Preconditioning for iterative computation of search directions within interior methods for constrained optimization

Santiago Akle, Michael Saunders

[1]ICME
Stanford University

ISMP, 2012

Abstract

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

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

The solution of a search direction requires solving KKT systems of the form $\begin{pmatrix} -H + D & A^T \\ A & D_2 \end{pmatrix}$ or $A\,(H - D)^{-1} A^T + D_2$ for which there is reason to use MINRES rather than CG (see PhD thesis of David Fong (2011)) [Fon11].
For the cases when A is given as an operator, iterative methods are essential. We review the importance of eigenvalue clustering when using Krylov subspace methods. We show that we can find systems with very bad conditioning for which these methods converge promptly.
We establish a relationship between systems modified by partial Cholesky preconditioning, deflated systems, and systems solved via the Schur complement method.

Overview

- PDCO
- Partial Cholesky preconditioning.
- The behavior of MINRES in badly scaled systems but with good clustering.
- Changes in the spectrum after rank one updates.
- Schur complements, partial Cholesky preconditioning, and deflation as sequences of rank one updates.

PDCO Primal-Dual Interior Method

$$\underset{x,r}{\text{minimize}}\ c^T x + \frac{1}{2}\|\gamma x\|^2 + \frac{1}{2}\|r\|^2 \tag{1}$$

$$\text{subject to } Ax + \delta r = b,\ \ x \geq 0, \tag{2}$$

γ and $\delta \simeq 10^{-4}$ for linear programs
$\delta = 1$ for non negative least squares.

PDCO [Sau02] is a MATLAB solver for such problems
A may be sparse matrix or a linear *operator*. Can use iterative methods to solve the linear systems that arise.

Primal-Dual Interior Method

PDCO solves a sequence of nonlinear equations

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

Where $X = diag(x)$ and $\mu \downarrow 0$.

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's search direction

Define $D^2 = \left(X^{-1}Z + \gamma^2 I\right)$

Posdef diagonal with big and small elements and increasingly ill-conditioned.

PDCO solves either

$$\min \left\| \begin{pmatrix} D^{-1}A^T \\ \delta I \end{pmatrix} \Delta y - \begin{pmatrix} D^{-1}r_4 \\ r1/\delta \end{pmatrix} \right\|,$$

or the Schur complement of the KKT matrix

$$K = AD^{-2}A^T + \delta^2 I.$$

When A is an operator access to the entries is expensive, full columns cost the same as entries.

Partial Cholesky preconditioner

We hope to use columns to form a preconditioner cheaply.
Partial Cholesky lends itself to the task [Gon11].
The preconditioner requires a subset of columns of K.
Let

$$\begin{pmatrix} K_1 & K_2 \\ K_2^T & K_3 \end{pmatrix} = \begin{pmatrix} R^T & \\ K_2^T R^{-1} & I \end{pmatrix} \begin{pmatrix} I & \\ & S \end{pmatrix} \begin{pmatrix} R & R^{-T} K_2 \\ & I \end{pmatrix},$$

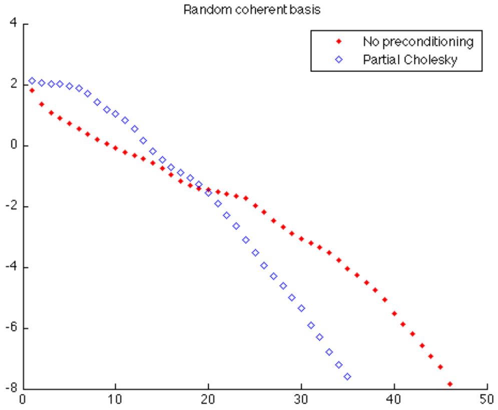
where $K_1 = R^T R$, and denote the partial Cholesky factor by

