University of Groningen

Flexible and deflated variants of the block shifted GMRES method

Sun, Dong-Lin; Huang, Ting-Zhu; Carpentieri, Bruno; Jing, Yan-Fei

Published in:
Journal of Computational and Applied Mathematics

DOI:
[10.1016/j.cam.2018.05.053](https://doi.org/10.1016/j.cam.2018.05.053)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Sun, D-L., Huang, T-Z., Carpentieri, B., & Jing, Y-F. (2019). Flexible and deflated variants of the block shifted GMRES method. *Journal of Computational and Applied Mathematics*, 345, 168-183.
<https://doi.org/10.1016/j.cam.2018.05.053>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Flexible and deflated variants of the block shifted GMRES method

Dong-Lin Sun^{a,b}, Ting-Zhu Huang^{a,*}, Bruno Carpentieri^c, Yan-Fei Jing^{a,*}

^a School of Mathematical Sciences/Institute of Computational Science, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, PR China

^b Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Nijenborgh 9, PO Box 407, 9747 AG Groningen, The Netherlands

^c Faculty of Computer Science, Free University of Bozen-Bolzano, piazza Domenicani 3, 39100 Bozen-Bolzano, Italy

ARTICLE INFO

Article history:

Received 6 April 2017

Received in revised form 30 September 2017

Keywords:

Block Krylov subspace methods

Shifted systems

Deflation

ABSTRACT

The solution of linear systems with multiple shifts and multiple right-hand sides given simultaneously is required in many large-scale scientific and engineering applications. In this paper we introduce new flexible and deflated variants of the shifted block GMRES method for this problem class. The proposed methods solve the whole sequence of linear systems simultaneously, detecting effectively the linear systems convergence and allowing the use of variable preconditioning which may be particularly useful in some applications. Numerical experiments are illustrated to show the overall significant robustness of the iterative method for solving general sparse multi-shifted and multiple right-hand-side systems, and in realistic PageRank calculations. To the best of our knowledge, this is the first Krylov subspace method that combines deflation techniques and variable preconditioning for solving sequences of multi-shifted linear systems with multiple right-hand sides simultaneously.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

We consider the solution by Krylov subspace methods of sequences of shifted linear systems with L shifts and p different right-hand sides given at once of the form

$$(A - \sigma_i I)X_i = B, \text{ with } i = 1, 2, \dots, L, \quad (1)$$

where the scalars $\{\sigma_i\}_{i=1}^L \subset \mathbb{C}$ are called shifts, $A - \sigma_i I \in \mathbb{C}^{n \times n}$ are large square nonsingular matrices of dimension n , I is the identity matrix of order n , and $B = [b^{(1)}, b^{(2)}, \dots, b^{(p)}]$ is the $n \times p$ matrix of right-hand sides $b^{(j)}$, $j = 1, \dots, p$. We assume that B is a full-rank matrix. This problem arises in the solution of many large-scale scientific and engineering applications, such as lattice quantum chromodynamics [1], Tikhonov–Phillips regularization [2], Newton trust region optimization [3] and PageRank computation [4]. Direct methods based on the LU-factorization of $A - \sigma_i I$ are computationally not very attractive to use as they require to recalculate a new factorization for each shift σ_i . Moreover, in some applications A may not be available explicitly. Iterative methods, namely the class of modern Krylov subspace methods, are matrix-free and thus they can solve the memory bottlenecks of direct methods [5–9]. Block variants of Krylov methods have shown important advantages for

* Corresponding authors.

E-mail addresses: donglin1220@126.com (D.-L. Sun), tingzhuang@126.com (T.-Z. Huang), bcarpentieri@gmail.com (B. Carpentieri), 00jyfvictory@163.com (Y.-F. Jing).

the solution of multiple right-hand-side systems over methods designed for solving systems with a single right-hand side. They can search over much larger spaces and compute matrix–vector products simultaneously [10–12].

The block GMRES method is an effective Krylov subspace solver for multiple right-hand side linear systems of the form

$$AX = B. \quad (2)$$

The algorithm computes an approximate solution $X_m \approx X$ of (2) over the block Krylov subspace

$$\mathbb{K}_m(A, R_0) = \text{span} (R_0, AR_0, \dots, A^{m-1}R_0), \quad (3)$$

where we denote by $R_0 = B - AX_0$ the initial block residual corresponding to the initial block guess X_0 of X . Let $R_0 = V_1 R$ be the QR-decomposition of R_0 , where $V_1 \in \mathbb{C}^{n \times p}$ is a matrix with orthonormal columns and $R \in \mathbb{C}^{p \times p}$ is an upper triangular full-rank matrix. Block GMRES computes an orthogonal basis for $\mathbb{K}_m(A, R_0)$ by running m steps of the block Arnoldi procedure based on a Modified Block Gram–Schmidt process started with matrix V_1 . The procedure is described in Algorithm 1. Let $\mathcal{V}_m = [V_1, V_2, \dots, V_{m+1}]$ be the block orthonormal sequence spanning the block Krylov subspace $\mathbb{K}_m(A, R_0)$. Then the approximate solution to system (2) can be written as $X_m = X_0 + \mathcal{V}_m Y_m$, where Y_m minimizes the residual $R_m = R_0 - A\mathcal{V}_m Y_m$. The expression of Y_m can be computed by using the block Arnoldi relation

$$A\mathcal{V}_m = \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m, \quad (4)$$

where $\mathcal{V}_m \in \mathbb{C}^{n \times mp}$ and $\mathcal{V}_{m+1} \in \mathbb{C}^{n \times (m+1)p}$ are such that $\mathcal{V}_{m+1}^H \mathcal{V}_{m+1} = I_{m+1}$, and $\tilde{\mathcal{H}}_m \in \mathbb{C}^{(m+1)p \times mp}$ has the following form:

$$\tilde{\mathcal{H}}_m = \begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,m} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,m} \\ 0_{p \times p} & H_{3,2} & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0_{p \times p} & 0_{p \times p} & 0_{p \times p} & H_{m+1,m} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_m \\ H_{m+1,m} E_m^H \end{bmatrix}. \quad (5)$$

Note that $\mathcal{H}_m \in \mathbb{C}^{mp \times mp}$ in (5) is a block Hessenberg matrix with $p \times p$ blocks $H_{i,j}$, and E_m is the matrix of the last p columns of I_{mp} . The minimization problem leads us to solving the small least-squares problem

$$Y_m = \underset{Y}{\operatorname{argmin}} \|ER - \tilde{\mathcal{H}}_m^H Y\|_F, \quad (6)$$

with $R_0 = V_1 R = \mathcal{V}_{m+1} E$, E is the $(m+1)p \times p$ matrix whose upper $p \times p$ principal block is an identity matrix.

Algorithm 1 Block Arnoldi with blockwise modified Gram–Schmidt orthogonalization.

- 1: Choose a unitary matrix V_1 of size $n \times p$.
 - 2: **for** $j = 1, 2, \dots, m$ **do**
 - 3: Compute $W_j = AV_j$
 - 4: **for** $i = 1, 2, \dots, j$ **do**
 - 5: $H_{i,j} = V_i^H W_j$
 - 6: $W_j = W_j - V_i H_{i,j}$
 - 7: **end for**
 - 8: $W_j = V_{j+1} H_{j+1,j}$ (reduced QR decomposition)
 - 9: **end for**
 - 10: Define $\mathcal{V}_{m+1} = [V_1, V_2, \dots, V_{m+1}]$, $\tilde{\mathcal{H}}_m = (H_{k,l})_{1 \leq k \leq m+1, 1 \leq l \leq m}$
-

For the solution of multi-shifted linear systems (1), it is possible to take advantage of the block shift-invariance property of the block Krylov subspace $\mathbb{K}_m(A, R_0)$, which can be written as

$$\mathbb{K}_m(A, R_0) = \mathbb{K}_m(A - \sigma_i I, R_0), \quad i = 1, \dots, L, \quad (7)$$

and compute a unique basis that can be employed for solving the whole sequence of shifted block systems. Indeed, Eq. (4) implies the new Arnoldi-like relation

$$(A - \sigma_i I)\mathcal{V}_m = \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m^i, \quad (8)$$

where

$$\tilde{\mathcal{H}}_m^i = \tilde{\mathcal{H}}_m - \sigma_i \begin{bmatrix} I_{mp \times mp} \\ 0_{p \times mp} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_m^i \\ H_{m+1,m} E_m^H \end{bmatrix}.$$

Instead of applying shifted Krylov subspace methods to the solution of each of the p multi-shifted linear systems independently [13–20], or block Krylov subspace methods to the solution of the L linear systems with multiple right-hand sides in sequence [10,12,21,22], it is much more efficient to use a shifted block method to solve all of the multi-shifted

linear systems simultaneously [23,24]. In this context we refer to the system $(A - \sigma_1 I)X_1 = B$ as the base shifted block system, and to $(A - \sigma_i I)X_i = B$ ($i = 2, \dots, k$) as the additional shifted block Krylov subspace solvers use much larger spaces to compute the solution than a conventional Krylov solver applied to each right-hand side one by one, the high costs due to the block orthogonalization can make these methods very expensive especially memory-wise (see [25], Table 2.4 for some figures of complexity). Therefore, as m increases, the block Arnoldi procedure is restarted periodically using the last computed approximation. Upon restarting, however, the block shift-invariance property does not hold anymore as the block residuals R_m^1 and R_m^i , ($i = 2, \dots, L$) of the base shifted block system and of the additional shifted block systems are in general no longer collinear. This means, in turn, that $\mathbb{K}_m(A - \sigma_1 I, R_m^1) \neq \mathbb{K}_m(A - \sigma_i I, R_m^i)$, $i = 2, \dots, L$, where we denote by $R_m^i = B - (A - \sigma_i I)X_0^i$, $i = 1, 2, \dots, L$ the block residual at iteration m of the i th shifted linear system.

