Experiments with iterative computation of search directions within interior methods for constrained optimization

Santiago Akle and Michael Saunders
Institute for Computational Mathematics and Engineering (ICME)
Stanford University

Twelfth Copper Mountain Conference
on Iterative Methods

Copper Mountain, Colorado
March 25–30, 2012

Outline

1. PDCO
2. Iterative solvers
3. colnorms, QR
4. Partial Cholesky or QR
5. Numerical results

Abstract

Our primal-dual interior-point optimizer PDCO has found many applications for optimization problems of the form

$$\min \ \varphi(x) \ \text{ st } \ Ax = b, \ \ l \le x \le u,$$

in which $\varphi(x)$ is convex and A is a sparse matrix or a linear operator. We focus on the latter case and the need for iterative methods to compute dual search directions from linear systems of the form

$$AD^2A^T\Delta y = r, \quad D \text{ diagonal and positive definite.}$$

Although the systems are positive definite, they do not need to be solved accurately and there is reason to use MINRES rather than CG (see PhD thesis of David Fong (2011)). When the original problem is regularized, the systems can be converted to least-squares problems and there is similar reason to use LSMR rather than LSQR. Also, D becomes increasingly ill-conditioned as the interior method proceeds and there is need for some kind of preconditioning, such as the partial Cholesky approach of Bellavia, Gondzio and Morini (2011).

We present numerical results on matters such as these.

PDCO Primal-Dual Interior Method

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & Ax \qquad = b, \quad x \geq 0, \end{array}$$

PDCO is a MATLAB solver for such problems

A may be a sparse matrix or a linear operator

PDCO Primal-Dual Interior Method

$$\begin{aligned} \underset{x,\,r}{\text{minimize}} \quad & c^T x + \tfrac{1}{2}\|\gamma x\|^2 + \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & Ax + \delta r = b, \qquad x \ge 0, \end{aligned}$$

γ and $\delta \approx 10^{-4}$ for linear programs
$\delta = 1$ for nonnegative least-squares

PDCO is a MATLAB solver for such problems
A may be a sparse matrix or a linear operator

Primal-Dual Interior Method

PDCO solves a sequence of nonlinear equations

$$\begin{aligned} Ax + \delta^2 y &= b \\ A^T y + z &= c + \gamma^2 x \\ Xz &= \mu e \end{aligned}$$

$$X = \mathrm{diag}(x) \qquad \mu \searrow 0$$

Newton's method:

$$\begin{pmatrix} A & \delta^2 I & \\ -\gamma^2 I & A^T & I \\ Z & & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}$$

PDCO search direction

Define $D^2 = (X^{-1}Z + \gamma^2 I)^{-1}$
Posdef diagonal with big and small elements

Solve either

$$\left(AD^2A^T + \delta^2 I\right)\Delta y = AD^2 r_4 + r_1$$

or

$$\min \left\| \begin{pmatrix} DA^T \\ \delta I \end{pmatrix} \Delta y - \begin{pmatrix} Dr_4 \\ r_1/\delta \end{pmatrix} \right\|^2$$

D changes each PDCO iteration
Increasingly ill-conditioned

Iterative solvers

spd systems
least squares

Iterative solvers for PDCO

$$\left(AD^2A^T + \delta^2 I\right)\Delta y = AD^2 r_4 + r_1 \qquad \text{CG or MINRES}$$

$$\min \left\| \begin{pmatrix} DA^T \\ \delta I \end{pmatrix} \Delta y - \begin{pmatrix} Dr_4 \\ r_1/\delta \end{pmatrix} \right\|^2 \qquad \text{LSQR or LSMR}$$

Iterative solvers for PDCO

$$\left(AD^2A^T + \delta^2 I\right)\Delta y = AD^2 r_4 + r_1 \qquad \text{CG or MINRES}$$

$$\min \left\| \begin{pmatrix} DA^T \\ \delta I \end{pmatrix} \Delta y - \begin{pmatrix} D r_4 \\ r_1/\delta \end{pmatrix} \right\|^2 \qquad \text{LSQR or LSMR}$$

Comparisons:

