1st Brazilian Workshop on Interior Point Methods

27-28 April, 2015 - Campinas, Brazil

An adaptive preconditioner for primal block-angular problems by an interior point method

Silvana Bocanegra and Jordi Castro,

UFRPE

UPC

E-mail: silvana@deinfo.ufrpe.br; jordi.castro@upc.edu;

Motivation

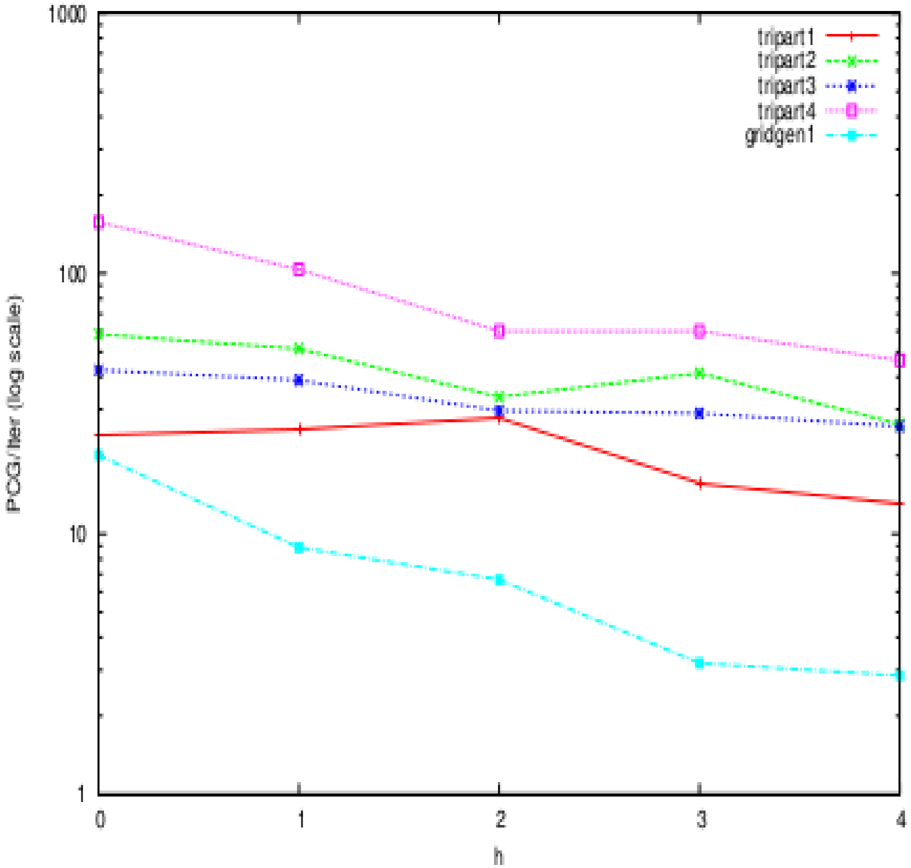
- Problems with block angular structure are common in many applications;
- The efficiency of Interior Point Methods (IPM) depends of the linear system solver used to compute the Newton direction;
- Preconditioned Iterative linear solvers may be more efficient to solve large-scale problems due to storage and time limitations;
- An efficient specialized IPM for primal block-angular problems solved the normal equations in two stages: Cholesky factorizations for the block constraints and a PCG for the linking constraint;