In order to exploit Eq. (7), the BGMRES-Sh method proposed by Wu, Wang and Jin [23] forces the shifted block residuals to be collinear at each restart, i.e., it imposes that

$$R_m^i = R_m^1 W_m^i, \quad i = 2, \dots, L, \quad (9)$$

where W_m^i are some $p \times p$ nonsingular matrices. Assuming that the initial shifted block residuals are collinear, $R_0^i = R_0^1 W_0^i$, $i = 2, \dots, L$ (note that this relation is satisfied by taking $X_0^i = 0$), the approximate solutions of the base shifted block system can be formulated as $X_m^1 = X_0^1 + \mathcal{V}_m Y_m^1$, where X_0^1 is the initial approximation and Y_m^1 minimizes over the corresponding Krylov subspace the base shifted block residual

$$\begin{aligned} R_m^1 &= R_0^1 - (A - \sigma_1 I) \mathcal{V}_m Y_m^1, \\ &= \mathcal{V}_{m+1} (ER - \tilde{\mathcal{H}}_m^1 Y_m^1), \end{aligned}$$

with $R_0^1 = B - (A - \sigma_1 I)X_0^1 = \mathcal{V}_{m+1} ER$ and E is the $(m+1)p \times p$ matrix whose upper $p \times p$ principal block is an identity matrix. According to (6), since $\mathcal{V}_{m+1}$ is orthonormal, Y_m^1 can be computed by solving the following minimization problem

$$Y_m^1 = \underset{Y^1}{\operatorname{argmin}} \|ER - \tilde{\mathcal{H}}_m^1 Y^1\|_F.$$

The additional shifted block residuals R_m^i are obtained from Eq. (9) and write

$$R_0^i - (A - \sigma_i I) \mathcal{V}_m Y_m^i = (R_0^1 - (A - \sigma_1 I) \mathcal{V}_m Y_m^1) W_m^i, \quad i = 2, 3, \dots, L \quad (10)$$

or equivalently, using the Arnoldi relation,

$$\begin{bmatrix} \tilde{\mathcal{H}}_m^i & \widehat{V} \end{bmatrix} \begin{bmatrix} Y_m^i \\ W_m^i \end{bmatrix} = ER W_0^i, \quad i = 2, 3, \dots, L. \quad (11)$$

In Eq. (11) we have set $R_{LS} = ER - \tilde{\mathcal{H}}_m^1 Y_m^1$, $\widehat{W}_m^i = \widehat{R} W_m^i$ and $\widehat{V}$, $\widehat{R}$ are the QR factors of the reduced QR decomposition of R_{LS} . The approximate solutions of the additional shifted block systems can be expressed as $X_m^i = X_0^i + \mathcal{V}_m Y_m^i$, ($i = 2, \dots, L$), where Y_m^i , ($i = 2, \dots, L$) is the solution of Eq. (11). The pseudocode of the BGMRES-Sh method is sketched in Algorithm 2.

One practical difficulty with the implementation and use of block Krylov methods, besides the large memory requirements and some lack of theoretical results, is the possible linear dependence of the various residuals that slows down or even prevent the convergence [12]. The problem is inherited also by the BGMRES-Sh method, as illustrated in the following example.

Example 1.1. Consider a bidiagonal matrix A of dimension $n = 1000$ with diagonal entries equal to $1, 2, 3, \dots, 1000$ and superdiagonal entries all equal to one, the set of shifts $\sigma = 0, -0.4, -2$ and two different $n \times 6$ right-hand sides matrices B_1 and B_2 . The columns of B_1 are linearly independent, randomly generated with normal distribution by the built-in MATLAB function *randn*, while three columns of B_2 are linearly dependent. The BGMRES-Sh convergence histories plotted in Fig. 1 were obtained using a search space of dimension $m = 90$, a convergence tolerance equal to $\epsilon = 10^{-6}$ and a maximum number of iterations equal to *maxiter* = 5000 for the stopping criterion and the initial block guess equal to $0 \in \mathbb{C}^{n \times p}$. The vertical axis displays the $\log_{10}$ of the maximum relative residual norms, and the horizontal axis the number of matrix-vector products. As it is shown, the BGMRES-Sh method converges well when the right-hand sides are linearly independent but may fail when the right-hand sides tend to be linearly dependent. This example is representative of the general observed trend in our runs. $\square$

In Section 2 we introduce a variant of the block shifted GMRES method based on initial deflation, called the DBGMRES-Sh method, that may overcome this computational difficulty for the solution of sequences of multi-shifted linear systems with multiple right-hand sides given at once. The numerical experiments reported in Sections 2.1.1–2.1.2 show a significant enhancement of robustness compared to the original BGMRES-Sh method. The results clearly establish the need for preconditioning. The problem is not trivial as only special preconditioners can maintain the shift invariance property of Krylov subspaces, like polynomial preconditioner [23], shift-and-invert preconditioner [24,26] and multi-shifted preconditioners [27–29]. In Section 3 we present a preconditioned DBGMRES-Sh method that can accommodate both fixed and variable preconditioning, and we report on numerical experiments in Section 3.2. Finally, we conclude the paper with some concluding remarks in Section 4.

Fig. 1. Convergence histories for Example 1.1 with different right-hand sides. (Left: the BGMRES-Sh method requires 2143 *Mvps* to solve the sequence of systems; Right: the BGMRES-Sh method fails to converge.)

Throughout this study, we denote by $\|\cdot\|_2$ the Euclidean norm and $\|\cdot\|_F$ the Frobenius norm of a matrix. The superscript H denotes the transpose conjugate of a vector or a matrix, $I_k \in \mathbb{C}^{k \times k}$ is the identity matrix of dimension k and $0_{i \times j} \in \mathbb{C}^{i \times j}$ is the zero rectangular matrix with i rows and j columns. The singular values of a matrix $C \in \mathbb{C}^{i \times j}$ are denoted by $\sigma_1(C) \geq \dots \geq \sigma_{\min(i,j)}(C)$. We adopt the matrix MATLAB notation: $R(i, j)$ denotes the (i, j) th entry of a matrix R and $R(1:m, 1:j)$ the submatrix of the first m rows and the first j columns of R .

Algorithm 2 BGMRES-Sh method.

- 1: *Start.* Let the p linearly independent right-hand sides be $B = [b^{(1)}, b^{(2)}, \dots, b^{(p)}]$. Choose n_m : the maximum size of the underlying block approximation Krylov subspace in each cycle, ϵ : the tolerance, *maxiter*: the maximum number of iterations, $\{\sigma_i\}_{i=1}^L \subset \mathbb{C}$: shifts.
- 2: Choose initial guesses $X_0^i \in \mathbb{C}^{n \times p}$ and set $W_0^i = I$ ($i = 1, \dots, L$).
- 3: Compute the initial shifted block residuals $R_0^i \in \mathbb{C}^{n \times p}$ as $R_0^i = B - (A - \sigma_i I)X_0^i$ ($i = 1, \dots, L$).
- 4: Compute the reduced QR decomposition of $R_0^i = V_1 R$.
- 5: *Solve the base shifted block linear system.* Apply standard block GMRES(m). It generates v_{m+1} , $\tilde{H}_m^1$ and forms the new approximate solutions $X_m^1 = X_0^1 + v_m Y_m^1$ where $Y_m^1 = \operatorname{argmin}_{Y^1} \|ER - \tilde{H}_m^1 Y^1\|_F$. Compute the residual vectors $R_m^1 = R_0^1 - v_{m+1} \tilde{H}_m^1 Y_m^1$. Check convergence, and proceed if not satisfied.
- 6: *Solve the additional shifted block linear systems.* Solve $\begin{bmatrix} \tilde{H}_m^i & \hat{V} \end{bmatrix} \begin{bmatrix} Y_m^i \\ \hat{W}_m^i \end{bmatrix} = ERW_0^i$, $i = 2, 3, \dots, L$, for Y_m^i and $\hat{W}_m^i$; Then form the approximate solutions $X_m^i = X_0^i + v_m Y_m^i$ and compute the residual vectors R_m^i as $R_m^i = R_0^i - \hat{R}^{-1} \hat{W}_m^i$.
- 7: Check convergence, and restart if convergence is not achieved, i.e., go to 4.

