Switching preconditioners in a hybrid approach for linear systems arising from interior point methods

Petra Maria Bartmeyer UNICAMP
Silvana Bocanegra UFRPE
Aurelio R. L. Oliveira
aurelio@ime.unicamp.br

University of Campinas

Preconditioning 2017

Summary

- Interior Point Methods
- Linear Systems Solution
- Splitting Preconditioner
- Controlled Cholesky Factorization
- Change of Preconditioner
- Numerical Experiments
- Conclusions

Standard LP

$$(\mathrm{P})\begin{cases}\min & c^T x\\ \text{s.t.} & A\ x = b\\ & \quad x \geq 0\end{cases}\qquad (\mathrm{D})\begin{cases}\max & y^T b\\ \text{s.t.} & A^T y + z = c\\ & \quad z \geq 0\\ & \quad y \in \mathbb{R}^m\end{cases}$$

$m < n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$.

KKT conditions

$$\begin{aligned} Ax - b &= 0\\ A^T y + z - c &= 0\\ XZe &= 0\\ (x,z) &\geq 0.\end{aligned}$$

$X = \mathrm{diag}(x)$, $Z = \mathrm{diag}(z)$ and $e = (1,\ldots,1)^T$.

Mehrotra predictor–corrector method

Given $(x, z) > 0$

- Affine-scaling direction

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta_a x^k \\ \Delta_a y^k \\ \Delta_a z^k \end{bmatrix} = \begin{bmatrix} r_p \\ r_d \\ r_a \end{bmatrix}$$

$r_p = b - Ax^k$, $r_d = c - A^T y^k - z^k$ and $r_a = -X^k Z^k e$.

- Centering-corrector direction $(\Delta_c x^k, \Delta_c y^k, \Delta_c z^k)$,

$$r_p = 0, \quad r_d = 0 \quad \text{and} \quad r_a = \sigma_k \mu_k e - \Delta_a X^k \Delta_a Z^k e \,,$$

σ_k is the centering parameter
$\Delta_a X^k = \mathrm{diag}(\Delta_a x^k)$ and $\Delta_a Z^k = \mathrm{diag}(\Delta_a z^k)$.

Augmented system

$$\begin{bmatrix} -\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_d - (X^k)^{-1} r_a \\ r_p \end{bmatrix}$$

$\Theta^{-1} = (X^k)^{-1} Z^k$.
Schur complement

$$(A\Theta A^T)\Delta y = A\Theta\left(r_d - (X^k)^{-1} r_a\right) + r_p \,.$$

Conjugate gradient method

$$M^{-1}(A\Theta A^T)M^{-T}\overline{y} = M^{-1}\left(A\Theta\left(r_d - (X^k)^{-1} r_a\right) + r_p\right),$$

$\bar{y} = M^T \Delta y$.

Linear system solution

$$\mathbf{A} = A\Theta A^T = \sum_{i=1}^{n} \theta_i A_i (A_i)^T = \sum_{i=1}^{n} \theta_i A_i (A_i)^T$$

1. Cholesky factorization.
2. Dense column implies dense $A\Theta A^T$.
3. Dense Cholesky factor.
4. Iterative methods.
5. $\theta_i = \frac{x_i}{z_i}$. $\quad x_i^* z_i^* = 0$.
6. Preconditioner.

Splitting Preconditioner: Features

- No Need to Compute $A\Theta A^T$.
- Designed for last IPM iterations.
  - Based on the behavior of diagonal matrix Θ.
  - Needs a permutation matrix P^n.
- Compute B by rectangular LU factorization of A.
- Designed for both Augmented and Schur complement systems.
- Bounded condition number.
- Works fine near a solution.
- It is not efficient far from a solution.

$$A = [B \;\; N]P^n, \quad P^n \Theta (P^n)^T = \begin{pmatrix} \Theta_B & 0 \\ 0 & \Theta_N. \end{pmatrix}$$

where P^n is a permutation matrix and B is a nonsingular matrix.

Splitting Preconditioner for the Augmented System:

$$P = \begin{pmatrix} \Theta^{1/2} & G \\ H & 0 \end{pmatrix}$$

$H = [I_m \; 0_{m \times n-m}]P^n$ and $G = H^t \Theta_B^{-1/2} B^{-1}$.