$$M^{\frac{1}{2}} = \begin{pmatrix} R & R^{-T} K_2 \\ & I \end{pmatrix}.$$

The symmetrically preconditioned matrix takes the form

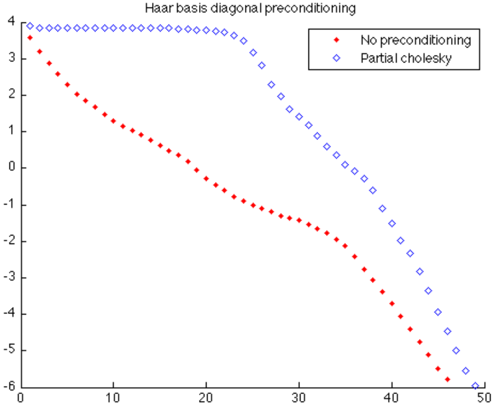
$$M^{-\frac{1}{2}^T} K M^{-\frac{1}{2}} = \begin{pmatrix} I & \\ & S \end{pmatrix}.$$

How good is partial Cholesky preconditioning?



Depends greatly on the system.

How good is partial Cholesky preconditioning?



Depends greatly on the system.

Example systems for numerical experiments

$$K = V \begin{pmatrix} \Lambda_1 & 0 & & \\ 0 & \Lambda_2 & & \\ & & \ddots & \\ & & & \Lambda_k \end{pmatrix} V^T$$

V is chosen as a random orthonormal matrix.

$$\Lambda_j = \begin{pmatrix} \lambda_j & & 0 \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}$$

Two eigenvalue system

$$n = 100,\ \lambda_1 = -10^{-10},\ \lambda_2 = 1,\ \kappa \simeq 10^{10}.$$

```
minres.m   SOL, Stanford University   Version of 06 Jul 2009
Solution of symmetric Ax = b or (A-shift*I)x = b

n      =    1000     shift =  0.00000000000000e+00
itnlim =   20000     rtol  =  1.00e-15

A is an explicit dense matrix

  Itn      x(1)       Compatible     LS        norm(A)   cond(A)  gbar/|A|

     1   2.29472e-02   5.619e-01   5.729e-01   1.2e+00   1.0e+00   4.1e-01
     2  -7.73470e+06   3.274e-15   9.799e-11   1.4e+00   5.1e+09  -9.8e-11
     3  -7.73470e+06   3.618e-16   5.697e-01   1.7e+00   7.1e+09   5.6e-01

istop   =  1               itn   =     3
Anorm   =  1.7393e+00      Acond =  7.0746e+09
rnorm   =  4.5003e-06      ynorm =  7.1519e+09
Arnorm  =  3.3171e-05
A solution to Ax = b was found, given rtol
```

Three eigenvalue system

$$n = 100,\ \lambda_1 = -10^{-20},\ \lambda_2 = 10^{-10},\ \lambda_3 = 1,\ \kappa \simeq 10^{20}.$$

```
minres.m    SOL, Stanford University    Version of 06 Jul 2009
Solution of symmetric Ax = b or (A-shift*I)x = b

n      =     120     shift =  0.00000000000000e+00
itnlim =    2400     rtol  =  1.00e-16

A is an explicit dense matrix

  Itn     x(1)       Compatible    LS         norm(A)  cond(A) gbar/|A|

     1  1.93969e-10  1.380e+00  1.000e+00  5.9e+09  1.0e+00   5.9e-01
     2  1.09210e+00  6.929e-11  1.017e-10  1.0e+10  4.8e+09  -8.9e-11
     3  1.09210e+00  6.929e-11  1.761e-05  1.0e+10  4.8e+09   4.3e-10
     4  1.39660e+05  4.050e-16  3.297e-12  1.4e+10  4.6e+10  -3.3e-12
     5  1.39660e+05  4.050e-16  3.459e-05  1.4e+10  4.6e+10   4.6e-09
     6  2.18305e+05  2.022e-16  6.546e-08  1.7e+10  4.6e+10  -6.5e-08
     7  2.18305e+05  2.022e-16  8.122e-10  1.7e+10  4.6e+10  -6.0e-17
     8  2.30375e+05  2.859e-16  1.160e-08  2.0e+10  4.6e+10   1.2e-08
     9  2.30375e+05  2.859e-16  3.996e-09  2.0e+10  4.6e+10  -4.9e-12

    10  2.35316e+05  2.254e-16  1.691e-04  2.2e+10  4.6e+10   1.7e-04
    11  2.35316e+05  2.254e-16  8.926e-14  2.2e+10  6.1e+11  -1.2e-16
    12  6.16006e+05  6.768e-17  1.130e-04  2.4e+10  6.1e+11   1.1e-04

istop  =  1             itn   =    12
Anorm  =  2.4495e+10    Acond =  6.0689e+11
rnorm  =  4.1165e+00    ynorm =  2.4831e+06
Arnorm =  1.3555e+07
A solution to Ax = b was found, given rtol
```