2. DBGMRRES-Sh, a deflated variant of the BGMRES-Sh method

Different approaches can be used to face the convergence problem due to approximate linear dependence of the right-hand-side vectors. One technique is called explicit reduction [12,30,31]. Before starting the iterative solution, linear dependency in the block right-hand side B or in the initial block residual R_0 can be detected by computing a singular value decomposition (SVD) [12,25,32] and filtering the smallest singular values according to a certain deflation tolerance to remove the useless information from the block Krylov subspace. Recently, Calandra, Gratton, Langou, et al. [32] have proposed a flexible variant of block GMRES method based on initial deflation (named as BFGMRSD) to detecting the possible convergence of a linear combination of the systems. Another approach known as Arnoldi deflation detects a possible rank deficiency occurring during the block Arnoldi procedure and removes the linearly dependent columns of the matrix W_j at step 8 of Algorithm 1) ensuring that V_{j+1} has full rank [12]. A rank-revealing QR algorithm can be used to determine the deficiency. In addition, it is well known that the convergence of Krylov subspace methods depends to a large degree on the distribution of eigenvalues. A different technique, also called deflation, that can greatly improve the convergence of a Krylov method computes spectral information of the linear system from the Arnoldi decomposition at restart, and use this information to “remove” the smallest eigenvalues [24].

The approach that we follow taken in this paper is similar to the deflation strategy proposed in [32]. First, we consider the base shifted block system. Given a QR-factorization of the scaled block true residual $R_0^1 D^{-1} = QT$, where $D = \operatorname{diag}(b_1, \dots, b_p) \in \mathbb{C}^{p \times p}$ is a nonsingular diagonal scaling matrix with diagonal entries $b_l = \|B(:, l)\|_2$, $1 \leq l \leq p$, we compute the SVD of T ,

$$T = U \Sigma W^H,$$

where $U \in \mathbb{C}^{p \times p}$, $W \in \mathbb{C}^{p \times p}$ are unitary and $\Sigma \in \mathbb{C}^{p \times p}$ is diagonal. Small singular values (i.e. small diagonal entries in the matrix Σ) would signal approximate linear dependency of the shifted residuals. We filter the largest p_d singular values based on the following requirement

$$\sigma_l(T) > \epsilon_d \text{ tol}, \quad l = 1, 2, \dots, p_d, \quad (12)$$

where $0 < \epsilon_d \leq 1$ and tol is the convergence threshold used in the stopping criterion of the DBGMRES-Sh method, arguing that the more accurate the iterative solution, the more right-hand sides we should solve for. Experiments in Section 2.1.1 show that an overall good choice for the deflation threshold parameters is $\epsilon_d = 1$. By decomposing Σ as

$$\Sigma = \begin{bmatrix} \Sigma_+ & 0 \\ 0 & \Sigma_- \end{bmatrix},$$

with $\Sigma_+ \in \mathbb{C}^{p_d \times p_d}$ and $\Sigma_- \in \mathbb{C}^{(p-p_d) \times (p-p_d)}$, the scaled block true residual can be written accordingly as

$$R_0^1 D^{-1} = QU_+ \Sigma_+ W_+^H + QU_- \Sigma_- W_-^H, \quad (13)$$

with $U_+ \in \mathbb{C}^{p \times p_d}$, $U_- \in \mathbb{C}^{p \times (p-p_d)}$, $W_+ \in \mathbb{C}^{p \times p_d}$ and $W_- \in \mathbb{C}^{p \times (p-p_d)}$. Upon the block residuals deflation, only the linear systems corresponding to the p_d most linearly independent columns of $R_0^1 D^{-1}$ are solved simultaneously. Therefore, the block Arnoldi procedure described in Algorithm 1 is started with the starting matrix V_1 to compute at step j the block basis $\mathcal{V}_{j+1} \in \mathbb{C}^{n \times (j+1)p_d}$ and block Hessenberg matrix $\tilde{\mathcal{H}}_j^1 \in \mathbb{C}^{(j+1)p_d \times jp_d}$. Then the block Arnoldi relation (4) reduces to the Arnoldi-like relation

$$(A - \sigma_1 I) \mathcal{V}_j = \mathcal{V}_{j+1} \tilde{\mathcal{H}}_j^1. \quad (14)$$

In [32], Calandra, et al. have proved that the approximate solution to the system (2) is the minimizer

$$\underset{Y}{\operatorname{argmin}} \|B - A(X_0 + \mathcal{Z}_j Y \Sigma_+ W_+^H D)\|_F \quad (15)$$

computed over the space $X_0 + \operatorname{range}(\mathcal{Z}_j Y \Sigma_+ W_+^H D)$. It is straightforward to extend the result to our DBGMRES-Sh method for the solution of shifted block linear systems as shown in the following proposition.

Proposition 1. *Minimizing the Frobenius norm of the base shifted block true residual $\|B - (A - \sigma_1 I)X^1\|_F$ over the space $X_0 + \operatorname{range}(\mathcal{V}_j Y \Sigma_+ W_+^H D)$ at iteration $j = 1, 2, \dots, m$ of a given cycle amounts to solving the reduced minimization problem*

$$\mathcal{P}_r^d : \quad Y_j^1 = \underset{Y^1}{\operatorname{argmin}} \|\mathcal{B}_j - \tilde{\mathcal{H}}_j^1 Y^1\|_F \quad (16)$$

with $\mathcal{B}_m = \begin{bmatrix} I \\ 0 \end{bmatrix}$.

Proof. We consider the reduced minimization problem $\mathcal{P}_r^d$:

$$\underset{Y^1}{\operatorname{argmin}} \|\mathcal{B}_j - \tilde{\mathcal{H}}_j^1 Y^1\|_F^2 = \underset{Y^1}{\operatorname{argmin}} \|\mathcal{V}_{j+1}(\mathcal{B}_j - \tilde{\mathcal{H}}_j^1 Y^1) \Sigma_+ W_+^H\|_F^2, \quad (17)$$

$$= \underset{Y^1}{\operatorname{argmin}} \|\mathcal{V}_{j+1} \mathcal{B}_j \Sigma_+ W_+^H - (A - \sigma_1 I) \mathcal{V}_j Y^1 \Sigma_+ W_+^H\|_F^2, \quad (18)$$

$$= \underset{Y^1}{\operatorname{argmin}} \|QU_+ \Sigma_+ W_+^H - (A - \sigma_1 I) \mathcal{V}_j Y^1 \Sigma_+ W_+^H\|_F^2. \quad (19)$$

The first equality follows by using some elementary properties and the unitarily invariant property of the Frobenius norm. The second equality is obtained by applying the Arnoldi-like relation (14). The last equality follows from the relation $V_1 = QU_+ = \mathcal{V}_{j+1} \mathcal{B}_j$.

By adding a term independent of Y^1 to the right-hand sides of Eq. (19), we can rewrite Eq. (16) as:

$$\underset{Y^1}{\operatorname{argmin}} \|\mathcal{B}_j - \tilde{\mathcal{H}}_j^1 Y^1\|_F^2 = \underset{Y^1}{\operatorname{argmin}} \|QU_+ \Sigma_+ W_+^H + QU_- \Sigma_- W_-^H - (A - \sigma_1 I) \mathcal{V}_j Y^1 \Sigma_+ W_+^H\|_F^2. \quad (20)$$

Finally, using Eq. (13) we have

$$\underset{Y^1}{\operatorname{argmin}} \|\mathcal{B}_j - \tilde{\mathcal{H}}_j^1 Y^1\|_F = \underset{Y^1}{\operatorname{argmin}} \|R_0^1 D^{-1} - (A - \sigma_1 I) \mathcal{V}_j Y^1 \Sigma_+ W_+^H\|_F. \quad \square$$

From Proposition 1, the m th iteration of the DBGMRES-Sh method builds approximations of the form

$$X_m^1 = X_0^1 + \mathcal{V}_m Y_m^1 \Sigma_+ W_+^H D. \quad (21)$$

The corresponding shifted block residual writes

$$\begin{aligned} R_m^1 D^{-1} &= R_0^1 D^{-1} - (A - \sigma_1 I) \mathcal{V}_m Y_m^1 \Sigma_+ W_+^H, \\ &= \mathcal{V}_{m+1} (\mathcal{B}_m - \tilde{\mathcal{H}}_m^1 Y_m^1) \Sigma_+ W_+^H, \\ &= \mathcal{V}_{m+1} R_{LS}, \end{aligned} \quad (22)$$

where we introduce the block quasi-residual $R_{LS} = (\mathcal{B}_m - \tilde{\mathcal{H}}_m^1 Y_m^1) \Sigma_+ W_+^H$. The stopping criterion used at line 15 of Algorithm 4 is $\|R_{LS}(:, s)\|_2 \leq \epsilon_q \text{ tol}$, with $0 < \epsilon_q < 1$, $1 \leq s \leq p$. The following proposition guides the choice of the stopping threshold ϵ_q in relation to the deflation threshold (ϵ_d).

Proposition 2. *If the stopping criterion $\|R_{LS}(:, s)\|_2 \leq \epsilon_q \text{ tol}$, with $0 < \epsilon_q < 1$ is satisfied, a sufficient condition to get*

$$\frac{\|R_j^1(:, s)\|_2}{\|B(:, s)\|_2} \leq \text{tol},$$

consists in choosing ϵ_q such that $\epsilon_q + \epsilon_d = 1$.

