Multiscale Optimization Methods and Applications

Robust Algorithms
for
Large Sparse Semidefinite Programming (SDP)

Saturday, Feb. 28 2004

Robust Solutions of Large Sparse Semidefinite Programming with Applications to the Nearest Euclidean Distance Matrix Problem

Henry Wolkowicz
Department of Combinatorics & Optimization
University of Waterloo

CONCORDIA CUM VERITATE

(with: Suliman Homidan, Hua Wei)

OUTLINE

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

Notation and Motivation

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

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

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

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

Illustration/Motivation on LP Case

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

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

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

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

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

stationary point of the Lagrangian / optimality conditions

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

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

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

overdetermined system in SDP case:

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

apply symmetrization 'undoes preconditioning'

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

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

Reduction/Block Elimination for the Normal Equations

Step 1 (Eliminate Δz):

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

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

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

Step 2 (Eliminate Δx):

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

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

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

- uniformly bounded condition number, e.g. Güler et al 1993
- structured singularity, e.g. 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 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, ... ■

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

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

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

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.

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

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

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

Summary: Path-following and NOT Interior-point

• staying interior is a heuristic for staying within a neighbourhood of the central path

• staying interior is required for numerical accuracy when solving the *current* ill-conditioned reduced systems

Nearest Euclidean Distance Matrix using SDP and True Primal-Dual Method

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

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

EDM $= \{D = (d_{ij}) \in \mathcal{S}^n : d_{ij} = \|x_i - x_j\|,$ 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:

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

(Public domain software packages are available)

Operator Notation: us2vec , us2Mat , svec , sMat

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

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

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

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

adjoint transformation sMat $^* =$ svec since:

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

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

with partition:

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

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

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

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:

$$\begin{array}{rcll} X & := & \mathrm{sMat}\,(x) \succeq 0 & \text{(primal feasibility)} \\ \Lambda & := & \mathcal{L}^* \left\{ H^{(2)} \circ (\mathcal{L}(X)) \right\} - C, \quad \Lambda \succeq 0 & \text{(dual feasibility)} \\ \Lambda X & := & 0 & \text{(complementary slackness)} \end{array}$$

We eliminate Λ. exact primal-dual feasibility maintained during iterations; full rank Jacobian at optimality.

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

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

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.

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

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

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

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:**

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

  - •• **Find initial strictly feasible points:** both $X^0, \Lambda^0 := (\mathcal{W}(X) - C) \succ 0; \mu > 0$
  - •• **Set initial parameters:**

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

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

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

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

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

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

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

•• **update**

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

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

• **end (while)**.

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

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.)

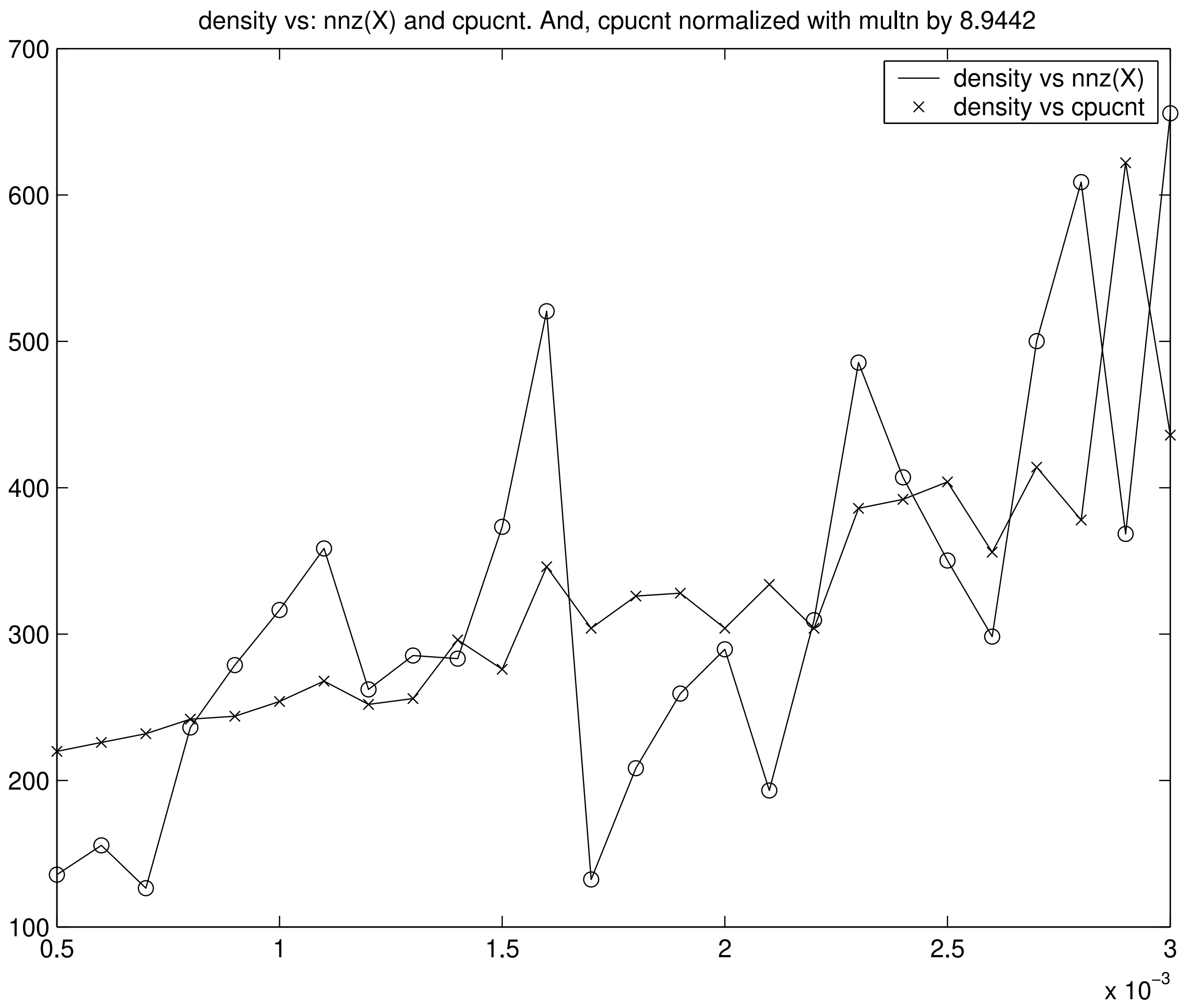

Figure 0.1: density .0005:.001:.003, CPU times and nnz(Λ), n=200

Conclusion

Gauss-Newton direction:

Advantages/Disadvantages:

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

Future:

- Need large sparse QR efficient as Cholesky
- predictor-corrector