Three eigenvalue system, with slight perturbation

$n = 100$, $\lambda_1 = -10^{-20} \pm 10^{-5}\lambda_1^2$, $\lambda_2 = 10^{-10} \pm 10^{-5}\lambda_2^2$, $\lambda_3 = 1 \pm 10^{-5}$, $\kappa \simeq 10^{20}$.

```
minres.m   SOL, Stanford University   Version of 06 Jul 2009
Solution of symmetric Ax = b or (A-shift*I)x = b

n      =   1200    shift =  0.00000000000000e+00
itnlim =  24000    rtol  =  1.00e-15

A is an explicit dense matrix

  Itn     x(1)       Compatible    LS        norm(A)  cond(A)  gbar/|A|

     1  1.09613e+00  2.395e-02  1.734e-02  3.4e+01  1.0e+00   1.0e-02
     2  1.22789e+00  1.389e-02  6.151e-06  3.4e+01  1.7e+03  -2.2e-09
     3  2.31312e+03  1.684e-06  7.316e-09  3.4e+01  2.3e+03   7.1e-09
     4  1.27284e+09  2.192e-12  2.774e-05  3.4e+01  2.3e+03  -2.8e-05
     5  1.31856e+09  2.080e-12  5.345e-03  3.4e+01  2.3e+03   1.4e-03
     6  1.31856e+09  2.079e-12  1.443e-06  3.4e+01  2.3e+03  -3.8e-10
     7  1.31856e+09  2.078e-12  3.596e-10  3.4e+01  2.4e+03   2.4e-13
     8  1.30998e+09  2.075e-12  5.893e-10  3.4e+01  2.4e+03  -5.9e-10
     9 -7.43433e+13  1.152e-16  2.461e-06  3.4e+01  2.4e+03   2.5e-06

istop  =  1               itn   =    9
Anorm  =  3.3938e+01      Acond =  2.4494e+03
rnorm  =  1.8724e+01      ynorm =  4.7877e+15
Arnorm =  1.6132e-03
A solution to Ax = b was found, given rtol
```

Eigenvalues after rank one update

- Let K be hermitian $n \times n$, let $y \in \Re^n \neq 0$.
- Let $B = K - yy^T$.
- Let x, λ be an eigen pair of B.
- Let $\rho(x) := y^T x \neq 0$.

$$
\begin{aligned}
&Bx = \lambda x && \text{(3a)}\\
\Rightarrow &Kx - \lambda x = y(y^T x) && \text{(3b)}\\
\Rightarrow &(K - \lambda I)^{-1} y = \rho^{-1} x && \text{(3c)}\\
\Rightarrow &y^T (K - \lambda I)^{-1} y = 1 && \text{(3d)}
\end{aligned}
$$

