A block minimum residual norm subspace solver for sequences of multiple left and right-hand side linear systems

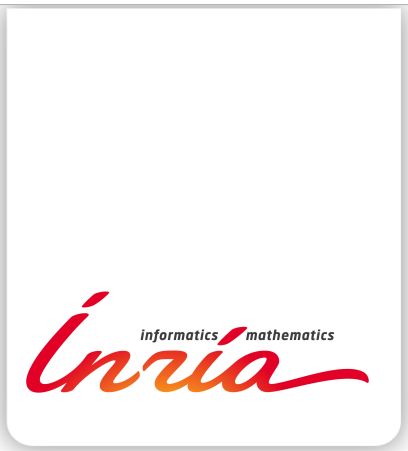
Luc Giraud, Yan-Fei Jing, Yanfei Xiang

A block minimum residual norm subspace solver for sequences of multiple left and right-hand side linear systems

L. Giraud, Y.-F. Jing, Y.-F. Xiang

RESEARCH REPORT
N° 9393
February 2021
Project-Teams HiePACS

ISSN 0249-6399 ISRN INRIA/RR--9393--FR+ENG



A block minimum residual norm subspace solver for sequences of multiple left and right-hand side linear systems

L. Giraud*, Y.-F. Jing†, Y.-F. Xiang* ‡

Project-Teams HiePACS

Research Report n° 9393 — February 2021 — 60 pages

Abstract: We are concerned with the iterative solution of linear systems with multiple right-hand sides available one group after another, including the case where there are massive number (like tens of thousands) of right-hand sides associated with a single matrix so that all of them cannot be solved at once but rather need to be split into chunks of possible variable sizes. For such sequences of linear systems with multiple left and right-hand sides, we develop a new recycling block generalized conjugate residual method with inner orthogonalization and inexact breakdown (IB-BGCRO-DR), which glues subspace recycling technique in GCRO-DR [SIAM J. Sci. Comput., 28(5) (2006), pp. 1651–1674] and inexact breakdown mechanism in IB-BGMRES [Linear Algebra Appl., 419 (2006), pp. 265–285] to guarantee this new algorithm could reuse spectral information for subsequent cycles as well as for the remaining linear systems to be solved. Related variant IB-BFGCRO-DR that suits for flexible preconditioning is devised to cope with constraints on some applications while also enabling mixed-precision calculation, which provides advantages in speed and memory usage over double precision as well as in perspective of emerging computing units such as the GPUs.

Key-words: Block subspace methods, augmentation, deflation, subspace recycling, Block-GMRES, inexact block rank deficiency.

* Inria, France

† School of Mathematical Sciences/Institute of Computational Science, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, P. R. China - supported by Science Challenge Project (TZ2016002-TZZT2019-B1.4), NSFC (12071062, 61772003), Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216), and Science Strength Promotion Programme of UESTC.

‡ Cerfacs, France

RESEARCH CENTRE
BORDEAUX – SUD-OUEST

200 avenue de la Vieille Tour
33405 Talence Cedex

Une méthode bloc de sous-espace minimisant la norme des résidus pour des séquence de systèmes linéaires

Résumé : Nous nous intéressons à la solution itérative de systèmes linéaires avec plusieurs second-membres disponibles un groupe après l'autre, y compris le cas où il y a un nombre massif (comme des dizaines de milliers) de second-membres associés à une seule matrice de sorte que tous ne peuvent pas être résolus en une fois mais doivent plutôt être divisés en morceaux de tailles variables possibles. Pour de telles séquences de systèmes linéaires à matrices et second-membres multiples, nous développons une nouvelle méthode de recyclage des résidus conjugués généralisés par blocs avec orthogonalisation interne et convergence partielle (IB-BGCRO-DR), qui exploite technique de recyclage subspatial dans GCRO-DR [SIAM J. Sci. Comput., 28(5) (2006), pp. 1651-1674] mécanisme de convergence partielle dans IB-BGMRES [Algèbre linéaire, 419 (2006), pp. 265-285] pour garantir que ce nouvel algorithme pourrait réutiliser les informations spectrales pour les cycles suivants ainsi que pour les systèmes linéaires restants à résoudre. La variante connexe IB-BFGCRO-DR qui convient au préconditionnement flexible est conçue pour faire face aux contraintes de certaines applications tout en permettant un calcul de précision mixte, ce qui présente des avantages en termes de vitesse et d'utilisation de la mémoire par rapport à la double précision ainsi que dans la perspective des unités de calcul émergentes telles que les GPU. En outre, nous discutons également des choix possibles lors de la construction d'un sous-espace de recyclage ainsi que de la manière d'exploiter le mécanisme de convergence partielle pour réaliser la flexibilité des politiques d'expansion de l'espace de recherche et surveiller les seuils de convergence individuels pour chaque second-membre. Comme effet secondaire, on peut également illustrer le fait que cette méthode peut être appliquée au cas des matrices constantes ou variant lentement. Enfin, nous démontrons les avantages numériques et informatiques de la combinaison de ces idées dans de tels algorithmes sur un ensemble d'exemples académiques simples.

Mots-clés : Méthode bloc de sous-espace, augmentation, déflation, recyclage de sous-espace, block-GMRES, convergence partielle.

Contents

1 Introduction

We consider the solution of a long sequence of slowly-changing families of general linear systems of the form:

$$A^{(\ell)}X^{(\ell)} = B^{(\ell)}, \quad \ell = 1, 2, \ldots, \tag{1}$$

where, associated with the ℓ^{th} family, $A^{(\ell)} \in \mathbb{C}^{n\times n}$ is a slowly-changing square nonsingular matrix of large dimension n along the family index ℓ, $B^{(\ell)} = [b^{(\ell,1)}, b^{(\ell,2)}, \ldots, b^{(\ell,p^{(\ell)})}] \in \mathbb{C}^{n\times p^{(\ell)}}$ are simultaneously given right-hand sides of full rank with $p^{(\ell)} \ll n$, and $X^{(\ell)} = [x^{(\ell,1)}, x^{(\ell,2)}, \ldots, x^{(\ell,p^{(\ell)})}] \in \mathbb{C}^{n\times p^{(\ell)}}$ are the solutions to be computed. Both the coefficient matrix $A^{(\ell)}$ and right-hand sides $B^{(\ell)}$ change from one family to the next, and the families of linear systems are typically available in sequence.

Many large scientific and industrial applications warrant the solution of such a sequence of families of linear systems, to name a few, such as full waveform inversion problems [37, 38], frequency response functions computation of a vibrating system over a given frequency range in engineering applications including structural dynamics and acoustics [22], parameter estimation involving parameter-dependent partial differential equations with multiple right-hand sides under realistic settings of Marine acquisition with different sources and receivers both moving [16], parametric macro-modeling problems of micro-electro-mechanical systems with the parametric model order reduction technique [11], electromagnetic wave propagation problems with the model reduction approach consisting of a multipoint Galerkin asymptotic waveform evaluation to automate the fast frequency sweep process [30], computations for diffuse optical tomographic imaging with multiple sources [19], Google PageRank model problems with multiple damping factors and multiple personalization vectors [44], electromagnetic scattering problems [17, 35], various source locations in seismic, parametric studies in general, and some other related time-dependent nonlinear inverse problems [18]. Specifically, large sparse complex non-Hermitian and nonsymmetric linear systems with thousands of right-hand sides are required to be solved when multiple wave sources are used in the solution of the acoustic full waveform inversion problem in geophysics associated with the phenomena of wave propagation through a heterogeneous model simulating the subsurface of Earth [20]. We refer to [1, 3, 7, 11–13, 21, 47] for more applications.

When solving sequences of linear systems as (1), attractive approaches are those that can exploit information generated during the solution of a given system to accelerate the convergence for the next ones. Deflated restarting is implementing a similar idea between the cycles in the generalized minimum residual norm method (GMRES) [31, 34, 43]; it is realized by using a deflation subspace containing a few approximate eigenvectors deemed to hamper the convergence of the Krylov subspace [23–25] methods. Another alternative technique is the subspace recycling proposed in the GCRO-DR [27] method, which is a combination of GMRES-DR [24] and GCRO [9] since it reuses the target approximated eigenvectors associated with deflated restarting technique in GMRES-DR within the GCRO framework. This latter method can reuse information accumulated in previous cycles as well as that accumulated in solving previous families. Given the multiple right-hand sides of (1) are simultaneously available, block Krylov subspace methods are often considered as the suitable candidates for their capability of sharing search subspaces that can be generated using BLAS3-like implementation [15]. A common issue in block Krylov subspace methods is the rank deficiency that might appear when expanding the residual spaces, which is caused by the convergence of individual or linear combination of solution vectors. Such rank deficiency problem could lead the block Arnoldi process to breakdown before the solution for all the right-hand sides are found. Although deflating the directions associated with almost converged solutions might appear as a natural way to deal with such issue, this often leads to eventually slowing down the convergence [21]. For the sake of balancing robustness and convergence rate, Robbé and Sadkane proposed an inexact breakdown detecting mechanism for the block GMRES algorithm (denoted by IB-BGMRES) [32], which could keep and reintroduce directions associated with the almost converged parts in next iterations if necessary. We refer to [2, 4, 32] for relevant works on inexact breakdown detection, as well as to [39–

42, 45] for related variants of block Krylov subspace methods for solving linear systems with multiple right-hand sides.

The contribution of this paper is to combine subspace recycling techniques of GCRO-DR [27] with the inexact breakdown mechanism of IB-BGMRES [32] to develop a new recycling block GCRO-DR variant with inexact breakdown detection. The remainder of this paper is organized as follows. We first recall the governing ideas of the minimum norm residual block Krylov subspace methods in Section 2. We also briefly discuss the inexact breakdown mechanism developed by Robbé and Sadkane [32]. Those two ingredients are exploited to develop the new algorithm IB-BGCRO-DR and its flexible preconditioning variant referred to as IB-BFGCRO-DR discussed in Section 2.4. Some additional numerical features to combine numerical and computation efficiency in the subspace expansion policy are discussed in Section 3. We also describe in this section how the subspace expansion can be adapted to accommodate different individual thresholds for each right-hand side. In Section 4 we show the benefit of combining the ideas mentioned above on a set of simple academic examples with both constant and slowly varying successive linear systems with multiple right-hand sides and finally make some concluding remarks in Section 5.

The symbol $||\cdot||_q$ denotes the Euclidean norm when $q = 2$ and the Frobenius norm when $q = F$. The superscript H denotes the transpose conjugate of a vector or a matrix, and the superscript $\dagger$ refers to the Moore-Penrose inverse. For convenience of the algorithm illustration and presentation, some MATLAB notation is used. A subscript j for a scalar or a matrix is used to indicate that the scalar or the matrix is obtained at iteration j. A matrix $C \in \mathbb{C}^{k\times\ell}$ consisting of k rows and ℓ columns sometimes is denoted as $C_{k\times\ell}$ explicitly. The identity and null matrices of dimension k are denoted respectively by I_k and 0_k or just I and 0 when the dimension is evident from the context. If $C \in \mathbb{C}^{k\times\ell}$, the singular values of C are denoted by $\sigma_1(C) \geq \ldots \geq \sigma_{min(k,\ell)}(C)$ in descent order. A positive integer m represents the maximal dimension of the underlying block approximation subspace in each cycle.

2 Block GCRO-DR with inexact breakdown

In this section, we first recall the subspace recycling techniques existing in minimum residual Krylov methods GCRO [9] and GCRO-DR [27]. For the sake of simplicity of exposure, the block formulation of GCRO-DR (BGCRO-DR) [26, 28, 29] is presented straightforwardly. Then, in the context of the block methods, the driving ideas of inexact breakdown mechanism as well as the corresponding block Arnoldi-like recurrence formula developed by Robbé and Sadkane [32] are discussed for handing the partial convergence of the solution vectors that might lead to rank deficiency.

For simplicity and notational convenience, we drop in the rest of this paper the superscript $^{(\ell)}$ in $B^{(\ell)}$ and $X^{(\ell)}$ when considering to solve the current ℓ^{th} family of linear systems in the entire sequence of families. We indicate the superscript for a family order explicitly when necessary. That is, suppose that the current ℓ^{th} family of linear systems to be solved is

$$AX = B, \tag{2}$$

where, $A \in \mathbb{C}^{n\times n}$ is the current square nonsingular matrix of dimension n, $B = [b^{(1)}, b^{(2)}, \ldots, b^{(p)}] \in \mathbb{C}^{n\times p}$ are the simultaneously given right-hand sides and $X = [x^{(1)}, x^{(2)}, \ldots, x^{(p)}] \in \mathbb{C}^{n\times p}$ are the solutions to be computed.

2.1 Block GCRO

The background of GCRO [9], a nested GMRES [33, 34] method based on GCR, is briefly reviewed first in the case of a single right-hand side and then extended to the block case. As described in [9], the

GCRO method relies on a given full-rank matrix $U_k \in \mathbb{C}^{n\times k}$, and a matrix C_k as the image of U_k with respect to the general matrix A satisfying the relations

$$AU_k = C_k, \tag{3}$$
$$C_k^H C_k = I_k. \tag{4}$$

For the solution of a single right-hand side linear system $Ax = b$ and a given initial guess x_0, the governing idea is to first define $x_1 \in x_0 + \mathrm{Range}(U_k)$ that minimizes the residual norm. From x_1 and its associated residual r_1, Arnoldi iterations are performed to enlarge the nested orthonormal basis of the residual spaces. The vector

$$x_1 = \underset{x\in x_0+\mathrm{Range}(U_k)}{\mathrm{argmin}} ||b - Ax||_2,$$

is defined by

$$x_1 = x_0 + U_k C_k^H r_0, \quad \text{and} \quad r_1 = (I - C_k C_k^H) r_0 \text{ so that } r_1 \in C_k^{\perp}.$$

Starting from the unitary vector $v_1 = r_1/\|r_1\|_2$, the Arnoldi procedure enables us to form an orthonormal basis $V_m = [v_1, ..., v_m]$ of the Krylov space $\mathcal{K}_m((I - C_k C_k^H)A, v_1) = \mathrm{span}(v_1, (I - C_k C_k^H)Av_1, ..., ((I - C_k C_k^H)A)^{m-1}v_1)$ that can be written in the matrix form as

$$(I - C_k C_k^H)AV_m = V_{m+1}\underline{H}_m. \tag{5}$$

Combining (3) and (5) in a matrix form allows us to write a relation very similar to an Arnoldi equality that reads

$$A\widehat{W}_m = \widehat{V}_{m+1}\underline{G}_m, \tag{6}$$

where $\widehat{W}_m = [U_k, V_m]$, $\widehat{V}_{m+1} = [\, C_k, V_{m+1}]$ and $\underline{G}_m = \begin{bmatrix} I_k & B_m \\ 0_{(m+1)\times k} & \underline{H}_m \end{bmatrix}$ with $\widehat{V}_{m+1}^H \widehat{V}_{m+1} = I_{m+1}$ and $B_m = C_k^H A V_m$. The minimum residual norm solution in the affine space $x_1 + \mathrm{Range}(\widehat{W}_m)$ can be written as $x_m = x_1 + \widehat{W}_m y_m$ where

$$y_m = \underset{y\in\mathbb{C}^{m+k}}{\mathrm{argmin}} \|c - \underline{G}_m y\|$$

and $c = \widehat{V}_{m+1}^H r_1 = (0_k, \|r_1\|_2, 0_{m-k})^T$ are the components of the residual associated with x_1 in the residual space $\widehat{V}_{m+1}$.

GCRO and GMRES, both belong to the family of residual norm minimization approach and rely on an orthonormal basis of the residual space. In addition to sharing the Arnoldi procedure to form part or all this basis, they do also share the property of "happy breakdown"; that is, if the search space cannot be enlarged because the new direction computed by the Arnoldi process is the null vector, then the solution is exactly found in the search space. This sharing of features does extend to the block context for the solution of linear system with multiple right-hand sides; in particular the inexact breakdown principle introduced in [32] in the context of block GMRES can be accommodated to block GCRO as discussed in the sequel. The purpose of the inexact breakdown mechanism is to handle in an elegant and effective way the loss of numerical rank of the search space basis, that turns out to be also a way to monitor the search space extension according to the final target accuracy.

The straightforward extension of the GCRO method in the block context can be briefly described. For ease of reading, we change the calligraphy of the notation but keep the same letters to denote the block counterpart of the quantities involved in the method. Starting from the block initial guess $X_0 = [x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(p)}] \in \mathbb{C}^{n\times p}$ and associated initial residual $R_0 = B - AX_0$ one can define

$$X_1 = \underset{X\in X_0+\mathrm{Range}(U_k)}{\mathrm{argmin}} ||B - AX||_F,$$

given by

$$X_1 = X_0 + U_k C_k^H R_0, \quad \text{and} \quad R_1 = (I - C_k C_k^H) R_0 \text{ so that } R_1 \in C_k^{\perp}. \tag{7}$$

For the sake of simplicity of exposure, we first assume that R_1 is of full rank and denote $R_1 = \mathbb{V}_1 \Lambda_1$ as its reduced QR-factorization. The orthonormal block $\mathbb{V}_1$ is then used to build the search space via m steps of block Arnoldi procedure depicted in Algorithm 1 to generate $\mathscr{V}_m = [\mathbb{V}_1, ..., \mathbb{V}_m]$ whose columns form an orthonormal basis of $\mathcal{K}_m((I - C_k C_k^H)A, \mathbb{V}_1) = \bigoplus_{t=1}^{p} \mathcal{K}_m((I - C_k C_k^H)A, v_1^{(t)})$. The block Arnoldi procedure leads to the matrix equality

$$(I - C_k C_k^H) A \mathscr{V}_m = \mathscr{V}_{m+1} \underline{\mathscr{H}}_m \tag{8}$$

where $\underline{\mathscr{H}}_m$ is a block Hessenberg matrix with (i, j) block defined by $H_{i,j}$. Similarly to the single

Algorithm 1 *Block Arnoldi procedure with deflation of the C_k space*

1: Given a nonsingular coefficient matrix $A \in \mathbb{C}^{n \times n}$, choose a unitary matrix $\mathbb{V}_1$ of size $n \times p$
2: **for** $j = 1, 2, \ldots, m$ **do**
3: Compute $\mathbb{W}_j = (I - C_k C_k^H) A \mathbb{V}_j$
4: **for** $i = 1, 2, \ldots, j$ **do**
5: $H_{i,j} = \mathbb{V}_i^H \mathbb{W}_j$
6: $\mathbb{W}_j = \mathbb{W}_j - \mathbb{V}_i H_{i,j}$
7: **end for**
8: $\mathbb{W}_j = \mathbb{V}_{j+1} H_{j+1,j}$ (reduced QR-factorization)
9: **end for**

right-hand side case, (3) and (8) can be gathered in a matrix form

$$A \widehat{\mathscr{W}_m} = \widehat{\mathscr{V}_{m+1}} \underline{\mathcal{G}}_m, \tag{9}$$

where $\widehat{\mathscr{W}_m} = [U_k, \mathscr{V}_m]$, $\widehat{\mathscr{V}_{m+1}} = [C_k, \mathscr{V}_{m+1}]$ and $\underline{\mathcal{G}}_m = \begin{bmatrix} I_k & \mathcal{B}_m \\ 0_{(m+1)p \times k} & \underline{\mathscr{H}}_m \end{bmatrix}$ with $\widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{V}}_{m+1} = I_{(m+1)\times p}$ and $\mathcal{B}_m = C_k^H A \mathscr{V}_m \in \mathbb{C}^{k \times mp}$ with $mp = m \times p$. The minimum residual norm solution in the affine space $X_1 + \text{Range}(\widehat{W}_m)$ can be written as $X_m = X_1 + \widehat{\mathscr{W}_m} Y_m$ where

$$Y_m = \underset{Y \in \mathbb{C}^{(mp+k)\times p}}{\text{argmin}} \|\mathcal{C} - \underline{\mathcal{G}}_m Y\|_F,$$

and $\mathcal{C} = \mathscr{V}_{m+1}^H R_1 = (0_{k\times p}, \Lambda_1^T, 0_{mp\times p})^T$, the columns of $\mathcal{C}$ are the components of the residual R_1 in the residual space $\mathscr{V}_{m+1}$.

2.2 Block GCRO with inexact breakdowns

When one solution or a linear combination of the solutions has converged to the target accuracy, the block-Arnoldi procedure implemented to build an orthonormal basis of $\mathcal{K}_j(A, R_0)$ needs to be modified to account for this partial convergence. This partial convergence is characterized by a numerical rank deficiency in the new p directions one attempts to introduce for enlarging the Krylov space. In [32], the authors present an elegant numerical variant that enables the detection of what is referred to as inexact breakdowns. In that approach the directions that have a low contribution to the residual block are discarded from the set of vectors used to expand the search space at the next iteration but these directions are kept and reintroduced in iterations afterwards if necessary. In this section, we try to give an insight

and the main equality required to derive the IB-BGMRES algorithm. We refer the reader to the original paper [32] for a detailed and complete description. For the sake of simplicity of exposure and easy cross-reading, we adopt most of the notations from [32].

Because when an inexact breakdown occurs, not all the space spanned by $\mathbb{W}_j$ is considered to build $\mathbb{V}_{j+1}$ in order to expand the space, a subscript j is added to denote its block number of columns. Let $p_1 = p$ and denote by p_{j+1} the column rank of the block orthonormal basis vector $\mathbb{V}_{j+1}$. Then $\mathbb{V}_{j+1} \in \mathbb{C}^{n\times p_{j+1}}, \mathbb{W}_j \in \mathbb{C}^{n\times p_j}$ and $H_{j+1,j} \in \mathbb{C}^{p_{j+1}\times p_j}$. As a consequence the dimension of the search space $\mathcal{K}_j(A, \mathbb{V}_1)$ considered at the j^{th} iteration is no longer necessarily equal to $j \times p$ but is equal to $n_j = \sum_{i=1}^{j} p_i$; that is, the sum of the column ranks of $\mathbb{V}_i's$ $(i = 1, \ldots, j)$. $\mathscr{V}_j = [C_k, \mathbb{V}_1, \ldots, \mathbb{V}_j] \in \mathbb{C}^{n\times(n_j+k)}$ $(\mathscr{V}_{j+1} = [\mathscr{V}_j, \quad \mathbb{V}_{j+1}])$ denotes the orthonormal basis of the residual space.

When no inexact breakdown has occurred $p_{j+1} = p_j = \ldots = p_1 = p$, the range of $\mathbb{W}_j$ has always been used to enlarge the search space and we obtain the block relation given by (9). To account for a numerical deficiency in the residual block $R_j = B - AX_j$ in a way that is described later, Robbé and Sadkane [32] proposed to split

$$\mathbb{W}_j = \mathbb{V}_{j+1} H_{j+1,j} + Q_j \tag{10}$$

so that the columns of Q_j and $\mathbb{V}_{j+1}$ are orthogonal to each other and only $\mathbb{V}_{j+1}$ is used to enlarge $\mathscr{V}_j$ to form $\mathscr{V}_{j+1}$. We can then extend Equation (9) into

$$A\widehat{\mathscr{W}_j} = \widehat{\mathscr{V}_j}\underline{\mathcal{G}}_j + [\mathcal{Q}_{j-1}, \quad \mathbb{W}_j], \tag{11}$$

where $\mathcal{Q}_{j-1} = [Q_1, \ldots, Q_{j-1}] \in \mathbb{C}^{n\times n_{j-1}}$ accounts for all the abandoned directions. The matrix $\mathcal{Q}_{j-1}$ is rank deficient, and it reduces to the zero matrix of $\mathbb{C}^{n\times n_{j-1}}$ as long as no inexact breakdown has occurred.