Proof. Using Eq. (20), the base shifted block true residual can be written as

$$\begin{aligned} R_m^1 &= \mathcal{V}_{m+1} (\mathcal{B}_m - \tilde{\mathcal{H}}_m^1 Y_m^1) \Sigma_+ W_+^H D + QU_- \Sigma_- W_-^H D, \\ &= \mathcal{V}_{m+1} R_{LS} D + QU_- \Sigma_- W_-^H D. \end{aligned}$$

Recalling the definition of $D = \text{diag}(b_1, \dots, b_p)$ with $b_s = \|B(:, s)\|_2$, $1 \leq s \leq p$, for each linear system at iteration j we have

$$\begin{aligned} \frac{\|R_j^1(:, s)\|_2}{\|B(:, s)\|_2} &\leq \|\mathcal{V}_{j+1} R_{LS}(:, s)\|_2 + \|QU_- \Sigma_- W_-^H(s, :)^H\|_2, \\ &\leq \|R_{LS}(:, s)\|_2 + \sigma_{p_d+1}(T), \quad 1 \leq s \leq p \end{aligned} \quad (23)$$

(the second inequality is easily verified from Eq. (13) and since $\mathcal{V}_{m+1}$ is orthonormal). Now assuming that deflation has occurred at iterations j of the DBGREMS-Sh method, according to Eq. (12) we have

$$\sigma_{p_d+1}(T) \leq \epsilon_d \text{ tol}.$$

The last two relations result in the following inequality

$$\frac{\|R_j^1(:, s)\|_2}{\|B(:, s)\|_2} \leq \|R_{LS}(:, s)\|_2 + \epsilon_d \text{ tol}. \quad (24)$$

Therefore, if the stopping criterion $\|R_{LS}(:, s)\|_2 \leq \epsilon_q \text{ tol}$ is satisfied, a sufficient condition to get

$$\frac{\|R_j^1(:, s)\|_2}{\|B(:, s)\|_2} \leq \text{tol},$$

consists in choosing $\epsilon_q \in (0, 1)$ such that $\epsilon_q + \epsilon_d = 1$. $\square$

After solving the base shifted block linear system, in order to take advantage of the shift-invariance property of the block Krylov subspace for the computation of the additional shifted block residuals, we force the block residuals to be collinear by imposing the condition

$$R_m^i = R_m^1 N_m^i, \quad i = 2, \dots, L, \quad (25)$$

where N_m^i are some $p \times p$ nonsingular matrices and $R_m^1 = \mathcal{V}_{m+1} R_{LS} D$. Here we set $X_0^i = 0$, $i = 1, 2, \dots, L$, i.e., $R_0^i = R_0^1 N_0^i$. A similar approach is used in other shifted GMRES-type methods [14,23,24,33]. At this stage we can use the search space generated for the base shifted block system to approximate the solutions of the additional shifted block systems. As in Eq. (21), we can write

$$X_m^i = X_0^i + \mathcal{V}_m Y_m^i \Sigma_+ W_+^H D, \quad (i = 2, 3, \dots, L). \quad (26)$$

Then we obtain the expression for the shifted block residuals

$$R_m^i = B - (A - \sigma_i I) X_m^i = R_0^i - \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m^i Y_m^i \Sigma_+ W_+^H D. \quad (27)$$

The collinear condition (25), and Eqs. (22) and (27) give

$$\begin{aligned} R_m^i &= R_m^1 N_m^i \\ \Leftrightarrow R_0^i - \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m^i Y_m^i \Sigma_+ W_+^H D &= \mathcal{V}_{m+1} R_{LS} D N_m^i \\ \Leftrightarrow R_0^1 N_0^i - \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m^i Y_m^i \Sigma_+ W_+^H D &= \mathcal{V}_{m+1} R_{LS} D N_m^i \\ \Leftrightarrow \mathcal{V}_{m+1} (\tilde{\mathcal{H}}_m^i Y_m^i \Sigma_+ W_+^H D + R_{LS} D N_m^i) &= R_0^1 N_0^i \\ \Leftrightarrow \mathcal{V}_{m+1} (\tilde{\mathcal{H}}_m^i Y_m^i \Sigma_+ W_+^H D + R_{LS} D N_m^i) &= \mathcal{V}_{m+1} \mathcal{B}_m \Sigma_+ W_+^H D N_0^i. \end{aligned}$$

Unknowns Y_m^i and N_m^i are easy to calculate by solving the following linear systems

$$\begin{bmatrix} \tilde{H}_m^i & R_{LS} \end{bmatrix} \begin{bmatrix} Y_m^i \Sigma + W_+^H D \\ DN_m^i \end{bmatrix} = B_m \Sigma + W_+^H DN_0^i, \quad i = 2, 3, \dots, L.$$

Finally, we can use these quantities to form the approximate solutions X_m^i of the additional shifted block linear systems (see Eq. (26)). The complete pseudocode of the DBGMRES-Sh method is sketched in Algorithm 3.

Algorithm 3 DBGMRES-Sh method.

- 1: *Start.* Let the p linearly independent right-hand sides be $B = [b^{(1)}, b^{(2)}, \dots, b^{(p)}]$. Choose m : the maximum size of the underlying block approximation Krylov subspace in each cycle, tol : the tolerance, ϵ_d : deflation threshold, ϵ_q : a quality of convergence threshold, $maxiter$: the maximum number of iterations, $\{\sigma_i\}_{i=1}^L \subset \mathbb{C}$: shifts.
 - 2: Choose initial guesses $X_0^i \in \mathbb{C}^{n \times p}$ and set $N_0^i = I$ ($i = 1, \dots, L$).
 - 3: Define the diagonal matrix $D \in \mathbb{C}^{p \times p}$ as $D = \text{diag}(d_1, \dots, d_p)$ with $d_l = \|B(:, l)\|_2$ for l such that $1 \leq l \leq p$.
 - 4: Compute the initial shifted block residuals $R_0^i \in \mathbb{C}^{n \times p}$ as $R_0^i = B - (A - \sigma_i I)X_0^i$ ($i = 1, \dots, L$).
 - 5: Compute the QR decomposition of $R_0^1 D^{-1} = QT$ with $Q \in \mathbb{C}^{n \times p}$ and $T \in \mathbb{C}^{p \times p}$.
 - 6: Compute the SVD of T as $T = U \Sigma W^H$.
 - 7: Select p_d singular values of T such that $\sigma_l(T) > \epsilon_d tol$ for all l such that $1 \leq l \leq p_d$.
 - 8: Define $V_1 \in \mathbb{C}^{n \times p_d}$ as $V_1 = QU(:, 1 : p_d)$.
 - 9: *Solve the base shifted block linear system.* Apply BGMRES method (Algorithm 4). It generates $\mathcal{V}_{m+1}$, $\tilde{H}_m^1$ and forms the new approximate solutions $X_m^1 = X_0^1 + \mathcal{V}_m Y_m^1 \Sigma + W_+^H D$ where $Y_m^1 = \text{argmin}_{Y^1} \|B_m - \tilde{H}_m^1 Y^1\|_F$. Check convergence through R_{LS} , and proceed if not satisfied.
 - 10: *Solve the additional shifted block linear systems.* Solve $\begin{bmatrix} \tilde{H}_m^i & R_{LS} \end{bmatrix} \begin{bmatrix} Y_m^i \Sigma + W_+^H D \\ DN_m^i \end{bmatrix} = B_m \Sigma + W_+^H DN_0^i, i = 2, 3, \dots, L$, for Y_m^i and N_m^i ; Then form the approximate solutions $X_m^i = X_0^i + \mathcal{V}_m Y_m^i \Sigma + W_+^H D$ and compute the residual vectors R_m^i as $R_m^i = R_m^1 N_m^i$.
 - 11: Check convergence, and restart if convergence is not achieved, i.e., go to 4.
-

Algorithm 4 BGMRES method.

