JSIAM Letters Vol.3 (2011) pp.5–8 ©2011 Japan Society for Industrial and Applied Mathematics

A convergence improvement of the BSAIC preconditioner by deflation

Ikuro Yamazaki[1], Hiroto Tadano[1], Tetsuya Sakurai[1] and Keita Teranishi[2]

[1] Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki 305-8573, Japan
[2] Cray, Inc., 380 Jackson St. Suite 210, St Paul, MN 55101, USA

E-mail *yamazaki@mma.cs.tsukuba.ac.jp*

Received May 31, 2010, Accepted September 16, 2010

Abstract

We have proposed a block sparse approximate inverse with cutoff (BSAIC) preconditioner for relatively dense matrices. The BSAIC preconditioner is effective for semi-sparse matrices which have relatively large number of nonzero elements. This method reduces the computational cost for generating the preconditioning matrix, and overcomes the performance bottlenecks of SAI using the blocked version of Frobenius norm minimization and the drop-threshold schemes (cutoff) for semi-sparse matrices. However, a larger parameter of cutoff leads to a less effective preconditioning matrix with a large number of iterations. We analyze this convergence deterioration in terms of eigenvalues, and describe a deflation-type method which improves the convergence.

Keywords linear system, preconditioning, sparse approximate inverse, deflation

Research Activity Group Algorithms for Matrix / Eigenvalue Problems and their Applications

1. Introduction

Linear systems

$$A\boldsymbol{x} = \boldsymbol{b},$$

where $A \in \mathbb{C}^{n\times n}$ is a semi-sparse matrix which is relatively dense, appear in nano-simulations. A sparse approximate inverse (SAI) technique is proposed as a parallel preconditioner for sparse matrices [1]. This preconditioner has a good parallel performance. However, the arithmetic costs of constructing the preconditioning matrix grow cubically with the number of nonzero entries per row. We have proposed a block sparse approximate inverse with cutoff (BSAIC) [2] preconditioner for such semi-sparse linear systems.

The BSAIC preconditioner can reduce the computational cost for constructing the approximate inverse matrix, and overcome the performance bottlenecks of SAI using the blocked version of Frobenius norm minimization and the cutoff strategy for semi-sparse matrices. A large cutoff parameter leads to a further decrease cost of constructing the approximate inverse matrix. Thus, we want to use a larger cutoff parameter as much as possible. However, a convergence of Krylov subspace methods preconditioned with BSAIC deteriorates when the cutoff parameter is large. In this paper, this deterioration of convergence is investigated in terms of eigenvalues, and a method of the convergence improvement is also presented.

This paper is organized as follows. In Section 2, our method, the BSAIC preconditioner, is described. We describe the convergence deterioration by large cutoff parameters, and how to improve this convergence deterioration and algorithms of the method in Section 3. In Section 4, the BSAIC preconditioner applied to the improving method is verified by numerical experiments, followed by the concluding remarks in Section 5.

2. Block SAI with Cutoff (BSAIC)

We describe the block SAI with cutoff (BSAIC) preconditioner. In the BSAIC preconditioner, the cutoff is applied to the coefficient matrix A in order to reduce the computational cost of least square problems which appear in block SAI. Firstly, the approximate coefficient matrix A_{c} is generated by the following cutoff:

$$A_{\mathrm{c}} = [\tilde{a}_{ij}], \quad \tilde{a}_{ij} = \begin{cases} a_{ij}, & (|a_{ij}| > \theta \text{ or } i = j), \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where θ is a nonnegative real value. After applying the cutoff, least square problems with the approximate matrix A_{c}:

$$\min_{M} \|A_{\mathrm{c}}M - I\|_{\mathrm{F}}^2 \approx \sum_{k=1}^{L} \min_{M_k} \|A_{\mathrm{c}}M_k - E_k\|_{\mathrm{F}}^2, \tag{2}$$

where l is a block size, $L = \lceil n/l \rceil$ and E_k is a submatrix of the identity matrix I such that $I = [E_1, E_2, \dots, E_L]$ are solved. The matrix $M = [M_1, M_2, \dots, M_L]$ is employed as the preconditioning matrix. The initial sparsity pattern M_0 of the preconditioning matrix is decided by the following:

$$\mathrm{spy}(M_0) = \mathrm{spy}(A_{\mathrm{c}}), \tag{3}$$

where "spy" denotes the sparsity pattern of a matrix.

