\mathcal{X}_d}^*\mathcal{X}_d \end{array} \right] \begin{pmatrix} \Delta s \\ \Delta y \end{pmatrix} = -\begin{pmatrix} {\mathcal{X}_u}^* + \mathcal{S}^* \\ {\mathcal{X}_d}^* \end{pmatrix} F_\mu$$

preconditioner based on partial Cholesky factorizations of the block diagonal positive definite operator:

$$\widetilde{P}^* \widetilde{P} = \left[\begin{array}{c|c} (\mathcal{X}_u{}^* + \mathcal{S}^*)\,(\mathcal{X}_u + \mathcal{S}) & 0 \\ \hline 0 & \mathcal{X}_d{}^* \mathcal{X}_d \end{array} \right].$$

$\widehat{P}^* \widehat{P}$ has the approximate factorization

$$\left[\begin{array}{c|c} \widehat{R}^T \widehat{R} & \\ \hline & D^2 \end{array} \right]$$

can exploit the special structure again

4 Numerical Tests

Pentium 4; MATLAB 6.5; 1 GIG RAM.

crossover heuristic: relative duality gap $< .1$.

Initially , zero out all elements $|A_{ij}| < \epsilon_4 = .5$.
In each iteration with (relative duality gap) $< .003$, set $\epsilon_4 \leftarrow (\epsilon_4 - .1)$

Stopping criteria (relative duality gap) $< \epsilon_1 = 1e - 10$.
(But - average accuracy attained $1e - 13$, q-quadratic convergence.)

