A proposal to improve the conditioning of the splitting preconditioner for linear systems from interior point methods

Cecilia Orellana Castro   Aurelio R. L. Oliveira

Applied Mathematics Department, State University of Campinas

1st BWIPM April-15

Primal-dual canalized linear programming problem

$$(P)\begin{cases} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x + s = u \\ & x, s \geq 0 \end{cases} \qquad (D)\begin{cases} \max & b^T y - u^T w \\ \text{s.t.} & A^T y - w \leq c \\ & w \geq 0 \\ & y \in \mathbb{R}^m \end{cases}$$

$$\min c^T x - \mu \sum_{i=1}^{n} \log x_i - \mu \sum_{i=1}^{n} \log s_i \quad s.a. \quad Ax = b, \quad x + s = u, \quad x, s > 0$$

Optimality conditions

1. $Ax = b$
2. $x + s = u \quad x, s > 0$
3. $A^T y + z - w = c \quad z, w > 0$
4. $SWe = \mu e$
5. $XZe = \mu e$

Search direction by first order approximation of KKT conditions

Consider $F:\mathbb{R}^{2n}_{+}\times\mathbb{R}^{m}\times\mathbb{R}^{2n}_{+}\longrightarrow\mathbb{R}^{4n+m}$

$$F\left(x,s,y,w,z\right)=\left(Ax-b,x+s-u,A^{t}y+z-w-c,XZe-\mu e,SWe-\mu e\right)$$

$$F(X)+J(X)(\Delta X)=0 \tag{1}$$

$$\begin{pmatrix} A & 0 & 0 & 0 & 0\\ I_n & I_n & 0 & 0 & 0\\ 0 & 0 & A^T & -I_n & I_n\\ Z & 0 & 0 & X & 0\\ 0 & W & 0 & 0 & S \end{pmatrix}\begin{pmatrix}\Delta x\\ \Delta s\\ \Delta y\\ \Delta w\\ \Delta z\end{pmatrix}=\begin{pmatrix} r_b\\ r_u\\ r_c\\ \mu e-XZe\\ \mu e-SWe\end{pmatrix} \tag{2}$$

$r_b=b-Ax,\quad r_u=u-x-s\quad$ e $\quad r_c=c+w-z-A^{T}y$

Augmented System and Normal Equations System

Augmented System

System of equations with symmetric indefinite matrix of size $m + n$

$$\begin{pmatrix} -D & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} r \\ h \end{pmatrix} \tag{3}$$

where $D = X^{-1}Z + S^{-1}W$,
$r = r_c - X^{-1}(\sigma\mu e - XZe) + S^{-1}(\sigma\mu e - SWe) - S^{-1}Wr_u$ e $h = r_b$.

Normal Equations System

System of equations with symmetric and positive definite matrix of size m.

$$AD^{-1}A^T\Delta y = h + AD^{-1}r \tag{4}$$

Considerations

$$AD^{-1}A^T = \sum_{i=1}^{n} d_i^{-1} A^i (A^i)^T = \sum_{i=1}^{n} \left(\frac{z_i}{x_i} + \frac{w_i}{s_i} \right)^{-1} A^i (A^i)^T$$

1. A single dense column A^j implies that $AD^{-1}A^T$ loses sparseness
2. The eigenvalues λ of $AD^{-1}A^T$ satisfy $\lambda \leq \sum_{i=1}^{n} d_i^{-1} \|A^i\|$

J. Gondzio showed in [2] that if all iterations are in a neighborhood of infeasible points $N_\beta = \left\{ (x, y, z) \in \mathcal{F} : \beta\mu \leq x_j z_j \leq \frac{1}{\beta}\mu \right\}$ then

$$\kappa(AD^{-1}A^T) \leq \kappa(A)^2 \mathcal{O}(\mu_k^{-2})$$

.

1. iterative methods are used.
2. the use of preconditioners is indispensable.