We overcome a performance bottleneck by using a blocked version of SAI with drop-threshold schemes to reduce the computational cost for constructing the approximate inverse matrix and to improve the convergence of Krylov subspace methods. However, a larger value of θ leads to a less effective preconditioning matrix with a large number of iterations, but a value of θ is preferred to be large as much as possible. In the next section, we describe this convergence deterioration and the improvement method.

3. Convergence improvement by deflation

We consider to solve preconditioned linear systems $(AM)(M^{-1}\boldsymbol{x}) = \boldsymbol{b}$, by some Krylov subspace methods. We investigate an eigenvalue distribution of AM. The block size l is fixed and the cutoff parameter θ is varied in BSAIC. The preconditioning matrix M approximates the inverse of matrix A, and AM is nearly equal to the identity matrix I when M is a good approximation to A^{-1}. Eigenvalues of AM are clustered around 1 when M is a good approximation to A^{-1}.

In the restarting GMRES (GMRES(m)) [3] method, the information concerning the eigenvalues around the origin is discarded at the restart. These small eigenvalues often slow the convergence. As GMRES iterations are performed, deflation-type schemes (e.g. GMRES-IR [4] and GMRES-DR [5]) calculate small approximate eigenvalues and corresponding eigenvectors. These eigenvectors are added to the Krylov space in a bid to speed convergence. An implicitly restarted GMRES (GMRES-IR) [4] proposed by Morgan is employed in Section 4.

In the GMRES-IR(m, k) method, we compute the eigenpairs of the eigenvalue problem from an Arnoldi process of length m. We then apply an implicitly restarting Arnoldi (IRA) [6] with the unwanted harmonic Ritz values [7] as shifts. The IRA method filters a chosen harmonic Ritz value away from the Arnoldi process. Here, small harmonic Ritz values are chosen, and k small eigenvalues near the origin can be deflated. Therefore, the convergence will be improved by this deflation. Fig. 1 shows the algorithm of GMRES-IR. Our experiments in Section 4 indicate the validity of the GMRES-IR method preconditioned with BSAIC.

4. Numerical experiments

In this section, firstly, the performance of the Krylov subspace method preconditioned with the BSAIC preconditioner corresponding to θ is verified. Secondly, we analyze the convergence deterioration by a larger value of θ and apply the improvement strategy to the BSAIC preconditioner. All experiments are carried out by MATLAB 7.4 on MacBook (CPU: Intel Core 2 Duo 2.26GHz, Memory: 4.0Gbytes, OS: Mac OS 10.6.3). The test problems are solved by the preconditioned GMRES(50) method. The stopping criterion for the relative residual is 10^{-10}. The initial guess $\boldsymbol{x}_0$ is set to $\boldsymbol{0}$ and all elements of $\boldsymbol{b}$ are set to 1. The notation #MVs means the number of matrix-vector products, and the dagger (†) means that the stopping criterion is not satisfied in 5,000 MVs.

The test matrix is derived from the computation of the molecular orbitals of an epidermal growth factor (EGF).