Fong and S	2011a,b
Fong thesis	2011

Backward errors for CG and MINRES on spd $Ax = b$

$$(A + E_k)x_k = b \qquad r_k = b - Ax_k$$

$$E_k = \frac{r_k x_k^T}{\|x_k\|^2} \qquad \|E_k\| = \frac{\|r_k\|}{\|x_k\|}$$

Backward errors for CG and MINRES on spd $Ax = b$

$$(A + E_k)x_k = b \qquad\qquad r_k = b - Ax_k$$

$$E_k = \frac{r_k x_k^T}{\|x_k\|^2} \qquad\qquad \|E_k\| = \frac{\|r_k\|}{\|x_k\|}$$

We know

$\|r_k\| \searrow$	for MINRES	(but not for CG)
$\|x_k\| \nearrow$	for CG	(Steihaug 1983)
$\|x_k\| \nearrow$	for MINRES	(Fong 2011)

Backward errors for CG and MINRES on spd $Ax = b$

$$(A + E_k)x_k = b \qquad r_k = b - Ax_k$$

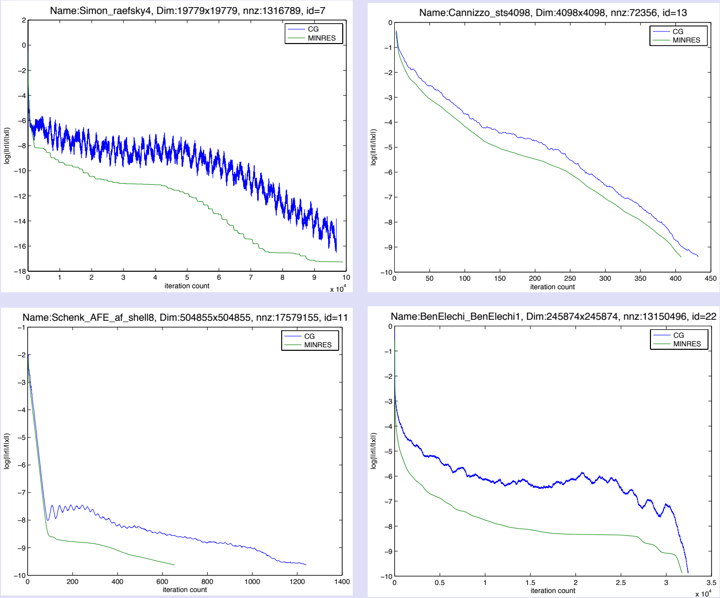
$$E_k = \frac{r_k x_k^T}{\|x_k\|^2} \qquad \|E_k\| = \frac{\|r_k\|}{\|x_k\|}$$

We know

$\|r_k\| \searrow$	for MINRES	(but not for CG)	
$\|x_k\| \nearrow$	for CG	(Steihaug	1983)
$\|x_k\| \nearrow$	for MINRES	(Fong	2011)

Plot $\log_{10} \|E_k\|$ for CG and MINRES
Data: Tim Davis's sparse matrix collection
Real spd examples $Ax = b$ that include b

Backward errors for CG and MINRES when $A \succ 0$



Backward errors for LSQR and LSMR

Stewart (1977):

$$\min \|(A + E_k)x_k - b\| \qquad r_k = b - Ax_k$$

$$E_k = -\frac{r_k r_k^T A}{\|r_k\|^2} \qquad \|E_k\| = \frac{\|A^T r_k\|}{\|r_k\|}$$

Backward errors for LSQR and LSMR

Stewart (1977):

$$\min \|(A + E_k)x_k - b\| \qquad r_k = b - Ax_k$$

$$E_k = -\frac{r_k r_k^T A}{\|r_k\|^2} \qquad \|E_k\| = \frac{\|A^T r_k\|}{\|r_k\|}$$

LSQR $\equiv$ CG on $A^T A x = A^T b$
LSMR $\equiv$ MINRES on $A^T A x = A^T b$
Hence $\|A^T r_k\| \searrow$ for LSMR (but not for LSQR)