- 1: Choose a convergence threshold tol , a deflation threshold ϵ_d , a quality of convergence threshold ϵ_q , the size of the restart m and the maximum number of iteration $itermax$.
 - 2: Choose initial guesses $X_0 \in \mathbb{C}^{n \times p}$.
 - 3: Define the diagonal matrix $D \in \mathbb{C}^{p \times p}$ as $D = \text{diag}(d_1, \dots, d_p)$ with $d_l = \|B(:, l)\|_2$ for l such that $1 \leq l \leq p$.
 - 4: Compute the initial shifted block residuals $R_0 \in \mathbb{C}^{n \times p}$ as $R_0 = B - AX_0$.
 - 5: **for** $iter = 1, \dots, itermax$ **do**
 - 6: Compute the reduced QR decomposition of $R_0 D^{-1} = QT$ with $Q \in \mathbb{C}^{n \times p}$ and $T \in \mathbb{C}^{p \times p}$.
 - 7: Compute the SVD of T as $T = U \Sigma W^H$.
 - 8: Select p_d singular values of T such that $\sigma_l(T) > \epsilon_d tol$ for all l such that $1 \leq l \leq p_d$.
 - 9: Define $V_1 \in \mathbb{C}^{n \times p_d}$ as $V_1 = QU(:, 1 : p_d)$.
 - 10: Let $\mathcal{B}_k = \begin{bmatrix} I_{p_d} \\ 0_{kp_d \times p_d} \end{bmatrix}, 1 \leq k \leq m$
 - 11: **for** $j = 1, \dots, m$ **do**
 - 12: Apply block Arnoldi process (Algorithm 1 ($M^{-1} = I$)). It generates $\mathcal{V}_{j+1}$, $\tilde{H}_j$ such that: $A\mathcal{V}_j = \mathcal{V}_{j+1} \tilde{H}_j$ with $\mathcal{V}_{j+1}^H \mathcal{V}_{j+1} = I_{(j+1)p_d}$.
 - 13: Solve the minimization problem $Y_j = \text{argmin}_{Y \in \mathbb{C}^{p_d \times p_d}} \|\mathcal{B}_j - \tilde{H}_j Y\|_F$;
 - 14: Compute $R_{LS} = (\mathcal{B}_j - \tilde{H}_j Y_j) \Sigma(1 : p_d, 1 : p_d) W(1 : p, 1 : p_d)^H$
 - 15: **if** $\|R_{LS}(:, l)\| \leq \epsilon_q tol, \forall l \mid 1 \leq l \leq p$ **then**
 - 16: Compute the new approximate solutions $X_j = X_0 + \mathcal{V}_j Y_j \Sigma(1 : p_d, 1 : p_d) W(1 : p, 1 : p_d)^H D$
 - 17: **end if**
 - 18: **end for**
 - 19: $X_m = X_0 + \mathcal{V}_m Y_m \Sigma(1 : p_d, 1 : p_d) W(1 : p, 1 : p_d)^H D$
 - 20: $R_m = B - AX_m$.
 - 21: Check convergence, and restart if convergence is not achieved, set $R_0 = R_m$ and $X_0 = X_m$, i.e., go to 5.
 - 22: **end for**
-

Example 2.1. We illustrate the effectiveness of the DBGMRES-Sh method for solving the same sequence of linear systems and using the same settings of Example 1.1. As it is shown in the left plot of Fig. 2, the DBGMRES-Sh method clearly outperforms

Fig. 2. Convergence histories for [Example 2.1](#). (Left: convergence comparison between the BGMRES-Sh and DBGMRRES-Sh methods with right-hand sides B_1 . Right: convergence comparison between the BGMRES-Sh and DBGMRRES-Sh methods with right-hand sides B_2).

Table 1

Set and characteristics of the test matrices used in our experiments.

Example	Name	Size	Nonzeros	Field
1	sherman4	1,104	3,786	Computational fluid dynamics
2	cdde1	961	4681	Computational fluid dynamics
3	pde2961	2,961	14,585	Two-dimensional/three-dimensional problem
4	Poisson3Da	13,514	352,762	Computational fluid dynamics problem
5	fv3	9,801	87,025	Two-dimensional/three-dimensional problem
6	light_in_tissue	29,282	406,084	Electromagnetics problem

the BGMRES-Sh method in terms of *mvps*. The enhanced robustness of the new method is evident in the right plot of [Fig. 2](#) where the BGMRES-Sh method fails to converge. We illustrate more experiments in the next Section [2.1](#).

2.1. Numerical experiments

In the experiments of this section, the right-hand sides matrix $B = [b^{(1)}, \dots, b^{(p)}] \in \mathbb{C}^{n \times p}$ is a full-rank matrix consisting of p linearly independent column vectors randomly generated with normal distribution. The iterative solution is started from the initial guess $0 \in \mathbb{C}^{n \times p}$ and the stopping criterion used is

$$\frac{\|R_m^i(:, j)\|_2}{\|B(:, j)\|_2} \leq \text{tol}, \quad i = 1, \dots, L, \quad j = 1, \dots, p.$$

The convergence of each shifted linear system is monitored by computing the Euclidean norm of the columns of the block shifted residual, normalized by the corresponding right-hand side. The experiments are run under Windows 7 and MATLAB version R2014a, on a PC with an Inter(R) Core(TM) i5-4590 CPU @3.30 GHz and 8 GB of memory. All the timings are expressed in seconds.

2.1.1. General test problems

First we analyze the convergence of the DBGMRRES-Sh method for solving a set of general sparse matrices extracted from the Sparse Matrix Collection available at the University of Florida [\[34\]](#), described in [Table 1](#). We compare the convergence of the BGMRES-Sh method by Wu, Wang and Jin [\[23\]](#) against the DBGMRRES-Sh method introduced in Section [2](#) in terms of number of matrix–vector products (*mvps*), maximum true relative residuals norms (*trr*) and CPU solution time in seconds (*cpu*). The number of right-hand sides p is set equal to 6, the maximum dimension m of the search space equal to 90, the convergence tolerance tol to 10^{-6} and the maximum number of iterations *maxiter* to 5000. Except specific mention, we use the set of shifts $\sigma_i = [-0.001, -1, -10, -100]$ and $\epsilon_d = 0.1$ as deflation threshold.

As it is shown in [Table 2](#), the DBGMRRES-Sh method outperformed the BGMRES-Sh method in terms of *mvps* and *cpu* time in all our runs, converging to the targeted accuracy also on problems where the BGMRES-Sh method failed to converge. We use the symbol “—” to indicate no convergence achieved. The effect of deflation is evident from the plots in [Fig. 3](#) displaying the convergence histories of the DBGMRRES-Sh versus the BGMRES-Sh methods on the four shifted systems for each Example 1–6. Three convergence curves are similar for the two solvers. This indicates that no deflation has occurred on those shifted systems. On the curve where the convergence of the DBGMRRES-Sh method is significantly faster, the block residuals have been filtered to remove the approximate linear dependence of some column vectors during the cycle. In [Fig. 4](#) we display the number of actual right-hand sides processed versus the number of *mvps* over the iterations for the base shifted block system in each example. The approximate linear dependence of the shifted block residuals is detected according to the given deflation tolerance $\epsilon_d \text{ tol}$ at each restart, and the number of linear systems (p_d) solved is effectively reduced. The deflation

Table 2Comparative convergence behavior of the BGMRES-Sh and DBGMRES-Sh methods using a convergence tolerance $tol = 10^{-6}$.

Example	Method	<i>Mvps</i>	<i>Trr</i> (max)	<i>Cpu</i>
1	BGMRES-Sh	2335	9.3537e-07 9.7205e-07 7.9608e-07 3.4125e-07	1.7033
	DBGMRES-Sh	769	8.7650e-07 9.5631e-07 7.6744e-07 3.4125e-07	0.6658
2	BGMRES-Sh	–	–	–
	DBGMRES-Sh	803	8.1512e-07 6.5489e-07 5.2958e-07 2.5028e-07	0.6340
3	BGMRES-Sh	2049	9.9870e-07 9.9832e-07 5.8794e-07 5.8980e-07	2.3294
	DBGMRES-Sh	1640	8.9221e-07 9.8093e-07 5.8794e-07 5.8980e-07	1.3051
4	BGMRES-Sh	961	9.7383e-07 3.5471e-07 1.7616e-07 1.024e-08	5.2088
	DBGMRES-Sh	731	7.8038e-07 3.5471e-07 1.7616e-07 1.024e-08	3.0121
5	BGMRES-Sh	2743	9.9338e-07 6.8714e-07 6.4288e-07 1.6284e-08	5.8829
	DBGMRES-Sh	1431	9.0468e-07 6.8714e-07 6.4288e-07 8.0636e-07	3.3733
6	BGMRES-Sh	3913	9.9849e-07 5.4145e-07 8.1138e-08 4.416e-07	56.7468
	DBGMRES-Sh	2596	7.7635e-07 5.4145e-07 9.0627e-07 4.416e-07	38.9835

process can enhance significantly the overall robustness of the iterative solver, as shown for example on Problem 2. When the number of right-hand sides increases, the gain is more evident. In the experiments shown in Table 3 we use $p = 4, 8$ and 16 right-hand sides, setting the restart parameter m equal to $5 \times p$, $10 \times p$ and $15 \times p$, respectively (for instance, $m = 20, 40, 60$ for $p = 4$). Again the DBGMRES-Sh method required the smallest number of *mvps*, especially for the case of small restart parameter and large right-hand sides number. Note that on Example 6 none of the tested methods converged to the targeted accuracy using $p = 16$ right-hand sides, establishing the need for preconditioning, an issue that will be discussed in Section 3. Finally, in Table 4 different deflation threshold parameters ϵ_d were tested for the DBGMRES-Sh method, satisfying the condition $\epsilon_q + \epsilon_d = 1$ according to Proposition 2. It turns out that $\epsilon_d = 1$ can be an overall good choice to improve the convergence. In the BGMRES-Sh method no deflation occurs, and thus $\epsilon_d = 0$.