Motivation

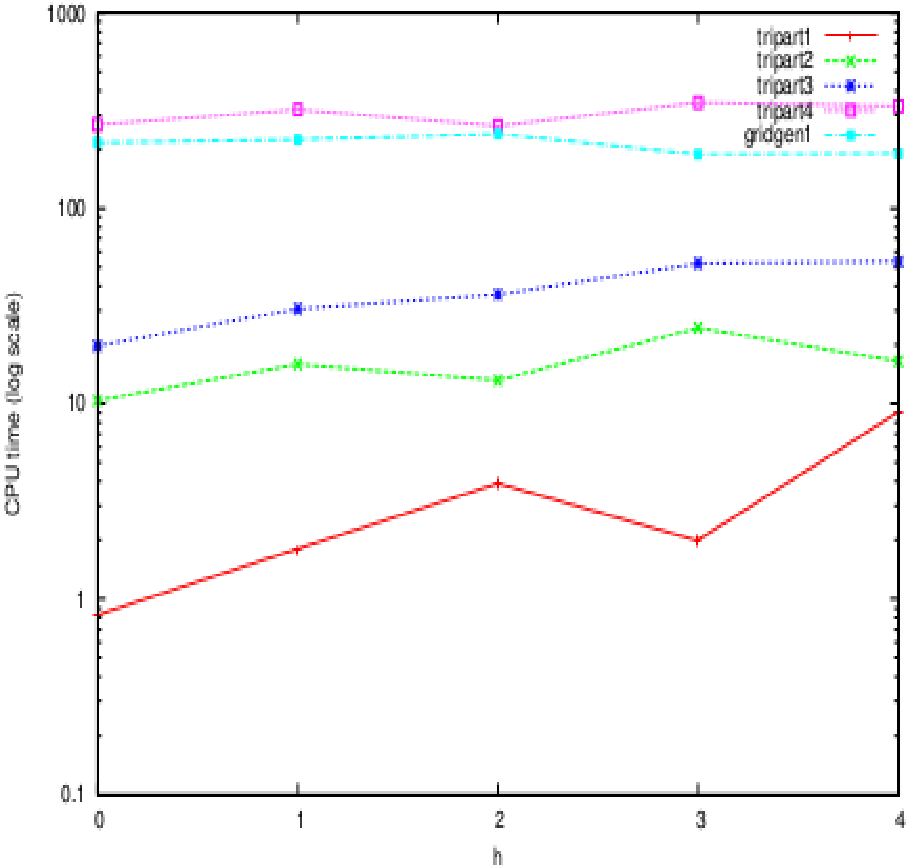
- ✓ The preconditioner used in second stage considers a few terms of an infinite power series which provides the inverse of the Schur complement of the normal equations

- ✓ The more terms of the power series, the more accurate the preconditioner, at the expense of increasing the running time of each PCG iteration

Motivation



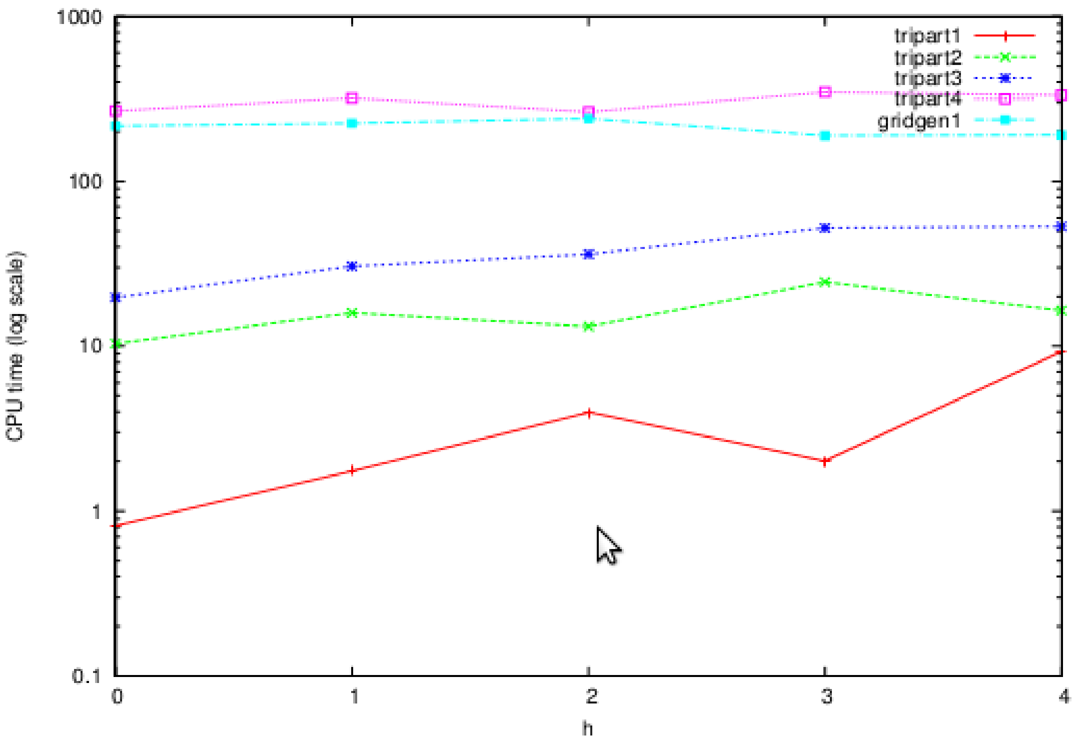
1000
100
10
1
PCG/iter (log scale)
0
1
2
3
4
h
tripart1
tripart2
tripart3
tripart4
gridgen1