Splitting Preconditioner $P^{-1} = D_B^{1/2} B^{-1}$

After permutation of the columns, consider

$A = [B, N]$, $D = \begin{pmatrix} D_B & 0 \\ 0 & D_N \end{pmatrix}$, then

$$\begin{aligned} P^{-1}\left(AD^{-1}A^T\right)P^{-T} &= D_B^{1/2} B^{-1} (B D_B^{-1} B^T + N D_N^{-1} N^T) B^{-T} D_B^{1/2} \\ &= I + (D_B^{1/2} B^{-1} N D_N^{-1/2})(D_N^{-1/2} N^T B^{-T} D_B^{1/2}) \\ &= I + W W^T \end{aligned}$$

Preconditioning Normal Equations $AD^{-1}A^T \Delta y = h + AD^{-1} r$

$$(I + W W^T)\widetilde{\Delta} y = D_B^{1/2} B^{-1} (h + AD^{-1} r)$$

Search direction using Splitting preconditioner

$$\Delta y = P^{-T}\widetilde{\Delta} y = B^{-T} D_B^{1/2} \widetilde{\Delta} y$$

Further, if $\Delta x = [\Delta x_B,\, \Delta x_N]^t$, $r = [r_B, r_N]^t$ e $r_y = h$

$$\Delta x_B = D_B^{-1}(B^T \Delta y - r_B)$$
$$\Delta x_N = D_N^{-1}(N^T \Delta y - r_N)$$

Considerations about the matrix $W = D_B^{1/2} B^{-1} N D_N^{-1/2}$

- Changes every iteration
- In the last iterations, $\mu = \frac{x^T z + s^T w}{2n} \approx 0$, that is $x_i z_i \approx 0$ e $s_i w_i \approx 0$, then $d_i \approx 0$ or $d_i \approx \infty$
- Obtain the matrix B has a high computational cost
- The Splitting preconditioner can maintain a matrix B of the k-th iteration in subsequent iterations

Heuristics for choosing the submatrix B

- Oliveira and Sorensen [3] suggest the choice of the first m columns l. i. of the matrix A considering the highest 1-norm of the AD^{-1} columns
- Velazco, et al. [4] did tests using 1-norm of the matrices $AD^{-1/2}$ and $AD^{-3/2}$ without significant improvements, however using the columns of AD^{-1} with the highest 2-norm showed many improvements.

Spectral analysis of Splitting preconditioner

Consider (λ, v) an eigenvalue and eigenvector of matrix $I + WW^T$

Then $v + WW^t v = \lambda v$

$$|\lambda| = 1 + \frac{\|W^t v\|^2}{\|v\|^2} \geq 1$$

$\|W^t v\| \leq \|W^t\| \|v\|$ implies that

$$|\lambda| = 1 + \frac{\|W^t v\|^2}{\|v\|^2} \leq 1 + \|W^t\|^2$$

Therefore, to limit the spectrum of $I + WW^T$ is necessary to limit $\|W^t\|$

Idea

To minimize $\|W^T\|$ putting conditions on the rows and columns of W, we use the Frobenius' norm.

$$\|W^T\|_F^2 = \sum_{j=1}^{m} \|W^j\|_2^2 \qquad (5)$$

where W^j is the j-th column of W

Proposition

Consider $W = D_B^{1/2} B^{-1} N D_N^{-1/2}$, if B is composed by the first m linearly independent columns A^j of A such that j corresponds to the highest $\|A^j d_j^{-\frac{1}{2}}\|_2$, then $\|W\|_F$ is minimized.

$$\|W\| \leq \|D_B^{1/2} B^{-1}\| \|N D_N^{-1/2}\|$$

Proof.

We must minimize $\|D_B^{1/2}B^{-1}\|$ and $\|ND_N^{-1/2}\|$.

1. Since $\|BD_B^{-1/2}\| \geq \dfrac{1}{\|D_B^{1/2}B^{-1}\|}$

   $$\max \|BD_B^{-1/2}\|_F \Rightarrow \max \frac{1}{\|D_B^{1/2}B^{-1}\|_F} \Leftrightarrow \min \|D_B^{1/2}B^{-1}\|_F$$

   Observe that

   $$AD^{-1/2} = [BD_B^{-1/2},\ ND_N^{-1/2}] \tag{6}$$

   By Frobenius' norm, to maximize $\|BD_B^{-1/2}\|_F$ we should choose columns of $AD^{-1/2}$ with the highest 2-norm. So B must be conformed by linearly independent columns A^j such that j corresponds to the highest $\|A^j d_j^{-\frac{1}{2}}\|_2$

2. When we choice columns of $AD^{-1/2}$ with lowest 2-norm, $\|ND_N^{-1/2}\|$ is minimized

Using (6), to maximize $\|BD_B^{-1/2}\|_F$ and to minimize $\|ND_N^{-1/2}\|_F$ are equivalent. □

Numerical Experiments

The hybrid preconditioner Bocanegra, et al. [6] with the change of phase proposed M. Velazco et al. [4] was integrated to the PCx code.