In order to characterize a minimum norm solution in the space spanned by $\widehat{\mathscr{W}_j}$ using Equation (11) we need to form an orthonormal basis of the space spanned by $[\widehat{\mathscr{V}_j}, \mathcal{Q}_{j-1}, \mathbb{W}_j]$. This is performed by first orthogonalizing $\mathcal{Q}_{j-1}$ against $\widehat{\mathscr{V}_j}$, that is $\widetilde{\mathcal{Q}}_{j-1} = (I - \widehat{\mathscr{V}_j}\widehat{\mathscr{V}_j}^H)\mathcal{Q}_{j-1}$. Because $\mathcal{Q}_{j-1}$ is of low rank so is $\widetilde{\mathcal{Q}}_{j-1}$ that can be written

$$\widetilde{\mathcal{Q}}_{j-1} = P_{j-1}G_{j-1} \text{ with } \left\{ \begin{array}{l} P_{j-1} \in \mathbb{C}^{n\times q_j} \text{ has orthonormal columns with } \widehat{\mathscr{V}_j}^H P_{j-1} = 0, \\ G_{j-1} \in \mathbb{C}^{q_j\times n_{j-1}} \text{ is of full rank with } q_j = p - p_j. \end{array} \right. \tag{12}$$

Next $\mathbb{W}_j$, that is already orthogonal to $\widehat{\mathscr{V}_j}$, is made to be orthogonal to P_{j-1} with $\mathbb{W}_j - P_{j-1}E_j$ where $E_j = P_{j-1}^H \mathbb{W}_j$; then one computes $\widetilde{W}_j D_j$ with $\widetilde{W}_j \in \mathbb{C}^{n\times p_j}$ and $D_j \in \mathbb{C}^{p_j\times p_j}$ the reduced QR-factorization of $\mathbb{W}_j - P_{j-1}E_j$. Eventually, the columns of the matrix $[\widehat{\mathscr{V}_j}, P_{j-1}, \widetilde{W}_j]$ form an orthonormal basis of the space spanned by $[\widehat{\mathscr{V}_j}, \mathcal{Q}_{j-1}, \mathbb{W}_j]$.

With this new basis Equation (11) writes

$$\begin{aligned} A[U_k, \mathscr{V}_j] = \quad & [C_k, \mathscr{V}_j] \begin{bmatrix} I & \mathcal{B}_j \\ 0 & \mathscr{L}_j \end{bmatrix} + \left[0_k, P_{j-1}G_{j-1}, \left[P_{j-1}, \widetilde{W}_j\right] \begin{bmatrix} E_j \\ D_j \end{bmatrix}\right] \\ = \quad & \left[C_k, \mathscr{V}_j, P_{j-1}, \widetilde{W}_j\right] \begin{bmatrix} I_k & & \mathcal{B}_j \\ & & \mathscr{L}_j \\ 0 & G_{j-1} & E_j \\ & 0 & D_j \end{bmatrix}, \end{aligned} \tag{13}$$

where $\mathscr{L}_j = \begin{bmatrix} H_{1,1} & \cdots & \cdots & \cdots & H_{1,j} \\ H_{2,1} & \ddots & & & \vdots \\ \mathbb{V}_3^H Q_1 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbb{V}_j^H Q_1 & \cdots & \mathbb{V}_j^H Q_{j-2} & H_{j,j-1} & H_{j,j} \end{bmatrix} \in \mathbb{C}^{n_j \times n_j}$ is no longer upper Hessenberg as soon as one inexact breakdown occurs, i.e., $\exists \ell,\ Q_\ell \neq 0$.

Equation (13) can be written in a more compact form as

$$A\widehat{\mathscr{W}_j} = \left[C_k, \mathscr{V}_j, [P_{j-1}, \widetilde{W}_j]\right] \underline{\mathscr{F}}_j, \tag{14}$$

so that the least-squares problem to be solved to compute the minimum norm solution associated with the generalized Arnoldi relation (13) becomes

$$Y_j = \underset{Y \in \mathbb{C}^{(k+n_j)\times p}}{\operatorname{argmin}} \left\| \Lambda_j - \underline{\mathscr{F}}_j Y \right\|_F, \tag{15}$$

with

$$\underline{\mathscr{F}}_j = \begin{bmatrix} I_k & & \mathcal{B}_j \\ & & \mathscr{L}_j \\ 0_{(n_j+p)\times k} & G_{j-1} & E_j \\ & 0 & D_j \end{bmatrix} = \begin{bmatrix} \mathscr{F}_j \\ \mathbb{H}_j \end{bmatrix} \in \mathbb{C}^{(k+n_j+p)\times(k+n_j)} \tag{16}$$

and $\Lambda_j = \begin{bmatrix} 0_{k\times p} \\ \Lambda_1 \\ 0_{n_j \times p} \end{bmatrix}$, where $\mathscr{F}_j = \begin{bmatrix} I_k & \mathcal{B}_j \\ 0_{(n_j+p)\times k} & \mathscr{L}_j \end{bmatrix} \in \mathbb{C}^{(k+n_j)\times(k+n_j)}$ and $\mathbb{H}_j = \begin{bmatrix} 0_{p\times k} & G_{j-1} & E_j \\ & 0 & D_j \end{bmatrix} \in \mathbb{C}^{p\times(k+n_j)}$.

The numerical mechanism to select $\mathbb{V}_{j+1}$ out of $\left[P_{j-1}, \widetilde{W}_j\right]$ follows the same ideas as discussed in [2, 32] in the context of block GMRES. The governing idea consists in building an orthonormal basis for the directions that contribute the most to the individual residual norms and make them larger than the target threshold $\epsilon^{(R)}$. Based on the SVD of the coordinate vector of the residual

$$\Lambda_j - \underline{\mathscr{F}}_j Y_j = \mathbb{U}_{1,L} \Sigma_1 \mathbb{U}_{1,R}^H + \mathbb{U}_{2,L} \Sigma_2 \mathbb{U}_{2,R}^H, \tag{17}$$

where Σ_1 contains the p_{j+1} singular values larger than the prescribed threshold $\epsilon^{(R)}$, they decompose $\mathbb{U}_{1,L} = \begin{pmatrix} \mathbb{U}_1^{(1)} \\ \mathbb{U}_1^{(2)} \end{pmatrix}$ in accordance with $\left[[C_k, \mathscr{V}_j], [P_{j-1}, \widetilde{W}_j]\right]$, that is $\mathbb{U}_1^{(1)} \in \mathbb{C}^{(k+n_j)\times p}$ and $\mathbb{U}_1^{(2)} \in \mathbb{C}^{p\times p}$. Because, the objective is to construct orthonormal basis we consider $[\mathbb{W}_1, \mathbb{W}_2]$ unitary so that $\mathrm{Range}(\mathbb{W}_1) = \mathrm{Range}(\mathbb{U}_1^{(2)})$. The new set of orthonormal candidate vectors to expand the search space

$$\mathbb{V}_{j+1} = \left[P_{j-1}, \widetilde{W}_j\right] \mathbb{W}_1 \tag{18}$$

is the one that contributes the most to the residual norms while

$$P_j = \left[P_{j-1}, \widetilde{W}_j\right] \mathbb{W}_2,$$

is the new set of orthogonal abandoned directions. Through this mechanism, directions that have been abandoned at a given iteration can be reintroduced, if the residual block has a large component along them. Furthermore, this selection strategy ensures that all the solutions have converged when p inexact

breakdowns have been detected; we refer to Section 3.2 for the discussion on how $\epsilon^{(R)}$ should be defined to ensure a convergence of all the solutions to a common or individual prescribed backward error. We do not give the details of the calculation and refer to [32] for a complete description, but only state that via this decomposition the main terms that appear in Equation (13) can be computed incrementally.

2.3 Subspace recycling policies

So far, we have not made any specific assumption on the definition of the deflation space U_k except that it has full column rank. In the context of subspace recycling, one key point is to specify what subspace is to be recycled at restart. At the cost of the extra storage of k vectors, block GCRO offers more flexibility than block GMRES in the choice of the recycled space. This extra storage, that enables us to remove the constraints to have the search space included in the residual space, allows us to consider any subspace to be deflated at restart. In particular any of the two classical alternatives, that are Rayleigh-Ritz (RR) or Harmonic-Ritz (HR) approaches, can be considered to compute approximated eigenvectors to define U_k and C_k at restart.

Definition 1. *Harmonic Ritz projection.*
Consider a subspace $\mathcal{W}$ of $\mathbb{C}^n$. Given a general nonsingular matrix $B \in \mathbb{C}^{n\times n}$, $\lambda \in \mathbb{C}$ and $y \in \mathcal{W}$, (λ, y) is a harmonic Ritz pair of B with respect to the space $\mathcal{W}$ if and only if

$$By - \lambda\, y \perp B\,\mathcal{W}$$

or equivalently,

$$\forall w \in \mathrm{Range}(B\,\mathcal{W}) \quad w^H\,(By - \lambda\, y) = 0.$$

The vector y is a harmonic Ritz vector associated with the harmonic Ritz value λ.

Definition 2. *Rayleigh-Ritz projection.*
Consider a subspace $\mathcal{W}$ of $\mathbb{C}^n$. Given a general nonsingular matrix $B \in \mathbb{C}^{n\times n}$, $\lambda \in \mathbb{C}$ and $y \in \mathcal{W}$, (λ, y) is a Rayleigh-Ritz pair of B with respect to the space $\mathcal{W}$ if and only if

$$By - \lambda\, y \perp \mathcal{W}$$

or equivalently,

$$\forall w \in \mathrm{Range}(\mathcal{W}) \quad w^H\,(By - \lambda\, y) = 0.$$

The vector y is a Rayleigh-Ritz vector associated with the Rayleigh Ritz value λ.

Once the maximum size of the space has been reached, we have

$$\begin{aligned}
A\widehat{\mathscr{W}}_m &= \widehat{\mathscr{V}}_{m+1}\underline{\mathscr{F}}_m = \left[C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]\underline{\mathscr{F}}_m, && (19)\\
X_m &= X_1 + \widehat{\mathscr{W}}_m Y_m,\\
R_m &= B - AX_m = \left[C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right](\Lambda_m - \underline{\mathscr{F}}_m Y_m), && (20)\\
Y_m &= \underset{Y\in\mathbb{C}^{(k+n_m)\times p}}{\mathrm{argmin}} \|\Lambda_m - \underline{\mathscr{F}}_m Y\|_F\,, \quad \Lambda_m = [0_{p\times k}, \Lambda_1^T, 0_{p\times n_m}]^T, && (21)
\end{aligned}$$

a restart procedure has to be implemented to possibly refine the spectral information to be recycled during the next cycle. Based on this equality we have to compute the approximated eigen-information as shown in Proposition 1 and then use it to define U_k^{new} and C_k^{new} as described in Theorem 1.

Proposition 1. *At the restart of IB-BGCRO-DR, the update of the deflated subspace for the next cycle relies on the computation of harmonic Ritz vectors $g_i^{(HR)} \in \mathrm{span}(\widehat{\mathscr{W}}_m)$, or Rayleigh Ritz vectors $g_i^{(RR)} \in \mathrm{span}(\widehat{\mathscr{W}}_m)$, of A with respect to $\widehat{\mathscr{W}}_m = [U_k \quad \mathscr{V}_m] \in \mathbb{C}^{n\times(k+n_m)}$.*

- *The harmonic Ritz pairs* $(\theta_i, \widehat{\mathscr{W}}_m g_i^{(HR)})$ *to be possibly used for the next restart satisfy*

$$\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_i^{(HR)} = \theta_j \underline{\mathscr{F}}_m^H \widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{W}}_m g_i^{(HR)}, \quad \text{for } 1 \le i \le n_m, \tag{22}$$

where $\widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{W}}_m = \begin{bmatrix} C_k^H U_k & 0_{k\times n_m} \\ \mathscr{V}_m^H U_k & I_{n_m} \\ P_{m-1}^H U_k & 0_{p\times n_m} \\ \widetilde{W}_m^H U_k & \end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times(k+n_m)}$.

- *The Rayleigh Ritz pairs* $(\theta_i, \widehat{\mathscr{W}}_m g_i^{(RR)})$ *to be possibly used for the next restart satisfy*

$$\widehat{\mathscr{W}}_m^H \widehat{\mathscr{V}}_{m+1} \underline{\mathscr{F}}_m g_i^{(RR)} = \theta_j \widehat{\mathscr{W}}_m^H \widehat{\mathscr{W}}_m g_i^{(RR)}, \quad \text{for } 1 \le j \le n_m \tag{23}$$

where $\widehat{\mathscr{W}}_m^H \widehat{\mathscr{V}}_{m+1} = \begin{bmatrix} U_k^H C_k & U_k^H \mathscr{V}_m & U_k^H P_{m-1} & U_k^H \widetilde{W}_m \\ 0_{n_m\times k} & I_{n_m} & 0_{n_m\times p} & \end{bmatrix} \in \mathbb{C}^{(k+n_m)\times(k+n_m+p)}$ *and* $\widehat{\mathscr{W}}_m^H \widehat{\mathscr{W}}_m = \begin{bmatrix} U_k^H U_k & U_k^H \mathscr{V}_m \\ \mathscr{V}_m^H U_k & I_{n_m} \end{bmatrix} \in \mathbb{C}^{(k+n_m)\times(k+n_m)}$.

Proof. The proofs basically rely on some matrix computations as shortly described below:

- According to Definition 1, each harmonic Ritz pair $(\theta_i, \widehat{\mathscr{W}}_m g_i^{(HR)})$ satisfies

$$\forall w \in \mathrm{Range}(A\widehat{\mathscr{W}}_m) \quad w^H \left(A\widehat{\mathscr{W}}_m g_i^{(HR)} - \theta_i \widehat{\mathscr{W}}_m g_i^{(HR)}\right) = 0,$$

  which is equivalent to

$$(A\widehat{\mathscr{W}}_m)^H \left(A\widehat{\mathscr{W}}_m g_i^{(HR)} - \theta_i \widehat{\mathscr{W}}_m g_i^{(HR)}\right) = 0. \tag{24}$$

  Using Equation (19) leads to

$$\left(\widehat{\mathscr{V}}_{m+1}\underline{\mathscr{F}}_m\right)^H \left(\widehat{\mathscr{V}}_{m+1}\underline{\mathscr{F}}_m g_i^{(HR)} - \theta_i \widehat{\mathscr{W}}_m g_i^{(HR)}\right) = 0. \tag{25}$$

  Because $\widehat{\mathscr{V}}_{m+1} = [C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]]$ generated at the end of each cycle is orthonormal, (25) becomes

$$\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_i^{(HR)} - \theta_i \underline{\mathscr{F}}_m^H \widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{W}}_m g_i^{(HR)} = 0,$$

  which is the same as formulation (22).

- Rayleigh Ritz pairs: using Definition 2 and similar arguments and matrix computation enable to derive the proof.

□

Depending on the region of the spectrum that is intended to be deflated (e.g., subspace associated with the smallest, largest eigenvalues in magnitude), a subset of k approximated eigenvectors is chosen among the n_m ones to define the space that will span U_k^{new}. Then, we describe in Theorem 1 the update of U_k^{new} and its image C_k^{new} with respect to A at restart of IB-BGCRO-DR.

Theorem 1. *At restart of the IB-BGCRO-DR, if we intend to deflate the space* $\mathrm{span}([U_k, \mathscr{V}_m]G_k^{(*)})$ *where* $G_k^{(*)} = \left[g_{i_1}^{(*)}, ..., g_{i_k}^{(*)}\right]$ *the set of vectors associated with the targeted eigenvalues, the matrices* U_k^{new} *and* C_k^{new} *to be used for the next cycle are defined by*

$$U_k^{new} = \widehat{\mathscr{W}}_m G_k^{(*)} R^{-1} = [U_k, \mathscr{V}_m] G_k^{(*)} R^{-1}, \tag{26}$$

$$C_k^{new} = \widehat{\mathscr{V}}_{m+1} Q = [C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m] Q, \tag{27}$$

where Q and R are the factors of the reduced QR-factorization of $\underline{\mathscr{F}}_m G_k^{(*)}$, which ensure that $AU_k^{new} = C_k^{new}$ and $(C_k^{new})^H C_k^{new} = I_k$ with $G_k^{(*)} = G_k^{(HR)}$ or $G_k^{(*)} = G_k^{(RR)}$.

Proof. Let $[Q, R]$ be the reduced QR-factorization of $\underline{\mathscr{F}}_m G_k^{(*)}$ and multiply by $G_k^{(*)}$ on the right both sides of Equation (19). It leads to $A\widehat{\mathscr{W}}_m G_k^{(*)} = \widehat{\mathscr{V}}_{m+1}\underline{\mathscr{F}}_m G_k^{(*)} = \widehat{\mathscr{V}}_{m+1}QR$, that is equivalent to $A\widehat{\mathscr{W}}_m G_k^{(*)} R^{-1} = \widehat{\mathscr{V}}_{m+1}\underline{\mathscr{F}}_m G_k^{(*)} R^{-1} = \widehat{\mathscr{V}}_{m+1}Q$ concluding the proof as $\mathrm{span}(\widehat{\mathscr{W}}_m G_k^{(*)} R^{-1}) = \mathrm{span}(\widehat{\mathscr{W}}_m G_k^{(*)})$ and $\widehat{\mathscr{V}}_{m+1}Q$ is the product of two matrices with orthonormal columns so are its columns. □

Corollary 1. *The residual at restart $R_1^{new} = R_m^{old} = B - AX_1^{new}$ with $X_1^{new} = X_m^{old}$ is orthogonal to C_k^{new}.*

Proof. $X_m^{old} = X_1 + \widehat{\mathscr{W}}_m Y_m$ where Y_m solve the least-squares problem (21) so that $(\Lambda_m - \underline{\mathscr{F}}_m Y_m) \in (\mathrm{Range}(\underline{\mathscr{F}}_m))^{\perp} = \mathrm{Null}(\underline{\mathscr{F}}_m^H)$. We also have $R_m^{old} = \widehat{\mathscr{V}}_{m+1}(\Lambda_m - \underline{\mathscr{F}}_m Y_m)$, consequently

$$\begin{aligned}
(C_k^{new})^H R_m^{old} &= \left(\widehat{\mathscr{V}}_{m+1}Q\right)^H \left(\widehat{\mathscr{V}}_{m+1}(\Lambda_m - \underline{\mathscr{F}}_m Y_m)\right) \\
&= \left(\widehat{\mathscr{V}}_{m+1}\underline{\mathscr{F}}_m G_k^{(*)} R^{-1}\right)^H \left(\widehat{\mathscr{V}}_{m+1}(\Lambda_m - \underline{\mathscr{F}}_m Y_m)\right) \\
&= R^{-H} G_k^{(*)\,H} \underbrace{\underline{\mathscr{F}}_m^H (\Lambda_m - \underline{\mathscr{F}}_m Y_m)}_{=0 \quad \text{because (21)}} = 0.
\end{aligned}$$

□

Next, we discuss the relationship between residuals of harmonic Ritz pairs and the linear system residuals at restart, that can be exploited to lower the cost of the calculation of $G_k^{(HR)}$ at the end of the next cycle. The residuals of the harmonic Ritz pairs can be formulated in a matrix form as

$$R_m^{(HR)} = A\widehat{\mathscr{W}}_m G_k^{(HR)} - \widehat{\mathscr{W}}_m G_k^{(HR)} \mathrm{diag}(\theta_1, \cdots, \theta_k). \tag{28}$$

The linear system residuals $R_m \in \mathbb{C}^{n\times p}$ and $R_m^{(HR)} \in \mathbb{C}^{n\times k}$ belong to the same subspace $\mathrm{Range}([C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m]) \in \mathbb{C}^{n\times(k+n_m+p)}$; because of the minimum residual norm property R_m is orthogonal to $\mathrm{Range}(A\widehat{\mathscr{W}}_m) \in \mathbb{C}^{n\times(k+n_m)}$, so is $R_m^{(HR)}$ by definition. Therefore the linear system residuals R_m and the residuals of harmonic Ritz vectors $R_m^{(HR)}$ are in $\mathrm{Range}(A\widehat{\mathscr{W}}_m)^{\perp} \cap \mathrm{Range}([C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m])$, that is a p-dimensional space since $\mathrm{Range}(A\widehat{\mathscr{W}}_m) \subset \mathrm{Range}([C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m])$. This means that it exists a matrix $\beta_{p\times k} \in \mathbb{C}^{p\times k}$ such that $R_m^{(HR)} = R_m \beta_{p\times k}$, that combines with (28) leads to:

$$A\widehat{\mathscr{W}}_m G_k^{(HR)} = \widehat{\mathscr{W}}_m G_k^{(HR)} \mathrm{diag}(\theta_1, \cdots, \theta_k) + R_m \beta_{p\times k}. \tag{29}$$

Proposition 2. *At the restart of IB-BGCRO-DR, if harmonic Ritz pairs are used to define recycling space, then the following relations hold:*

$$\mathrm{Range}(U_k^{new}) \quad \subset \quad \mathrm{Range}([C_k^{new}, \mathbb{V}_1]), \tag{30}$$

where $\mathbb{V}_1\,\Lambda_1$ is the reduced QR-factorization of $R_1^{new} = R_m^{old} = \mathbb{V}_1\Lambda_1$.

Proof. Substituting (26) into $C_k^{new} = AU_k^{new}$ and from (29), we have

$$\begin{aligned}
C_k^{new} &= A\widehat{\mathscr{W}}_m G_k^{(HR)} R^{-1}, \\
&= \widehat{\mathscr{W}}_m G_k^{(HR)} \mathrm{diag}(\theta_1, \cdots, \theta_k) R^{-1} + R_m \beta_{p\times k} R^{-1}, \\
&= \widehat{\mathscr{W}}_m G_k^{(HR)} \mathrm{diag}(\theta_1, \cdots, \theta_k) R^{-1} + \mathbb{V}_1 \Lambda_1 \beta_{p\times k} R^{-1},
\end{aligned}$$

so that

$$[C_k^{new}\,,\,\mathbb{V}_1] \;=\; [\widehat{\mathscr{W}_m}G_k^{(HR)}\,,\,\mathbb{V}_1]\begin{bmatrix}\mathrm{diag}(\theta_1,\cdots,\theta_k)R^{-1} & 0_{k\times p}\\ \Lambda_1\beta_{p\times k}R^{-1} & I_p\end{bmatrix}. \tag{31}$$

By (26) we also have: $U_k^{new}=\widehat{\mathscr{W}_m}G_k^{(HR)}R^{-1}$. That shows that

$$\mathrm{Range}(U_k^{new})\subset\mathrm{Range}([\widehat{\mathscr{W}_m}G_k^{(HR)}\,,\,\mathbb{V}_1])=\mathrm{Range}([C_k^{new}\,,\,\mathbb{V}_1]).$$

□

Corollary 2. *For all the restarts but the first, if harmonic Ritz vectors are used to define the deflated space, the generalized eigenvalue problem to be solved simplifies and reads:*

$$\underline{\mathscr{F}}_m^H\underline{\mathscr{F}}_m g_i^{(HR)}=\theta_i\,\mathscr{F}_m^H\mathscr{T}_m g_i^{(HR)}\quad \textit{for } 1\le j\le n_m, \tag{32}$$

where $\mathscr{T}_m=\widehat{\mathscr{V}}_{m+1}^H\widehat{\mathscr{W}_m}(1:k+n_m,:)=\begin{bmatrix} C_k^HU_k & 0_{k\times n_m}\\ \mathbb{V}_1^HU_k & \begin{bmatrix} I_{p_1} & 0_{p_1\times(n_m-p_1)}\end{bmatrix}\\ 0_{(n_m-p_1)\times k} & \begin{bmatrix} 0_{(n_m-p_1)\times p_1} & I_{n_m-p_1}\end{bmatrix}\end{bmatrix}\in\mathbb{C}^{(k+n_m)\times(k+n_m)}$.