2.1.2. PageRank application

The PageRank model uses graph theory to rank data relevance. The problem can be described as the one of modeling the behavior of a random Web surfer who keeps visiting Web pages by clicking on hyperlinks available on the currently visited page or by jumping to an external page chosen at random. The mathematics underlying the PageRank problem are entirely general and thus can be applied to any graph or network, for example those arising in Complex Engineered Systems, Chemistry, Neuroscience, Biology and Bioinformatics to name only a few (see e.g. [35] for a comprehensive discussion on the widespread use of this model in science). Mathematically, the problem can be written as

$$(I - \alpha_i P^T)X_i = B, \text{ with } i = 1, 2, \dots, L,$$

where $0 < \alpha_i < 1$ are the damping parameters which assign the weights to the nodes of the graph, $B = [b_1, \dots, b_p]$ is the $n \times p$ matrix of personalization vectors b_j ($j = 1, \dots, p$) which represent the probability distribution to visit an unlinked node (i.e., of jumping to an external page). The columns of X_i are the unknown ranking vectors. In some instances one needs to solve a finite number of deterministic PageRank problems, each for a different combination of damping factors and personalization vectors (see e.g. [36]), leading to the solution of sequences of shifted linear systems with multiple right-hand sides.

We run experiments with the DBGMRES-Sh method on a set of Web matrices available from the University of Florida Sparse Matrix Collection [34], listed in Table 5. Large damping factors α close to one (e.g., $\alpha > 0.9$) may give a more accurate ranking than small ones [37]; however, higher damping factors typically slow down the iterative solution requiring many more iterations to converge [38]. In our experiments we set $\alpha = 0.99, 0.97, 0.95, 0.90, 0.85$, the columns of B are $b_j = \text{ones}(n, 1)/n$, $j = 1, \dots, p$, the number of right-hand sides is set equal to $p = 10, 15, 20$ and the maximum number of iterations allowed to achieve convergence is $\text{maxiter} = 6000, 9000, 12000$. As it is shown in Table 6, the DBGMRES-Sh method costs less *mvps* than the BGMRES-Sh method and it converges on the web-Stanford problem using $p = 10$ and $p = 15$, whereas the BGMRES-Sh method fails to converge. Convergence histories showing this trend are plotted in Fig. 5.

3. FDBGMRES-Sh, a flexible deflated variant of the BGMRES-Sh method

As it is shown in the previous section, the DBGMRES-Sh method can outperform other shifted block Krylov subspace methods; however, it may still suffer from slow convergence or could even fail to converge in some cases. When this happens, the convergence can be accelerated by preconditioning the original sequence of linear systems so that the transformed coefficient matrices have most of the eigenvalues clustered close to one. The difficulty of preconditioning shifted linear systems is that the shift-invariant property of the Krylov subspace may not be preserved under transformation. Therefore some special preconditioners need to be used, like polynomial preconditioners [23], shift-and-invert preconditioners [24,26] and multi-shifted preconditioners [27–29]. Additionally, in some computational settings the use of a constant preconditioner may be restrictive; for example in domain decomposition theory approximate solvers are often applied in the interior of the

Fig. 3. Convergence histories for Examples 1–6. The different curves for the convergence of the shifted block Krylov solvers correspond to different shifts.

domain [39, Sect. 4.3]. In this case, variable preconditioning that use a different operator at each iteration is an attractive option to consider [27,28,32,40–43]. This consideration prompts the need to develop flexible Krylov subspace methods for solving shifted systems that can accommodate the use of variable preconditioning, like in [27], Saibaba, Bakhos and Kitanidis.

3.1. Combining the DBGMRRES-Sh method with variable preconditioning

We consider a right preconditioner for the DBGMRRES-Sh method of the form $A_\tau = A - \tau I$, so that the right preconditioned system writes as

$$(A - \sigma_i I)(A - \tau I)^{-1} \tilde{X}_i = B, \quad i = 1, \dots, L,$$

where $X_i = (A - \tau I)^{-1} \tilde{X}_i$. From the identity

$$(A - \sigma_i I)(A - \tau I)^{-1} = I + (\tau - \sigma_i)(A - \tau I)^{-1},$$

it follows that A_τ maintains the shift-invariance property $\mathbb{K}_m((A - \sigma_i I)(A - \tau I)^{-1}, B) = \mathbb{K}_m((A - \tau I)^{-1}, B)$. A basis for the Krylov subspace $\mathbb{K}_m((A - \tau I)^{-1}, B)$ can be computed by running the block Arnoldi algorithm on the matrix $(A - \tau I)^{-1}$ starting

Fig. 4. Effect of deflation: number of right-hand sides processed during the convergence of the DBGMRRES-Sh method for solving the base shifted block system in Examples 1–6.

from the matrix B [24]. An Arnoldi-like factorization can be derived from the following relations

$$\mathcal{Z}_m = \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m, \quad (28)$$

$$(A - \tau I) \mathcal{Z}_m = \mathcal{V}_m, \quad (29)$$

with $\mathcal{V}_m^H \mathcal{V}_m = I$ by multiplying both sides of (28) with $(\tau - \sigma_i)$ and adding it to (29):

$$(A - \sigma_i I) \mathcal{Z}_m = \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m^i, \quad (30)$$

where

$$\tilde{\mathcal{H}}_m^i = \begin{bmatrix} I \\ 0 \end{bmatrix} + (\tau - \sigma_i) \tilde{\mathcal{H}}_m.$$

We can extend our DBGMRRES-Sh method by employing a variable preconditioner $A_{\tau_k} = A - \tau_k I$, for $k = 1, \dots, m$ at each iteration instead of using a fixed τ ,

$$(A - \tau_k I) \mathcal{Z}_k^F = \mathcal{V}_k, \quad k = 1, \dots, m.$$

Table 3

Number of matrix–vector products for the standard GMRES-Sh, BGMRES-Sh and DBGMRES-Sh methods with a different number of right-hand sides and different combinations of restart parameters, using a convergence tolerance $tol = 10^{-6}$.

m ($p = 4$)	Method	Example					
		1	2	3	4	5	6
20	GMRES-Sh	4135	–	–	1844	–	–
	BGMRES-Sh	–	–	1945	1493	–	–
	DBGMRES-Sh	1208	2064	1219	860	2299	4987
40	GMRES-Sh	3653	–	1684	1297	3656	–
	BGMRES-Sh	2829	–	1325	889	2597	3725
	DBGMRES-Sh	866	883	1069	637	1369	2696
60	GMRES-Sh	3102	–	1371	1055	2586	3868
	BGMRES-Sh	1655	–	1272	718	1833	2613
	DBGMRES-Sh	794	665	1073	562	1107	1997
m ($p = 8$)	Method	Example					
		1	2	3	4	5	6
40	GMRES-Sh	–	–	4697	3940	–	–
	BGMRES-Sh	–	–	4129	2993	–	–
	DBGMRES-Sh	1441	1446	2070	1379	2967	–
80	GMRES-Sh	–	–	3358	2533	–	–
	BGMRES-Sh	3821	–	2713	1618	–	–
	DBGMRES-Sh	1039	965	2015	1031	1931	4215
120	GMRES-Sh	–	–	2728	2030	–	–
	BGMRES-Sh	2026	–	2689	1113	3537	–
	DBGMRES-Sh	736	913	2115	913	1710	3229
m ($p = 16$)	Method	Example					
		1	2	3	4	5	6
90	GMRES-Sh	–	–	–	–	–	–
	BGMRES-Sh	–	–	–	4925	–	–
	DBGMRES-Sh	1806	1701	3594	2077	3828	–
160	GMRES-Sh	–	–	–	–	–	–
	BGMRES-Sh	–	–	–	3489	–	–
	DBGMRES-Sh	1367	1427	3494	1733	2890	–
240	GMRES-Sh	–	–	–	4079	–	–
	BGMRES-Sh	1521	–	–	1889	–	–
	DBGMRES-Sh	982	1272	3600	1461	2557	–

Table 4

Number of matrix–vector products of the DBGMRES-Sh method for different deflation thresholds ϵ_d .

Example	BGMRES-Sh	DBGMRES-Sh		
		$\epsilon_d = 0.01$	$\epsilon_d = 0.1$	$\epsilon_d = 1$
1	2335	853	769	721
2	–	892	803	721
3	2049	1810	1640	1531
4	961	834	731	721
5	2743	1618	1431	1261
6	3913	2991	2596	2251

Table 5

Characteristics of the Web matrices used in our PageRank calculations.

Name	Size	Nonzeros	Description
enron	69,244	276,143	Enron email network
email-EuAll	265,214	420,045	Email network from a EU research institution
web-Stanford	281,903	2,312,497	Web graph of Stanford.edu
wikipedia-20051105	1,634,989	19,753,078	Wikipedia pages

In the flexible block Arnoldi process, Eqs. (28)–(30) write as

$$\mathcal{Z}_m^F = \mathcal{V}_{m+1} \tilde{\mathcal{H}}_m, \quad (31)$$

$$A \mathcal{Z}_m^F - \mathcal{Z}_m^F T_m = \mathcal{V}_m, \quad (32)$$

Table 6

Number of matrix–vector products for the BGMRES-Sh and DBGMRES-Sh methods with a different number of right-hand sides, for solving a set of multi-shifted linear systems arising in PageRank computation using a convergence tolerance $tol = 10^{-6}$.