$\min \|Ax - b\|$

Measure of convergence

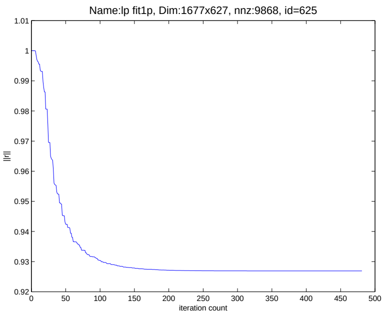
- $r_k = b - Ax_k$
- $\|r_k\| \to \|\hat{r}\|$, $\|A^T r_k\| \to 0$

$\min \|Ax - b\|$

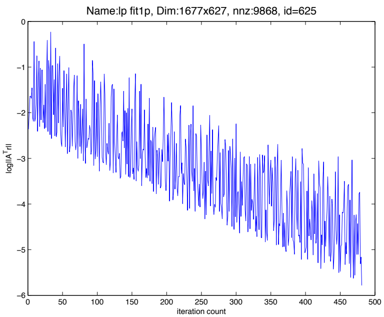
Measure of convergence

- $r_k = b - Ax_k$
- $\|r_k\| \to \|\hat{r}\|$, $\|A^T r_k\| \to 0$

LSQR $\|r_k\|$



LSQR $\log \|A^T r_k\|$

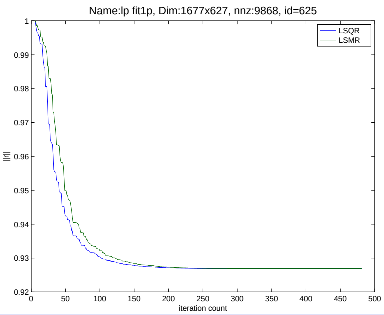


$\min \|Ax - b\|$

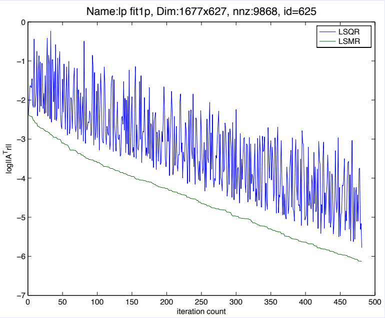
Measure of convergence

- $r_k = b - Ax_k$
- $\|r_k\| \to \|\hat{r}\|$, $\|A^T r_k\| \to 0$

$\|r_k\|$



$\log \|A^T r_k\|$



Two linear algebra tools

`colnorms.m`

Householder QR

`colnorms.m`

$$C = \boxed{\phantom{\begin{matrix} x \\ x \\ x \\ x \\ x \end{matrix}}} = [c_1, \dots, c_n] \qquad (m \times n)$$

- Estimates all $\|c_j\|$ from p products $C^T v$
- $p \approx \sqrt{n}$
- Each $v =$ `randn(m,1)`

If $A = C^T C$, we estimate $\mathrm{diag}(A)$ from $\texttt{colnorms}(C)^2$

Householder QR

$$W = \boxed{\phantom{\begin{matrix}W\\ \\ \\ \\ \\ \end{matrix}}} = Q\begin{pmatrix}R\\0\end{pmatrix} = \begin{pmatrix}Y & Z\end{pmatrix}\begin{pmatrix}R\\0\end{pmatrix} = YR \qquad (n \times k)$$

$Q =$ product of Householder transformations

$$Y = Q\begin{pmatrix}I\\0\end{pmatrix} \qquad Z = Q\begin{pmatrix}O\\I\end{pmatrix}$$

Q, Y, Z are fast operators if k is small

- Choose W to be approximate eigenvectors (say)
- Householder QR orthogonalizes W, represents full Q

Partial Cholesky or partial QR

Two ideas for spd $Hx = b$

Part direct, part iterative (hybrid!)

- Schur complement CG — Axelsson 1994
- Partial Cholesky preconditioning — Bellavia, Gondzio, Morini 2011

Two ideas for spd $Hx = b$

Part direct, part iterative (hybrid!)