Proof. The simplified right-hand side of (32) is deduced from the (2×1) block structure of $\underline{\mathscr{F}}$ in Equation (16) and the equality $\widehat{\mathscr{V}}_{m+1}^H\widehat{\mathscr{W}_m}=[\mathscr{T}_m\quad O_p]^T$ that comes from $\mathrm{Range}(U_k^{new})\subset\mathrm{Range}([C_k^{new}\,,\,\mathbb{V}_1])\subset\mathrm{Range}(\widehat{\mathscr{V}}_{m+1})=\mathrm{Range}([C_k^{new}\,,\,\mathbb{V}_1],\mathbb{V}_2,\ldots,\mathbb{V}_m,P_{m-1},\widetilde{W}_m])$, $\forall i\in\{2,..,m\}\quad \mathbb{V}_i^HU_k=0_{p_i\times k}$ and $([P_{m-1},\widetilde{W}_m])^HU_k=0_{p\times k}$. □

2.4 A variant suited for flexible preconditioning

All what has been described in the previous sections, naturally extends to the preconditioned case where, for right preconditioning with a fixed preconditioner M, the central equality writes

$$A[U_k,M\mathscr{V}_m]=\left[C_k,\mathscr{V}_m,[P_{m-1},\widetilde{W}_m]\right]\underline{\mathscr{F}}_m, \tag{33}$$

the least-squares problem to be solved to compute the minimum norm solution becomes

$$Y_m=\underset{Y\in\mathbb{C}^{(k+n_m)\times p}}{\mathrm{argmin}}\;\|\Lambda_m-\underline{\mathscr{F}}_mY\|_F\,,$$

and the solution is

$$X_m=X_1+[U_k,M\mathscr{V}_m]Y_m.$$

If we denote $\mathscr{M}_j$ a (possibly nonlinear) nonsingular preconditioning operator at iteration j and $\mathscr{M}_j(\mathbb{V}_j)$ denotes the action of $\mathscr{M}_j$ on a block vector $\mathbb{V}_j$, Equation (33) translates to

$$A[U_k,\mathscr{Z}_m]=\left[C_k,\mathscr{V}_m,[P_{m-1},\widetilde{W}_m]\right]\underline{\mathscr{F}}_m\ \text{with}\ \mathscr{Z}_m=[\mathscr{M}_1(\mathbb{V}_1),...,\mathscr{M}_m(\mathbb{V}_m)]\,,$$

that writes in a more compact form

$$A\widehat{\mathscr{Z}_m}=\widehat{\mathscr{V}}_{m+1}\underline{\mathscr{F}}_m\ \text{with}\ \widehat{\mathscr{Z}_m}=[U_k,\mathscr{Z}_m]\ \text{and}\ \widehat{\mathscr{V}}_{m+1}=\left[C_k,\mathscr{V}_m,[P_{m-1},\widetilde{W}_m]\right]. \tag{34}$$

The solution update is $X_m=X_1+[U_k,\mathscr{Z}_m]Y_m$. To keep the notation simple, we choose to keep the notation for quantities that have the same meaning as in the non-flexible case but of course will have different values.

In the context of flexible preconditioning many strategies for defining harmonic Ritz vectors can be envisioned for GCRO-DR. Among those considered in [5], we follow the one with a lower computational cost required in solving the generalized eigenvalue problem, referred to as Strategy C in [5]. Furthermore, it also allows us to obtain very similar properties in the flexible case to the ones we have exposed in the non-preconditioned case shown in Section 2.3. We refer to Appendix A for two other strategies for approximating target eigen-information. The proposition below indicates that with an appropriated definition of the harmonic Ritz vectors, all the properties of IB-BGCRO-DR extend to the flexible situation.

Proposition 3. *At the end of a cycle of the IB-BFGCRO-DR algorithm, if the deflation space is built on the harmonic Ritz vectors* $g_i^{(HR)} \in \mathrm{span}(\mathcal{W}_m)$ *of* $A\widehat{\mathscr{Z}_m}\mathcal{W}_m^{\dagger}$ *with respect to* $\mathcal{W}_m = [\mathcal{W}_k \quad \mathscr{V}_m] \in \mathbb{C}^{n\times(k+n_m)}$:

1. *The harmonic Ritz pairs* $(\theta_i, \mathcal{W}_m g_i^{(HR)})$
   - *for the first restart satisfy*
   $$\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_i^{(HR)} = \theta_j \underline{\mathscr{F}}_m^H \widehat{\mathscr{V}}_{m+1}^H \mathcal{W}_m g_i^{(HR)}, \quad \text{for } 1 \le i \le n_m, \tag{35}$$
   where $\widehat{\mathscr{V}}_{m+1}^H \mathcal{W}_m = \begin{bmatrix} C_k^H \mathcal{W}_k & 0_{k\times n_m} \\ \mathscr{V}_m^H \mathcal{W}_k & I_{n_m} \\ P_{m-1}^H \mathcal{W}_k & \\ \widetilde{W}_m^H \mathcal{W}_k & 0_{p\times n_m} \end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times(k+n_m)}$,
   - *for the subsequent restart, we have* $\widehat{\mathscr{V}}_{m+1}^H \mathcal{W}_m = [\mathscr{T}_m \quad O_p]^T$ *with*
   $$\mathscr{T}_m = \begin{bmatrix} C_k^H \mathcal{W}_k & 0_{k\times n_m} \\ \mathbb{V}_1^H \mathcal{W}_k & \begin{bmatrix} I_{p_1} & 0_{p_1\times(n_m-p_1)} \end{bmatrix} \\ 0_{(n_m-p_1)\times k} & \begin{bmatrix} 0_{(n_m-p_1)\times p_1} & I_{n_m-p_1} \end{bmatrix} \end{bmatrix} \in \mathbb{C}^{(k+n_m)\times(k+n_m)}, \tag{36}$$
   and Equation (35) *can be recast in*
   $$\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_i^{(HR)} = \theta_i \, \mathscr{F}_m^H \mathscr{T}_m g_i^{(HR)} \quad \text{for } 1 \le j \le n_m. \tag{37}$$
2. *At restart, if* $G_k^{(HR)} = \left[g_{i_1}^{(HR)}, \ldots, g_{i_k}^{(HR)}\right]$ *are associated with the* k *targeted eigenvalues, the matrices* $\mathcal{W}_k^{new}$, U_k^{new} *and* C_k^{new} *to be used for the next cycle are defined by*
   $$\mathcal{W}_k^{new} = \mathcal{W}_m G_k^{(HR)} R^{-1} = [\mathcal{W}_k, \mathscr{V}_m] G_k^{(HR)} R^{-1}, \tag{38}$$
   $$U_k^{new} = \widehat{\mathscr{Z}_m} G_k^{(HR)} R^{-1} = [U_k, \mathscr{Z}_m] G_k^{(HR)} R^{-1}, \tag{39}$$
   $$C_k^{new} = \widehat{\mathscr{V}}_{m+1} Q = [C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m] Q, \tag{40}$$
   where Q *and* R *are the factors of the reduced QR-factorization of* $\underline{\mathscr{F}}_m G_k^{(HR)}$ *that ensures* $AU_k^{new} = C_k^{new}$ *with* $(C_k^{new})^H C_k^{new} = I_k$.
3. *The residual at restart* $R_1^{new} = R_m^{old} = B - AX_1^{new}$ *with* $X_1^{new} = X_m^{old}$ *is orthogonal to* C_k^{new}.
4. *At all the restarts but the first, we have the following relation that holds*
   $$\mathrm{Range}(\mathcal{W}_k^{new}) \quad \subset \quad \mathrm{Range}([C_k^{new}, \mathbb{V}_1]), \tag{41}$$
   where $\mathbb{V}_1 \, \Lambda_1$ *is the reduced QR-factorization of* $R_1^{new} = R_m^{old} = \mathbb{V}_1 \Lambda_1$.

Proof. The proof essentially follows the same arguments as the ones developed for IB-BGCRO-DR, and we refer the reader to the Appendix B for the technical details. □

We also mention that a closely related numerical technique that extend IB-BGMRES-DR in the flexible preconditioning context can be derived. We refer to Appendix C where the resulting algorithm named IB-BFGMRES-DR is detailed and its properties are described.

3 Additional numerical features to complement the inexact breakdown

3.1 Subspace expansion policies

Thanks to the inexact breakdown mechanism, the abandoned directions at a given iteration might be reintroduced in a subsequent one, thereby we can trade on the original policy and select for the subspace expansion only a subset of those defined by the SVD decomposition of the least-squares residuals defined by (17). In particular, it might be relevant to choose a block size p^{CB} that is suited to best cope with the computational features on a given platform rather than selecting the numerical block size p_{j+1} defined as the number of singular values larger than the prescribed threshold ϵ^R. In that respect, we consider a subspace expansion policy so that the block size at the end of step j is defined as $p_{j+1}^{CB} = \min(p^{CB}, p_{j+1})$. We refer this variant as Inexact Breakdown Block GCRO-DR with computational blocking (denoted by IB-BGCRO-DR-CB).

3.2 Monitoring individual convergence thresholds

A classical stopping criterion for the solution of a linear system $Ax = b$ is based on backward error analysis and consists in stopping the iteration when

$$\eta_b(x_j) = \frac{\|b - Ax_j\|_2}{\|b\|_2} \leq \varepsilon. \tag{42}$$

When we have to solve for a family of right-hand sides $B = [b^{(1)}, ..., b^{(p)}]$, we can first scale all the right-hand sides so that they are all of norm one (i.e., form $\tilde{b}^{(i)} = b^{(i)}/\|b^{(i)}\|_2$, iterate for the scaled right-hand sides $\tilde{B} = [\tilde{b}^{(1)}, ..., \tilde{b}^{(p)}]$ until all the residual norms $\|\tilde{b}^{(i)} - A\tilde{x}_j^{(i)}\|_2 \leq \varepsilon$ and scale back the computed solutions to get the one initially sought $x^{(i)} = \tilde{x}_j^{(i)}\|b^{(i)}\|_2$; this latter will comply with (42). For the sake of simplicity of exposure and without loss of generality, we will assume in the rest of this section that the set of right-hand sides are of norm one.

The governing idea in the inexact breakdown mechanism is to only select the directions associated with singular values larger than $\epsilon^{(R)}$, so that, when there is no more candidate direction for extending the search space, i.e., $p_{j+1} = 0$ all the solutions are computed at the target accuracy $\epsilon^{(R)}$. Setting $\epsilon^{(R)} = \varepsilon$ ensures that all the solutions complied with the stopping criterion (42). This relies on the following inequalities:

$$\|b^{(i)} - Ax_j^{(i)}\|_2 \leq \|B - AX_j\|_2 = \|\Lambda_j - \underline{\mathscr{F}}_j Y_j\|_2 \leq \epsilon^{(R)}. \tag{43}$$

The occurrence of p inexact breakdowns is a sufficient condition that ensures the convergence of the p solution vectors, but the convergence might happen before and a more classic stopping criterion can be accommodated at a very low computational cost. One can also check the convergence by looking at the norm of the least-squares residuals, that are easy to compute. Let $Q_j^{LS} R_j^{LS}$ be the (full) QR-factorization of $\underline{\mathscr{F}}_j$ (i.e., Q_j^{LS} is unitary) , then

$$\Lambda_j - \underline{\mathscr{F}}_j Y_j = Q_j^{LS} \begin{pmatrix} 0_{(n_j+k)\times p} \\ R_j^{\ell s} \end{pmatrix}, \tag{44}$$

where $R_j^{\ell s} \in \mathbb{C}^{p\times p}$ are the last p rows of $(Q_j^{LS})^H \Lambda_j$ so that $\|b^{(i)} - Ax_j^{(i)}\|_2 = \|R_j^{\ell s}(:, i)\|_2$. Those residual norm calculations allow the definition of a stopping criterion for the block algorithm based on the individual convergence criterion of each right-hand side defined by (43).

In some applications all the solutions associated with a block of right-hand sides do not need to be solved with the same accuracy. In that context, the subspace expansion policy and the stopping criterion should be adapted. Let $\epsilon^{(i)}$ denotes the target accuracy for the solution associated with the right-hand side $b^{(i)}$. We define the candidate direction by applying the candidate directions policy to the scaled residual matrix $(\Lambda_j - \underline{\mathscr{F}}_j Y_j) D_\epsilon$ where $D_\epsilon = \text{ diag } ((\epsilon^{(1)})^{-1}, (\epsilon^{(2)})^{-1}, \cdots, (\epsilon^{(p)})^{-1})$ with $\epsilon^{(R)} = 1$. When p inexact breakdowns have occurred, because the 2-norm of a matrix is an upper bound of the 2-norm of its columns (that are the residual scaled by their associated stopping criterion threshold), we have

$$\left\| b^{(i)} - A x_m^{(i)} \right\|_2 (\epsilon^{(i)})^{-1} \quad \leq \quad \| (\Lambda_1 - \underline{\mathscr{F}}_m Y_m)\, D_\epsilon \|_2 \leq 1 \text{ for } \forall i \in \{1, \cdots, p\}.$$

As discussed previously the individual convergence of the solutions can be checked at each iteration by computing the norm of the columns of the $(p \times p)$ least-squares residual $R_j^{\ell s} D_\epsilon$.

3.3 Remarks on some computational and algorithmic aspects

On the computational point of view, a few remarks can be made for the practical implementation of the described numerical methods. For the sake of conciseness of this paper, we do not give the full technical details of what we briefly expose below but rather refer for each of them to a particular section in the appendix where a complete proof is given. The points we wanted to make are:

1. Note that both techniques for monitoring individual convergence thresholds and the subspace expansion policies discussed in Section 3.1 and 3.2 could be applied to any other block minimum residual norm methods equipped with the inexact breakdown mechanism such as the IB-BGMRES and IB-BGMRES-DR algorithms.

2. The full $Q_j^{LS} R_j^{LS}$-factorization involved in the solution of the least-squares problems (44) enables the cheap calculation of the SVD for the residual block (17). This observation applies naturally to the IB-BGMRES and IB-BGMRES-DR algorithms (we refer to Appendix E for the details).

3. For the sake of exposure, we made the assumption that the initial residual block R_1 is of full rank. In practice, this constraint can be removed by applying already to $\mathbb{V}_1$ the candidate search space expansion policy based on the SVD of the R factor of the reduced QR factorization of R_1. We refer to the Appendix F for the details for a detailed exposure of the resulting algorithm. In particular, a consequence of the occurrence of an inexact breakdown in R_1 is that the right-hand side of the least-squares problems needs to be updated at each iteration and not simply expanded with a zero block. The pseudocode for IB-BGCRO-DR with inexact breakdown detection in R_1 and updated right-hand side of the least-squares problems for constant and slowly-changing left-hand sides with massive number of right-hand sides are presented in Appendix I and J, respectively.

4 Numerical experiments

In the following sections we illustrate the different numerical features of the novel algorithm introduced above. For the sake of comparison, for some of the experiments we also display results with closely related block methods such as BGCRO-DR [28,29,36,46] or IB-BGMRES-DR [2]. All the numerical experiments have been run using a Matlab prototype, so that the respective performances of the algorithms are evaluated in term of number of matrix-vector products, denoted as $mvps$ (and preconditioner applications in the preconditioned case) required to converge. For all block methods, the stopping criterion is that the p individual normwise backward errors satisfy $\eta_{b^{(i)}}(x_s^{(i)}) = \frac{||b^{(i)} - A x_s^{(i)}||_2}{||b^{(i)}||_2} < \varepsilon \ (i = 1, \ldots, p)$

with respect to the approximate solution $x_s^{(i)}$, or the $mvps$ exceeds the allowed maximal number (referred to as $maxMvps$).

For each set of block of right-hand sides, referred to as a family, the block initial guess is equal to $0 \in \mathbb{C}^{n\times p}$, where p is the number of right-hand sides. The block of right-hand side $B = [b^{(1)}, b^{(2)}, \ldots, b^{(p)}] \in \mathbb{C}^{n\times p}$ is composed of p linearly independent vectors generated randomly (using the same seed when block methods are compared). While any part of the spectrum could be considered to define the recycled space we consider for all the experiments the approximated eigenvectors associated with the k smallest approximated eigenvalues in magnitude. The maximum dimension of the search space in each cycle is set to be $m_d = 15 \times p$, the targeted backward error is $\varepsilon = 10^{-8}$ and $maxMvps = 2000 \times p$ for each solver run. To illustrate the potential benefit of IB-BGCRO-DR when compared to another block solver, we consider the overall potential gain when solving a sequence of ℓ families defined as

$$\text{Gain } (\ell) = \frac{\sum_{s=1}^{\ell} \#mvps \text{ (method)}^{(s)}}{\sum_{s=1}^{\ell} \#mvps \text{ (IB-BGCRO-DR)}^{(s)}}. \tag{45}$$

4.1 Ritz versus harmonic Ritz subspace recycling policies

To illustrate the flexibility of subspace recycling in IB-BGCRO-DR as discussed in Section 2.3, both the harmonic Ritz (HR) and Rayleigh Ritz (RR) projections are considered to construct the recycled subspace; the associated algorithms are referred to as IB-BGCRO-DR(HR) and IB-BGCRO-DR(RR). Following the spirit of the test examples considered in [24] we consider bidiagonal matrices of size 5000 with lower diagonal unity so that their spectrum is defined by their diagonal entries; we denote them Matrix 1 and Matrix 2. Matrix 1 has diagonal entries $0.1, 1, 2, 3, \ldots, 4999$ and Matrix 2 has diagonal entries $10.1, 10.2, \ldots, 20, 21, \ldots, 4920$. We consider experiments with a family size $p = 20$, the size of the recycled space $k = 30$ and the maximal dimension of the search space $m_d = 300$. In Figure 1 we display some experimental results. The graphs on the left give the envelope of the convergence histories of the p backward errors as a function of the number of matrix-vector products ($mvps$) for the first three families. On the right graphs we depict the number of matrix-vector products for each of the 30 families. For Matrix 1, one can observe that the HR-projection does capture a space that slows down the initial convergence once the first family has been solved; that is, for families 2 and 3 the converge histories do not exhibit anymore any plateau. On that example the RR-projection does not capture a recycled space that helps much the convergence as the three convergence histories exhibit very similar pattern. For Matrix 2, both RR and HR projections work pretty much the same. In Table 1, we report the total required $mvps$ for the two matrix examples for 3 and 30 families. Those results do not attempt to highlight that one projection is superior to the other one, but simply illustrate the flexibility of the GCRO approach to accommodate both. The selection or discussion of the best suited projection method is out of the scope of this paper.

In the rest of this paper, only the HR projection is considered to build recycling subspace used in the GCRO-DR like methods. Besides, the bidiagonal Matrix 1 is chosen as the constant left-hand sides in following Subsection 4.2- 4.5, in which the related parameters are likewise set to be $p = 20$, $k = 30$ and $m_d = 300$.

4.2 Comparison with IB-BGMRES-DR

In order to illustrate the difference between BGCRO-DR, IB-BGCRO-DR and IB-BGMRES-DR. The convergence histories of two consecutive families are displayed in the left plot of Figure 2. Several observations can be made. Because IB-BGCRO-DR and BGCRO-DR do not have a deflation space to start with for the first family, the convergence histories of the three solvers overlap as long as the IB-mechanism does not detect any partial convergence. At this point, the convergence rate of IB-BGCRO-

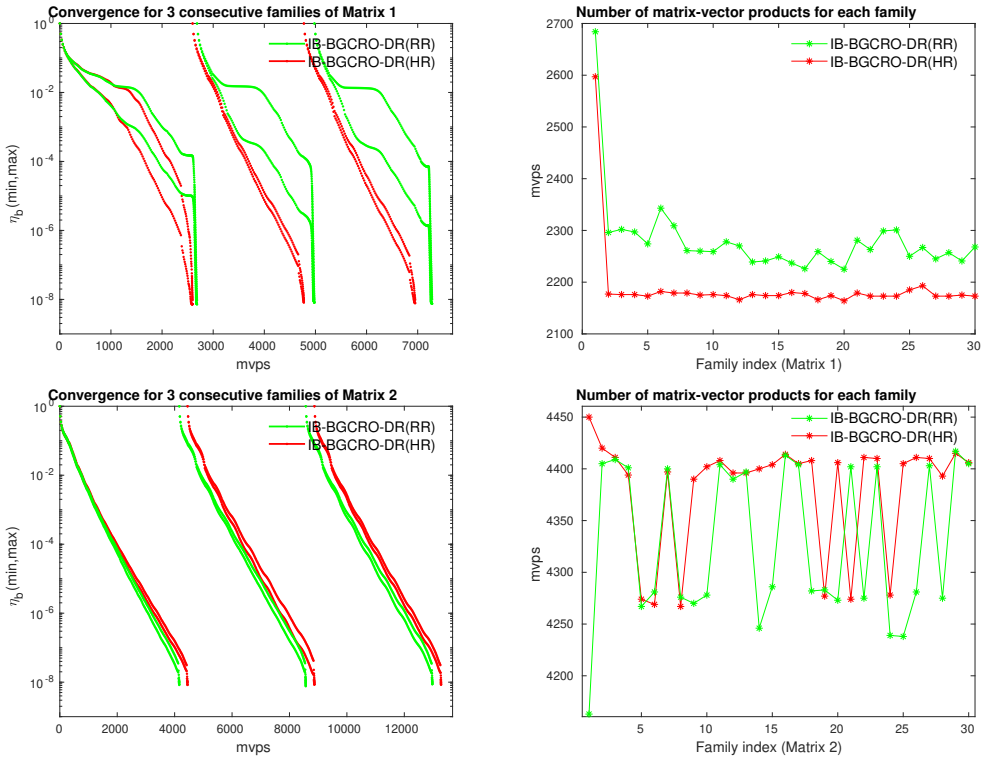


Figure 1: History of bidiagonal Matrix 1 and Matrix 2 ($p = 20$, $m_d = 300$ and $k = 30$). Left: convergence histories of the largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 3 consecutive families. Right: consumed number of $mvps$ versus family index.

Family number	Matrix	Method	$mvps$
3	Matrix 1	IB-BGCRO-DR(HR)	**6950**
		IB-BGCRO-DR(RR)	7280
30	Matrix 1	IB-BGCRO-DR(HR)	**65686**
		IB-BGCRO-DR(RR)	68421
3	Matrix 2	IB-BGCRO-DR(HR)	13281
		IB-BGCRO-DR(RR)	**12977**
30	Matrix 2	IB-BGCRO-DR(HR)	131401
		IB-BGCRO-DR(RR)	**129865**

Table 1: Numerical results of IB-BGCRO-DR with recycling subspace generated by RR or HR-projection for Matrix 1 and Matrix 2 with $p = 20$, $m_d = 300$ and $k = 30$.

DR and IB-BGMRES-DR becomes faster (in terms of matrix-vector products) than that of BGCRO-DR, and the two convergence histories mostly overlap as the two IB techniques remain mathematically equivalent. For the second and subsequent families, the capability to start with a deflation space shows its benefit for BGCRO-DR and IB-BGCRO-DR. Because IB-BGMRES-DR has to capture again this

spectral information it needs a few restarts before it finds the spectral information again and refines it in its subsequent search spaces; eventually it exhibits a convergence rate similar to the BGCRO-DR counterpart.

For the sake of comparison and to illustrate the benefit of the IB mechanism we also display the converge histories of BGCRO-DR that always requires more matrix-vector products compared to its IB counterpart. Those extra matrix-vector products mostly concur to improve the solution quality for some right-hand sides beyond the targeted accuracy. To further highlight the effect of the IB mechanism, we report in the right plot the size of search space expansion as a function of the iterations. Because BGCRO-DR does not implement the IB mechanism, the search space is increased by $p = 20$ at each iteration. For the two other block solvers, the block size monotonically decreases down to 1. Because the IB mechanism is implemented on R_1 in IB-BGCRO-DR, the block size does not increase at restart. By construction, IB-BGMRES-DR implements the IB mechanism at restart so that the same observation applies.