If

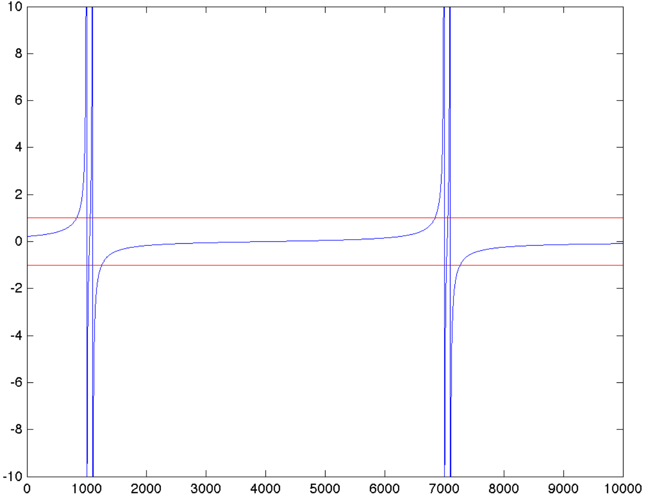
$$K = VDV^T \text{ is the eigensystem and} \tag{4}$$
$$\hat{y} = Vy \text{ then} \tag{5}$$
$$1 = y^T(K - \lambda I)y = \hat{y}^T(D - \lambda I)^{-1}\hat{y}, \tag{6}$$

which is equivalent to the secular equation:

$$\phi(\lambda) = \sum_i^n \frac{\hat{y}_i^2}{d_i - \lambda} = 1.$$

Eigenvalues after rank one update

If $d_i < d_{i+1}$ are consecutive eigenvalues of K, then $\phi(\lambda)$ will have a root at $d_i < \lambda < d_{i+1}$ strictly between the eigenvalues.



Deflated systems

Let $K \in \Re^{n \times n}$ hermitian, and let $\mathcal{W}$ be a subspace $\mathcal{W} \subseteq \Re^n$. the deflation of K w.r.t. $\mathcal{W}$ is given by:

$$B = K - KW(W^T KW)^{-1} W^T K,$$

where W is any basis for $\mathcal{W}$.

0.4.1. Claim. *The selection of the basis will not affect the resulting system.*

0.4.2. Claim. *Given $\mathcal{W}_1 \subseteq \mathcal{W}_2 \subseteq \cdots \subseteq \mathcal{W}_k$ nested subspaces, there exist $W_1 \subseteq W_2, \ldots, W_k$ nested basis, such that $W^T KW = D$ is diagonal.*

Deflated systems

Assume $W^T K W = D$ is diagonal, then

$$B = K - \sum_i^k K w_i \frac{1}{d_i} w_i^T K = K - \sum_i^k \pm y_i y_i^T,$$

B is formed from by a sequence of rank one updates to K.

Schur complement and partial Cholesky as deflation

Let $\mathcal{W}$ be the span $\{e_1, \ldots, e_k\}$

$$\begin{aligned} B &= \begin{pmatrix} K_1 & K_2 \\ K_2^T & K_3 \end{pmatrix} - \begin{pmatrix} K_1 \\ K_2^T \end{pmatrix} K_1^{-1} \begin{pmatrix} K_1 & K_2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \\ & K_2 - K_2^T K_1^{-1} K_2 \end{pmatrix} = \begin{pmatrix} 0 & \\ & S \end{pmatrix}. \end{aligned}$$

With S the Schur complement of K_1.
The partial Cholesky preconditioned matrix was

$$M^{-\frac{1}{2}^T} K M^{-\frac{1}{2}} = \begin{pmatrix} I & \\ & S \end{pmatrix}.$$