- Schur complement CG — Axelsson 1994
- Partial Cholesky preconditioning — Bellavia, Gondzio, Morini 2011

Transform first: $Q^T HQy = Q^T b$
Q = Householder QR on W
W = 10, 20, 50, . . . approximate eigenvectors of H

- Two-level (subspace splitting) Schur complement CG
  Hanke and Vogel 1999
- Subspace preconditioned LSQR — Jacobsen, Hansen, and S 2003
  = partial QR equivalent of partial Cholesky

Partial Cholesky preconditioning for spd $Hx = b$

$$Q = [Y \ Z], \qquad Y \text{ is } n \times k \text{ for small } k$$

- BGM 2011: $Q =$ permutation (approx diagonal pivoting)
- Our experiments: Q is from Householder QR on $W \approx k$ eigenvectors

$$Q^T H Q = \begin{pmatrix} L_1 & \\ L_2 & I \end{pmatrix} \begin{pmatrix} I & \\ & S \end{pmatrix} \begin{pmatrix} L_1^T & L_2^T \\ & I \end{pmatrix}$$

$$M = \begin{pmatrix} L_1 & \\ L_2 & I \end{pmatrix} \begin{pmatrix} I & \\ & \tilde{S} \end{pmatrix} \begin{pmatrix} L_1^T & L_2^T \\ & I \end{pmatrix}$$

$\tilde{S} = \mathrm{diag}(S)$ or `colnorms` gives $\tilde{S} \approx \mathrm{diag}(S)$

Numerical Results

PDCO on LPs, satellite image

LP degen3	1503	$\times$	2604
LP fit2p	3000	$\times$	13525
Satellite	16384	$\times$	16384

MINRES on

$$\left(AD^2A^T + \delta^2 I\right)\Delta y = AD^2 r_4 + r_1$$

at PDCO iteration 1, middle, end

Think of systems as $Hx = b \quad \rightarrow \quad QHQ^T y = Q^T b$

Preconditioners constructed from $k = 50$ cols of partial Cholesky

$Q =$ permutation for diagonal pivoting

or Householder QR on k approx eigenvectors

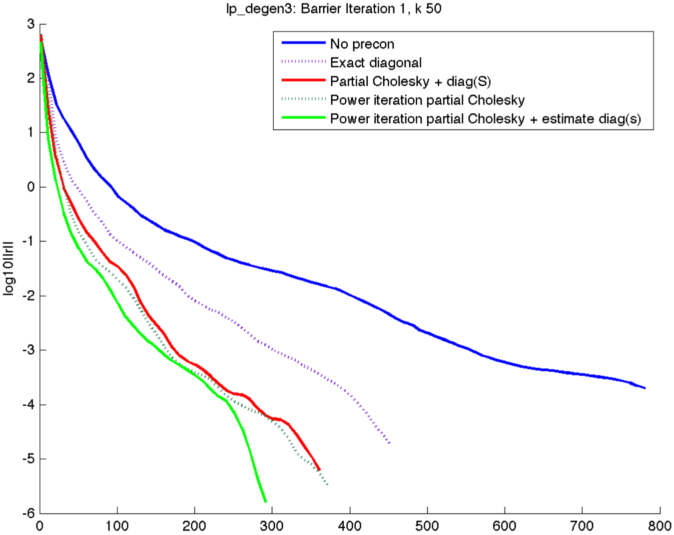
MINRES on $\left(AD^2A^T + \delta^2 I\right)\Delta y = AD^2 r_4 + r_1$ at PDCO iteration 1, middle, end

Think of systems as $Hx = b \quad \rightarrow \quad QHQ^T y = Q^T b$
Preconditioners constructed from $k = 50$ cols of partial Cholesky

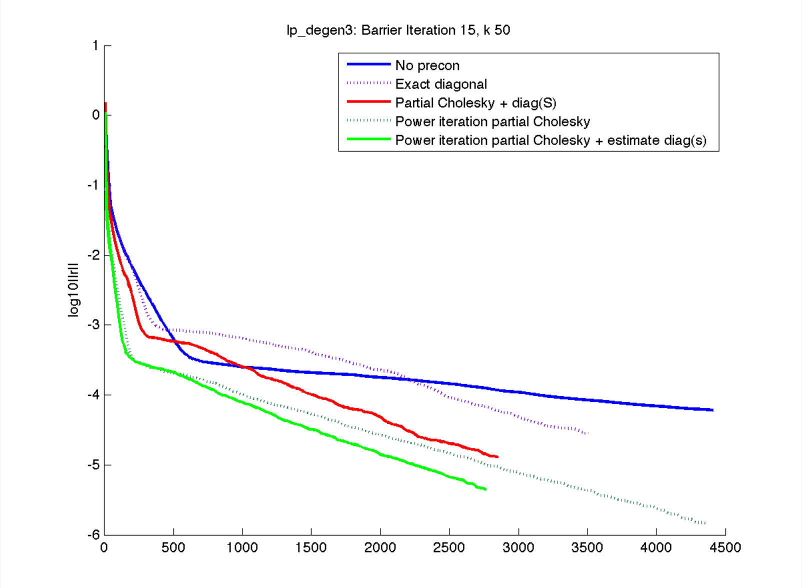
Q = permutation for diagonal pivoting
or Householder QR on k approx eigenvectors