Algorithm GMRES-IR(m, k) method

1: **Compute** $p = m - k$ and $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$
2: **Compute** $\beta = \|\boldsymbol{r}_0\|_2$ and $\boldsymbol{v}_1 = \boldsymbol{v}_0/\beta$
3: **Compute** V_{m+1}, $\bar{H}_m$ with Arnoldi method
4: **Compute** $\boldsymbol{y}$, the minimizer of $\|V_{m+1}^{\top}\boldsymbol{r}_0 - \bar{H}_m\boldsymbol{y}\|_2$, and $\boldsymbol{x}_m = \boldsymbol{x}_0 + V_m\boldsymbol{y}$
5: **If** satisfied **Stop**, **else** proceed
6: **Compute** the harmonic Ritz values $\tilde{\theta}_1, \dots, \tilde{\theta}_m$
7: **Sort** $|\tilde{\theta}_1\| \geq \cdots \geq \|\tilde{\theta}_m|$
8: **Set** shift $\tilde{\theta}_1, \dots, \tilde{\theta}_p$
9: **Update** V_{k+1} and $\bar{H}_k$ with IRA method
10: **Go to 3**, and resume the Arnoldi method from step $k+1$

Fig. 1. Algorithm of the GMRES-IR(m, k) method.

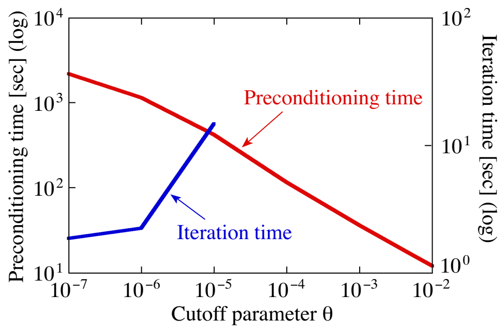


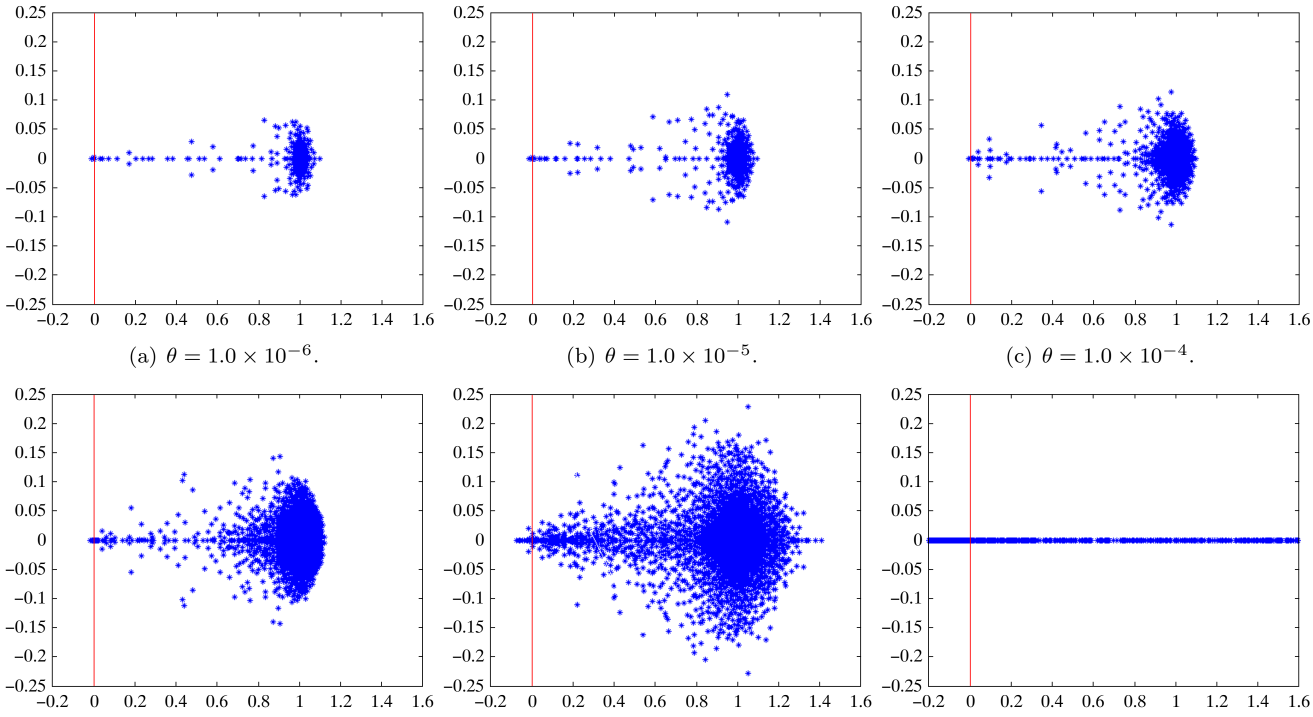
Fig. 2. The computational time of GMRES(50) with BSAIC corresponding to θ for EGF.