1000
100
10
1
0.1
CPU time (log scale)
0
1
2
3
4
h
tripart1
tripart2
tripart3
tripart4
gridgen1

Motivation

- Is possible to find a number of terms used on each interior point iteration to provides better results?



Instance	k	m	n
tripart1	16	3294	33774
tripart2	16	13301	135941
tripart3	20	25541	329161
tripart4	35	38004	869814
gridgen1	320	329831	985191

Outline

- We propose an adaptive scheme for update the number of terms used in the preconditioner at each interior point iteration. This scheme is based on Ritz Values

- Ritz values can thus be used to estimate the spectral radius of a certain matrix in the power series, which measures the efficiency of the preconditioner.

- Preliminary numerical experiments are provide to both multicommodity flows and the minimum congestion problems.

The specialized block-angular IPM

A block angular-problem can be written in this general formulation

$$\min \quad \sum_{i=o}^{k} ({c^i}^T x^i + {x^i}^T Q_i x^i)$$

$$\text{s.t.} \quad \begin{bmatrix} N_1 & & & & \\ & N_2 & & & \\ & & \ddots & & \\ & & & N_k & \\ L_1 & L_2 & \dots & L_k & I \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^k \\ x^0 \end{bmatrix} = \begin{bmatrix} b^1 \\ b^2 \\ \vdots \\ b^k \\ b^0 \end{bmatrix}$$

$$0 \leqslant x^i \leqslant u^i \quad i = 1, \dots, k.$$

- One of the most efficient IPMs for block-angular problems solves the normal equations exploiting the block structure

$$A\Theta A^T = \begin{bmatrix} N_1\Theta_1 N_1^T & & & N_1\Theta_1 L_1^T \\ & \ddots & & \vdots \\ & & N_k\Theta_k N_k^T & N_k\Theta_k L_k^T \\ L_1\Theta_1 N_1^T & \dots & L_k\Theta_k N_k^T & \Theta_0 + \sum_{i=1}^{k} L_i\Theta_i L_i^T \end{bmatrix}.$$

$$\begin{bmatrix} B & C \\ C^T & D \end{bmatrix} \begin{bmatrix} \Delta y_1 \\ \Delta y_2 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}.$$

Cholesky + PCG

(COAP 2007)