A summary of the $mvps$ and the number of block iterations (referred to as its) is given in Table 2 that shows the benefit of using IB-BGCRO-DR.

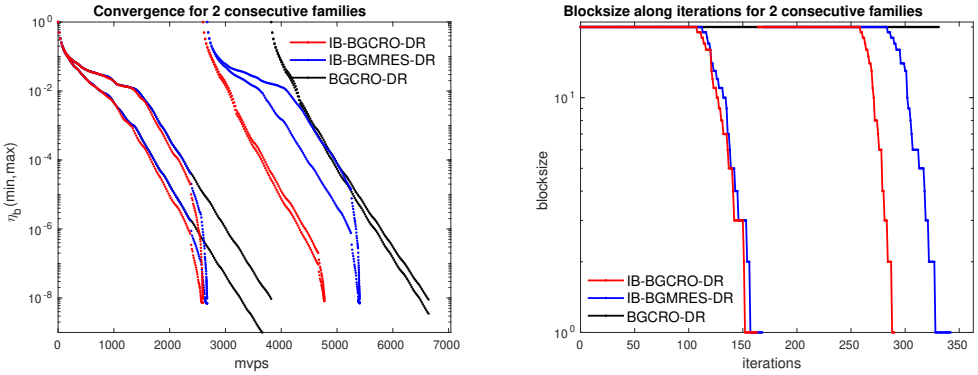


Figure 2: History for Section 4.2. Comparison IB-BGCRO-DR with BGCRO-DR and IB-BGMRES-DR by solving bidiagonal Matrix 1 ($p = 20$, $m_d = 300$ and $k = 30$). Left: convergence histories of the largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 2 consecutive families. Right: varying blocksize comparison along iterations.

Family number	Method	$mvps$	its
2	BGCRO-DR	6640	332
	IB-BGMRES-DR	5407	343
	IB-BGCRO-DR	**4774**	**291**

Table 2: Numerical results in both terms of $mvps$ and its for Matrix 1 ($p = 20$, $m_d = 300$ and $k = 30$).

4.3 Convergence analysis with different target accuracies

In this section we illustrate how the target accuracy interplays with the quality of the extracted spectral information.

The convergence histories of solving three successive families with accuracy $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-8}$ are described in Figure 3, from which it is observed that the benefits of inexact breakdown detection is significative especially when solving the first family with lower accuracy, like $10^{-2}, 10^{-3}$ or 10^{-4}. Besides it is interesting to notice the different curve shapes of IB-BGCRO-DR displayed among the accuracy

10^{-2}, 10^{-3} and $10^{-4}, 10^{-8}$. For the former two, with lower accuracy, it seems that the inexact breakdown detection mechanism prevents IB-BGCRO-DR to capture a deflation space enabling us to have as smooth and fast convergence for the subsequent families similar to what BGCRO-DR exhibits. For the first family, the absence of IB mechanism in regular BGCRO-DR leads to the exploration of spaces that are not important for the linear system solution (at least for the level of targeted accuracy) but relevant to capture useful spectral information that will significant speed-up its convergence for the subsequent families.

For $\epsilon = 10^{-4}$, IB-BGCRO-DR needs the solution of two families to capture all the relevant spectral information and to exhibit the same convergence rate as BGCRO-DR. While, the first family is enough for $\epsilon = 10^{-8}$.

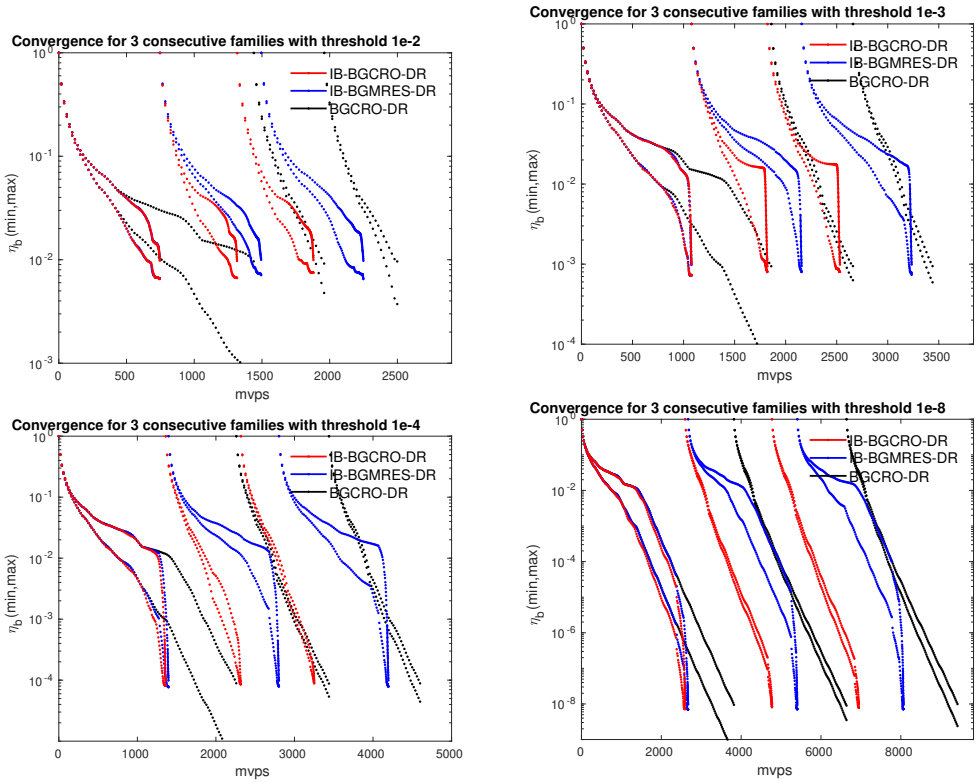


Figure 3: History of Section 4.3 for the behavior in case of different target accuracy (10^{-2}, 10^{-3}, 10^{-4} and 10^{-8}). Convergence history of IB-BGCRO-DR, BGCRO-DR and IB-BGMRES-DR on the families constructed by bidiagonal Matrix 1 with parameters setting as $p = 20$, $m_d = 300$ and $k = 30$.

4.4 Subspace expansion policies

As discussed in Section 3.1, only a subset of the candidate directions exhibited by the IB mechanism can be eventually selected to expand the search space at the next block iteration; we denote this maximum size p^{CB} and refer to this variant as IB-BGCRO-DR-CB where the CB stands for Computational Blocking. In Table 3 we show the effect of this algorithmic parameter on $mvps$ and its for the solution of 3 and 30 families with Matrix 1 when p^{CB} varies from 1 to 15 for a number of right-hand sides $p = 20$. It can be seen that smaller p^{CB} is, the smaller $mvps$, but larger its. While reported only on one example this

trend has been observed in all our numerical experiments. Depending on the computational efficiency or cost of the matrix-vector products with respect to the computation weight of the least-squares solution and SVD of the least squares residuals, this gives opportunities to monitor the overall computational efficiency of the complete solution.

Family number	Method	$mvps$	its
3	IB-BGCRO-DR	6950	**416**
	IB-BGCRO-DR-CB ($p^{CB} = 15$)	6849	460
	IB-BGCRO-DR-CB ($p^{CB} = 10$)	6856	661
	IB-BGCRO-DR-CB ($p^{CB} = 5$)	6859	1297
	IB-BGCRO-DR-CB ($p^{CB} = 1$)	**6844**	6370
30	IB-BGCRO-DR	65686	**3818**
	IB-BGCRO-DR-CB ($p^{CB} = 1$)	**64700**	60131

Table 3: Numerical results of IB-BGCRO-DR and IB-BGCRO-DR-CB for $p^{CB} = 1, 5, 10, 15$ in terms of $mvps$ and its for Section 4.4, where the involved parameters for bidiagonal Matrix 1 are set to be $p = 20$, $m_d = 300$ and $k = 30$.

We notice that this subspace expansion policy also applies to IB-BGMRES-DR, and we refer to Figure 9 and Table 8 of Appendix L for an illustration for this block solver.

4.5 Solution with individual convergence thresholds

To illustrate this feature, we consider a family of p right-hand sides and a convergence threshold 10^{-4} for the first $p/2$ and 10^{-8} for the last $p/2$ ones. To illustrate the benefit of this feature we compare with calculations where all the right-hand sides are solved with the most stringent accuracy, that is 10^{-8}. We display in the left part of Figure 4, the convergence histories for 3 successive families. The variant that controls the individual threshold is denoted as IB-BGCRO-DR-VA where VA stands for Variable Accuracy. It can be seen that the numerical feature works well and that the envelope of the backward errors has the expected shape, that is, the minimum backward error goes down to 10^{-8} while the maximum ones (associated with the first $p/2$ solutions) only goes down to 10^{-4}. If we compare the convergence history of IB-BGCRO-DR and IB-BGCRO-DR-VA, it can be seen that the slope of IB-BGCRO-DR-VA is deeper than that of IB-BGCRO-DR once the first $p/2$ solutions have converged; at this point IB-BGCRO-DR-VA somehow focuses the new directions (produced by the matrix-vector products used for the x-axis) to reduce the residual norms of the remaining $p/2$ solutions that have not yet converged. The plot on the right of Figure 4 shows the computational gain induced by the individual control of the accuracy compared to the situation where all the right-hand sides would have been solved the most stringent one if this feature had not been designed. In this case the individual monitoring of the convergence saves around 30 % of the matrix-vector products. Those results are summarized in Table 4.

Similar to previous subsections, we refer to Figure 10 and Table 9 of Appendix M for an illustration of extending such individual control to the block solver IB-BGMRES-DR that can also accommodate this feature.

4.6 Behavior on sequences of slowly varying left-hand sides problems

The example used in this section is from a finite element fracture mechanics problem that is fully documented in [27, Section 4.1]. Over 2000 linear systems of size 3988×3988 need to be solved in order to capture the fracture progression, and among them 151 linear systems $400 - 550$ representing a typical subset of the fracture progression in which many cohesive elements break are examined in [27].

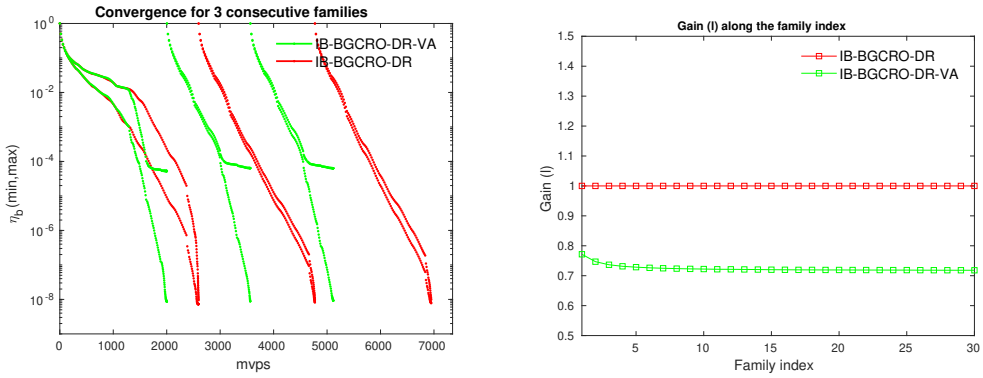


Figure 4: Comparison of IB-BGCRO-DR to IB-BGCRO-DR-VA on families with Matrix 1 ($p = 20$, $m_d = 300$ and $k = 15$). Left: convergence histories of the largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 3 consecutive families. Right: Gain (ℓ) of IB-BGCRO-DR-VA to IB-BGCRO-DR versus family index.

Family number	Method	$mvps$	its
3	IB-BGCRO-DR	6948	416
	IB-BGCRO-DR-VA	**5118**	**395**
30	IB-BGCRO-DR	65641	3820
	IB-BGCRO-DR-VA	**47141**	**3567**

Table 4: Numerical results of IB-BGCRO-DR with fixed/varying target accuracy for each right-hand side in terms of $mvps$ and its for Section 4.5, where the coefficient matrix is bidiagonal Matrix 1 with involving parameters defined as $p = 20$, $m_d = 300$ and $k = 30$.

The solution of these linear systems have been investigated using both GCRO-DR and GCROT, and we refer to [10] for a comprehensive experimental analysis. For our numerical experiments we borrow the ten linear systems numbered $400 - 409$ of this *FFEC* collection. For each set of linear system we select the matrix and the corresponding right-hand sides that we expand to form a block of $p = 20$ right-hand sides by appending random linearly independent vectors.

We display the convergence histories for solving first 3 consecutive families of linear systems in the left plot of Figure 5. For the solution of the first block of right-hand sides, the observations on the IB and DR mechanisms discussed in Section 4.2 apply. Even though the matrix has changed, the recycled spectral information computed for the previous matrix still enable a faster convergence at the beginning of the solution of the next one. For the solution of the first family the convergence histories of the three methods fully overlap until the first inexact breakdown occurs, as until this step the three methods are mathematically equivalent. For the subsequent families, it can be seen that the sequence of matrices are close enough to ensure that the recycled space from one system to the next still makes benefit to the convergence as IB-BGCRO-DR and BGCRO-DR converge faster than IB-BGMRES-DR at the initial stage. The benefit of the IB mechanism is also illustrated on that example as IB-BGCRO-DR still outperforms BGCRO-DR. The overall benefit in term of $mvps$ saving is illustrated in the right plot on a sequence of 10 linear systems, where the saving is close to 14% with respect to IB-BGMRES-DR and more than 67 % with respect to BGCRO-DR. Those results are summarized in Table 5.

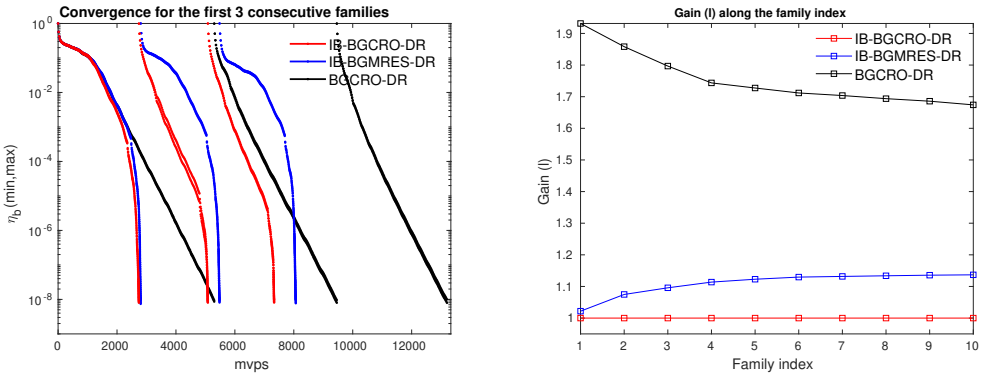


Figure 5: Convergence results of IB-BGCRO-DR, BGCRO-DR and IB-BGMRES-DR on a sequence of slowly varying left-hand sides built on FFEC with $p = 20$, $m_d = 300$ and $k = 15$.

Family number	Method	$mvps$	its
3	BGCRO-DR	13200	660
	IB-BGMRES-DR	8068	622
	IB-BGCRO-DR	**7363**	**523**
10	BGCRO-DR	39515	1969
	IB-BGMRES-DR	26832	2022
	IB-BGCRO-DR	**23603**	**1611**

Table 5: Numerical results in terms of $mvps$ and its for Section 4.6, in which the involving parameters for FFEC are set to be $p = 20$, $m_d = 300$ and $k = 15$.

4.7 A variant suited for flexible preconditioning

In this section, we illustrate the numerical behavior of the flexible variant IB-BFGCRO-DR that we have derived in Section 2.4 and make comparison with closely related variants namely BFGCRO-DR (a straightforward block extension of FGCRO-DR [6]) and IB-BFGMRES-DR. We refer to Appendix C for a detailed description of IB-BFGMRES-DR that is although novel.

We consider a representative quantum chromodynamics (QCD) matrix from the University of Florida sparse matrix collection [8]. It is the conf5.4-00l8x8-0500 matrix denoted as B of size 49152×49152 with the critical parameter $\kappa_c = 0.17865$ as a model problem. Thirty families of linear systems are constructed that are defined as $A^{(\ell)} = I - \kappa_c(\ell)B$ with $0 \leq \kappa_c(\ell) < \kappa_c$ and $\ell = 1, 2, \ldots, 30$. We use the Matlab function **linspace**$(0.1780, 0.1786, 30)$ to generate the parameters $\kappa_c(\ell)$ for the sequence of left-hand side matrices and observe that those matrices have the same eigenvectors associated with shifted eigenvalues. A sequence of $p = 12$ successive canonical basis vectors are chosen to be the block of right-hand sides for a given left-hand side matrix following [27, Section 4.3] so that the complete set of the right-hand sides for the ℓ linear systems reduces to the first $p \times \ell$ columns of the identity matrix. This choice could be supported by the fact that the problem of numerical simulations of QCD on a four-dimensional space-time lattice for solving QCD ab initio (cf. [27, Section 4.3]) has a 12×12 block structure, and then a system with 12 right-hand sides related to a single lattice site is often of interest to solve.

The flexible preconditioner is defined by a 32-bit $ILU(0)$ factorization of the matrix involved in the linear system. In a 64-bit calculation framework, the preconditioning consists in casting the set of

directions to be preconditioned in 32-bit format, performing the forward/backward substitution in 32-bit calculation and casting back the solutions in 64-bit arithmetic. The rounding applied to the vectors has a nonlinear effect that makes the preconditioner nonlinear.

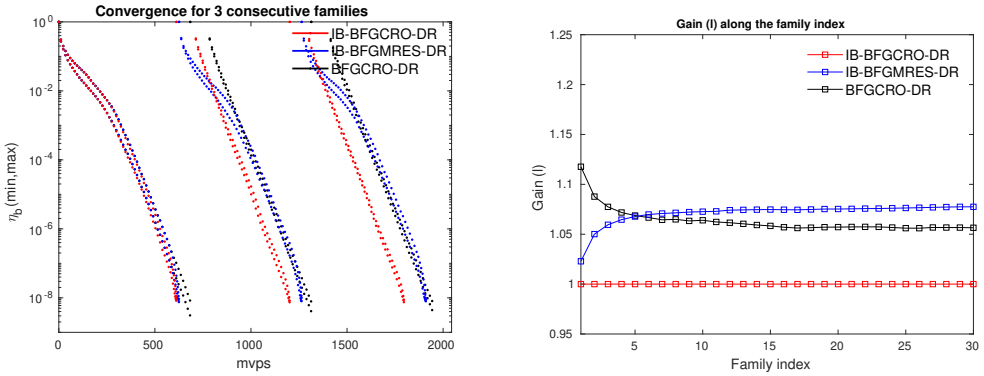


Figure 6: Behavior of flexible block solver variant on families of QCD matrices with $p = 12$, $m_d = 180$ and $k = 90$. Left: convergence histories of the largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 3 consecutive families. Right: Gain (l) of the block methods with respect to IB-BFGCRO-DR along family index.

Family number	Method	$mvps$	its
3	BFGCRO-DR	1944	147
	IB-BFGMRES-DR	1911	177
	IB-BFGCRO-DR	**1797**	**145**
30	BFGCRO-DR	18774	1347
	IB-BFGMRES-DR	19147	1779
	IB-BFGCRO-DR	**17770**	**1327**
60	BFGCRO-DR	37494	2682
	IB-BFGMRES-DR	38363	3570
	IB-BFGCRO-DR	**35642**	**2655**

Table 6: Numerical results in terms of $mvps$ and its for Section 4.7, in which the involving parameters for QCD matrix are set to be $p = 12$, $m_d = 15 \times p = 180$ and $k = 90$.

For those experiments, we attempt to favor the recycling of the space, because the matrices share the same invariant space, so that we choose a relative large value for k that is $k = m_d/2$. We report in the left plot of Figure 6, the convergence histories of the three flexible block variants. Similarly to what has already been observed previously the convergences are very similar on the first family and only differ when the IB mechanism becomes active mostly in the last restart. For the second and third families, one can see that IB-BFGCRO-DR and BFGCRO-DR have identical convergence speed. One can observe a shift in the convergence histories between the end of the solution of one family and the beginning of the next one for both IB-BFGCRO-DR and BFGCRO-DR. This shift is due to the extra k matrix-vector products that have to be performed when the matrix changes in order to adapt the deflation space as follows

1. compute $A^{(\ell+1)}U_k^{(\ell)} = \tilde{C}_k$
2. compute the reduced QR factorization of $\tilde{C}_k = C_k^{(\ell+1)}R$

3. update the basis of the deflation space $U_k^{(\ell+1)} = U_k^{(\ell)} R^{-1}$ so that $A^{(\ell+1)} U_k^{(\ell+1)} = C_k^{(\ell+1)}$.

Because k is large, we can clearly see the shift in the left plot of Figure 6. On the second and third family IB-BFGMRES-DR has the same convergence history as for the solution of the first one. For this parameter selection on those examples, it can be noticed that the dominating effect on the convergence improvement is due to the space recycling and not the IB mechanism as BFGCRO-DR outperforms IB-BFGMRES-DR gradually. This observation is highlighted in the right plot of Figure 6, where the benefit of using IB-BFGCRO-DR rather than BFGCRO-DR does shrink when compared to previous experiments and is only about 8%. This example also illustrate the benefit of combining the two numerical features IB and subspace recycling as IB-BGCRO-DR is the method that requires the less matrix-vector products (and preconditioning applications) as well as the less iterations as it is summarized in Table 6 for various numbers of linear system families.

5 Concluding remarks

In this paper, we develop new variants of block GCRO-DR methods denoted as IB-BGCRO-DR. We demonstrate that these new solvers combine the nice numerical features of both inexact breakdown detection mechanism and subspace recycling strategy through extensive numerical experiments for solving linear systems with constant or slowing-varying left-hand sides and massive number of right-hand sides. We discuss the multiple choices for constructing the recycling subspace used in the GCRO-DR like methods and illustrate that the performance of IB-BGCRO-DR with the recycling subspace built by the commonly used Harmonic Ritz projection is not always superior to that with Rayleigh Ritz projection. Based on the inexact breakdown mechanism, we present the flexibility of the search space expansion policy that can be accommodated to find a trade-off between the computational and numerical efficiency of the solver. We also introduce a technique for monitoring individual convergence thresholds for each right-hand. To comply with mixed-precision calculation, the flexible preconditioning variant is also proposed, which would be of interest for emerging computing platforms where mixed-precision calculation could be a way to reduce data movement that is foreseen as one of the major bottleneck to reach high performance.

Acknowledgments

We would like to thank Matthieu Simonin (who developped a C++ implementation of the solver - available on https://gitlab.inria.fr/solverstack/fabulous/ - in the framework of the PRACE 6IP project) for his reading and comments on earlier version of this document.

References

[1] J. Abels and P. Benner. CAREX–A collection of benchmark examples for continuous-time algebraic Riccati equations (Version 2.0). SLICOT Working Note 1999-14, Mathematic and Informatik, Universitä Bremen, 1999.

[2] E. Agullo, L. Giraud, and Y.-F. Jing. Block GMRES method with inexact breakdowns and deflated restarting. *SIAM J. Matrix Anal. Appl.*, 35(4):1625–1651, 2014.

[3] P. Benner and L.-H. Feng. Recycling Krylov subspaces for solving linear systems with successively changing right-hand sides arising in model reduction. In P. Benner, M. Hinze, and E. J. W. ter Maten, editors, *Model Reduction for Circuit Simulation*, volume 74 of *Lecture Notes in Electrical Engineering*, pages 125–140. Springer Netherlands, 2011.

[4] H. Calandra, S. Gratton, R. Lago, X. Vasseur, and L. M. Carvalho. A modified block flexible GMRES method with deflation at each iteration for the solution of non-Hermitian linear systems with multiple right-hand sides. *SIAM J. Sci. Comput.*, 35:S345–S367, 2013.