The size of A is 4,505 and the number of nonzero elements is 5,254,215 (25.89%). In this example, the block size l of BSAIC is set to 30.

The computational time of GMRES(50) with preconditioned BSAIC corresponding to θ for EGF is reported in Fig. 2. Our BSAIC preconditioner can solve this problem faster than SAI and block SAI. However, as Fig. 2 indicates, GMRES(50) preconditioned with BSAIC does not converge when θ is larger than 10^{-5}. Fig. 2 also shows that the cutoff parameter θ is preferred to be large as much as possible (e.g. $\theta = 10^{-3}$) for the preconditioning time. We investigate the slow down of the convergence in terms of eigenvalue distributions of AM.

Figs. 3(a), 3(b), ..., 3(e) show eigenvalue distributions of AM corresponding to $\theta = 10^{-6}, 10^{-5}, \dots, 10^{-2}$, respectively. Fig. 3(f) shows the eigenvalue distribution of A. The red line in Fig. 3 denotes a zero eigenvalue. In Figs. 3(a)–3(e), the eigenvalue distributions of AM are clustered around 0 as θ becomes larger. The eigenvalue ditribution of A in Fig. 3(f) is expanded and clustered around 0 more than that of AM. It is predicted that the coefficient matrix A is ill-conditioned. This clustering of eigenvalues is one of the key reasons for the convergence deterioration. Therefore, we apply the BSAIC preconditioner and GMRES-IR, which deflates the smallest eigenvalues, to these linear equations, and we improve the convergence of Krylov subspace methods.

Table 1 shows the results of GMRES-IR without preconditioner and GMRES-IR preconditioned with ILU(0) [8] and ILUT [8]. The ε in Table 1 denotes the threshold of ILUT. The GMRES-IR method does not converge ex-



(a) $\theta = 1.0 \times 10^{-6}$. (b) $\theta = 1.0 \times 10^{-5}$. (c) $\theta = 1.0 \times 10^{-4}$.

(d) $\theta = 1.0 \times 10^{-3}$. (e) $\theta = 1.0 \times 10^{-2}$. (f) Eigenvalue distribution of A.

Fig. 3. Eigenvalue distributions of AM and A for EGF.

Table 1. Results of preconditioned GMRES-IR(50, 25) for EGF.

Preconditioner	#MVs	Wall clock time [sec]		
		Precond.	Iter.	Total
None	†	—	—	—
ILU(0)	†	11.04	—	—
ILUT($\varepsilon = 10^{-2}$)	†	20.51	—	—
ILUT($\varepsilon = 10^{-3}$)	74	33.49	16.76	50.25

Table 2. Results of BiCGSTAB, GMRES(50) and GMRES-IR preconditioned with BSAIC ($l = 30, \theta = 1.0 \times 10^{-3}$) for EGF.

Krylov	#MVs	Wall clock time [sec]			
		Cutoff	Precond.	Iter.	Total
BiCGSTAB	†	0.59	36.89	—	—
GMRES(50)	†	0.59	36.89	—	—
IR(50, 5)	†	0.59	36.89	—	—
IR(50, 10)	331	0.59	36.89	15.91	53.39
IR(50, 15)	262	0.59	36.89	12.55	50.03
IR(50, 20)	233	0.59	36.89	10.90	48.38
IR(50, 25)	226	0.59	36.89	10.83	48.30

Table 3. Results of BiCGSTAB, GMRES(50) and GMRES-IR preconditioned with BSAIC ($l = 30, \theta = 5.0 \times 10^{-3}$) for EGF.