- **Stage I:** Cholesky factorization for each block :

$$B\Delta y_1 = (g_1 - C\Delta y_2).$$

- **Stage II:** PCG for Linking constrains

$$(D - C^T B^{-1} C)\Delta y_2 = (g_2 - C^T B^{-1} g_1)$$

Power Series Preconditioner:

$$(D - C^T B^{-1} C)\Delta y_2 = (g_2 - C^T B^{-1} g_1)$$

Power series preconditioner: M^{-1}

$$(D - C^T B^{-1} C)^{-1} = \left(\sum_{i=0}^{\infty} (D^{-1}(C^T B^{-1} C))^i \right) D^{-1}$$

$$\begin{aligned}
M^{-1} &= D^{-1} && \text{if } h = 0, \\
M^{-1} &= (I + D^{-1}(C^T B^{-1} C))D^{-1} && \text{if } h = 1. \\
M^{-1} &= (I + D^{-1}(C^T B^{-1} C) + D^{-1}(C^T B^{-1} C)^2)D^{-1} && \text{if } h = 2.
\end{aligned}$$

Power Series Preconditioner:

The effectiveness of this preconditioner depends on the spectral radius of

$$D^{-1}(C^T B^{-1} C)$$

which is in [0,1). The farther away from 1, the better is the preconditioner.

Estimating the spectral radius

A procedure to estimate spectral radius of $D^{-1}(C^T B^{-1} C)$ was recently introduced for h = 0.

Proposition 1. *Let v be the eigenvector of matrix* $I - D^{-1}(C^T B^{-1} C)$ *associated with the eigenvalue* λ*. Then, v is eigenvector of* $D^{-1}(C^T B^{-1} C)$ *associated to eigenvalue* $1 - \lambda$*.*

Corollary 1. *Let* $\lambda_{min} \geqslant 0$ *be the minimum eigenvalue of* $I - D^{-1}(C^T B^{-1} C)$*. Therefore, the spectral radius of* $D^{-1}(C^T B^{-1} C)$ *is* $1 - \lambda_{min}$*.*

Estimating the spectral radius

The espectral radius of the preconditioned matrix can be estimated from de solution of the system by PCG using the relation of Lanczos and CG.

The eigenvalues of the tridiagonal matrices

$$T_k = \begin{pmatrix} \gamma_1 & \eta_2 & & & \\ \eta_2 & \gamma_2 & \eta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \eta_{k-1} & \gamma_{k-1} & \eta_k \\ & & & \eta_k & \gamma_k \end{pmatrix},$$

$k = 1, \ldots, l$, converge to the eigenvalues of the preconditioned matrix of the system solved by PCG as the number of PCG iterations approaches l.

Estimating the spectral radius

Given $x_0,\ r_0 = b - Mx_0,\ \rho_0 = r_0,\ k = 1$

while $r_k \neq 0$ **and** $k < k_{\max}$

$$\alpha_{k-1} = \left(\frac{\|r_{k-1}\|^2}{(\rho_{k-1}, M\rho_{k-1})} \right)$$

$$x_k = x_{k-1} + \alpha_{k-1}\rho_{k-1}$$

$$r_k = r_{k-1} - \alpha_{k-1}M\rho_{k-1}$$

$$\beta_{k-1} = \left(\frac{\|r_k\|^2}{\|r_{k-1}\|^2} \right)$$

$$\rho_k = r_k + \beta_{k-1}\rho_{k-1}$$

$$k := k + 1$$

end_while.

$$\gamma_k = \frac{1}{\alpha_{k-1}} + \frac{\beta_{k-1}}{\alpha_{k-2}},$$

$$\beta_0 = 0, \qquad \alpha_{-1} = 0,$$

$$\eta_{k+1} = -\frac{\sqrt{\beta_k}}{\alpha_{k-1}}.$$

Estimating the spectral radius

Extreme eigenvalues of the preconditioned matrix are well approximated already during early PCG iterations.