[5] L. M. Carvalho, S. Gratton, R. Lago, and X. Vasseur. A flexible generalized Conjugate Residual method with inner orthogonalization and deflated restarting. Technical Report TR/PA/10/10, CERFACS, Toulouse, France, 2010.

[6] L. M. Carvalho, S. Gratton, R. Lago, and X. Vasseur. A flexible generalized conjugate residual method with inner orthogonalization and deflated restarting. *SIAM J. Matrix Anal. Appl.*, 32:1212–1235, 2011.

[7] R. Cockett. The block Conjugate Gradient for multiple right hand sides in a direct current resistivity inversion, 2009. http://www.row1.ca/s/pdfs/courses/BlockCG.pdf.

[8] T. A. Davis and Y. Hu. The University of Florida Sparse Matrix Collection. *ACM Trans. Math. Softw.*, 38:1:1–1:25, 2011.

[9] E. de Sturler. Nested Krylov methods based on GCR. *J. Comput. Appl. Math.*, 67:15–41, 1996.

[10] E. de Sturler. Truncation strategies for optimal Krylov subspace methods. *SIAM J. Numer. Anal.*, 36:864–889, 1999.

[11] L.-H. Feng, P. Benner, and J. G. Korvink. Subspace recycling accelerates the parametric macro-modeling of MEMS. *Internat. J. Numer. Methods Engrg.*, 94(1):84–110, 2013.

[12] P. F. Fischer. Projection techniques for iterative solution of $Ax = b$ with successive right-hand sides. *Computer Methods in Applied Mechanics and Engineering*, 163:193–201, 1998.

[13] L. Giraud, S. Gratton, and E. Martin. Incremental spectral preconditioners for sequences of linear systems. *Appl. Numer. Math.*, 57(11-12):1164–1180, 2007.

[14] L. Giraud, S. Gratton, X. Pinel, and X. Vasseur. Flexible GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 32:1858–1878, 2010.

[15] M. H. Gutknecht. Block Krylov space methods for linear systems with multiple right-hand sides: An introduction. In I. S. Duff, A. H. Siddiqi, and O. Christensen, editors, *Modern Mathematical Models, Methods and Algorithms for Real World Systems*, pages 420–447. Anamaya Publishers, New Delhi, India, 2006.

[16] E. Haber, M. Chung, and F. Herrmann. An effective method for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM J. Optimization*, 22(3):739–757, 2012.

[17] Y.-F. Jing, B. Carpentieri, and T.-Z. Huang. Experiments with Lanczos biconjugate A-orthonormalization methods for MoM discretizations of Maxwell's equations. *Progress In Electromagnetics Research*, 99:427–451, 2009.

[18] V. Kalantzis. Solving linear systems with multiple right-hand sides: Kernel for scientific computing and parallel computing challengues. Joint HP-SEE LinkSCEEM-2 and PRACE HPC Summer Training LinkSCEEM-2, University of Patras, Greece, 2011.

[19] M. Kilmer and E. de Sturler. Recycling subspace information for diffuse optical tomography. *SIAM J. Sci. Comput.*, 27(6):2140–2166, 2006.

[20] R. Lago. A Study on Block Flexible Iterative Solvers with Applications to Earth Imaging Problem in Geophysics. Ph.D. Dissertation TH/PA/06/03, CERFACS, University of Toulouse, France, 2013.

[21] J. Langou. Iterative Methods for Solving Linear Systems with Multiple Right-Hand Sides. Ph.D. dissertation TH/PA/03/24, CERFACS, Toulouse, France, 2003.

[22] K. Meerbergen and Z.-J. Bai. The Lanczos method for parameterized symmetric linear systems with multiple right-hand sides. *SIAM J. Matrix Anal. Appl.*, 31(4):1642–1662, 2010.

[23] R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM J. Matrix Anal. Appl.*, 16:1154–1171, 1995.

[24] R. B. Morgan. GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 24(1):20–37, 2002.

[25] R. B. Morgan. Restarted block GMRES with deflation of eigenvalues. *Appl. Numer. Math.*, 54(2):222–236, 2005.

[26] M. L. Parks. The Iterative Solution of a Sequence of Linear Systems Arising from Nonlinear Finite Element Analysis. Ph.D. Dissertation UIUCDCS-R-2005-2497, University of Illinois at Urbana-Champaign, 2005.

[27] M. L. Parks, E. de Sturler, G. Mackey, D.D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM J. Sci. Comput.*, 28(5):1651–1674, 2006.

[28] M. L. Parks and K. M. Soodhalter. Block GCRO-DR. in Belos package of the Trilinos C++ Library, 2011.

[29] M. L. Parks, K. M. Soodhalter, and D. B. Szyld. A block recycled GMRES method with investigations into aspects of solver performance, 2016. http://arxiv.org/abs/1604.01713.

[30] Z. Peng, Y. Shao, and J.-F. Lee. Advanced model order reduction technique in real-life IC/package design. In *Electrical Design of Advanced Packaging Systems Symposium (EDAPS), 2010 IEEE*, pages 1–4, Dec. 2010.

[31] L. G. Ramos, R. Kehl, and R. Nabben. Projections, deflation, and multigrid for nonsymmetric matrices. *SIAM J. Matrix Anal. Appl.*, 41(1):83–105, 2020.

[32] M. Robbé and M. Sadkane. Exact and inexact breakdowns in the block GMRES method. *Linear Algebra Appl.*, 419:265–285, 2006.

[33] Y. Saad. *Iterative Methods for Sparse Linear Systems, 2nd ed.* SIAM, Philadelphia, 2003.

[34] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving non-symmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.

[35] C.F. Smith, A.F. Peterson, and R. Mittra. A conjugate gradient algorithm for the treatment of multiple incident electromagnetic fields. *IEEE Transactions on Antennas and Propagation*, 37 (11):1490–1493, 1989.

[36] K. Soodhalter. Krylov Subspace Methods with Fixed Memory Requirements: Nearly Hermitian Linear Systems and Subspace Recycling. Ph.D. dissertation, Temple University, 2012.

[37] F. Sourbier, S. Operto, J. Virieux, P. Amestoy, and J. Y. L. Excellent. *FWT2D*: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data, part 1: Algorithm. *Comput. Geosci.*, 35:487–495, 2009.

[38] F. Sourbier, S. Operto, J. Virieux, P. Amestoy, and J. Y. L. Excellent. *FWT2D*: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data, part 2: Numerical examples and scalability analysis. *Comput. Geosci.*, 35:496–514, 2009.

[39] D.-L. Sun, B. Carpentieri, T.-Z. Huang, and Y.-F. Jing. A spectrally preconditioned and initially deflated variant of the restarted block GMRES method for solving multiple right-hand sides linear systems. *Internat. J. Mech. Sci.*, 144:775–787, 2018.

[40] D.-L. Sun, T.-Z. Huang, B. Carpentieri, and Y.-F. Jing. Flexible and deflated variants of the block shifted GMRES method. *J. Comput. Appl. Math.*, 345:168–183, 2019.

[41] D.-L. Sun, T.-Z. Huang, B. Carpentieri, and Y.-F. Jing. A new shifted block GMRES method with inexact breakdowns for solving multi-shifted and multiple right-hand sides linear systems. *J. Sci. Comput.*, 78:746–769, 2019.

[42] D.-L. Sun, T.-Z. Huang, Y.-F. Jing, and 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.*, 25, 2018. e2148. https://doi.org/10.1002/nla.2148.

[43] A. Tajaddini, G. Wu, F. S. Movahed, and N. Azizizadeh. Two new variants of the simpler block GMRES method with vector deflation and eigenvalue deflation for multiple linear systems. *J. Sci. Comput.*, 86, 2021.

[44] G. Wu, Y.-C. Wang, and X.-Q. Jin. A preconditioned and shifted GMRES algorithm for the PageRank problem with multiple damping factors. *SIAM J. Sci. Comput.*, 34(5):A2558–A2575, 2012.

[45] Y.-F. Xiang, Y.-F. Jing, and T.-Z. Huang. A new projected variant of the deflated block conjugate gradient method. *J. Sci. Comput.*, 80:1116–1138, 2019.

[46] F. Xue and H. C. Elman. Fast inexact subspace iteration for generalized eigenvalue problems with spectral transformation. *Linear Algebra Appl.*, 435:601–622, 2011.

[47] R.-Y. Yu, E. de Sturler, and D. D. Johnson. A block iterative solver for complex non-Hermitian systems applied to large-scale, electronic-structure calculations. Technical report UIUCDSC-R-2002-2299 and UILU-ENG-2002-1742, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 2002.

A Other two alternatives to compute the approximate eigen-information

Proposition 4. *(Strategy A [5]) At the end of a cycle of the IB-BFGCRO-DR algorithm, if the deflation space is built on the harmonic Ritz vectors $g_i^{(HR)} \in \mathrm{span}(\widehat{\mathscr{Z}_m})$ of A with respect to $\widehat{\mathscr{Z}_m} = [U_k \quad \mathscr{Z}_m] \in \mathbb{C}^{n\times(k+n_m)}$:*

1. *The harmonic Ritz pairs $(\theta_i, \widehat{\mathscr{Z}_m} g_i^{(HR)})$ for the each restart satisfy*

$$\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_i^{(HR)} = \theta_j \underline{\mathscr{F}}_m^H \widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{Z}_m} g_i^{(HR)}, \quad \text{for } 1 \le i \le n_m, \tag{46}$$

   - *for the first restart:* $\widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{Z}_m} = \left[C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]^H [U_k, \mathscr{Z}_m] \in \mathbb{C}^{(k+n_m+p)\times(k+n_m)}$,
   - *for the subsequent restart:*

$$\widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{Z}_m} = \begin{bmatrix} C_k^H U_k & C_k^H \mathscr{Z}_m \\ \begin{bmatrix} \mathbb{V}_1^H U_k \\ 0_{(n_m-p_1)\times k} \end{bmatrix} & \mathscr{V}_m^H \mathscr{Z}_m \\ 0_{p\times k} & \begin{bmatrix} P_{m-1}^H \mathscr{Z}_m \\ \widetilde{W}_m^H \mathscr{Z}_m \end{bmatrix} \end{bmatrix}. \tag{47}$$

2. *At restart, if $G_k^{(HR)} = \left[g_{i_1}^{(HR)}, \ldots, g_{i_k}^{(HR)}\right]$ are associated with the k targeted eigenvalues, the matrices U_k^{new} and C_k^{new} to be used for the next cycle are defined by*

$$U_k^{new} = \widehat{\mathscr{Z}_m} G_k^{(HR)} R^{-1} = [U_k, \mathscr{Z}_m] G_k^{(HR)} R^{-1}, \tag{48}$$

$$C_k^{new} = \widehat{\mathscr{V}}_{m+1} Q = [C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m] Q, \tag{49}$$

   where Q and R are the factors of the reduced QR-factorization of $\underline{\mathscr{F}}_m G_k^{(HR)}$ that ensures $AU_k^{new} = C_k^{new}$ with $(C_k^{new})^H C_k^{new} = I_k$.

3. *The residual at restart $R_1^{new} = R_m^{old} = B - AX_1^{new}$ with $X_1^{new} = X_m^{old}$ is orthogonal to C_k^{new}.*

4. *At all the restarts but the first, we have the following relation that holds*

$$\mathrm{Range}(U_k^{new}) \quad \subset \quad \mathrm{Range}([C_k^{new}, \mathbb{V}_1]), \tag{50}$$

   where $\mathbb{V}_1\ \Lambda_1$ is the reduced QR-factorization of $R_1^{new} = R_m^{old} = \mathbb{V}_1 \Lambda_1$.

Proof. The proofs basically relay on some matrix computations as shortly described below:

- According to Definition 1, each harmonic Ritz pair $(\theta_i, \widehat{\mathscr{Z}_m} g_i^{(HR)})$ satisfies

$$\forall w \in \mathrm{Range}(A\widehat{\mathscr{Z}_m}) \quad w^H \left(A\widehat{\mathscr{Z}_m} g_i^{(HR)} - \theta_i\, \widehat{\mathscr{Z}_m} g_i^{(HR)}\right) = 0, \tag{51}$$

  which equivalently becomes

$$(A\widehat{\mathscr{Z}_m})^H \left(A\widehat{\mathscr{Z}_m} g_i^{(HR)} - \theta_i\, \widehat{\mathscr{Z}_m} g_i^{(HR)}\right) = 0. \tag{52}$$

  Using Equation (34) leads to

$$\left(\widehat{\mathscr{V}}_{m+1} \underline{\mathscr{F}}_m\right)^H \left(\widehat{\mathscr{V}}_{m+1} \underline{\mathscr{F}}_m g_i^{(HR)} - \theta_i\, \widehat{\mathscr{Z}_m} g_i^{(HR)}\right) = 0. \tag{53}$$

Because $\widehat{\mathscr{V}}_{m+1} = [C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]]$ generated at the end of cycle is orthonormal, (53) becomes

$$\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_i^{(HR)} - \theta_i \underline{\mathscr{F}}_m^H \widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{Z}}_m g_i^{(HR)} = 0,$$

which is the same as formulation (46). The corresponding simplified form of $\widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{Z}}_m$ in the right hand side of (46) established in the second and subsequent cycles could be deduced by the to-be-proved formula (50), which means $\mathrm{Range}(U_k^{new}) \subset \mathrm{Range}([C_k^{new}, \mathbb{V}_1]) \subset \mathrm{Range}(\widehat{\mathscr{V}}_{m+1}) = \mathrm{Range}([C_k^{new}, \mathbb{V}_1], \mathbb{V}_2, \ldots, \mathbb{V}_m, P_{m-1}, \widetilde{W}_m]), \forall i \in \{2, .., m\} \quad \mathbb{V}_i^H U_k = 0_{p_i \times k}$ and $([P_{m-1}, \widetilde{W}_m])^H U_k = 0_{p \times k}$.

- $[Q, R]$ is the reduced QR-factorization of $\underline{\mathscr{F}}_m G_k^{(HR)}$ and multiply by $G_k^{(HR)}$ on the right both sides of Equation (34). It leads to $A\widehat{\mathscr{Z}}_m G_k^{(HR)} = \widehat{\mathscr{V}}_{m+1} \underline{\mathscr{F}}_m G_k^{(HR)} = \widehat{\mathscr{V}}_{m+1} QR$, that is equivalent to $A\widehat{\mathscr{Z}}_m G_k^{(HR)} R^{-1} = \widehat{\mathscr{V}}_{m+1} \underline{\mathscr{F}}_m G_k^{(HR)} R^{-1} = \widehat{\mathscr{V}}_{m+1} Q$ that concludes the proof as $\widehat{\mathscr{V}}_{m+1} Q$ is the product of two matrices with orthonormal columns so are its columns.
- The same process for proving Corollary 1.
- Prove (50): $\mathrm{Range}(U_k^{new}) \subset \mathrm{Range}([C_k^{new}, \mathbb{V}_1])$. According to (52), the residuals of the harmonic Ritz pairs can be formulated in a matrix form as

$$R_m^{(HR)} = A\widehat{\mathscr{Z}}_m G_k^{(HR)} - \widehat{\mathscr{Z}}_m G_k^{(HR)} \,\mathrm{diag}(\theta_1, \cdots, \theta_k),$$

from which and the relationship between residual of harmonic Ritz pairs and the residual at restart: $R_m^{(HR)} = R_m \beta_{p \times k}$ proved in Section 2.3, we obtain

$$A\widehat{\mathscr{Z}}_m G_k^{(HR)} = \widehat{\mathscr{Z}}_m G_k^{(HR)} \,\mathrm{diag}(\theta_1, \cdots, \theta_k) + R_m \beta_{p \times k}. \tag{54}$$

Using (48) into $C_k^{new} = AU_k^{new}$ and from (54), we have

$$\begin{aligned} C_k^{new} &= A\widehat{\mathscr{Z}}_m G_k^{(HR)} R^{-1}, \\ &= \widehat{\mathscr{Z}}_m G_k^{(HR)} \,\mathrm{diag}(\theta_1, \cdots, \theta_k) R^{-1} + R_m \beta_{p \times k} R^{-1}, \\ &= \widehat{\mathscr{Z}}_m G_k^{(HR)} \mathrm{diag}(\theta_1, \cdots, \theta_k) R^{-1} + \mathbb{V}_1 \Lambda_1 \beta_{p \times k} R^{-1}, \end{aligned}$$

so that

$$[C_k^{new}, \mathbb{V}_1] = [\widehat{\mathscr{Z}}_m G_k^{(HR)}, \mathbb{V}_1] \begin{bmatrix} \mathrm{diag}(\theta_1, \cdots, \theta_k) R^{-1} & 0_{k \times p} \\ \Lambda_1 \beta_{p \times k} R^{-1} & I_p \end{bmatrix}, \tag{55}$$

By (48) we also have: $U_k^{new} = \widehat{\mathscr{Z}}_m G_k^{(HR)} R^{-1}$. That shows that

$$\mathrm{Range}(U_k^{new}) \subset \mathrm{Range}([\widehat{\mathscr{Z}}_m G_k^{(HR)}, \mathbb{V}_1]) = \mathrm{Range}([C_k^{new}, \mathbb{V}_1]).$$

□

Proposition 5. *(Strategy B [5]) At the end of a cycle of the IB-BFGCRO-DR algorithm, if the deflation space is built on the harmonic Ritz vectors $g_i^{(HR)} \in \mathrm{span}(\widehat{\mathscr{V}}_m)$ of $A\widehat{\mathscr{Z}}_m \widehat{\mathscr{V}}_m^H$ with respect to $\widehat{\mathscr{V}}_m = [C_k \quad \mathscr{V}_m] \in \mathbb{C}^{n \times (k + n_m)}$:*

1. *The harmonic Ritz pairs $(\theta_i, \widehat{\mathscr{V}}_m g_i^{(HR)})$ for the each restart satisfy*

$$\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_i^{(HR)} = \theta_i \mathscr{F}_m^H g_i^{(HR)} \quad \text{for } 1 \le i \le n_m, \tag{56}$$

2. *At restart, if* $G_k^{(HR)} = \left[g_{i_1}^{(HR)}, \ldots, g_{i_k}^{(HR)}\right]$ *are associated with the* k *targeted eigenvalues, the matrices* U_k^{new} *and* C_k^{new} *to be used for the next cycle are defined by*

$$U_k^{new} = \widehat{\mathscr{Z}_m} G_k^{(HR)} R^{-1} = [U_k, \mathscr{Z}_m] G_k^{(HR)} R^{-1}, \tag{57}$$

$$C_k^{new} = \widehat{\mathscr{V}}_{m+1} Q = [C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m] Q, \tag{58}$$

where Q *and* R *are the factors of the reduced QR-factorization of* $\underline{\mathscr{F}}_m G_k^{(HR)}$ *that ensures* $AU_k^{new} = C_k^{new}$ *with* $(C_k^{new})^H C_k^{new} = I_k$.

3. *The residual at restart* $R_1^{new} = R_m^{old} = B - AX_1^{new}$ *with* $X_1^{new} = X_m^{old}$ *is orthogonal to* C_k^{new}.

Proof. Given the proof essentially follows the same arguments as the ones developed for Proposition 4 or 3, the details are omitted here. □

Although the Strategy A depicted in Proposition 4 is the most efficient way among the possible three strategies described in [5] for approximating the eigen-information of the coefficient matrix A, the computational cost of the last n_m columns of $\widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{Z}_m}$ as shown in the right-hand side of Equation (47) is too heavy especially with larger n_m. Therefore, another possible alternatives are considered to reduce the computational cost of solving such general eigen-solving problem. Inspired from the way of computing eigen-information under the context of flexible GMRES with deflated restarting (FGMRES-DR) as shown in [14, Proposition 1], Strategy B shown in Proposition 5 is described for the IB-BFGCRO-DR, while which turns out to be not that suitable under the GCRO-DR context by numerical results shown in Table 7. Thus, the Strategy C is devised and described in Proposition 3, which has the same sense as Strategy A but with a lower computational cost of solving the general eigen-solving problem thanks to the structure of $\widehat{\mathscr{V}}_{m+1}^H \mathcal{W}_m = [\mathscr{T}_m \quad O_p]^T$ with the simplified form $\mathscr{T}_m$ as shown in Equation (36). From Table 7, it is easy to observed that the numerical result of IB-BFGCRO-DR with Strategy C is approximate to that with Strategy A through the later one costs the fewest $mvps$ and its.

Family number	Method	$mvps$	its
3	IB-BFGCRO-DR (Strategy A)	**1750**	**139**
	IB-BFGCRO-DR (Strategy B)	2006	171
	IB-BFGCRO-DR (Strategy C)	1797	145

Table 7: Numerical results of IB-BFGCRO-DR with three kinds of strategies in terms of $mvps$ and its, in which the involving parameters for QCD matrix are set to be $p = 12$, $m_d = 15 \times p = 180$ and $k = 90$.

B Proof of Proposition 3

Proof. The proofs basically relay on some matrix computations as shortly described below:

- According to Definition 1, each harmonic Ritz pair $(\theta_i, \mathcal{W}_m g_i^{(HR)})$ satisfies

$$\forall w \in \mathrm{Range}(A\widehat{\mathscr{Z}_m}\mathcal{W}_m^{\dagger}\mathcal{W}_m) \quad w^H\,(A\widehat{\mathscr{Z}_m}\mathcal{W}_m^{\dagger}\mathcal{W}_m g_i^{(HR)} - \theta_i\,\mathcal{W}_m g_i^{(HR)}) = 0. \tag{59}$$

  Because $\mathcal{W}_m$ is initially set to be equal to $\mathscr{V}_m$ and then is updated by (38), which has full column rank, taking a left inverse for the Moore-Penrose inverse of $\mathcal{W}_m$ makes $\mathcal{W}_m^{\dagger}\mathcal{W}_m = I$. Therefore, the second formula of (59) equivalently becomes

$$(A\widehat{\mathscr{Z}_m})^H\,(A\widehat{\mathscr{Z}_m} g_i^{(HR)} - \theta_i\,\mathcal{W}_m g_i^{(HR)}) = 0. \tag{60}$$

  Using Equation (34) leads to

$$\left(\widehat{\mathscr{V}}_{m+1}\underline{\mathscr{Z}}_m\right)^H\,\left(\widehat{\mathscr{V}}_{m+1}\underline{\mathscr{Z}}_m g_i^{(HR)} - \theta_i\,\mathcal{W}_m g_i^{(HR)}\right) = 0. \tag{61}$$

  Because $\widehat{\mathscr{V}}_{m+1} = [C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]]$ generated at the end of each cycle is orthonormal, (61) becomes

$$\underline{\mathscr{Z}}_m^H\underline{\mathscr{Z}}_m g_i^{(HR)} - \theta_i\underline{\mathscr{Z}}_m^H\widehat{\mathscr{V}}_{m+1}^H\mathcal{W}_m g_i^{(HR)} = 0,$$

  which is the same as formulation (35). The corresponding alternative form (37) with simplified right-hand side established in the second and subsequent cycles could be deduced by the (2×1) block structure of $\underline{\mathscr{Z}}$ in Equation (16) and the equality $\widehat{\mathscr{V}}_{m+1}^H\mathcal{W}_m = [\mathscr{T}_m \quad O_p]^T$ that comes from the to-be-proved formula (41), which means $\mathrm{Range}(\mathcal{W}_k^{new}) \subset \mathrm{Range}([C_k^{new}, \mathbb{V}_1]) \subset \mathrm{Range}(\widehat{\mathscr{V}}_{m+1}) = \mathrm{Range}([C_k^{new}, \mathbb{V}_1], \mathbb{V}_2, \ldots, \mathbb{V}_m, P_{m-1}, \widetilde{W}_m])$, $\forall i \in \{2, .., m\}$ $\mathbb{V}_i^H\mathcal{W}_k = 0_{p_i\times k}$ and $([P_{m-1}, \widetilde{W}_m])^H\mathcal{W}_k = 0_{p\times k}$.