Schur complement $S = C^T C$ approximated by
diag(S) or colnorms(C)

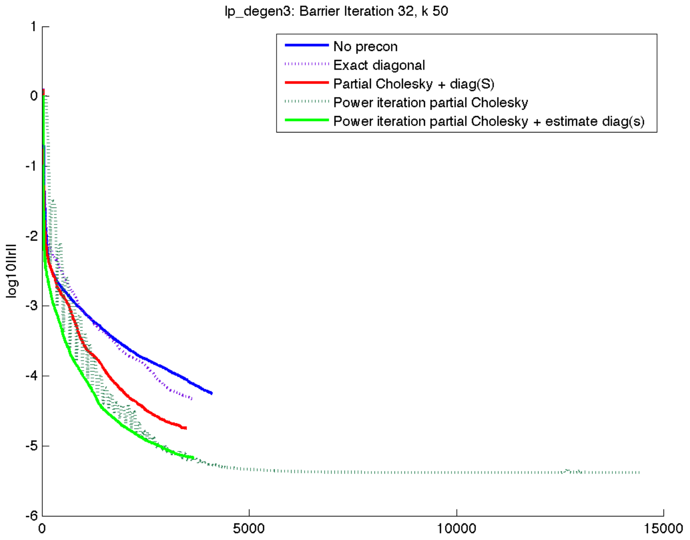
LP degen3, PDCO itn 1



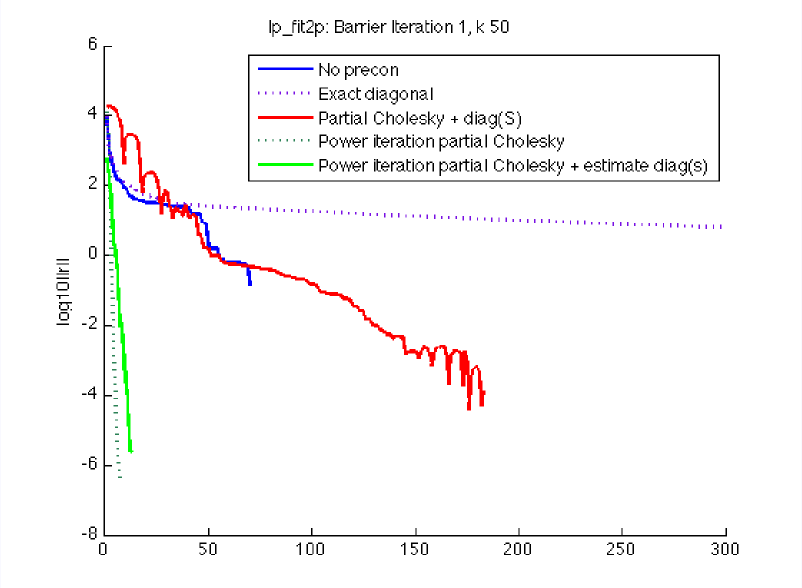
LP degen3, PDCO itn 15



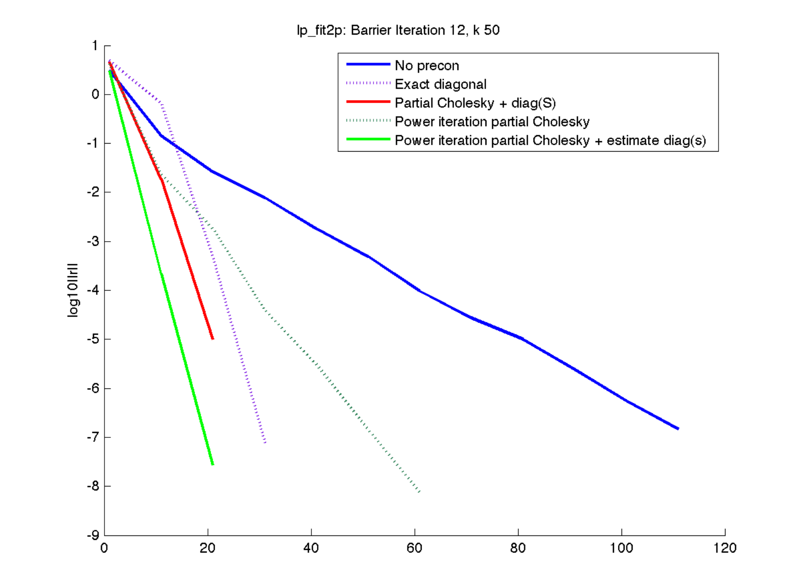
LP degen3, PDCO itn 32



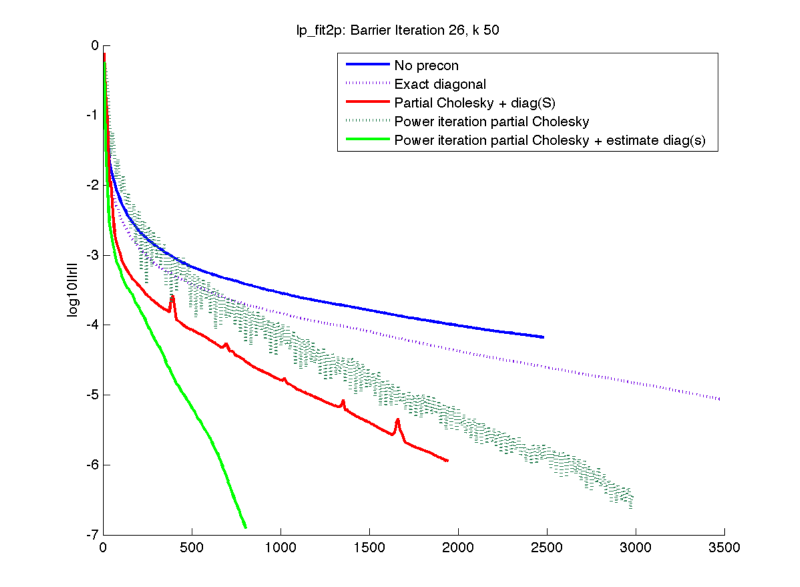
LP fit2p, PDCO itn 1



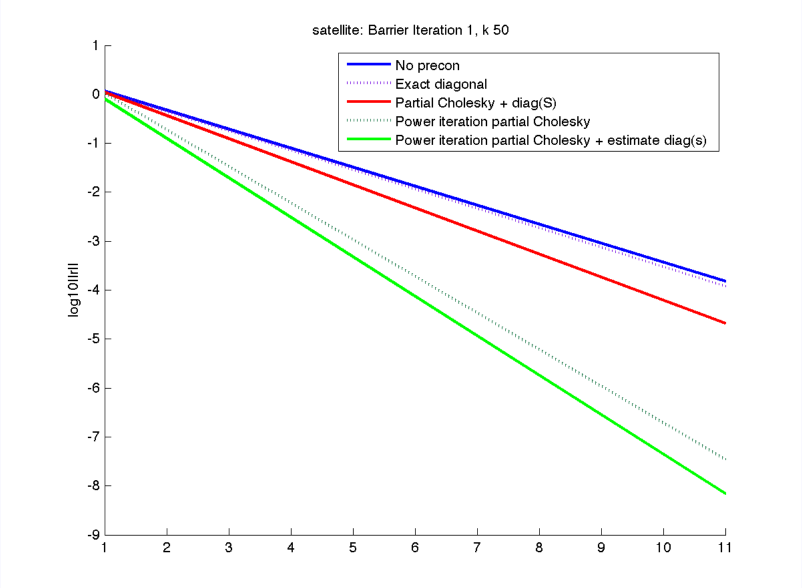
LP fit2p, PDCO itn 12



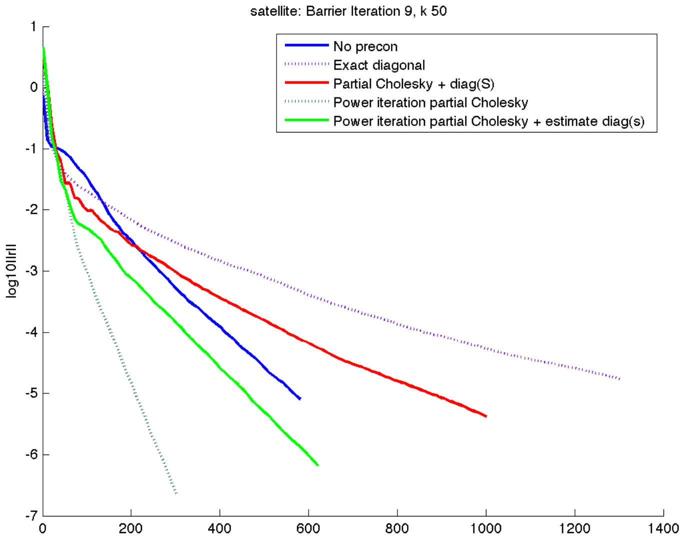
LP fit2p, PDCO itn 26



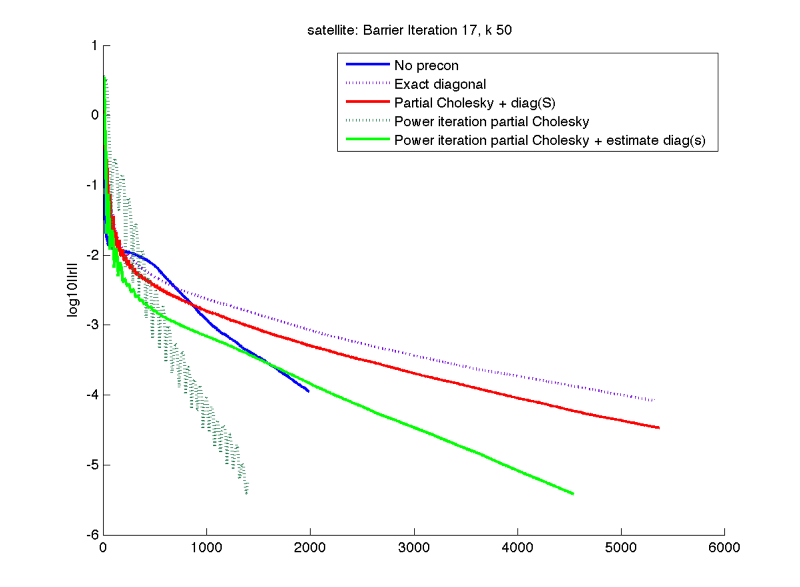
Satellite, PDCO itn 1