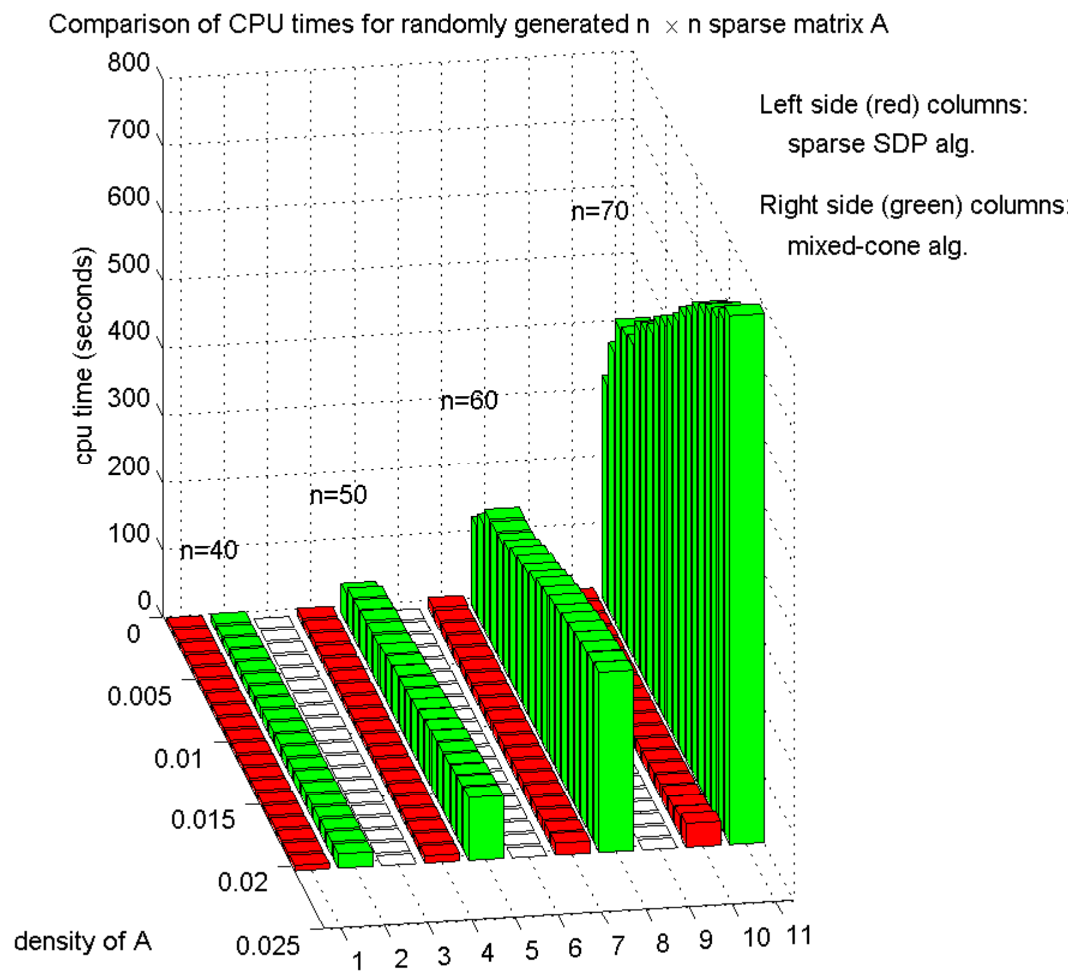

Figure 4.1: comparison: 2nd order cone (SeDuMi) (avrg 10 each density)