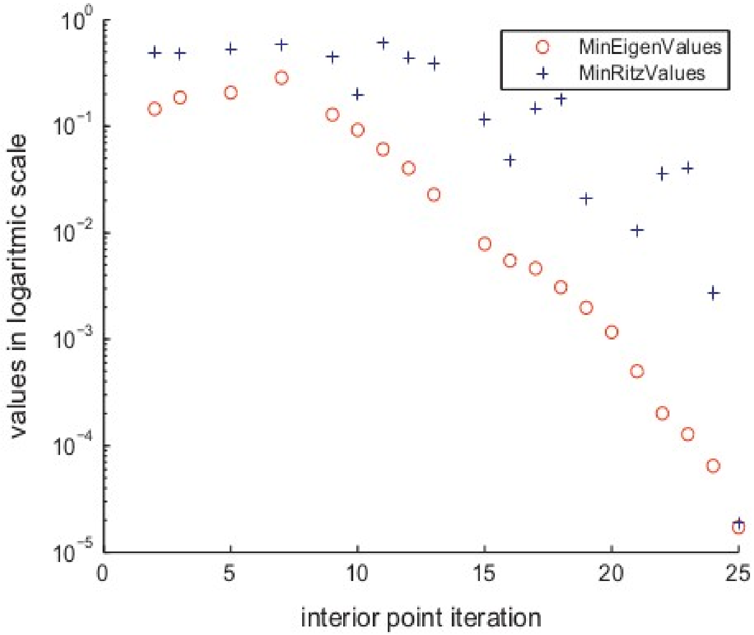
- ✓ The smallest eigenvalue $\lambda_{\min}$ is used to estimate the spectral radius of the preconditioned matrix

$$I - D^{-1}(C^T B^{-1} C).$$

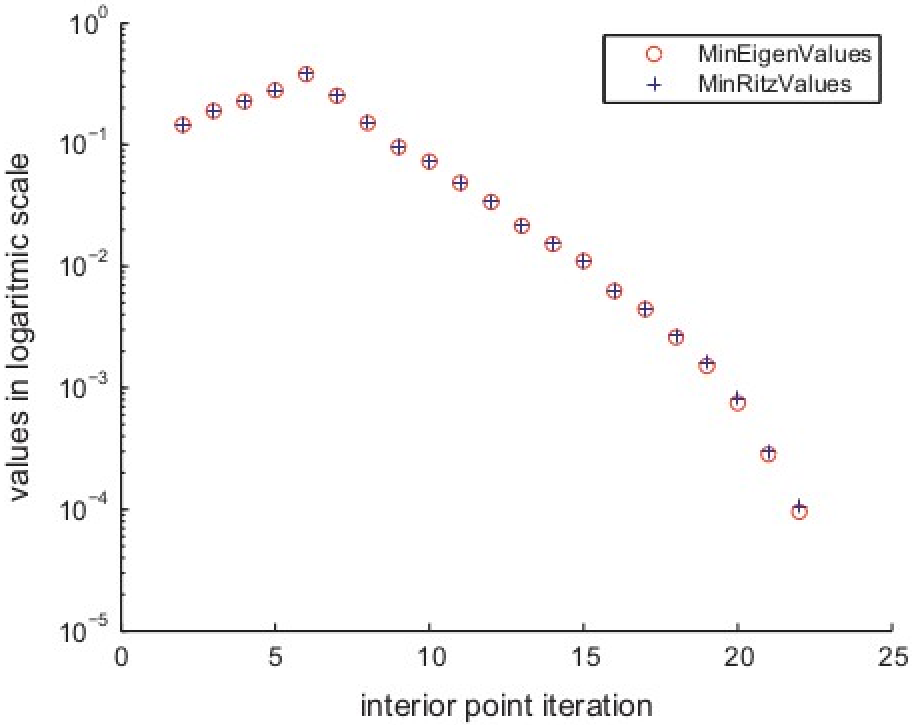
Ritz Values - Examples

Proximity between the Ritz values and the eigenvalues

m32_32_12



(a) Original PCG tolerance.



(b) Tighter PCG tolerance: 10^{-12}.