Example	Method	$p = 10$	$p = 15$	$p = 20$
enron	BGMRES-Sh	541	1506	2511
	DBGMRES-Sh	532	899	1754
email-EuAll	BGMRES-Sh	981	2446	4841
	DBGMRES-Sh	785	1525	3755
web-Stanford	BGMRES-Sh	–	–	–
	DBGMRES-Sh	5394	8927	–
wikipedia-20051105	BGMRES-Sh	2043	3862	7797
	DBGMRES-Sh	1979	3691	5691

Fig. 5. Convergence histories of the BGMRES-Sh and the DBGMRES-Sh methods on the Web matrices in PageRank computation. In these experiments we use $p = 20$ right-hand sides and $maxiter = 12000$.

$$(A - \sigma_i I)z_m^F = v_{m+1} \tilde{\mathcal{H}}_m^i, \quad i = 1, \dots, L, \quad (33)$$

where

$$T_m = \text{diag}(\tau_1, \dots, \tau_m) \quad \text{and} \quad \tilde{\mathcal{H}}_m^i = \begin{bmatrix} I \\ 0 \end{bmatrix} + (T_m - \sigma_i I) \tilde{\mathcal{H}}_m.$$

In Algorithm 1 we use preconditioners of the form $M_j^{-1} = A_{\tau_j}^{-1}$, we compute the preconditioned block vectors $Z_j = M_j^{-1}V_j$ at step $j = 1, \dots, m$; finally, we replace step 3 with $W_j = AZ_j$.

The flexible DBGMRES-Sh (shortly, FDBGMRES-Sh) algorithm seeks approximate solutions of the form $X_m^i = X_0 + z_m^F y_m^i \Sigma + W_+^H D$ in the subspace $X_0 + \text{span}(z_m^F y_m^i \Sigma + W_+^H D)$, which is no longer a Krylov subspace though, similarly to the procedure proposed in Section 2. Note that the analysis and results of the DBGMRES-Sh are still valid for the FDBGMRES-Sh

Table 7Convergence behavior of DBGMRES-Sh, PDBGMRES-Sh and FDBGMRES-Sh for different numbers of restart with $p = 16$.

Methods	$m = 40$		$m = 90$		$m = 160$	
	<i>Mvps</i>	<i>Cpu</i>	<i>Mvps</i>	<i>Cpu</i>	<i>Mvps</i>	<i>Cpu</i>
DBGMRES-Sh	–	–	–	–	–	–
PDBGMRES-Sh	234	154.6554	233	154.5543	160	227.2221
	–	–	–	–	219	152.2389
FDBGMRES-Sh	199	86.1151	180	80.8542	160	121.5863

Table 8

Convergence behavior of the DBGMRES-Sh, PDBGMRES-Sh and FDBGMRES-Sh methods with different number of right-hand sides combinations of different restart parameters.

Methods	$p = 6, m = 90$		$p = 10, m = 100$		$p = 16, m = 110$	
	<i>Mvps</i>	<i>Cpu</i>	<i>Mvps</i>	<i>Cpu</i>	<i>Mvps</i>	<i>Cpu</i>
DBGMRES-Sh	–	–	–	–	–	–
	54	4.5044	85	8.2102	129	12.8048
PDBGMRES-Sh	–	–	–	–	–	–
	406	26.4689	607	38.7868	–	–
FDBGMRES-Sh	51	3.7037	80	4.8010	117	7.5282

method. In practice, only a few values of τ_k close to σ_i could be selected, as the use of m different preconditioners may become very expensive when m is large; see [27,28] for more details.

3.2. Numerical experiments with variable preconditioning

We illustrate the benefits of using variable preconditioning with the DBGMRES-Sh method by comparing, on the same set of matrix problems listed in Table 1, the convergence of the flexible deflated block shifted GMRES method (FDBGMRES-Sh), the right preconditioned deflated block shifted GMRES method (PDBGMRES-Sh) and the unpreconditioned deflated block shifted GMRES method (DBGMRES-Sh). We focus our analysis on Example 6 using $p = 16$ right-hand sides as it is shown in Table 3 that the DBGMRES-Sh method cannot converge on this problem even using a large restart m .

The choice of the preconditioners' shifts τ_k is important in our method. Inspired by [21], in our experiments we choose τ_k near the shifts of interest as this strategy has been shown to yield often fast convergence rates. However, we acknowledge that the choice of τ_k for multi-shifted linear systems is still an open problem that deserves closer attention; e.g., see [44] for a strategy to select the optimal parameter τ for a single shift-and-invert preconditioner applied to the solution of shifted linear systems. For the PDBGMRES-Sh method, we choose a right preconditioner in shifted form $A - \tau I$ with shifts $\tau = -0.002, -0.006, -2$. For the restart value $m = 90$, the preconditioner is computed by choosing $\tau_k = -0.002$ for $1 \leq k \leq 20$, $\tau_k = -0.006$ for $21 \leq k \leq 30$ and $\tau_k = -2$ for $31 \leq k \leq 40$. For the value $m = 100$, the preconditioner is computed by choosing $\tau_k = -0.002$ for $1 \leq k \leq 40$, $\tau_k = -0.006$ for $41 \leq k \leq 80$ and $\tau_k = -2$ for $81 \leq k \leq 90$. Finally, for $m = 160$, the preconditioner is setup by selecting $\tau_k = -0.002$ for $1 \leq k \leq 90$, $\tau_k = -0.006$ for $91 \leq k \leq 120$ and $\tau_k = -2$ for $121 \leq k \leq 160$. In our experiments, we use a sparse direct solver (the built-in MATLAB's '\ operator) to invert the preconditioned systems at each step.

From the results reported in Table 7, the following conclusions can be drawn

- (1) the FDBGMRES-Sh method is more robust than the DBGMRES-Sh method also on problems where the latter fails to converge;
- (2) the PDBGMRES-Sh method is effective if $\tau \approx \sigma_i$; however, it may still fail in some cases;
- (3) the FDBGMRES-Sh method outperforms the PDBGMRES-Sh method in terms of *mvps* and *cpu* time. For $m = 160$, the FDBGMRES-Sh and PDBGMRES-Sh methods have the same *mvps*, however the FDBGMRES-Sh need significantly less *cpu* time;
- (4) the FDBGMRES-Sh method can converge using small restart values, hence at low memory costs. Higher restart values can enable us to reduce further the number of *mvps* but increase the overall *cpu* time. as more inner systems $(A - \tau_k I)\mathcal{Z}_k^F = \mathbf{v}_k$ need to be solved.

Next, we consider an additional experiment on Example 3; the results are reported in Table 8. In this experiment we set 20 shift parameters $\sigma_i = -0.001 * i$, $1 \leq i \leq 10$ and $\sigma_i = -(1.0 + 0.001 * i)$, $11 \leq i \leq 20$. We select a few different τ_k 's in the FDBGMRES-Sh method matching the full range of parameters σ_i . We set different values of $\tau = -0.018, -0.0018, -1$ to be used respectively in the PDBGMRES-Sh method. For the setting $p = 6, m = 90$, we consider three preconditioners computed by choosing $\tau_k = -0.018$, $1 \leq k \leq 20$, $\tau_k = -0.0018$, $21 \leq k \leq 60$ and $\tau_k = -1$, $61 \leq k \leq 90$; for the setting $p = 10, m = 100$, the three preconditioners are setup by using $\tau_k = -0.018$, $1 \leq k \leq 40$, $\tau_k = -0.0018$, $41 \leq k \leq 80$

and $\tau_k = -1$, $81 \leq k \leq 100$; finally, for the setting $p = 16$, $m = 110$, the preconditioners are computed by choosing $\tau_k = -0.018$, $1 \leq k \leq 40$, $\tau_k = -0.0018$, $41 \leq k \leq 80$ and $\tau_k = -1$, $81 \leq k \leq 110$. As it is shown in Table 8, the FDBGMRES-Sh method is still more robust than the DBGMRES-Sh method and cheaper than the PDBGMRES-Sh method.

4. Conclusions

A new flexible and deflated variant of the shifted block GMRES method is presented. The algorithm does a good job in solving the whole sequence of linear systems with multiple shifts and multiple right-hand sides simultaneously, and in detecting the linear systems convergence. In addition, it allows to use variable preconditioning. Numerical experiments show that the new method can outperform the BGMRES-Sh and the conventional GMRES-Sh methods for solving a set of general sparse matrix problems as well as in PageRank calculations. Therefore, it can be a viable solver to use in applications. To the best of our knowledge, the proposed FDBGMRES-Sh method is the first algorithm that combines deflation techniques and variable preconditioning for solving sequences of multi-shifted linear systems with multiple right-hand sides. Identifying the best deflation strategy in this context is still an open question and will be a topic of interest for a future study.

Acknowledgments

This research is supported by NSFC (61772003). The first author is also funded by the University of Groningen Ubbo Emmius scholarship, the Netherlands.