- $[Q, R]$ is the reduced QR-factorization of $\underline{\mathscr{Z}}_m G_k^{(HR)}$ and multiply by $G_k^{(HR)}$ on the right both sides of Equation (34). It leads to $A\widehat{\mathscr{Z}_m}G_k^{(HR)} = \widehat{\mathscr{V}}_{m+1}\underline{\mathscr{Z}}_m G_k^{(HR)} = \widehat{\mathscr{V}}_{m+1}QR$, that is equivalent to $A\widehat{\mathscr{Z}_m}G_k^{(HR)}R^{-1} = \widehat{\mathscr{V}}_{m+1}\underline{\mathscr{Z}}_m G_k^{(HR)}R^{-1} = \widehat{\mathscr{V}}_{m+1}Q$ that concludes the proof as $\widehat{\mathscr{V}}_{m+1}Q$ is the product of two matrices with orthonormal columns so are its columns.

- The same process for proving Corollary 1.

- Prove (41): $\mathrm{Range}(\mathcal{W}_k^{new}) \subset \mathrm{Range}([C_k^{new}, \mathbb{V}_1])$. According to (60), the residuals of the harmonic Ritz pairs can be formulated in a matrix form as

$$R_m^{(HR)} = A\widehat{\mathscr{Z}_m}G_k^{(HR)} - \mathcal{W}_m G_k^{(HR)}\,\mathrm{diag}(\theta_1, \cdots, \theta_k),$$

  from which and the relationship between residuals of harmonic Ritz pairs and the linear system residuals at restart: $R_m^{(HR)} = R_m\beta_{p\times k}$ proved in Section 2.3, we obtain

$$A\widehat{\mathscr{Z}_m}G_k^{(HR)} = \mathcal{W}_m G_k^{(HR)}\,\mathrm{diag}(\theta_1, \cdots, \theta_k) + R_m\beta_{p\times k}. \tag{62}$$

  Using (39) into $C_k^{new} = AU_k^{new}$ and from (62), we have

$$\begin{aligned} C_k^{new} &= A\widehat{\mathscr{Z}_m}G_k^{(HR)}R^{-1}, \\ &= \mathcal{W}_m G_k^{(HR)}\,\mathrm{diag}(\theta_1, \cdots, \theta_k)R^{-1} + R_m\beta_{p\times k}R^{-1}, \\ &= \mathcal{W}_m G_k^{(HR)}\mathrm{diag}(\theta_1, \cdots, \theta_k)R^{-1} + \mathbb{V}_1\Lambda_1\beta_{p\times k}R^{-1}, \end{aligned}$$

so that

$$[C_k^{new}\,,\,\mathbb{V}_1] \;=\; [\mathcal{W}_m G_k^{(HR)}\,,\,\mathbb{V}_1] \begin{bmatrix} \mathrm{diag}(\theta_1,\cdots,\theta_k)R^{-1} & 0_{k\times p} \\ \Lambda_1\beta_{p\times k}R^{-1} & I_p \end{bmatrix}. \tag{63}$$

By (38) we also have: $\mathcal{W}_k^{new} = \mathcal{W}_m G_k^{(HR)} R^{-1}$. That shows that

$$\mathrm{Range}(\mathcal{W}_k^{new}) \subset \mathrm{Range}([\mathcal{W}_m G_k^{(HR)}\,,\,\mathbb{V}_1]) = \mathrm{Range}([C_k^{new}\,,\,\mathbb{V}_1]).$$

□

C IB-BFGMRES-DR: Block flexible GMRES with inexact breakdowns and deflated restarting

C.1 Block flexible Arnoldi with inexact breakdowns

Starting from an orthonormal block vector $\mathbb{V}_1$ obtained from the reduced QR-factorization of the initial residual [1] $R_0 = B - AX_0 = \mathbb{V}_1\Lambda_1$, Algorithm 2 describes details about the block flexible Arnoldi process used to construct a pair of orthonormal basis. In the no exact breakdown situation, i.e., $p_{j+1} = p_j = \ldots = p_1 = p$, the whole columns of $\mathbb{W}_j$ in step 10 of Algorithm 2 have been used to enlarge the search space, and then the block Arnoldi relation at the jth iteration is obtained as

$$A\mathscr{Z}_j = \mathscr{V}_j\mathscr{H}_j + [0_{n\times n_{j-1}}, \quad \mathbb{W}_j] = \mathscr{V}_{j+1}\underline{\mathscr{H}}_j, \tag{64}$$

in which $\mathscr{Z}_j = [\mathscr{M}_1(\mathbb{V}_1), ..., \mathscr{M}_m(\mathbb{V}_j)]$, $\mathscr{V}_j = [\mathbb{V}_1, ..., \mathbb{V}_j] \in \mathbb{C}^{n\times n_j}$ ($n_j = j \times p$) contains orthonormal columns and $\underline{\mathscr{H}}_j = \begin{bmatrix} \mathscr{H}_j \\ 0\ldots 0 \ H_{j+1,j} \end{bmatrix} \in \mathbb{C}^{n_{j+1}\times n_j}$ composed by square matrices $H_{j+1,j} \in \mathbb{C}^{p_j\times p_j}$ ($p_j = p$) is a block upper Hessenberg matrix. The minimum residual norm solution in the affine space $X_0 + \mathrm{Range}(\mathscr{Z}_j)$ can be written as $X_j = X_0 + \mathscr{Z}_j Y_j$ where

$$Y_j = \underset{Y\in\mathbb{C}^{n_j\times p}}{\mathrm{argmin}} \|\tilde{\Lambda}_j - \underline{\mathscr{H}}_j Y\|_F$$

and $\tilde{\Lambda}_j = \mathscr{V}_{j+1}^H R_0 = (\Lambda_1, 0_{n_j\times p})^T$, the columns of $\tilde{\Lambda}_j$ are the components of the individual initial residual in the residual space $\mathscr{V}_{j+1}$.

Algorithm 2 *BLOCK FLEXIBLE ARNOLDI PROCEDURE WITH BLOCKWISE MODIFIED GRAM-SCHMIDT ORTHOGONALIZATION:*

1: Given a nonsingular coefficient matrix $A \in \mathbb{C}^{n\times n}$, choose a unitary matrix V_1 of size $n \times p$
2: **for** $j = 1, 2, \ldots, m$ **do**
3: Choose a (possibly nonlinear) preconditioning operator $\mathscr{M}_j$
4: $Z_j = \mathscr{M}_j(\mathbb{V}_j)$
5: Compute $\mathbb{W}_j = AZ_j$
6: **for** $i = 1, 2, \ldots, j$ **do**
7: $H_{i,j} = \mathbb{V}_i^H \mathbb{W}_j$
8: $\mathbb{W}_j = \mathbb{W}_j - \mathbb{V}_i H_{i,j}$
9: **end for**
10: $\mathbb{W}_j = \mathbb{V}_{j+1} H_{j+1,j}$ (reduced QR-factorization)
11: **end for**

When an inexact breakdown occurs up to iteration j in Algorithm 2, the dimension of the approximation space $\mathrm{Range}(\mathscr{Z}_j)$ generated at the jth iteration is no longer equal to $j \times p$ but equal to $n_j = \sum_{i=1}^{j} p_i$ with $n_j < j \times p$. According to the inexact breakdown detecting mechanism in IB-BGMRES [32], the block flexible Arnoldi with inexact breakdowns [2] and equation (10) developed by Robbé and Sad-

[1] Out of simplicity, the initial residual R_0 in here is assumed to be of full column rank, while such assumption could be removed by introducing inexact breakdown detection in R_0 as the contents described in Appendix F.

[2] The block flexible Arnoldi with inexact breakdowns is obtained by changing the step 4 of Algorithm 3 of Appendix D into
4: Orthogonalize $A\mathscr{M}_j(\mathbb{V}_j)$ against previous block orthonormal vector $\mathscr{V}_j = [\mathbb{V}_1, \ldots, \mathbb{V}_j]$ as

$$\mathscr{L}_{1,1:j} = \mathscr{V}_j^H(A\mathscr{M}_j(\mathbb{V}_j)),\ W_j = A\mathscr{M}_j(\mathbb{V}_j) - \mathscr{V}_j\mathscr{L}_{1,1:j},\ \text{where } \mathscr{L}_{1,1:j} \text{ is a block column matrix.}$$

kane [32], the Equation (64) could be extended into

$$A\mathscr{Z}_j = \mathscr{V}_j\mathscr{H}_j + [\mathcal{Q}_{j-1}, \quad \mathbb{W}_j], \tag{65}$$

where $\mathcal{Q}_{j-1} = [Q_1, \ldots, Q_{j-1}] \in \mathbb{C}^{n\times n_{j-1}}$ is rank deficient and accounts for all the abandoned directions.

In order to characterize a minimum norm solution in the space spanned by $\mathscr{Z}_j$ using Equation (65) we need to form an orthonormal basis of the space spanned by $[\mathscr{V}_j, \mathcal{Q}_{j-1}, \mathbb{W}_j]$. This is performed by first orthogonalizing $\mathcal{Q}_{j-1}$ against $\mathscr{V}_j$, that is $\widetilde{\mathcal{Q}}_{j-1} = (I - \mathscr{V}_j\mathscr{V}_j^H)\mathcal{Q}_{j-1}$. Because $\mathcal{Q}_{j-1}$ is of low rank so is $\widetilde{\mathcal{Q}}_{j-1}$ that can be written as formula (12). Next $\mathbb{W}_j$, that is already orthogonal to $\mathscr{V}_j$, is made to be orthogonal to P_{j-1} with $\mathbb{W}_j - P_{j-1}E_j$ where $E_j = P_{j-1}^H\mathbb{W}_j$; then one computes $\widetilde{W}_jD_j$ the reduced QR-factorization of $\mathbb{W}_j - P_{j-1}E_j$. Eventually, the columns of the matrix $[\mathscr{V}_j, P_{j-1}, \widetilde{W}_j]$ form an orthonormal basis of the space spanned by $[\mathscr{V}_j, \mathcal{Q}_{j-1}, \mathbb{W}_j]$.

With this new basis Equation (65) writes

$$A\mathscr{Z}_j \quad = \quad \left[\mathscr{V}_j, [P_{j-1}, \widetilde{W}_j]\right]\underline{\tilde{\mathscr{H}}}_j, \tag{66}$$

where $\underline{\tilde{\mathscr{H}}}_j = \begin{bmatrix} \mathscr{L}_j \\ \tilde{\mathbb{H}}_j \end{bmatrix} \in \mathbb{C}^{(n_j+p)\times n_j}$ with $\tilde{\mathbb{H}}_j = \begin{bmatrix} G_{j-1} & E_j \\ 0 & D_j \end{bmatrix} \in \mathbb{C}^{p\times n_j}$ and $\mathscr{L}_j \in \mathbb{C}^{n_j\times n_j}$ owns the same details as described in formula (13), which is no longer a block upper Hessenberg as shown in the right-hand sides of (64) as soon as inexact breakdown occurs, i.e., $\exists \ell\ Q_\ell \neq 0$.

The numerical mechanism to select V_{j+1} out of $\left[P_{j-1}, \widetilde{W}_j\right]$ follows the same ideas as discussed in [2, 32] within the context of block GMRES. The governing idea consists in building the orthonormal basis for the directions that contribute the most to the individual residual norms and make them larger than the target threshold $\epsilon^{(R)}$. Based on the SVD of the coordinate vector of the least-square residual $\tilde{\Lambda}_j - \underline{\tilde{\mathscr{H}}}_jY_j = \mathbb{U}_{1,L}\Sigma_1\mathbb{V}_{1,R}^H + \mathbb{U}_{2,L}\Sigma_2\mathbb{V}_{2,R}^H$ where Σ_1 contains the singular values larger than the prescribed threshold $\epsilon^{(R)}$, they decompose $\mathbb{U}_{1,L} = \begin{pmatrix} \mathbb{U}_1^{(1)} \\ \mathbb{U}_1^{(2)} \end{pmatrix}$ in accordance with $\left[\mathscr{V}_j, [P_{j-1}, \widetilde{W}_j]\right]$, that is $\mathbb{U}_1^{(1)} \in \mathbb{C}^{n_j\times p}$ and $\mathbb{U}_2^{(2)} \in \mathbb{C}^{p\times p}$. Because, the objective is to construct orthonormal basis we consider $[\mathbb{W}_1, \mathbb{W}_2]$ unitary so that $\mathrm{Range}(\mathbb{W}_1) = \mathrm{Range}(\mathbb{U}_1^{(2)})$. The new set of orthonormal vectors selected to expand the search space as formula (18), which contributes the most to the residual. We do not give the detailed calculation and refer to [32] for a complete description, but only state that via this decomposition the main terms that appear in Equation (66) can be computed incrementally by an alternative formulation:

$$A\mathscr{Z}_j \quad = \quad \mathscr{V}_{j+1}\underline{\mathscr{L}}_j + \widetilde{\mathcal{Q}}_j, \tag{67}$$

with $\underline{\mathscr{L}}_j = \begin{bmatrix} \mathscr{L}_j \\ V_{j+1}\mathcal{Q}_{j-1}\ \mathrm{H_{j+1,j}} \end{bmatrix}$, where $\mathscr{L}_j = \begin{bmatrix} & H_{1,j} \\ \underline{\mathscr{L}}_{j-1} & \vdots \\ & H_{j,j} \end{bmatrix}$, the last block row of $\underline{\mathscr{L}}_j$ at next iteration $(j+1)$ is given by $\underline{\mathscr{L}}_{j+1,:} = \mathbb{W}_1^H\mathbb{H}_j$. The last block column of $\mathscr{L}_{j+1}$ results from the block flexible Arnoldi orthogonalization. The new compressed form of the abandoned direction $\widetilde{\mathcal{Q}}_j$ is given by the new orthonomal set of vectors

$$P_j = \left[P_{j-1}, \widetilde{W}_j\right]\mathbb{W}_2, \tag{68}$$

and the complementary part of V_{j+1} and their components in the space spanned by P_j are $G_j = \mathbb{W}_2^H\mathbb{H}_j$.

Consequently, in one cycle of IB-BFGMRES-DR, once the maximum size of the space has been reached, we have

$$\begin{aligned} A\mathscr{Z}_m &= \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] \underline{\tilde{\mathscr{F}}}_m, && (69)\\ A\mathscr{Z}_m &= \mathscr{V}_{m+1}\underline{\mathscr{L}}_m + \widetilde{\mathcal{Q}}_m, && (70)\\ X_m &= X_0 + \mathscr{Z}_m Y_m, && (71)\\ R_m &= \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] \left(\tilde{\Lambda}_m - \underline{\tilde{\mathscr{F}}}_m Y_m\right), && (72)\\ Y_m &= \underset{Y\in\mathbb{C}^{n_m\times p}}{\operatorname{argmin}} \left\|\tilde{\Lambda}_m - \underline{\tilde{\mathscr{F}}}_m Y\right\|_F, \tilde{\Lambda}_m = [\Lambda_1^T, 0_{p\times n_m}]^T. \end{aligned}$$

C.2 Harmonic Ritz vectors and residuals

We first illustrate how to compute the harmonic Ritz vectors used for deflation as described in Proposition 6 and then discuss the relation between the linear system residuals and the residuals of harmonic Ritz vectors at the restart of IB-BFGMRES-DR.

Proposition 6. *At the end of a cycle of IB-BFGMRES-DR, the updating of deflated restarting used in next cycle relies on the computation of k harmonic Ritz vectors $Y_k = \mathscr{V}_m G_k^{(HR)}$ of $A\mathscr{Z}_m\mathscr{V}_m^H$ with respect to* $\mathrm{Range}(\mathscr{V}_m)$*, where each harmonic Ritz pair $(\theta_j, \mathscr{V}_m g_j^{(HR)})$ computed at the end of cycle is supposed to satisfy*

$$(\mathscr{L}_m + \mathscr{L}_m^{-H}\tilde{\mathbb{H}}_m^H\tilde{\mathbb{H}}_m)g_j^{(HR)} = \theta_j\, g_j^{(HR)} \quad \textit{for } 1 \le j \le k, \tag{73}$$

where $\mathscr{L}_m \in \mathbb{C}^{n_m\times n_m}$ and $\tilde{\mathbb{H}}_m \in \mathbb{C}^{p\times n_m}$.

Proof. According to Definition 1, each harmonic Ritz pair $(\theta_j, \mathscr{V}_m g_j^{(HR)})$ satisfies

$$\forall w \in \mathrm{Range}(A\mathscr{Z}_m\mathscr{V}_m^H\mathscr{V}_m) \quad w^H\,(A\mathscr{Z}_m\mathscr{V}_m^H\mathscr{V}_m g_j^{(HR)} - \theta_j\,\mathscr{V}_m g_j^{(HR)}) = 0,$$

which is equivalent to

$$(A\mathscr{Z}_m)^H\,(A\mathscr{Z}_m g_j^{(HR)} - \theta_j\,\mathscr{V}_m g_j^{(HR)}) = 0,$$

by the orthonormality of $\mathscr{V}_m$. Substituting (69) into the above equation yields

$$(\left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] \underline{\tilde{\mathscr{F}}}_m)^H\,(\left[\mathscr{V}_j, [P_{m-1}, \widetilde{W}_m]\right] \underline{\tilde{\mathscr{F}}}_m\, g_j^{(HR)} - \theta_j\,\mathscr{V}_m\, g_j^{(HR)}) = 0. \tag{74}$$

Because of the structure of $\underline{\tilde{\mathscr{F}}}_m$ and the orthonormality of $[\mathscr{V}_m, P_{m-1}, \widetilde{W}_m]$, (74) becomes

$$(\mathscr{L}_m^H\,\mathscr{L}_m + \tilde{\mathbb{H}}_m^H\tilde{\mathbb{H}}_m)g_j^{(HR)} = \theta_j\,\mathscr{L}_m^H g_j^{(HR)}, \tag{75}$$

which completes the proof since $\mathscr{L}_m$ is assumed to be nonsingular. □

Assume $R_m^{LS} = \left(\tilde{\Lambda}_m - \underline{\tilde{\mathscr{F}}}_m Y_m\right) \in \mathbb{C}^{(n_m+p)\times p}$, the residual of linear system presented in (72) could be simplified as

$$R_m = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] R_m^{LS} \in \mathbb{C}^{n\times p}. \tag{76}$$

Denote the corresponding residual of harmonic vectors as R_m^{har} similarly, which owns form as

$$R_m^{har} = A\mathscr{Z}_m G_k^{(HR)} - \mathscr{V}_m G_k^{(HR)}\,\mathrm{diag}(\theta_1, \ldots, \theta_k) \in \mathbb{C}^{n\times k}. \tag{77}$$

Given that both R_m and R_m^{har} are resided in the subspace $\mathrm{Range}(\left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]) \in \mathbb{C}^{n\times(n_m+p)}$ and are orthogonal to the same subspace $\mathrm{Range}(A\mathscr{Z}_m) \in \mathbb{C}^{n\times n_m}$. Therefore, the residuals of linear system R_m and the residuals of harmonic Ritz vectors R_m^{har} are in the same p-dimensional space denoted as $\mathrm{Range}(A\mathscr{Z}_m)^{\perp} \cap \mathrm{Range}(\left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right])$, which means there exists a $\beta_{p\times k} \in \mathbb{C}^{p\times k}$ such that $R_m^{har} = R_m\beta_{p\times k}$. According to (76) and (77), such collinear relationship between the linear system residuals and residuals of harmonic Ritz vectors could be further described as the following formula

$$A\mathscr{Z}_m G_k^{(HR)} = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] \underline{G} \begin{bmatrix} diag(\theta_1, \cdots, \theta_k) \\ \beta_{p\times k} \end{bmatrix}, \tag{78}$$

where $G_k^{(HR)} = [g_1^{(HR)}, \ldots, g_k^{(HR)}] \in \mathbb{C}^{n_m\times k}$, $\underline{G} = \begin{bmatrix} G_k^{(HR)} & R_m^{LS} \\ 0_{p\times k} & \end{bmatrix} \in \mathbb{C}^{(n_m+p)\times(k+p)}$, $\beta_{p\times k} = (\beta_1, \cdots, \beta_k) \in \mathbb{C}^{p\times k}$ and $\beta_i \in \mathbb{C}^p$ $(1 \le i \le k)$. Based on (66) and the orthonormality of $\left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]$, relation (78) can be also expressed as

$$\underline{\tilde{\mathscr{F}}}_m G_k^{(HR)} = \underline{G} \begin{bmatrix} diag(\theta_1, \cdots, \theta_k) \\ \beta_{p\times k} \end{bmatrix}, \tag{79}$$

which is the block form of Equation (3.4) shown in [2, Lemma3.3].