Two eigenvalue system no preconditioning

$n = 100,\ \lambda_1 = 10^{-5}\ \lambda_2 = 1,\ \kappa = 10^5.$

```
minres.m   SOL, Stanford University   Version of 06 Jul 2009
Solution of symmetric Ax = b or (A-shift*I)x = b

n      =      100     shift =  0.00000000000000e+00
itnlim =     2000     rtol  =  1.00e-11

A is an explicit dense matrix

  Itn     x(1)       Compatible     LS         norm(A)   cond(A)  gbar/|A|

    1  -1.12210e-01   6.284e-01   6.306e-01   1.1e+00   1.0e+00   4.5e-01
    2  -5.21418e+03   1.445e-15   1.064e-05   1.3e+00   5.0e+04  -1.1e-05

istop   =  1               itn   =     2
Anorm   =  1.3266e+00      Acond =  5.0179e+04
rnorm   =  1.1795e-10      ynorm =  6.1530e+04
Arnorm  =  8.6861e-06
A solution to Ax = b was found, given rtol
```

Two eigenvalue system with preconditioning, $k = 25$

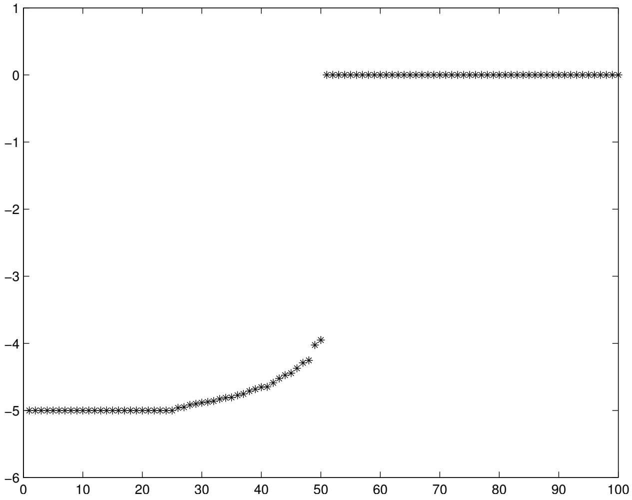


Figure: Spectrum of the partial Cholesky preconditioned matrix

Two eigenvalue system with preconditioning, $k = 25$

$n = 100$, $\lambda_1 = 10^{-5}$ $\lambda_2 = 1$, $\kappa = 10^5$.

Itn	x(1)	Compatible	LS	norm(A)	cond(A)	gbar/|A|
1	1.41643e+00	1.065e−01	3.527e−01	2.0e+00	1.0e+00	2.5e−01
2	7.50373e+03	6.974e−06	4.172e−05	2.1e+00	7.5e+03	−3.9e−05
3	8.58631e+03	2.569e−06	1.263e−05	2.1e+00	2.6e+04	1.1e−05
4	7.13714e+03	1.346e−06	1.506e−05	2.1e+00	2.6e+04	−1.3e−05
5	7.13498e+03	1.343e−06	2.467e−02	2.1e+00	2.6e+04	1.6e−03
6	7.64939e+03	4.854e−07	2.068e−04	2.4e+00	2.6e+04	−2.1e−04
7	8.52454e+03	2.023e−07	1.549e−05	2.4e+00	2.6e+04	1.4e−05
8	7.92197e+03	8.392e−08	1.225e−05	2.4e+00	2.6e+04	−1.1e−05
9	7.92186e+03	8.387e−08	1.239e−02	2.4e+00	2.6e+04	4.1e−04
10	7.87576e+03	3.785e−08	3.787e−04	2.6e+00	2.6e+04	−3.8e−04
11	7.93076e+03	1.263e−08	9.655e−06	2.6e+00	2.8e+04	9.2e−06
12	7.92581e+03	3.680e−09	8.467e−06	2.6e+00	3.3e+04	−8.1e−06
13	7.92581e+03	3.676e−09	1.690e−02	2.6e+00	3.3e+04	7.7e−04
14	7.92487e+03	1.204e−09	2.033e−04	2.7e+00	3.3e+04	−2.0e−04
15	7.92555e+03	4.154e−10	8.185e−06	2.7e+00	3.3e+04	7.7e−06
16	7.92662e+03	9.242e−11	8.590e−06	2.7e+00	3.3e+04	−8.4e−06
17	7.92662e+03	9.138e−11	5.100e−02	2.8e+00	3.3e+04	7.4e−03
18	7.92655e+03	2.880e−11	5.410e−05	2.9e+00	3.3e+04	−5.4e−05
19	7.92645e+03	6.807e−12	6.858e−06	2.9e+00	3.5e+04	6.7e−06

```
istop  =  1               itn   =    19
Anorm  =  2.9247e+00      Acond =  3.5315e+04
rnorm  =  1.4794e-06      ynorm =  7.4312e+04
Arnorm =  1.2555e-10
A solution to Ax = b was found, given rtol
```