In the Splitting preconditioner the basis B changes when $8 * n_g \geq m$.

PCxm is the ordering of basis B which is based on M. Velazco. heuristic

PCxc is the ordering of basis B given by the presented proposition.

Problems

	Iterations IPM		**Time**		**Iterations PCG**		**Size**	
Prob	PCxm	PCxc	PCxm	PCxc	PCxm	PCxc	Line	Column
25fv47	29	**26**	1.80	**1.35**	5122	2951	825	1571
bnl1	40	40	0.75	0.76	2814	2634	643	1175
chr22b	29	29	19.33	17.94	938	909	5587	5335
chr25a	29	29	42.94	**40.03**	2785	2964	8149	7825
cre-a	27	27	7.67	7.65	176	187	3516	4067
cre-b	43	43	43.40	42.29	108	166	9648	72447
cre-c	27	27	5.83	5.02	151	155	3068	3678
cre-d	42	42	28.41	27.33	79	133	8926	69980
els19	31	31	44.24	**35.49**	3243	3212	4350	9937

Problems

Prob	Iterations IPM		Time		Iterations PCG		Size	
	PCxm	PCxc	PCxm	PCxc	PCxm	PCxc	Line	Column
ex01	28	28	0.41	0.34	1448	1036	246	1379
ex02	46	**37**	0.95	0.67	6431	3717	238	1378
ex05	39	39	5.82	4.92	2332	2290	833	6980
ex09	**45**	52	**52.14**	54.95	11862	15541	1846	16422
ganges	18	18	0.63	0.63	326	383	1309	1681
ken13	29	29	93.80	92.35	33	34	28632	42659
ken18	41	41	1040.20	**1011.89**	409	504	105127	154699
maros	40	**25**	2.31	**1.10**	13813	5086	840	1443
nesm	31	31	1.57	1.28	4943	3729	662	2923
nug15	–	–	–	–	–	–	6330	22275
qap15	–	–	–	–	–	–	6330	22275
rou20	24	24	757.49	**420.04**	1470	1889	7359	33840
scr15	24	24	7,66	6.61	2013	1768	2234	4635
scr20	21	21	60.08	**55.45**	1534	2369	5079	12180
ste36a	37	37	14078.34	**5523,68**	14128	13793	27686	109653
stocfor2	21	21	1.13	1.17	467	477	2157	2031
stocfor3	32	32	87.90	87.51	5110	5154	16675	15695

Conclusions

- The property to maintain a matrix B of the k-th iteration in subsequent iterations is a cheaper strategy.

$$\text{If} \quad P_k^{-1} = D_B^{1/2} B^{-1} \quad \text{then} \quad P_{k+1}^{-1} = \hat{D}_B^{1/2} B^{-1}$$

  However, $I + \hat{W}\hat{W}^T$ has not the best set of columns and therefore its performance is not very good compared to the k-th iteration.

- PCx*c* improves the total time in most problems, some problems such as **ken18**, **rou20** and **ste36a**, the differance is very remarkable.
  1. In some iterations of IPM were required fewer iterations of PCG method.
  2. Fewer rearrangements of the matrix B were necessary

- Save time on reordering columns of B carries more iterations in the PCG method, for this reason there are problems such as **ken18** with more PCG method iterations while it is solved in less time.

- F.F.Campos. Analysis of conjugate Gradients - type methods for solving linear equations. Phd tesis,Oxford UniversityComputing Laboratory, Oxford, 1995.
- J. Gondzio, Matrix-Free Interior Point Method, Computational Optimization and Applications March 2012, Volume 51, Issue 2, pp 457-480.
- A.R L.Oliveira, D.C. Sorensen, A new class of preconditioners for large-scale linear systems from interior point methods for linear programming, Linear Algebra and Its Applications 394(2005), pp. 1-24.
- M.I. Velazco, A.R.L. Oliveira, and F.F. Campos, Heuristics for implementation of a hybrid preconditioner for interior-point methods, Pesqui. Oper. vol.31 no.3 Rio de Janeiro Sept./Dec. 2011.
- S. J. Wright, Primal Dual Interior Point Methods, SIAM Publications, SIAM, Philadelphia,PA, USA, 1997
- Bocanegra S, Campos F. F. Oliveira A. R. L. 2007. Using a hybrid preconditioner for solving large-scale linear systems arising from

THANK YOU