Satellite, PDCO itn 9



Satellite, PDCO itn 17



References

- O. Axelsson (1994),
  Iterative Solution Methods,
  Cambridge Univ Press
- M. Hanke and C. R. Vogel (1999),
  Two-level preconditioners for regularized inverse problems,
  Numer Math 83, 385–402
- M. Jacobsen, P. C. Hansen, and M. A. Saunders (2003),
  Subspace preconditioned LSQR for discrete ill-posed problems,
  BIT 43:5, 975–989
- S. Bellavia, J. Gondzio, and B. Morini (Jul 2011),
  New matrix-free preconditioner for sparse least-squares problems,
  Report ERGO-11-010, University of Edinburgh
- D. C.-L. Fong and M. A. Saunders (Aug 2011),
  LSMR: An iterative algorithm for sparse least-squares problems,
  SISC 33, 2950–2971
- D. C.-L. Fong and M. A. Saunders (Dec 2011),
  CG versus MINRES: An empirical comparison,
  for SQU Journal for Science
- D. C.-L. Fong (Dec 2011), *Minimum-Residual Methods for Sparse Least-Squares Using Golub-Kahan Bidiagonalization,*
  PhD thesis, ICME, Stanford University

Special thanks to Sven Leyffer

Conferences really do promote action and creativity!