Two eigenvalue system preconditioned, $k = 50$

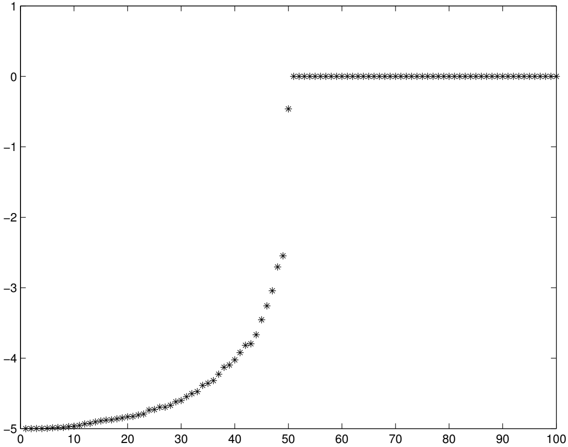


Figure: Spectrum of the partial Cholesky preconditioned matrix with 50 steps

Iterations 51, residual norm rnorm $\simeq 2^{-6}$.

Limiting the damage

Find an approximate invariant space (cheaply) and rotate. Let $Q = [WZ]$ with $Q^T Q = I$ and W, Z basis for invariant subspaces, then

$$Q^T A Q \simeq \begin{pmatrix} Y^T K Y & 0 \\ 0 & Z^T K Z \end{pmatrix}.$$

Then Cholesky preconditioning will yield

$$M^{-\frac{1}{2}} Q^T K Q M^{-\frac{1}{2}} \simeq \begin{pmatrix} I & 0 \\ 0 & Z^T K Z \end{pmatrix}$$

Rotated two eigenvalue system preconditioned, $k = 25$

$n = 100$, $\lambda_1 = 10^{-5}$, $\lambda_2 = 10$.

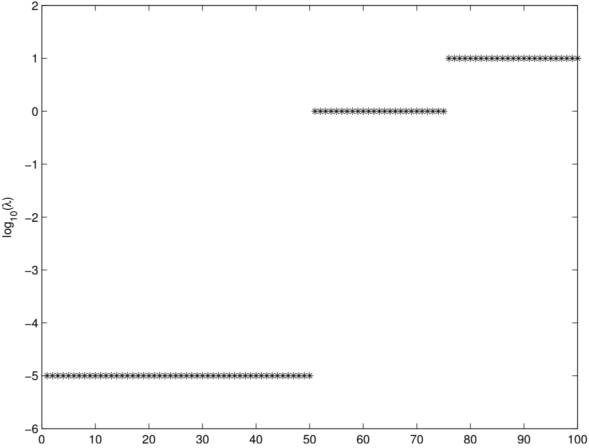


Figure: Spectrum of the partial Cholesky preconditioned matrix with 25 steps

Thank you!

Bibliography I

David Fong.
Minimum-residual methods for sparse least-squares using golub-kahan bidiagonalization.
PhD. Thesis ICME, Stanford University, 2011.

Jacek Gondzio.
Interior Point Methods 25 Years Later .
pages 1–33, 2011.

Michael Saunders.
Pdco: Primal-dual method for optimization with convex objectives.
2002.