Krylov	#MVs	Wall clock time [sec]			
		Cutoff	Precond.	Iter.	Total
BiCGSTAB	†	0.56	17.09	—	—
GMRES(50)	†	0.56	17.09	—	—
IR(50, 5)	†	0.56	17.09	—	—
IR(50, 10)	†	0.56	17.09	—	—
IR(50, 15)	†	0.56	17.09	—	—
IR(50, 20)	351	0.56	17.09	15.15	32.80
IR(50, 25)	329	0.56	17.09	15.85	32.50

cept ILUT($\varepsilon = 10^{-3}$). GMRES-IR preconditioned with ILUT($\varepsilon = 10^{-3}$) has good convergence. However, ILUT does not have good parallel efficiency such as SAI.

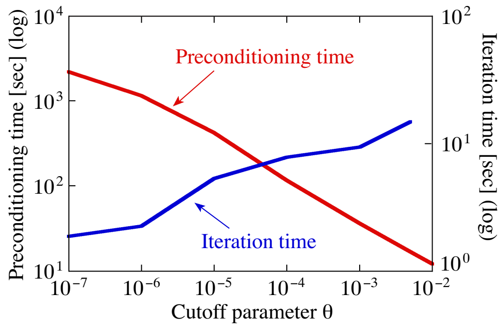


Fig. 4. The computational time of GMRES-IR(50, 25) with BSAIC corresponding to θ for EGF.

Tables 2 and 3 show the results for EGF with $\theta = 1.0\times 10^{-3}$ and 5.0×10^{-3}, respectively. When BiCGSTAB [9] and GMRES(50) are used, the stopping criterion is not satisfied in both Tables 2 and 3. In Table 2, the GMRES-IR method preconditioned with BSAIC converges except GMRES-IR(50, 5). As a result, the GMRES-IR(50, 25) method converges faster than other Krylov subspace methods. Table 3 shows that each of GMRES-IR(50, 20) and GMRES-IR(50, 25) converges, and GMRES-IR(50, 25) converges faster than any other method. Fig. 4 shows that a larger value of θ can be applied by using GMRES-IR. The convergence is dependent not only on the cutoff parameter θ but also on the restart value m and the number of deflated eigenvalues k. Thus, we need to set an appropriate m and k. Morgan also mentioned that the choice of m and k changes the convergence in [4].

Tables 4 and 5 show the real part of harmonic Ritz values of GMRES-IR(50, 25) and the real part of small eigenvalues of AM, respectively. In Table 4, the param-

Table 4. The harmonic Ritz values of GMRES-IR(50, 25) and the eigenvalues of AM ($l = 30, \theta = 1.0 \times 10^{-3}$). Underlines indicate the correct digits.

	Re(H. R.)	Re(eig(AM))
λ_1	0.000001191238635	0.000001191238641
λ_2	0.000225851028469	0.000225851028462
λ_3	−0.000307394803250	−0.000307394803254
λ_4	0.002204006095033	0.002204006095033
λ_5	−0.003517837137617	−0.003517837137617

Table 5. The harmonic Ritz values of GMRES-IR(50, 25) and the eigenvalues of AM ($l = 30, \theta = 5.0 \times 10^{-3}$). Underlines indicate the correct digits.

	Re(H. R.)	Re(eig(AM))
λ_1	−0.000072402852854	−0.000072402852260
λ_2	0.000197498551705	0.000197498552292
λ_3	−0.001027046005229	−0.001027046005286
λ_4	0.002560480805997	0.002560480809970
λ_5	0.002632881319376	0.002632881318118

Table 6. The number of eigenvalues of AM around 0 ($l = 30$).