$(\varepsilon^i = \max\{0.95\varepsilon^{i-1}, \min_\varepsilon\}, \ \min_\varepsilon = 10^{-8}, \ \text{and} \ \varepsilon^0 = 10^{-2}$

Estimating the spectral radius

The previous estimations have been extended to consider any number h ≥ 0 of terms in the power series preconditioner:

Proposition 2. *Let* $M^{-1} = \left(\sum_{i=0}^{h}(D^{-1}(C^T B^{-1} C))^i\right) D^{-1}$ *be the preconditioner with* h *terms of the power series. And let* λ_{min} *be the smallest eigenvalue of the preconditioned matrix* $M^{-1}S$*. Then the spectral radius of* $D^{-1}(C^T B^{-1} C)$ *is*

$$\rho =^{h+1} \sqrt{(1 - \lambda_{min})}$$

Dynamic update of the h terms

Start: h=0 :

If h < 5 and

1. *The estimated spectral radius* $(\tilde{\rho} = ^{h+1}\sqrt{(1 - \tilde{\lambda}_{min})}) > 0.9$

2. *The number of PCG iterations for solving the linear system reaches 0.10l, where l is the dimension of* $I - D^{-1}(C^T B^{-1} C)$.

h=h+1

Restart h if solution is worse $\mu < 0.001$ *and* $\mu_{i+1} > \mu_i$.

Numerical Experiments

Multicomodity flow problems (oriented)

Instances	Constraints	Variables
tripart1	3294	33774
tripart2	13301	135941
tripart3	25541	329161
tripart4	38004	869814
gridgen1	329831	985191

Dynamic h terms: h in [0;5]

Original: h = 0 for all iterations

Instance	CPU (s)	it	PCG (it)	h_0	h_max	CPU (s)	it	PCG (it)
tripart1	1,52	56	993	0	5	**0,87**	51	1260
tripart2	**7,77**	73	2208	0	5	23,02	117	10427
tripart3	56,14	108	7997	0	5	**21,51**	78	3363
tripart4	**186,7**	122	5938	0	5	fail		
gridgen1	**159,49**	207	1989	0	5	224,65	195	3938

Numerical Experiments

Gridgen 1

IP_iteration	PCG_it eration	terms: h	Spectral radius
168	9	0	0
169	2	0	0
170	10	0	0
171	386	0	0
172	161	1	0,99
173	2	2	0,75
174	1	2	0,15
175	1	2	0,10
176	1	2	0,5
177	1	2	0,53
178	10	2	0,99
179	1	3	0,7
180	1	3	0,73
181	2	3	0,84
182	4	3	0,95
183	1	4	0,44
184	1	4	0,28
185	1	4	0,51
186	1	4	0,48
187	2	4	0,94
189	1	5	0,44
190	2	5	0,81
191	4	5	0,98

Numerical Experiments

Minimmum congestion problems

Instances	Constraints	Variables
M32-32	2449	33533
M64-64	5564	67962
M128-64	11640	155742
M128-128	19867	314243
M256-256	71891	1139467
M512-64	470075	634143
M512-128	79765	1249145

Dynamic h terms: h in [0;5]

Original: h = 0 for all iterations

Instance	Dynamic h terms: CPU (s)	it	PCG (it)	h_0	h_max	Original: CPU (s)	it	PCG (it)
M32-32	0,83	95	317	0	0	0,84	95	317
M64-64	1,92	94	178	0	0	1,94	94	178
M128-64	6,28	100	240	0	0	6,3	100	240
M128-128	15,59	100	234	0	0	15,54	100	234
M256-256	198,04	121	1319	0	0	198,03	121	1319
M512-64	130,37	123	2676	0	0	129,78	123	2676
M512-128	155,38	102	123	0	0	155,4	102	123

Conclusions

- ✓ An adaptive selection of the number of terms in the power series preconditioner seems a good strategy to improve the performance of this preconditioner .

- ✓ Future tasks: Apply this adaptative scheme to test another classes of problems.