C.3 Flexible block GMRES with inexact breakdowns at restart

In this subsection, the forthcoming Theorem 2 will be presented to illustrate that the flexible Arnoldi relation with inexact breakdowns described in (66) and (67) (or in (69) and (70)) still hold at restart. Firstly, let us denote $\underline{G} = Q_{\underline{G}} R_{\underline{G}}$ the reduced QR-factorization of $\underline{G}$ shown in (79) and the reduced factors could be partitioned as

$$Q_{\underline{G}} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \\ 0_{p\times k} & \end{bmatrix} \in \mathbb{C}^{(n_m+p)\times(k+p)}, \tag{80}$$

$$R_{\underline{G}} = \begin{bmatrix} \Theta_1 & \Theta_2 \\ 0_{p\times k} & \end{bmatrix} \in \mathbb{C}^{(n_m+p)\times(k+p)}, \tag{81}$$

with $\Gamma_1 = Q_{\underline{G}}(1:n_m, 1:k)$, $\Gamma_2 = Q_{\underline{G}}(:, k+1:k+p)$, $\Theta_1 = R_{\underline{G}}(1:n_m, 1:k)$, $\Theta_2 = R_{\underline{G}}(:, k+1:k+p)$ and

$$G_k^{(HR)} = \Gamma_1\Theta_1, \tag{82}$$

$$R_m^{LS} = Q_{\underline{G}}\Theta_2. \tag{83}$$

Theorem 2. *At each restart of block flexible GMRES with inexact breakdowns and deflated restarting, the initial block-flexible-Arnoldi-like relation (66) and (67) still hold in exact arithmtic as*

$$A\mathscr{Z}_1^{new} = \left[\mathscr{V}_1^{new}, [P_0, \widetilde{W}_1]^{new}\right] \underline{\tilde{\mathscr{F}}}_1^{new}, \tag{84}$$

$$A\mathscr{Z}_1^{new} = \mathscr{V}_2^{new}\,\underline{\mathscr{L}}_1^{new} + \widetilde{\mathcal{Q}}_1^{new}, \tag{85}$$

$$R_0^{new} = R_m = \left[\mathscr{V}_1^{new}, [P_0, \widetilde{W}_1]^{new}\right] \tilde{\Lambda}_1^{new} \text{ and } \tilde{\Lambda}_1^{new} = \Theta_2, \tag{86}$$

with

$$\begin{aligned}
\mathscr{Z}_1^{new} &= \mathscr{Z}_m\Gamma_1,\ \left[\mathscr{V}_1^{new}, [P_0, \widetilde{W}_1]^{new}\right] = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] Q_{\underline{G}},\\
\mathscr{V}_1^{new} &= \mathscr{V}_m\Gamma_1,\ [P_0, \widetilde{W}_1]^{new} = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] \Gamma_2,\\
\underline{\tilde{\mathscr{F}}}_1^{new} &= \begin{bmatrix} \mathscr{L}_1^{\mathrm{new}} \\ \tilde{\mathbb{H}}_1^{new} \end{bmatrix} \textit{ and } \mathscr{L}_1^{new} = \Gamma_1^H \mathscr{L}_m\Gamma_1,\ \tilde{\mathbb{H}}_1^{new} = \Gamma_2^H \underline{\tilde{\mathscr{F}}}_m\Gamma_1,\\
\mathbb{V}_2^{new} &= [P_0, \widetilde{W}_1]^{new}\mathbb{W}_1^{new\,H},\ \mathscr{V}_2^{new} = [\mathscr{V}_1^{new}\ \mathbb{V}_2^{new}],\\
\mathscr{L}_{2,:}^{new} &= \mathbb{W}_1^{new\,H}\tilde{\mathbb{H}}_1^{new},\ \underline{\mathscr{L}}_1^{new} = \begin{bmatrix} \mathscr{L}_1^{new} \\ \mathscr{L}_{2,:}^{new} \end{bmatrix},\\
P_1^{new} &= [P_0, \widetilde{W}_1]^{new}\mathbb{W}_2^{new},\ G_1^{new} = \mathbb{W}_2^{new\,H}\tilde{\mathbb{H}}_1^{new},\ \widetilde{\mathcal{Q}}_1^{new} = P_1^{new}G_1^{new},
\end{aligned}$$

where $\mathbb{W}_1^{new}$ *and* $\mathbb{W}_2^{new}$ *satisfy*

$$\mathrm{Range}(\mathbb{W}_1^{new}) = \mathrm{Range}(\mathbb{U}_1^{new(2)}) \textit{ with } \mathbb{U}_{1,L}^{new} = \begin{bmatrix} \mathbb{U}_1^{new(1)} \\ \mathbb{U}_1^{new(2)} \end{bmatrix} \textit{ and } [\mathbb{W}_1^{new}\ \mathbb{W}_2^{new}] \textit{ is unitary}$$

with

$$\tilde{\Lambda}_1^{new} - \underline{\tilde{\mathscr{F}}}_1^{new} Y_1^{new} = \mathbb{U}_{1,L}^{new}\Sigma_1^{new}\mathbb{V}_{1,R}^{new\,H} + \mathbb{U}_{2,L}^{new}\Sigma_2^{new}\mathbb{V}_{2,R}^{new\,H},$$

where $\sigma_{min}(\Sigma_1^{new}) \geq \epsilon^{(R)} \geq \sigma_{max}(\Sigma_2^{new})$, *the SVD to detect inexact breakdown in the restarting block residual where*

$$Y_1^{new} = \underset{Y\in\mathbb{C}^{n_1\times p}}{\mathrm{argmin}} \left\| \tilde{\Lambda}_1^{new} - \underline{\tilde{\mathscr{F}}}_1^{new} Y \right\|_F.$$

Proof. Starting from the relationship between residual and harmonic Ritz vectors as shown in (78), let's substitute $\underline{G}$ by these reduced factors $Q_{\underline{G}}$ in (80) and $R_{\underline{G}}$ in (81) obtained by its reduced QR-factorization and change $G_k^{(HR)}$ by relation (82), then we have

$$A\mathscr{Z}_m\Gamma_1 = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] Q_{\underline{G}}R_{\underline{G}} \begin{bmatrix} diag(\theta_1, \cdots, \theta_k) \\ \beta_{p\times k} \end{bmatrix} \Theta_1^{-1}$$

by the nonsingularity of Θ_1, which could be rewritten as

$$A\mathscr{Z}_m\Gamma_1 = \left[\mathscr{V}_m\Gamma_1, [\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]]\Gamma_2\right] R_{\underline{G}} \begin{bmatrix} diag(\theta_1, \cdots, \theta_k) \\ \beta_{p\times k} \end{bmatrix} \Theta_1^{-1} \tag{87}$$

because of the partition of $Q_{\underline{G}}$ shown in (80). Then, repeating the same processes described above, the corresponding formula (79) could also be reformed as

$$\underline{\tilde{\mathscr{F}}}_m\Gamma_1 = Q_{\underline{G}}R_{\underline{G}} \begin{bmatrix} diag(\theta_1, \cdots, \theta_k) \\ \beta_{p\times k} \end{bmatrix} \Theta_1^{-1},$$

from which, we have

$$R_{\underline{G}} \begin{bmatrix} diag(\theta_1, \cdots, \theta_k) \\ \beta_{p\times k} \end{bmatrix} \Theta_1^{-1} = Q_{\underline{G}}^H \underline{\tilde{\mathscr{F}}}_m\Gamma_1.$$

According to the structure of $Q_{\underline{G}}$ and $\mathscr{F}_m$ as shown in (80) and (66), we obtain

$$R_{\underline{G}} \begin{bmatrix} diag(\theta_1, \cdots, \theta_k) \\ \beta_{p\times k} \end{bmatrix} \Theta_1^{-1} = \begin{bmatrix} \Gamma_1^{\mathrm{H}}\mathscr{L}_{\mathrm{m}}\Gamma_1 \\ \Gamma_2^H \underline{\tilde{\mathscr{F}}}_m\Gamma_1 \end{bmatrix}. \tag{88}$$

If we denote

$$\begin{aligned}
\mathscr{Z}_1^{new} &= \mathscr{Z}_m\Gamma_1,\ \mathscr{V}_1^{new} = \mathscr{V}_m\Gamma_1,\ [P_0, \widetilde{W}_1]^{new} = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]\Gamma_2, \\
\mathscr{L}_1^{new} &= \Gamma_1^H\mathscr{L}_m\Gamma_1,\ \tilde{\mathbb{H}}_1^{new} = \Gamma_2^H\underline{\tilde{\mathscr{F}}}_m\Gamma_1, \underline{\tilde{\mathscr{F}}}_1^{new} = \begin{bmatrix} \mathscr{L}_1^{\text{new}} \\ \tilde{\mathbb{H}}_1^{new} \end{bmatrix},
\end{aligned}$$

and substitute (88) into (87), then (84) is proven.

Next, show that equality (85) holds. Given $[\mathbb{W}_1^{new}, \mathbb{W}_2^{new}]$ is unitary, we have

$$[P_0, \widetilde{W}_1]^{new} = [P_0, \widetilde{W}_1]^{new}[\mathbb{W}_1^{new\,H}\mathbb{W}_1^{new} + \mathbb{W}_2^{new\,H}\mathbb{W}_2^{new}],$$

and substituting this into (84) gives

$$\begin{aligned}
A\mathscr{Z}_1^{new} &= \left[\mathscr{V}_1^{new}, [P_0, \widetilde{W}_1]^{new}[\mathbb{W}_1^{new\,H}\mathbb{W}_1^{new} + \mathbb{W}_2^{new\,H}\mathbb{W}_2^{new}]\right] \begin{bmatrix} \mathscr{L}_1^{\text{new}} \\ \tilde{\mathbb{H}}_1^{new} \end{bmatrix}, \\
&= \mathscr{V}_1^{new}\mathscr{L}_1^{new} + [P_0, \widetilde{W}_1]^{new}[\mathbb{W}_1^{new\,H}\mathbb{W}_1^{new} + \mathbb{W}_2^{new\,H}\mathbb{W}_2^{new}]\tilde{\mathbb{H}}_1^{new}, \\
&= \mathscr{V}_1^{new}\mathscr{L}_1^{new} + [P_0, \widetilde{W}_1]^{new}\mathbb{W}_1^{new\,H}\mathbb{W}_1^{new}\tilde{\mathbb{H}}_1^{new} + [P_0, \widetilde{W}_1]^{new}\mathbb{W}_2^{new\,H}\mathbb{W}_2^{new}\tilde{\mathbb{H}}_1^{new}, \\
&= \mathscr{V}_1^{new}\mathscr{L}_1^{new} + \mathscr{V}_2^{new}\mathscr{L}_{2,:}^{new} + P_1^{new}G_1^{new}, \\
&= [\mathscr{V}_1^{new}\, V_2^{new}] \begin{bmatrix} \mathscr{L}_1^{new} \\ \mathscr{L}_{2,:}^{new} \end{bmatrix} + P_1^{new}G_1^{new},
\end{aligned}$$

which is relation (85).

From relations (76) and (83), at restart we have

$$\begin{aligned}
R_0^{new} &= R_m = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] R_m^{LS} \\
&= \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right] Q_{\underline{G}}\Theta_2 = \left[\mathscr{V}_1^{new}, [P_0, \widetilde{W}_1]^{new}\right] \tilde{\Lambda}_1^{new}.
\end{aligned}$$

This complete the proof. □

D Block Arnoldi with inexact breakdown detection after the initial residuals

Algorithm 3 BLOCK ARNOLDI USING R-CRITERION TO DETECT INEXACT BREAKDOWNS.

1: Assuming $B = [b^{(1)}, \cdots, b^{(p)}] \in \mathbb{C}^{n\times p}$ is of full rank, choose the initial block guess X_0, and compute the corresponding nonsingular initial block residual $R_0 = B - AX_0$.

2: Form initial unitary matrix $\mathbb{V}_1$ from initial block residual $R_0 = \mathbb{V}_1\Lambda_1$ with reduced QR-factorization. Let $P_0 = 0, G_0 = 0$ and $\underline{\mathscr{L}}_0 = [\,]$. Choose a targeted backward error ε and set the corresponding $\epsilon^{(R)}$ (or if $\left\|b^{(i)}\right\|_2 = 1$, then $\epsilon^{VA} = [\epsilon^{(1)}, \cdots, \epsilon^{(p)}]$ and $D_\epsilon = \text{ diag } ((\epsilon^{VA})^{-1})$ in step 7. *iii*.

3: **for** $j = 1, 2, \ldots, m$ **do**

4: Orthogonalize $A\mathbb{V}_j$ against previous block orthonormal vector $\mathscr{V}_j = [\mathbb{V}_1, \ldots, \mathbb{V}_j]$ as

$$\mathscr{L}_{1,1:j} = \mathscr{V}_j^H(A\mathbb{V}_j),\ W_j = A\mathbb{V}_j - \mathscr{V}_j\mathscr{L}_{1,1:j},\ \text{where } \mathscr{L}_{1,1:j} \text{ is a block column matrix.}$$

5: Set $\mathscr{L}_j = \begin{bmatrix}\underline{\mathscr{L}}_{j-1}, & \mathscr{L}_{1,1:j}\end{bmatrix} \in \mathbb{C}^{n_j\times n_j}$.

6: Orthogonalize W_j against P_{j-1} and carry out its reduced QR-factorization as

$$E_j = P_{j-1}^H W_j, \quad \widetilde{W}_j D_j = W_j - P_{j-1}E_j.$$

7: Compute Y_j the solution of the least-squares problem: $\min\limits_{Y\in\mathbb{C}^{n_j\times p}} \left\|\tilde{\Lambda}_j - \underline{\tilde{\mathscr{F}}}_j Y\right\|_F$ with

$$\tilde{\Lambda}_j = \begin{pmatrix}\Lambda_1\\ 0\end{pmatrix} \in \mathbb{C}^{(n_j+p)\times p}, \underline{\tilde{\mathscr{F}}}_j = \begin{pmatrix}\mathscr{L}_{\mathrm{j}}\\ \tilde{\mathbb{H}}_j\end{pmatrix} \in \mathbb{C}^{(n_j+p)\times n_j}, \text{ and } \tilde{\mathbb{H}}_j = \begin{pmatrix}\mathrm{G_{j-1}} & \mathrm{E_j}\\ 0 & \mathrm{D_j}\end{pmatrix} \in \mathbb{C}^{p\times n_j}.$$

Based on different criteria to carry out the singular value decomposition algorithm to detect inexact breakdowns in residuals (i.e., implement 7.i., 7.ii. or 7.iii. distinguished by different color for the forward step)

7.i. **if** based on Section 2.2, **then**:$(\tilde{\Lambda}_j - \underline{\tilde{\mathscr{F}}}_j Y_j) = \mathbb{U}_{1,L}\Sigma_1\mathbb{V}_{1,R}^H + \mathbb{U}_{2,l}\Sigma_2\mathbb{V}_{2,R}^H$, with $\sigma_{\min}(\Sigma_1) \geq \epsilon^{(R)} > \sigma_{\max}(\Sigma_2)$.

7.ii. **if** computational blocking based on Section 3.1 is considered in subspace expansion policy, **then**:
$(\tilde{\Lambda}_j - \underline{\tilde{\mathscr{F}}}_j Y_j) = \mathbb{U}_{1,L_{CB}}\Sigma_{1_{CB}}\mathbb{V}_{1,R_{CB}}^H + \mathbb{U}_{2,L_{CB}}\Sigma_{2_{CB}}\mathbb{V}_{2,R_{CB}}^H$, with $\sigma_{\min}(\Sigma_{1_{CB}}) \geq \epsilon^{(R)} > \sigma_{\max}(\Sigma_{2_{CB}})$.

Then, $\mathbb{U}_1 = \mathbb{U}_{1,L_{CB}}(:, 1 : p_j^{CB})$ with $p_j^{CB} = min(p^{CB}, size(\Sigma_{1_{CB}}, 2))$ and $1 \leq p^{CB} \leq p$.

7.iii. **if** individual convergence criterion based on Section 3.2 is considered for each right-hand side, **then**:
$(\tilde{\Lambda}_j - \underline{\tilde{\mathscr{F}}}_j Y_j)D_\epsilon = \mathbb{U}_{1,L}\Sigma_1\mathbb{V}_{1,R}^H + \mathbb{U}_{2,L}\Sigma_2\mathbb{V}_{2,R}^H$, where $\sigma_{\min}(\Sigma_1) \geq \epsilon^{(R)} > \sigma_{\max}(\Sigma_2)$ with $\epsilon^{(R)} = 1$.

Compute $\mathbb{W}_1$ and $\mathbb{W}_2$ such that

$\text{Range}(\mathbb{W}_1) = \text{Range}(\mathbb{U}_1^{(2)})$ with $\mathbb{U}_{1,L} = \begin{pmatrix}\mathbb{U}_1^{(1)}\\ \mathbb{U}_1^{(2)}\end{pmatrix}$ and $[\mathbb{W}_1, \quad \mathbb{W}_2]$ is unitary.

Compute orthonormal matrices $\mathbb{V}_{j+1}$ and P_j, the last block row matrix $\mathscr{L}_{j+1,:}$ of $\underline{\mathscr{L}}_j$, and G_j as

$$\mathbb{V}_{j+1} = \begin{bmatrix}P_{j-1}, & \widetilde{W}_j\end{bmatrix}\mathbb{W}_1, \quad P_j = \begin{bmatrix}P_{j-1}, & \widetilde{W}_j\end{bmatrix}\mathbb{W}_2, \quad \mathscr{L}_{j+1,:} = \mathbb{W}_1^H\tilde{\mathbb{H}}_j, \quad G_j = \mathbb{W}_2^H\tilde{\mathbb{H}}_j.$$

8: Set $\underline{\mathscr{L}}_j = \begin{pmatrix}\mathscr{L}_j\\ \mathscr{L}_{j+1,:}\end{pmatrix}$.

9: **end for**

E The SVD decomposition of the residual block involved and the solution of the least-squares problem solution

The Inexact Breakdown mechanism (IB) allows to extract from the residual spaces new directions to expand the search space at the next iteration of the block method. The selection consists in extracting the directions that contribute the most to the residual block and is based on the SVD of the least-squares residuals. In this section, we detail how the solution of the least-squares problem (15) enables to compute easily and cheaply the SVD of the associated residual block. The least-squares problems that writes

$$Y_j = \underset{Y \in \mathbb{C}^{(k+n_j)\times p}}{\operatorname{argmin}} \left\| \Lambda_j - \underline{\mathscr{F}}_j Y \right\|_F , \tag{89}$$

is solved using a full QR-factorization of $\underline{\mathscr{F}}_j = Q_j^{LS} R_j^{LS}$ so that the residual of the least-squares problem is given by (44) that allows to compute it without forming Y_j:

$$\begin{aligned} \Lambda_j - \underline{\mathscr{F}}_j Y_j &= \Lambda_j - Q_j^{LS} R_j^{LS} Y_j \\ &= Q_j^{LS} \left((Q_j^{LS})^H \Lambda_j - R_j^{LS} Y_j \right) \\ &= Q_j^{LS} \begin{pmatrix} 0_{n_j \times p} \\ R_j^{\ell s} \end{pmatrix} \end{aligned}$$

where $R_j^{\ell s} \in \mathbb{C}^{p\times p}$ are the last p rows of $(Q_j^{LS})^H \Lambda_j$. The SVD of $R_j^{\ell s}$ can be written

$$R_j^{\ell s} = U_{\ell s} \Sigma V_{\ell s}^H ,$$

so that the SVD of the least-squares residual is

$$\Lambda_j - \underline{\mathscr{F}}_j Y_j = \underbrace{Q_j^{LS} \begin{pmatrix} 0_{(n_j+k)\times p} & I_{n_j+k} \\ U_{\ell s} & 0_{p\times(n_j+k)} \end{pmatrix}}_{\text{Unitary}} \begin{pmatrix} \Sigma \\ 0_{(n_j+k)\times p} \end{pmatrix} V_{\ell s}^H .$$

F Inexact breakdown detection in the initial residuals R_1

In that case, only a candidate subspace of the space spanned by R_1 will be selected to define the first search space but abandoned directions should be kept in the basis of the residual space. This has two main consequences:

1. The first iteration needs some extra attention to setup P_0 and G_0 defined by (12).
2. A consequence of having abandoned directions in the first search space is that the projection of the initial residual in the residual space, that defines the right-hand side of the least-squares problem solved at each block iteration, will not longer have the nested block structure that is expanded by a $p \times p$ zero block at each block iteration as presented in (21).

Let consider the reduced SVD of R_1 denoted

$$R_1 = [\mathbb{V}_1, P_0] \begin{bmatrix} \Sigma_{p_1} & \\ & \Sigma_{q_1} \end{bmatrix} \mathbb{V}_{R_1}^H = [\mathbb{V}_1, P_0] \hat{\Lambda}_1$$

where Σ_{p_1} contains the p_1 singular values of R_1 larger than $\epsilon^{(R)}$, and Σ_{q_1} the ones lower than $\epsilon^{(R)}$ with $p_1 + q_1 = p$. Consequently $\mathscr{V}_1 = \mathbb{V}_1$ will serve to span the first search space and P_0 will be abandoned for this first block iteration that will be run as follows.

1. we form $W_1 = A\mathbb{V}_1$ and orthogonalize it against the set of orthonormal vectors that are part of the residual space $[C_k, \mathbb{V}_1, P_0]$ by computing $\mathcal{B}_1 = C_k^H W_1$, $\mathscr{L}_{1,1} = \mathscr{V}_1^H W_1$ and $E_1 = P_0^H W_1$.
2. we compute $\bar{W}_1$ the resulting block that writes $\bar{W}_1 = W_1 - C_k\mathcal{B}_1 - \mathscr{V}_1\mathscr{L}_{1,1} - P_0 E_1$ and $\bar{W}_1 = \widetilde{W}_1 D_1$ its reduced QR-factorization.
3. In matrix form this also writes

$$W_1 = A\mathbb{V}_1 = \left[C_k, \mathbb{V}_1, P_0, \widetilde{W}_1\right] \begin{bmatrix} \mathcal{B}_1 \\ \mathscr{L}_{1,1} \\ E_1 \\ D_1 \end{bmatrix}$$

   So that we have the first Arnoldi-like relation

$$A[U_k, \mathbb{V}_1] = [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1] \underline{\mathscr{F}}_1$$

   with

$$\underline{\mathscr{F}}_1 = \begin{bmatrix} I_k & \mathcal{B}_1 \\ & \mathscr{L}_{1,1} \\ 0_{(n_1+p)\times k} & E_1 \\ & D_1 \end{bmatrix}$$

4. Next, we have to define the minimum norm solution $X_2 = X_1 + [U_k, \mathbb{V}_1]Y$ and notice that R_1 belongs to the space $[C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]$ where its components in this orthogonal basis are given by $[C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H R_1$, we have

$$\begin{aligned} \|B - AX_2\|_F &= \|R_1 - A\,[U_k, \mathscr{V}_1]\,Y\|_F \\ &= \|R_1 - [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]\underline{\mathscr{F}}_1 Y\|_F \\ &= \|[C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H R_1 - \underline{\mathscr{F}}_1 Y\|_F \\ &= \|[C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [\mathscr{V}_1, P_0]\hat{\Lambda}_1 - \underline{\mathscr{F}}_1 Y\|_F \end{aligned}$$

so that the right-hand side of the least-squares reads

$$\Lambda_1 = [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [\mathscr{V}_1, P_0]\hat{\Lambda}_1 = \begin{bmatrix} 0_{k\times p_1} & 0_{k\times q_1} \\ I_{p_1} & 0_{p_1\times q_1} \\ 0_{q_1\times p_1} & I_{q_1} \\ 0_{p_1\times p_1} & 0_{p_1\times q_1} \end{bmatrix} \hat{\Lambda}_1 \in \mathbb{C}^{(k+p_1+p)\times p}.$$

5. Compute Y_1 the solution of the first new least-squares problem

$$Y_1 = \underset{Y\in\mathbb{C}^{(k+n_1)\times p}}{\operatorname{argmin}} \|\Lambda_1 - \underline{\mathscr{F}}_1 Y\|_F .$$

As in regular IB-BGCRO-DR iteration carry out the singular value decomposition algorithm to detect inexact breakdown in the new residual block

$$(\Lambda_1 - \underline{\mathscr{F}}_1 Y_1) = \mathbb{U}_{1,L}\Sigma_1 \mathbb{V}_{1,R}^H + \mathbb{U}_{2,L}\Sigma_2 \mathbb{V}_{2,R}^H, \text{ where } \sigma_{\min}(\Sigma_1) \geq \epsilon^{(R)} > \sigma_{\max}(\Sigma_2).$$

Compute $\mathbb{W}_1$ and $\mathbb{W}_2$ such that $\mathrm{Range}(\mathbb{W}_1) = \mathrm{Range}(\mathbb{U}_1^{(2)}) \in \mathbb{C}^{p\times p_2}$ with $\mathbb{U}_{1,L} = \begin{pmatrix} \mathbb{U}_1^{(1)} \\ \mathbb{U}_1^{(2)} \end{pmatrix} \in \mathbb{C}^{(k+n_1+p)\times p_2}$ and $[\mathbb{W}_1, \quad \mathbb{W}_2]$ is unitary and $\mathbb{W}_2 \in \mathbb{C}^{p\times q_2}$ with $p_2 + q_2 = p$.
Compute new orthonormal matrices $\mathbb{V}_2$ and P_1, the last block row matrix $\mathscr{L}_{2,:}$ of $\underline{\mathscr{L}}_1$, and G_1 as

$$\mathbb{V}_2 = [P_0, \widetilde{W}_1]\mathbb{W}_1 \in \mathbb{C}^{n\times p_2},\ P_1 = [P_0, \widetilde{W}_1]\mathbb{W}_2 \in \mathbb{C}^{n\times q_2}, \mathscr{L}_{2,:} = \mathbb{W}_1^H \mathbb{H}_1 \in \mathbb{C}^{p_2\times p_1},\ G_1 = \mathbb{W}_2^H \mathbb{H}_1 \in \mathbb{C}^{q_2\times p_1}.$$

6. Set $\underline{\mathscr{L}}_1 = \begin{pmatrix} \mathscr{L}_1 \\ \mathscr{L}_{2,:} \end{pmatrix} \in \mathbb{C}^{(p_1+p_2)\times p_1} = \mathbb{C}^{n_2\times p_1}$.

Refer to Algorithm 4 for the corresponding pseudocode form of these contents described in above item 1-6. The subsequent iterations will then be run by IB-BGCRO-DR as discussed in the core of the report.

Whenever an inexact breakdown is detected in R_1, some of its directions, namely P_0 are abandoned and put into the residual space for the first iteration. The beauty of the IB mechanism is that the abandoned directions at a given iteration can be re-introduced in some subsequent iterations. One of the consequences, is that the last q_1 columns of the least-squares residual will evolve from one iteration to the next, depending on how P_0 directions will be re-introduced in the search space along the iterations. There is still a way to incrementally update it as discussed in the next proposition.