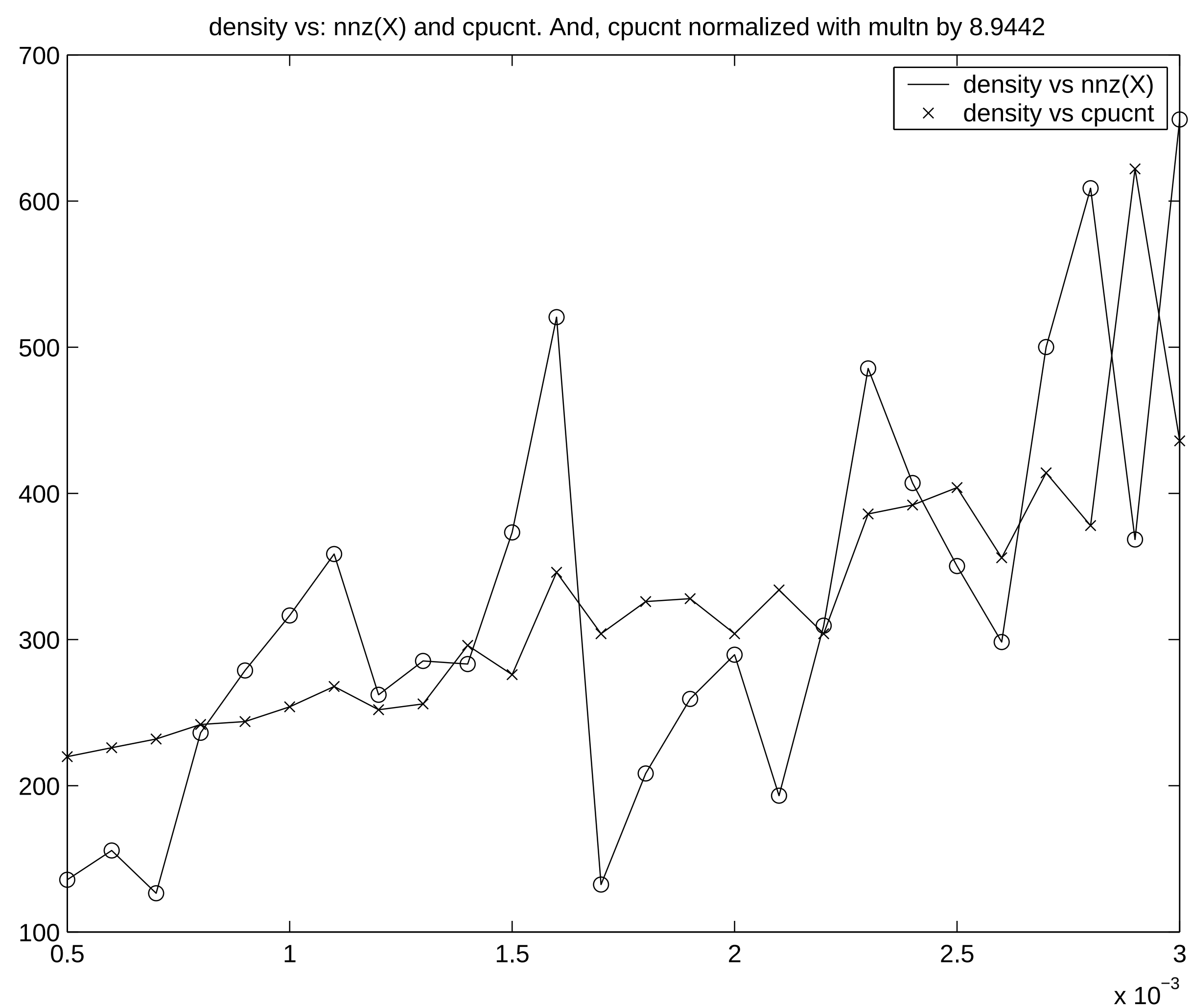

Figure 4.2: density .0005:.001:.003, CPU times and nnz(X), n=200

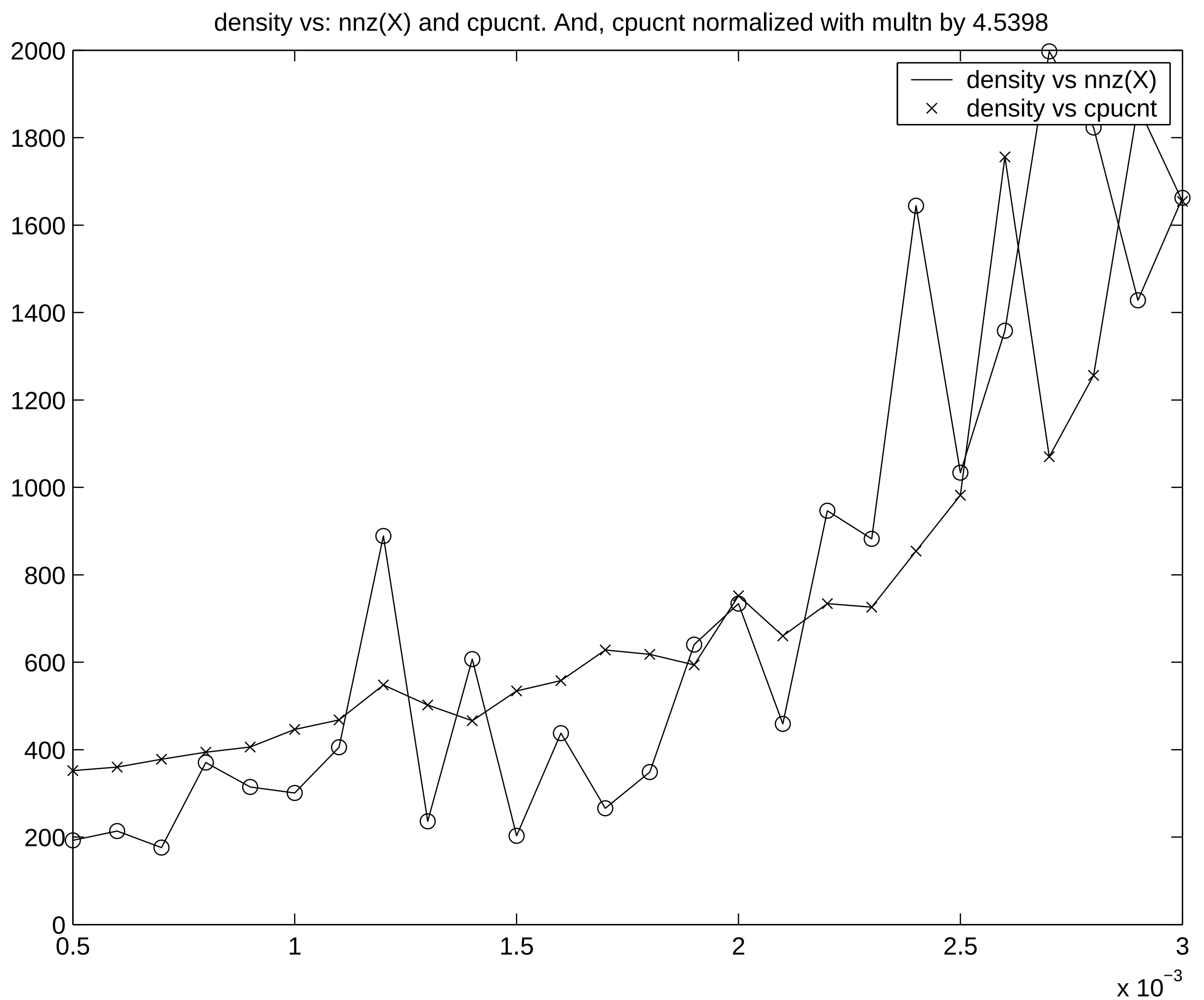

Figure 4.3: density .0005:.001:.003 vs CPU times and nnz(X), n=300

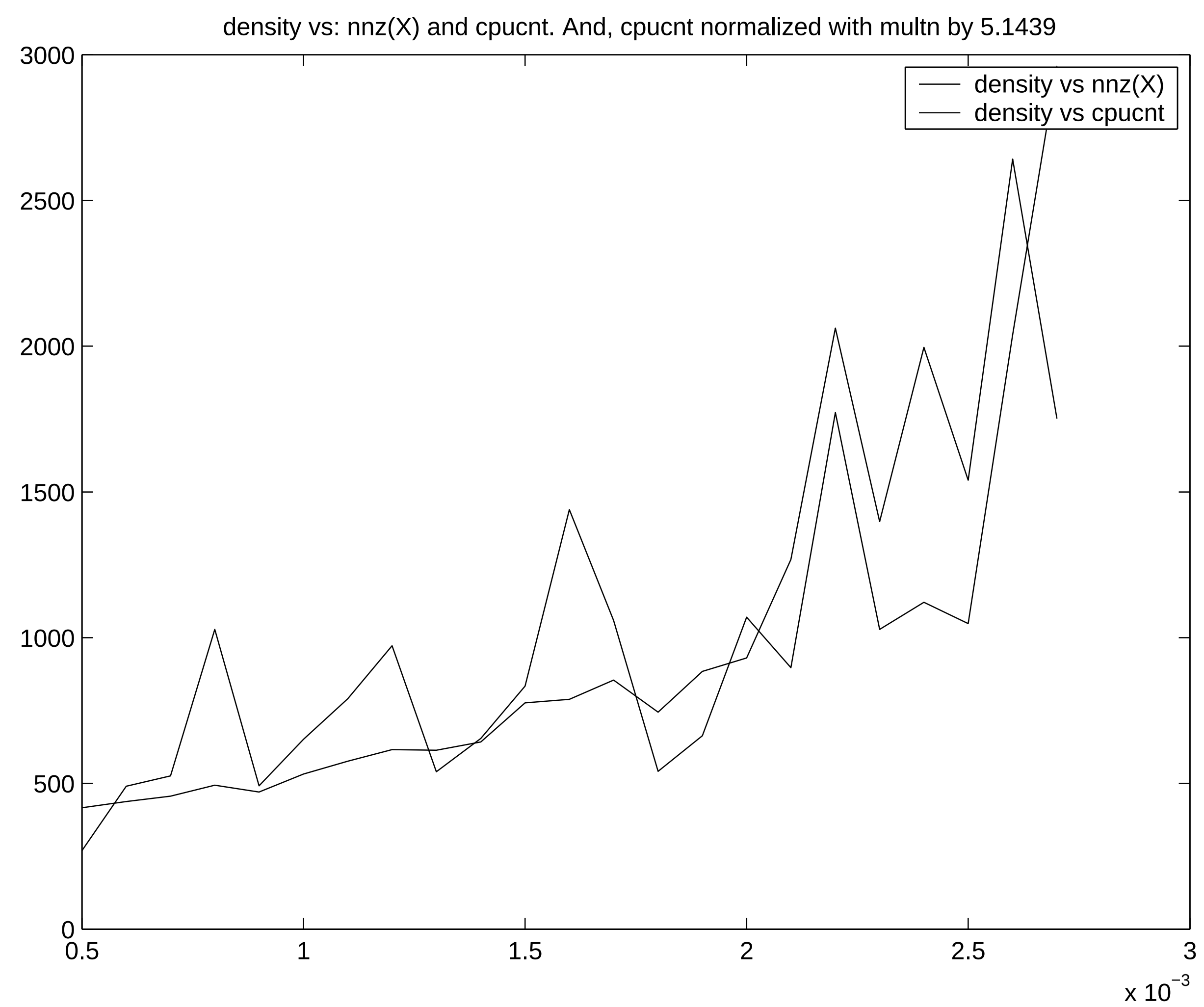

Figure 4.4: density .0005:.001:.027 vs CPU times and nnz(X), n=350

```
random A dim n =  100
density  0.008
number of nonzeros  80
noiter   -log10(relgapp)    step        sigmaa     optval        cputime.dir      cputimechol    lsqr.iter  nnz-X
   1      2.79e-02         0.95             1      5.2083e+01    4.2000e-01       4.0000e-02         9        322
   2      6.38e-01         0.95         0.715      5.0593e+01    5.5000e-01       1.3000e-01        10        322
   3      1.23e+00         0.95         0.715      5.0125e+01    7.8000e-01       1.2000e-01        18        322
*******CROSSover started at: 0.05856
   4      2.85e+00            1             0      5.0026e+01    1.0900e+00       1.1000e-01        33        322
   5      3.79e+00            1             0      5.0008e+01    1.5300e+00       1.1000e-01        49        322
   6      4.48e+00            1             0      5.0003e+01    1.6100e+00       1.2000e-01        52        322