θ	$\#(\lvert d\rvert < 10^{-1})$	$\#(\lvert d\rvert < 10^{-2})$	$\#(\lvert d\rvert < 10^{-3})$
1.0×10^{-6}	6	2	0
1.0×10^{-5}	10	3	0
1.0×10^{-4}	18	6	2
1.0×10^{-3}	26	9	3
5.0×10^{-3}	39	11	2

eters of BSAIC are set at $l = 30$ and $\theta = 1.0 \times 10^{-3}$. In Table 5, the parameters of BSAIC are set at $l = 30$ and $\theta = 5.0 \times 10^{-3}$. "Re" and "H.R." denote a real part and a harmonic Ritz value, respectively. The MATLAB command **eig** is used to calculate the eigenvalues of AM. Both Tables 4 and 5 show that the harmonic Ritz values approximate the eigenvalues of AM well. Hence, small eigenvalues of AM are deflated, and Tables 2 and 3 also show that the GMRES-IR method improves convergence more than any other Krylov subspace method.

The number of eigenvalues of AM around 0 corresponding to θ is reported in Table 6. The block size l is fixed at 30. $\#(|d| < value)$ in Table 6 denotes the number of absolute eigenvalues which are less than $value$. When $\theta = 1.0 \times 10^{-6}$ and 1.0×10^{-5} are used, $\#(|d| < 10^{-3})$ is zero and the GMRES(50) method with BSAIC converges in Fig. 2. However, when θ which is larger than 10^{-5} is used, $\#(|d| < 10^{-3})$ is not zero and GMRES(50) with BSAIC does not converge in Fig. 2. Thus, a larger value of θ increases the number of eigenvalue of AM around 0 and eventually deteriorates the convergence of Krylov subspace methods.

5. Conclusions

We proposed a method to improve the convergence of the BSAIC preconditioner using the deflation of small eigenvalues. Our BSAIC preconditioner reduces the constructing cost of the approximate inverse M for semi-sparse matrices. However, a larger value of cutoff parameter θ increases iteration counts and makes convergence difficult. We investigate this convergence deterioration with respect to eigenvalue distributions of AM. As a result, a larger value of θ leads the eigenvalue distribution of AM to be expanded and clustered around 0. This cluster of small eigenvalues makes the convergence slow, and thus the deflation-type Krylov subspace methods improve the convergence.

In future work, we will try to find an automatic procedure for selecting the cutoff parameter θ, the restart count m and the number of small eigenvalue k. We also apply for large scale problems.

Acknowledgments

This research was supported in part by a Grant-in-Aid for Scientific Research of Ministry of Education, Culture, Sports, Science and Technology, Japan (Grant Nos. 21246018 and 21105502).

References

[1] E. Chow and Y. Saad, Approximate inverse preconditioners via sparse-sparse iterations, SIAM J. Sci. Comput., **19** (1998), 995–1023.

[2] I. Yamazaki, M. Okada, H. Tadano, T. Sakurai and K. Teranishi, A block sparse approximate inverse with cutoff preconditioner for semi-sparse linear systems derived from Molecular Orbital calculations, JSIAM Letters, **2** (2010), 41–44.

[3] Y. Saad and M. H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput., **7** (1986), 856–869.

[4] R. B. Morgan, Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations, SIAM J. Matrix Anal. Appl., **21** (2000), 1112–1135.

[5] R. B. Morgan, GMRES with deflated restarting, SIAM J. Sci. Comput., **24** (2002), 20–37.

[6] D. C. Sorensen, Implicit application of polynomial filters in a k-step Arnoldi method, SIAM J. Matrix Anal. Appl., **13** (1992), 357–385.

[7] C. C. Paige, B. N. Parlett and H. A. Van der Vorst, Approximate solution and eigenvalue bounds from Krylov subspaces, Numer. Lin. Alg. Appl., **2** (1995), 115–133.

[8] Y. Saad, Iterative methods for sparse linear systems, SIAM, Philadelphia, 2003.

[9] H. A. van der Vorst, BiCGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems, SIAM J. Sci. Stat. Comput., **13** (1992), 631–644.