Splitting Preconditioner for the Normal Equations System:

$$P = \Theta_B^{-1/2} B^{-1}$$

$$PA\Theta A^T P^T = I + \Theta_B^{-1/2} B^{-1} N \Theta_N N^T B^{-T} \Theta_B^{-1/2}$$

Controlled Cholesky Factorization (CCF)

- Positive defined matrices
- The filling η can be controlled

Preconditioner construction:

$$A\Theta A^T = LL^T$$

Controlled Cholesky Factorization (CCF)

- Positive defined matrices
- The filling η can be controlled

Preconditioner construction:

$$A\Theta A^T = LL^T = \tilde{L}\tilde{L}^T + R$$

R is a remainder matrix
L is Cholesky factor $\tilde{L}$ is the controlled Cholesky factor

Controlled Cholesky Factorization (CCF)

- Positive defined matrices
- The filling η can be controlled

Preconditioner construction:

$$A\Theta A^T = LL^T = \tilde{L}\tilde{L}^T + R$$

R is a remainder matrix
L is Cholesky factor $\tilde{L}$ is the controlled Cholesky factor

Let $E = L - \tilde{L}$, the CCF minimizes $\|E\|_F$.
$\tilde{L}^{-1}(A\Theta A^T)\tilde{L}^{-T} = (I + \tilde{L}^{-1}E)(I + L^{-1}E)^T$.
When $\tilde{L} \approx L$, $E \approx 0$ then $\tilde{L}^{-1}(A\Theta A^T)\tilde{L}^{-T} \approx I$.

Controlled Cholesky Factorization

$$\min \|E\|_F^2 = \min \sum_{j=1}^{m} \left(\sum_{k=1}^{m_j+\eta} |l_{i_k j} - \tilde{l}_{i_k j}|^2 + \sum_{k=m_j+\eta+1}^{m} |l_{i_k j}|^2 \right) .$$

m_j number of nonzero entries below the diagonal
η number of extra entries allowed per column

CCF

Chose $m_j + \eta$ entries of $\tilde{L}_j$ by value.

- $\eta = -m$: diagonal preconditioner
- $\eta = 0$: incomplete Cholesky factorization
- $\eta = m$: full Cholesky factorization

Diagonal Failure - restart.

Hybrid approach

Two phases

1. General preconditioner
2. Specific Preconditioner

Two phases

1. Controlled Cholesky Factorization Preconditioner
2. η starts small
3. Increase η as needed
4. Splitting Preconditioner

Change of phases

Heuristic A

Heuristic proposal by [2]:

- *IterCG* conjugate gradient iterations
- *m* matrix dimension
- γ duality gap

---

if $\gamma < 10^{-6}\gamma_0$ *or (2* iterCG* $\leq$ *m and* $\eta = \eta_{max}$*)* **then**
    Phase exchange
end
if $4 * iterCG \geq m$ *and* $\eta < \eta_{max}$ **then**
    $\eta = \eta + 10$
end

Heuristic B

Heuristic proposal by [3]:

if $6 * iterCG \geq m$ **then**
    if $\eta < \eta_{max}$ **then**
        $\eta \leftarrow \eta + 10$ **else**
            Phase exchange
        end
    end
end

Heuristic proposal

- *IterCG* conjugate gradient iterations
- *m* matrix dimension
- Estimate $\kappa(A\Theta A^T)$ using Ritz values from CG
  - Norm 2
  - Ky Fan *k* norm

Heuristic proposal

- *IterCG* conjugate gradient iterations
- *m* matrix dimension
- Estimate $\kappa(A\Theta A^T)$ using Ritz values from CG
  - Norm 2
  - Ky Fan *k* norm

```
if κ > 10^β then
    if κ > c ∗ 10^γ then
        Phase exchange
    end
    if 3*iterCG > m then
        η = η + 10
    end
end
```

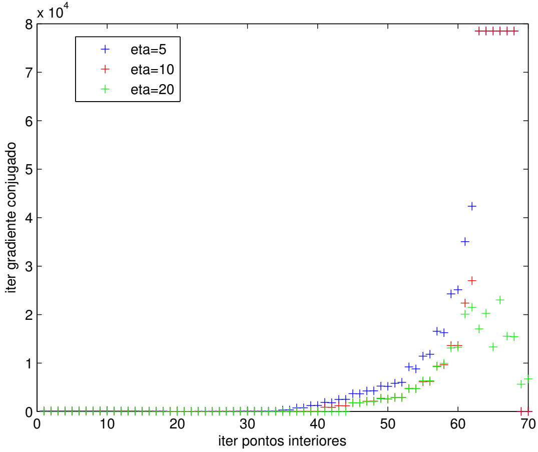
$c = 5$, $\gamma = 5$ and $\beta = 5$

Quality of the approximation

- Comparison between Ritz values estimation and Matlab *eig* function
- Small instances
- Results for $\eta = 5$ and conjugate gradient tolerance $= 10^{-8}$

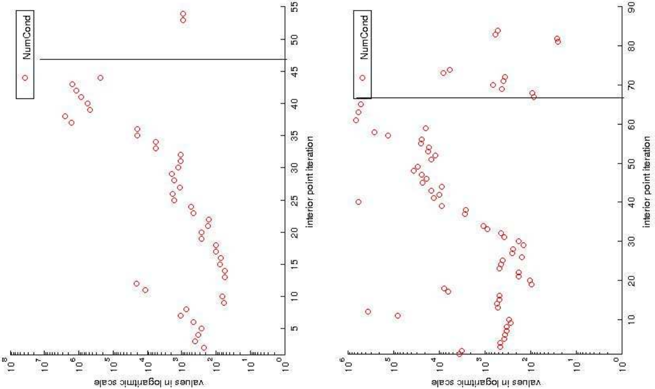
Best cases

Relative	*error*		
Instance	*Size*	*Largest eigenvalue*	*Smallest eigenvalue*
Afiro	27x51	0.2%	0.09%
*scsd*1	77x760	2.86%	6.62%
*sc*50*a*	49x77	0.94%	0.42%