References

- [1] A. Frommer, B. Nöckel, S. Güsken, T. Lippert, K. Schilling, Many masses on one stroke: Economic computation of quark propagators, *Internat. J. Modern Phys. C* 06 (05) (1995) 627–638.
- [2] A. Frommer, P. Maass, Fast CG-based methods for Tikhonov-Phillips regularization, *SIAM J. Sci. Comput.* 20 (5) (1999) 1831–1850.
- [3] M.R. Celis, J.E. Dennis, R.A. Tapia, A trust region strategy for nonlinear equality constrained optimization, in: P. Boggs, R.H. Byrd, R.B. Schnabel (Eds.), *Proceedings SIAM Conference on Numerical Optimization*, SIAM, Philadelphia, PA, 1985, pp. 71–82.
- [4] T.H. Haveliwala, Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search, *IEEE Trans. Knowl. Data Eng.* 15 (4) (2003) 784–796.
- [5] Y. Saad, M.H. Schultz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Comput.* 7 (3) (1986) 856–869.
- [6] R. Fletcher, *Conjugate Gradient Methods for Indefinite Systems*, in: *Lecture Notes in Mathematics*, vol. 506, Springer, New York, 1976, pp. 73–89.
- [7] D.-L. Sun, Y.-F. Jing, T.-Z. Huang, B. Carpentieri, A quasi-minimal residual variant of the BiCORSTAB method for nonsymmetric linear systems, *Comput. Math. Appl.* 67 (10) (2014) 1743–1755.
- [8] Y.-F. Jing, T.-Z. Huang, Y. Zhang, L. Li, G.-H. Cheng, Z.-G. Ren, Y. Duan, T. Sogabe, B. Carpentieri, Lanczos-type variants of the COCR method for complex nonsymmetric linear systems, *J. Comput. Phys.* 228 (17) (2009) 6376–6394.
- [9] X.-M. Gu, M. Clemens, T.-Z. Huang, L. Li, The SCBiCG class of algorithms for complex symmetric linear systems with applications in several electromagnetic model problems, *Comput. Phys. Comm.* 191 (2015) 52–64.
- [10] E. Agullo, L. Giraud, Y.-F. Jing, Block GMRES method with inexact breakdowns and deflated restarting, *SIAM J. Matrix Anal. Appl.* 35 (4) (2014) 1625–1651.
- [11] Y. Saad, *Iterative Methods for Sparse Linear Systems*, second ed., SIAM, Philadelphia, 2003.
- [12] M.H. Gutknecht, Block Krylov space methods for linear systems with multiple right-hand sides: An introduction, in: A.H. Siddiqi, I.S. Duff, O. Christensen (Eds.), *Modern Mathematical Models. Methods and Algorithms for Real World Systems*, Anamaya Publishers, New Delhi, 2006, pp. 420–447.
- [13] L. Du, T. Sogabe, S.-L. Zhang, IDR(s) for solving shifted nonsymmetric linear systems, *J. Comput. Appl. Math.* 274 (2015) 35–43.
- [14] A. Frommer, U. Glässner, Restarted GMRES for shifted linear systems, *SIAM J. Sci. Comput.* 19 (1) (1998) 15–26.
- [15] A. Frommer, BiCGStab(l) for families of shifted linear systems, *Computing* 70 (2) (2003) 87–109.
- [16] Y.-F. Jing, T.-Z. Huang, Restarted weighted full orthogonalization method for shifted linear systems, *Comput. Math. Appl.* 57 (9) (2009) 1583–1591.
- [17] K.M. Soodhalter, D.B. Szyld, F. Xue, Krylov subspace recycling for sequences of shifted linear systems, *Appl. Numer. Math.* 81 (2014) 105–118.
- [18] V. Simoncini, Restarted full orthogonalization method for shifted linear systems, *BIT Numer. Math.* 43 (2) (2003) 459–466.
- [19] Y.-F. Jing, P. Yuan, T.-Z. Huang, A simpler GMRES and its adaptive variant for shifted linear systems, *Numer. Linear Algebra Appl.* 24 (1) (2017). <http://dx.doi.org/10.1002/nla.2076>.
- [20] X.M. Gu, T.Z. Huang, L. Li, H.B. Li, T. Sogabe, M. Clemens, Quasi-minimal residual variants of the COCG and COCR methods for complex symmetric linear systems in electromagnetic simulations, *IEEE Trans. Microw. Theory Tech.* 62 (12) (2014) 2859–2867.
- [21] K. Meerbergen, Z. Bai, The Lanczos method for parameterized symmetric linear systems with multiple right-hand sides, *SIAM J. Matrix Anal. Appl.* 31 (4) (2010) 1642–1662.
- [22] W.E. Boyse, A.A. Seidl, A block QMR method for computing multiple simultaneous solutions to complex symmetric systems, *SIAM J. Sci. Comput.* 17 (1) (1996) 263–274.
- [23] G. Wu, Y.-C. Wang, X.-Q. Jin, A preconditioned and shifted GMRES algorithm for the pagerank problem with multiple damping factors, *SIAM J. Sci. Comput.* 34 (5) (2012) A2558–A2575.
- [24] D.-L. Sun, T.-Z. Huang, Y.-F. Jing, B. Carpentieri, A block GMRES method with deflated restarting for solving linear systems with multiple shifts and multiple right-hand sides, *Numer. Linear Algebra Appl.* (2018) e2148. <http://dx.doi.org/10.1002/nla.2148>.
- [25] X. Pinel, A perturbed two-level preconditioner for the solution of three-dimensional heterogeneous helmholtz problems with applications to geophysics, 2010.
- [26] K. Meerbergen, The solution of parametrized symmetric linear systems, *SIAM J. Matrix Anal. Appl.* 24 (4) (2003) 1038–1059.
- [27] A.K. Saibaba, T. Bakhos, P.K. Kitanidis, A flexible krylov solver for shifted systems with application to oscillatory hydraulic tomography, *SIAM J. Sci. Comput.* 35 (6) (2013) A3001–A3023.
- [28] G.-D. Gu, X.-L. Zhou, L. Lin, A flexible preconditioned Arnoldi method for shifted linear systems, *J. Comput. Math.* 25 (5) (2007) 522–530.
- [29] T. Bakhos, P. Kitanidis, S. Ladenheim, A.K. Saibaba, D. Szyld, Multipreconditioned GMRES for shifted systems, *SIAM J. Sci. Comput.* 39 (5) (2017) S222–S247.
- [30] J. Langou, *Iterative methods for solving linear systems with multiple right hand sides*, 2003.

- [31] M. Robbé, M. Sadkane, Exact and inexact breakdowns in the block GMRES method, *Linear Algebra Appl.* 419 (1) (2006) 265–285.
- [32] H. Calandra, S. Gratton, J. Langou, X. Pinel, X. Vasseur, Flexible variants of block restarted GMRES methods with application to geophysics, *SIAM J. Sci. Comput.* 34 (2) (2012) A714–A736.
- [33] D. Darnell, R.B. Morgan, W. Wilcox, Deflated GMRES for systems with multiple shifts and multiple right-hand sides, *Linear Algebra Appl.* 429 (10) (2008) 2415–2434.
- [34] T.A. Davis, Y. Hu, The university of florida sparse matrix collection, in: *NA Digest*, Vol. 92, 1994.
- [35] D.F. Gleich, Pagerank beyond the web, *SIAM Rev.* 57 (2015) 321–363.
- [36] P.G. Constantine, D.F. Gleich, Random alpha pagerank, *Internet Math.* 6 (2009) 189–236.
- [37] A.N. Langville, C.D. Meyer, Deeper inside pagerank, *Internet Math.* 1 (2005) 335–380.
- [38] B.-Y. Pu, T.-Z. Huang, C. Wen, A preconditioned and extrapolation-accelerated GMRES method for pagerank, *Appl. Math. Lett.* 37 (2014) 95–100.
- [39] A. Toselli, O.B. Widlund, *Domain Decomposition Methods: Algorithms and Theory*, Vol. 34, Springer, 2005.
- [40] B. Carpentieri, I.S. Duff, L. Giraud, G. Sylvand, Combining fast multipole techniques and an approximate inverse preconditioner for large electromagnetism calculations, *SIAM J. Sci. Comput.* 27 (3) (2005) 774–792.
- [41] L.M. Carvalho, S. Gratton, R. Lago, X. Vasseur, A flexible generalized conjugate residual method with inner orthogonalization and deflated restarting, *SIAM J. Matrix Anal. Appl.* 32 (4) (2011) 1212–1235.
- [42] L. Giraud, S. Gratton, X. Pinel, X. Vasseur, Flexible GMRES with deflated restarting, *SIAM J. Sci. Comput.* 32 (4) (2010) 1858–1878.
- [43] D.B. Szyld, J.A. Vogel, FQMR: A flexible quasi-minimal residual method with inexact preconditioning, *SIAM J. Sci. Comput.* 23 (2) (2001) 363–380.
- [44] M. Baumann, M.B. van Gijzen, An efficient two-level preconditioner for multi-frequency wave propagation problems. DIAM Technical Report, 17- 03 (2017) 1–17.