   7      5.27e+00            1             0      5.0002e+01    1.5700e+00       1.1000e-01        51        322
   8      6.53e+00            1             0      5.0002e+01    1.4200e+00       1.0000e-01        44        322
   9      8.93e+00            1             0      5.0002e+01    6.1000e-01       1.1000e-01        12        322
  10      1.37e+01            1             0      5.0002e+01    8.7000e-01       9.0000e-02        23        322
 final norm(XSy) is: 1.8144e-14
 rel. min of eigs is: -4.2533e-15
```

```
start with  elements abs(A) > 0.4
and number of nonzeros in A: 24
noiter   -log10(relgapp)    step        sigmaa        optval    cputime.dir      cputimechol    lsqr.iter  nnz-X
   1       2.52e-01         0.855            1     5.1840e+01    3.8000e-01       3.0000e-02          9        126
   2       5.39e-01         0.95        0.7435     5.0492e+01    2.7000e-01       3.0000e-02          8        126
   3       1.16e+00         0.95         0.715     5.0110e+01    4.3000e-01       2.0000e-02         17        126
*******CROSSover started at: 0.069726
   4       2.80e+00            1             0     5.0020e+01    5.7000e-01       3.0000e-02         25        126
   5       3.77e+00            1             0     5.0004e+01    6.0000e-01       2.0000e-02         27        126
perturbing A - using elements abs(A) > 0.3
increasing number of nonzeros in A to:30
   6       4.40e+00            1             0     5.0001e+01    6.7000e-01       2.0000e-02         29        132
perturbing A - using elements abs(A) > 0.2
increasing number of nonzeros in A to:46
   7       5.01e+00            1             0     5.0000e+01    1.0200e+00       3.0000e-02         45        166
perturbing A - using elements abs(A) > 0.1
increasing number of nonzeros in A to:66
   8       3.89e+00            1             0     5.0001e+01    1.1100e+00       5.0000e-02         43        234
perturbing A - using elements abs(A) > 2.7756e-17
increasing number of nonzeros in A to:80
   9       4.76e+00            1             0     5.0003e+01    8.6000e-01       8.0000e-02         26        322
  10       5.72e+00            1             0     5.0002e+01    9.7000e-01       9.0000e-02         26        322
  11       7.35e+00            1             0     5.0002e+01    1.3200e+00       1.1000e-01         38        322
  12       1.06e+01            1             0     5.0002e+01    6.0000e-01       9.0000e-02         12        322
 final norm(XSy) is: 2.6748e-11
 rel. min of eigs is: -4.9538e-13
```

5 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