Instance ken-18 - relation between η and the CG iteration

Numerical Experiments

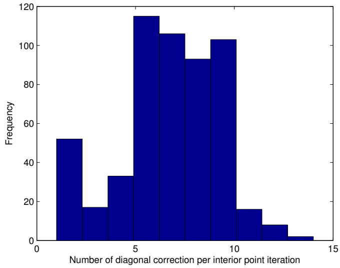


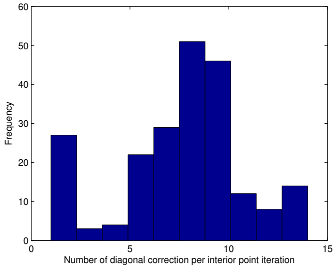
Instances 25fv47(left) and ken-18 (right)
Relation between condition number and interior point iteration

Numerical experiments

- Modified PCx
- 86 instances tested
- Comparison among the heuristics and two norms
  - Heuristic 1 - proposed in [1]
  - Heuristic 2 - proposed in [3]
  - Ritz Values - norm 2
  - Ritz Values - Ky Fan k norm

Diagonal Corrections in the CCF





Diagonal corrections per iteration - Heuristic A and new Heuristic norm 2

Numerical results

Corrections	Heuristic A	Heuristic B	Heuristic 2-norm	Heuristic KF-norm
1 to 5	102	401	34	25
6 to 10	314	288	96	72
more than 10	129	146	79	49
Total	545	848	209	138

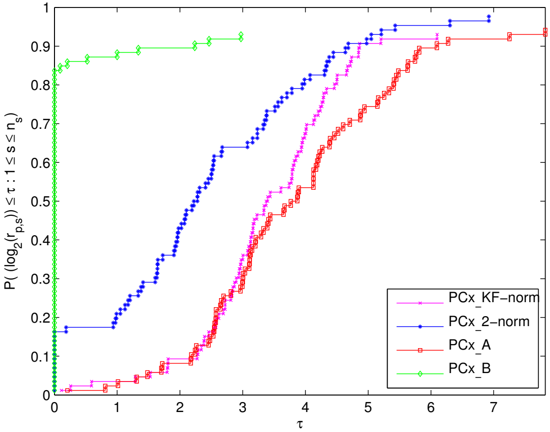
Number of ipm iterations that required diagonal corrections

Numerical results

	Heuristic A	Heuristic B	Heuristic 2-norm	Heuristic KF-norm
Optimal solution	81	80	84	80
Phase exchange	40	47	74	75
CG iterations	2 070 769	681 771	422 467	829 540
Time (seconds)	4903	1674	546	1030

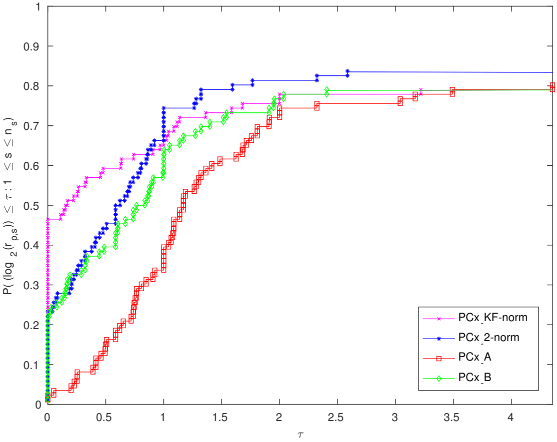
Results for all heuristics - Iterations and Time

Numerical Experiments



Performance profile CG iterations

Numerical Experiments



Performance profile solution time

Conclusions

- Hybrid preconditioner
- Change of phases
- Use of Ritz Values achieves very good results
- κ_2 obtained better performance
  - Fast
  - Robust
- Combine Ritz Values with PL information

Reference

[1] Silvana Bocanegra. *Algoritmos de Newton-Krylov precondicionados para métodos de pontos interiores*. PhD thesis, 2005.

[2] Silvana Bocanegra, Frederico F Campos, and Aurelio R L Oliveira. Using a hybrid preconditioner for solving large-scale linear systems arising from interior point methods. *Computational Optimization and Applications*, 36(2-3):149–164, 2007.

[3] Marta Inez Velazco, Aurelio R L Oliveira, and Frederico Ferreira Campos. A note on hybrid preconditioners for large-scale normal equations arising from interior-point methods. *Optimization Methods & Software*, 25(2):321–332, 2010.