Proposition 7. *If an inexact breakdown is detected in R_1, that is*

$$R_1 = [\mathbb{V}_1, P_0] \begin{bmatrix} \Sigma_{p_1} & \\ & \Sigma_{q_1} \end{bmatrix} \mathbb{V}_{R_1}^H = [\mathbb{V}_1, P_0]\hat{\Lambda}_1,$$

where Σ_{p_1} contains the p_1 singular values of R_1 larger than $\epsilon^{(R)}$, and Σ_{q_1} the ones lower than $\epsilon^{(R)}$ with $p_1 + q_1 = p$.

At each iteration of IB-BGCRO-DR, the new least-squares problem reads

$$Y_{j+1} = \underset{Y\in\mathbb{C}^{(k+n_{j+1})\times p}}{\operatorname{argmin}} \left\|\Lambda_{j+1} - \underline{\mathscr{F}}_{j+1} Y\right\|_F, \quad \Lambda_{j+1} \in \mathbb{C}^{(k+n_{j+1}+p)\times p}, \quad j = 0, 1, 2, \cdots \tag{90}$$

with the updated right-hand sides being

$$\Lambda_{j+1} = \begin{bmatrix} 0_{k\times p_1} & 0_{k\times q_1} \\ \begin{bmatrix} I_{p_1} \\ 0_{(n_j+p-p_1)\times p_1} \end{bmatrix} & \Phi_{j+1} \\ 0_{p_{j+1}\times p_1} & 0_{p_{j+1}\times q_1} \end{bmatrix} \hat{\Lambda}_1 \text{ with } \Phi_1 = \begin{bmatrix} 0_{p_1\times q_1} \\ I_{q_1} \end{bmatrix} \in \mathbb{C}^{p\times q_1}$$

and

$$\Phi_{j+1} = \left[\begin{array}{c} \Phi_j(1:n_j,:) \\ [\mathbb{W}_1, \mathbb{W}_2]^H \left[\begin{array}{c} \Phi_j(n_j+1:n_j+q_j,:) \\ 0_{p_j \times q_1} \end{array} \right] \end{array} \right] \in \mathbb{C}^{(n_j+p)\times q_1} \textit{ for } j = 1, 2, \cdots,$$

where $q_j = p - p_j$ *and* $[\mathbb{W}_1, \mathbb{W}_2]$ *is unitary as defined in step 7 of Algorithm 3.*

Proof. The initial residuals with inexact breakdown detection at restart could be described as

$$\begin{aligned} R_1 &= [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1][C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H R_1 = [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1][C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [\mathbb{V}_1, P_0]\hat{\Lambda}_1 \\ &= [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]\Lambda_1 \text{ with } \Lambda_1 = \left[\begin{array}{cc} 0_{k\times p_1} & 0_{k\times q_1} \\ I_{p_1} & 0_{p_1\times q_1} \\ 0_{q_1\times p_1} & I_{q_1} \\ 0_{p_1\times p_1} & 0_{p_1\times q_1} \end{array} \right] \hat{\Lambda}_1 \in \mathbb{C}^{(k+n_1+p)\times p}, \end{aligned}$$

That can also be written

$$\Lambda_1 = \left[\begin{array}{cc} 0_{k\times p_1} & 0_{k\times q_1} \\ I_{p_1} & \Phi_1 \\ 0_{q_1\times p_1} & \\ 0_{p_1\times p_1} & 0_{p_1\times q_1} \end{array} \right] \hat{\Lambda}_1.$$

The right-hand sides of the least-squares problem at iteration $(j+1)$ for $j = 1, 2, \cdots$, is defined by

$$\begin{aligned} \Lambda_{j+1} &= [C_k, \mathscr{V}_{j+1}, [P_j, \widetilde{W}_{j+1}]]^H R_1 = [C_k, \mathscr{V}_j, V_{j+1}, [P_j, \widetilde{W}_{j+1}]]^H R_1 \\ &= \left[C_k, \mathscr{V}_j, [P_{j-1}, \widetilde{W}_j]\mathbb{W}_1, [P_{j-1}, \widetilde{W}_j]\mathbb{W}_2, \widetilde{W}_{j+1}\right]^H R_1 \\ &= \left[C_k, \mathscr{V}_j, [P_{j-1}, \widetilde{W}_j][\mathbb{W}_1, \mathbb{W}_2], \widetilde{W}_{j+1}\right]^H [\mathbb{V}1, P_0]\hat{\Lambda}_1 \\ &= \left[\begin{array}{cc} C_k^H \mathbb{V}_1 & C_k^H P_0 \\ \mathscr{V}_j^H \mathbb{V}_1 & \mathscr{V}_j^H P_0 \\ [V_{j+1}, P_j]^H \mathbb{V}_1 & [\mathbb{W}_1, \mathbb{W}_2]^H \left[P_{j-1}, \widetilde{W}_j\right]^H P_0 \\ \widetilde{W}_{j+1}^H \mathbb{V}_1 & \widetilde{W}_{j+1}^H P_0 \end{array} \right] \hat{\Lambda}_1 = \left[\begin{array}{cc} 0_{k\times p_1} & 0_{k\times q_1} \\ \left[\begin{array}{c} I_{p_1} \\ 0_{(n_j-p_1)\times p_1} \end{array} \right] & \Phi_j(1:n_j,:) \\ 0_{p\times p_1} & [\mathbb{W}_1, \mathbb{W}_2]^H \left[\begin{array}{c} P_{j-1}^H \\ \widetilde{W}_j^H \end{array} \right] P_0 \\ 0_{p_{j+1}\times p_1} & 0_{p_{j+1}\times q_1} \end{array} \right] \\ &= \left[\begin{array}{cc} 0_{k\times p_1} & 0_{k\times q_1} \\ \left[\begin{array}{c} I_{p_1} \\ 0_{(n_j-p_1)\times p_1} \end{array} \right] & \Phi_j(1:n_j,:) \\ 0_{p\times p_1} & [\mathbb{W}_1, \mathbb{W}_2]^H \left[\begin{array}{c} \Phi_j(n_j+1:n_j+q_j,:) \\ 0_{p_j\times q_1} \end{array} \right] \\ 0_{p_{j+1}\times p_1} & 0_{p_{j+1}\times q_1} \end{array} \right] \hat{\Lambda}_1 = \left[\begin{array}{cc} 0_{k\times p_1} & 0_{k\times q_1} \\ \left[\begin{array}{c} I_{p_1} \\ 0_{(n_j+p-p_1)\times p_1} \end{array} \right] & \Phi_{j+1} \\ 0_{p_{j+1}\times p_1} & 0_{p_{j+1}\times q_1} \end{array} \right] \end{aligned}$$

where $\Phi_{j+1} \in \mathbb{C}^{(n_j+p)\times q_1}$ for $j = 1, 2, \cdots$. □

Note that we denote the new solver with constant right-hand sides of least-squares problem as shown in (15) or (21) as IBa-BGCRO-DR, where IBa stands for carrying out Inexact Breakdown after initial iteration (or without IB in the initial residuals), for contrasting with IB-BGCRO-DR that with updating right-hand sides of least-squares problem as shown in (90). The pseudocode for IBa-BGCRO-DR for constant and slowly-changing left-hand sides with massive number of right-hand sides are presented in Algorithm 5 in Appendix G and Algorithm 6 in Appendix H, respectively.

Figure 7 displays the results of adding the performance of IBa-BGCRO-DR to Figure 2 in Section 4.2 to illustrate the benefit of introducing IB in the initial residuals.

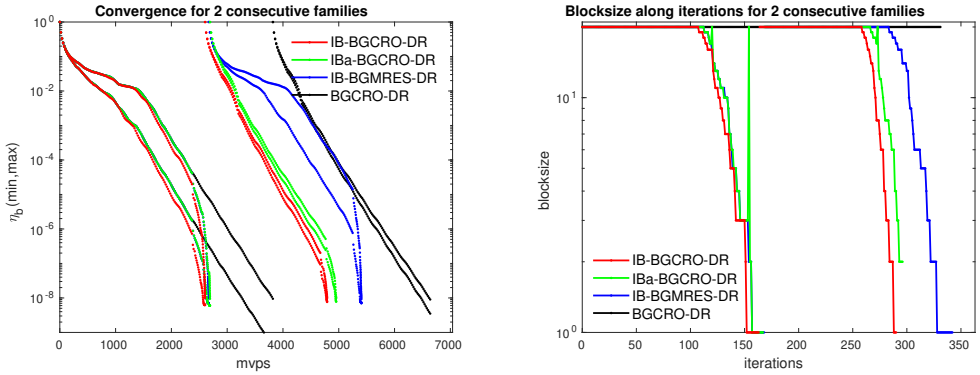


Figure 7: Comparison of IB-BGCRO-DR with IBa-BGCRO-DR, BGCRO-DR and IB-BGMRES-DR by solving bidiagonal Matrix 1 ($p = 20$, $m_d = 300$ and $k = 30$). Left: convergence histories of the largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 2 consecutive families. Right: varying blocksize comparison along iterations.

G IBa-BGCRO-DR without IB in the initial residuals R_1 for constant left-hand side and massive number of right-hand sides

Algorithm 4 SUBSPACE RECYCLING WITH INEXACT BREAKDOWN DETECTION IN INITIAL RESIDUALS R_1.

1: Let $\underline{\mathscr{L}}_0 = [\,]$ and the reduced SVD of R_1 as $R_1 = [\mathbb{V}_1, P_0] \begin{bmatrix} \Sigma_{p_1} & \\ & \Sigma_{q_1} \end{bmatrix} \mathbb{V}_{R_1}^H = [\mathbb{V}_1, P_0]\hat{\Lambda}_1$, where $\mathbb{V}_1 \in \mathbb{C}^{n\times p_1}$ and $P_0 \in \mathbb{C}^{n\times q_1}$ with $p_1 + q_1 = p$ and $C_k \perp [\mathbb{V}_1, P_0]$.

2: Orthogonalize $A\mathbb{V}_1$ against C_k as $W_1 = (I - C_k C_k^H)A\mathbb{V}_1$. Then orthogonalize W_1 against previous block orthonormal vector $\mathscr{V}_1 = [\mathbb{V}_1]$ as

$$\mathscr{L}_{1,1} = \mathscr{V}_1^H(W_1) = \mathscr{V}_1^H(A\mathbb{V}_1) \in \mathbb{C}^{n_1\times p_1},\ W_1 = W_1 - \mathscr{V}_1\mathscr{L}_{1,1} = A\mathbb{V}_1 - C_k C_k^H A\mathbb{V}_1 - \mathscr{V}_1\mathscr{L}_{1,1} \in \mathbb{C}^{n\times p_1}.$$

Set $\mathscr{L}_1 = [\underline{\mathscr{L}}_0, \mathscr{L}_{1,1}] \in \mathbb{C}^{n_1\times n_1} = \mathbb{C}^{p_1\times p_1}$.

3: Orthogonalize W_1 against P_0 and carry out its reduced QR-factorization as

$$E_1 = P_0^H W_1 \in \mathbb{C}^{q_1\times p_1}, \quad \widetilde{W}_1 D_1 = W_1 - P_0 E_1 \text{ with } \widetilde{W}_1 \in \mathbb{C}^{n\times p_1}, D_1 \in \mathbb{C}^{p_1\times p_1}.$$

4: Arnoldi relation at the end of first step

$$A[U_k, \mathbb{V}_1] = [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]\underline{\mathscr{F}}_1$$

with

$$\underline{\mathscr{F}}_1 = \begin{bmatrix} I_k & \mathcal{B}_1 \\ & \mathscr{L}_{1,1} \\ 0_{(n_1+p)\times k} & E_1 \\ & D_1 \end{bmatrix} \in \mathbb{C}^{(k+n_1+p)\times(k+n_1)}, \quad \mathcal{B}_1 = C_k^H A\mathbb{V}_1, \mathbb{H}_1 = \begin{bmatrix} E_1 \\ D_1 \end{bmatrix} \in \mathbb{C}^{p\times p_1}$$

5: Compute $R_1 = [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]\Lambda_1$ with $\Lambda_1 = \begin{bmatrix} 0_{k\times p_1} & 0_{k\times q_1} \\ I_{p_1} & 0_{p_1\times q_1} \\ 0_{q_1\times p_1} & I_{q_1} \\ 0_{p_1\times p_1} & 0_{p_1\times q_1} \end{bmatrix} \hat{\Lambda}_1 \in \mathbb{C}^{(k+n_1+p)\times p}$

6: Compute Y_1 the solution of the first new least-squares problem

$$Y_1 = \underset{Y\in\mathbb{C}^{(k+n_1)\times p}}{\operatorname{argmin}} \|\Lambda_1 - \underline{\mathscr{F}}_1 Y\|_F\,.$$

Carry out the singular value decomposition algorithm to detect inexact breakdown in first new block residuals

$$(\Lambda_1 - \underline{\mathscr{F}}_1 Y_1) = \mathbb{U}_{1,L}\Sigma_1\mathbb{V}_{1,R}^H + \mathbb{U}_{2,L}\Sigma_2\mathbb{V}_{2,R}^H, \text{ where } \sigma_{\min}(\Sigma_1) \geq \epsilon^{(R)} > \sigma_{\max}(\Sigma_2).$$

Compute $\mathbb{W}_1$ and $\mathbb{W}_2$ such that $\mathrm{Range}(\mathbb{W}_1) = \mathrm{Range}(\mathbb{U}_1^{(2)}) \in \mathbb{C}^{p\times p_2}$ with $\mathbb{U}_{1,L} = \begin{pmatrix} \mathbb{U}_1^{(1)} \\ \mathbb{U}_1^{(2)} \end{pmatrix} \in \mathbb{C}^{(k+n_1+p)\times p_2}$ and $[\mathbb{W}_1,\ \mathbb{W}_2]$ is unitary and $\mathbb{W}_2 \in \mathbb{C}^{p\times q_2}$ with $p_2 + q_2 = p$.
Compute new orthonormal matrices $\mathbb{V}_2$ and P_1, the last block row matrix $\mathscr{L}_{2,:}$ of $\underline{\mathscr{L}}_1$, and G_1 as

$$\mathbb{V}_2 = [P_0, \widetilde{W}_1]\mathbb{W}_1 \in \mathbb{C}^{n\times p_2},\ P_1 = [P_0, \widetilde{W}_1]\mathbb{W}_2 \in \mathbb{C}^{n\times q_2}, \mathscr{L}_{2,:} = \mathbb{W}_1^H\mathbb{H}_1 \in \mathbb{C}^{p_2\times p_1},\ G_1 = \mathbb{W}_2^H\mathbb{H}_1 \in \mathbb{C}^{q_2\times p_1}.$$

7: Set $\underline{\mathscr{L}}_1 = \begin{pmatrix} \mathscr{L}_1 \\ \mathscr{L}_{2,:} \end{pmatrix} \in \mathbb{C}^{(p_1+p_2)\times p_1} = \mathbb{C}^{n_2\times p_1}$.

Algorithm 5 IBa-BGCRO-DR FOR CONSTANT LEFT-HAND SIDE AND MASSIVE NUMBER OF RIGHT-HAND SIDES.

Suppose the ℓth family of linear systems is currently to be solved. Let the current p linearly independent right-hand sides be $B = [b^{(1)}, b^{(2)}, \ldots, b^{(p)}]$ given simultaneously. Choose the maximal dimension m of the underlying block approximation subspace in each cycle, k the desired number of approximate targeted eigenvectors, ε the targeted backward error, $X_0 = [x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(p)}]$ the initial block guess. Let $r_0^{(i)} = b^{(i)} - Ax_0^{(i)},\ i = 1, \ldots, p$. Denote $R_0 = [r_0^{(1)}, r_0^{(2)}, \ldots, r_0^{(p)}]$ the initial full-rank block residual. The recast problems are $A(x^{(i)} - x_0^{(i)}) = r_0^{(i)},\ i = 1, \ldots, p$. Set $n_{cycle} = 1$.
if an $(\ell-1)$st family of linear systems has been solved, that is if C_k, U_k, $\widehat{U}_k$ and $\mathcal{D}_k$ are defined from solving a previous (the $(\ell-1)$st) family of linear systems **then**
  $X_1 = X_0 + U_k C_k^H R_0, \quad$ and $R_1 = R_0 - C_k C_k^H R_0$
else
  Form the initial unitary matrix V_1 from the initial block residual $R_0 = V_1\Lambda_1$ with reduced QR-factorization. Implement an initial cycle of IB-BGMRES within m matrix-vector products.
  $X_m = X_0 + \mathcal{V}_m Y_m$
  $R_m = \left[\mathcal{V}_m, P_{m-1}, \widetilde{W}_m\right]\left(\tilde{\Lambda}_m - \underline{\tilde{\mathcal{F}}}_m Y_m\right)$. Check convergence, and proceed if not satisfied.
  Compute k harmonic Ritz vectors of A with respect to $\mathrm{Range}(\mathcal{V}_m)$ and store them in $\widetilde{Y}_k$. That is, compute k eigenvectors g_j of $\mathcal{L}_m + \mathcal{L}_m^{-H}\mathbb{H}_m^H\mathbb{H}_m$ associated with the smallest magnitude eigenvalues θ_j and store in G_k, such that $\widetilde{Y}_k = \mathcal{V}_m G_k$.
  Let $[Q, R]$ be the reduced QR-factorization of $\underline{\tilde{\mathcal{F}}}_m G_k$.
  $\mathcal{C}_k = \left[\mathcal{V}_m, P_{m-1}, \widetilde{W}_m\right] Q, \quad \mathcal{U}_k = \mathcal{V}_m G_k R^{-1}$
  Set $X_1 = X_m$ and $R_1 = R_m$.
end if
if $n_{cycle} = 1$ **then**
  if an $(\ell-1)$st family of linear systems has been solved **then**
    Form the initial unitary matrix V_1 from the initial block residual $R_1 = V_1\Lambda_1$ with reduced QR-factorization.
  else
    Let $[Q, R]$ be the reduced QR-factorization of $\tilde{\Lambda}_m - \underline{\tilde{\mathcal{F}}}_m Y_m$. Set $V_1 = \left[\mathcal{V}_m, P_{m-1}, \widetilde{W}_m\right] Q$ and $\Lambda_1 = R$.
  end if
else
  Let $[Q, R]$ be the reduced QR-factorization of $\Lambda_m - \underline{\mathcal{F}}_m Y_m$. Set $V_1 = \widehat{\mathcal{V}}_{m+p} Q$ and $\Lambda_1 = R$.
end if
$n_{cycle} = n_{cycle} + 1$
if the current family is the first one being solved **then**
  To reduce unnecessary ill-conditioning of the rightmost matrix in (14) or (19), let $\mathcal{D}_k$ be a diagonal scaling matrix such that $\widehat{U}_k = U_k\mathcal{D}_k$, where the columns of $\widehat{U}_k$ have unit norm [27, Section 2.4].
end if
Perform the block Arnoldi with inexact breakdowns as in Algorithm 3 within $m_d - k$ matrix-vector products with the linear operator $(I - C_k C_k^H)A$ and solve the least-squares problem as $\min \|\Lambda_m - \underline{\mathcal{F}}_m Y\|_F$ for $Y_m \in \mathbb{C}^{(k+n_m)\times p}$ with right-hand sides $\Lambda_m = \begin{bmatrix} 0_{k\times p} \\ \Lambda_1 \\ 0_{n_m\times p} \end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times p}$, generating Y_m, R_m^{LS}, $\mathcal{V}_m$, $\left[\mathcal{V}_m, P_{m-1}, \widetilde{W}_m\right]$, $\underline{\tilde{\mathcal{F}}}_m$, and $\mathcal{B}_m = C_k^H A\mathcal{V}_m$.
$\widehat{\mathcal{W}}_m = [\widehat{U}_k \quad \mathcal{V}_m], \quad \widehat{\mathcal{V}}_{m+1} = [C_k, \mathcal{V}_m, P_{m-1}, \widetilde{W}_m]$
$\underline{\mathcal{F}}_m = \begin{bmatrix} \mathcal{F}_m \\ \mathbb{H}_m \end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times(k+n_m)}$, $\mathcal{F}_m = \begin{bmatrix} \mathcal{D}_k & \mathcal{B}_m \\ 0_{(n_m+p)\times k} & \mathcal{L}_m \end{bmatrix} \in \mathbb{C}^{(k+n_m)\times(k+n_m)}$ and $\mathbb{H}_m = \begin{bmatrix} 0_{p\times k} & G_{m-1} & E_m \\ & 0 & D_m \end{bmatrix} \in \mathbb{C}^{p\times(k+n_m)}$
$X_m = X_1 + \widehat{\mathcal{W}}_m Y_m$
$R_m = \widehat{\mathcal{V}}_{m+1} R_m^{LS}$. Check convergence, and go to 37 if satisfied or else proceed.
if the current family is the first one being solved **then**
  Compute k harmonic Ritz vectors of A with respect to $\mathrm{Range}(\widehat{\mathcal{W}}_m) = [\widehat{U}_k, \mathcal{V}_m]$ and store them in $\widetilde{Y}_k$. That is, keep k eigenvectors g_j from solving the generalized eigenvalue problem $\underline{\mathcal{F}}_m^H\underline{\mathcal{F}}_m g_j = \theta_j \underline{\mathcal{F}}_m^H \widehat{\mathcal{V}}_{m+1}^H \widehat{\mathcal{W}}_m g_j$ associated with the smallest magnitude eigenvalues θ_j and store in G_k, such that $\widetilde{Y}_k = \widehat{\mathcal{W}}_m G_k$.
  Let $[Q, R]$ be the reduced QR-factorization of $\underline{\mathcal{F}}_m G_k$.
  $C_k = \widehat{\mathcal{V}}_{m+1} Q, \quad U_k = \widehat{\mathcal{W}}_m G_k R^{-1}$
end if
Restart with $X_1 = X_m$ and $R_1 = R_m$, i.e., go to 13.
Retain C_k, U_k, $\widehat{U}_k$ and $\mathcal{D}_k$ for the next (i.e., $(\ell+1)$st) family of linear systems.

H IBa-BGCRO-DR without IB in the initial residuals R_1 for slowly-changing left-hand sides and massive number of right-hand sides

Algorithm 6 IBa-BGCRO-DR FOR SLOWLY-CHANGING LEFT-HAND SIDES AND MASSIVE NUMBER OF RIGHT-HAND SIDES.

Suppose the ℓth family of linear systems is currently to be solved. Let the current p linearly independent right-hand sides be $B = [b^{(1)}, b^{(2)}, \ldots, b^{(p)}]$ given simultaneously. Choose the maximal dimension m of the underlying block approximation subspace in each cycle, k the desired number of approximate targeted eigenvectors, ε the targeted backward error, $X_0 = [x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(p)}]$ the initial block guess. Let $r_0^{(i)} = b^{(i)} - Ax_0^{(i)},\ i = 1, \ldots, p$. Denote $R_0 = [r_0^{(1)}, r_0^{(2)}, \ldots, r_0^{(p)}]$ the initial full-rank block residual. The recast problems are $A(x^{(i)} - x_0^{(i)}) = r_0^{(i)},\ i = 1, \ldots, p$. Set $n_{cycle} = 1$.
if $\widetilde{Y}_k$ is defined from solving a previous (the $(\ell-1)$st) family of linear systems **then**
  Let $[Q, R]$ be the reduced QR-factorization of $A\widetilde{Y}_k$.
  $C_k = Q, \quad U_k = \widetilde{Y}_k R^{-1}$
  $X_1 = X_0 + U_k C_k^H R_0$
  $R_1 = R_0 - C_k C_k^H R_0$
else
  Form the initial unitary matrix V_1 from the initial block residual $R_0 = V_1\Lambda_1$ with reduced QR-factorization. Implement an initial cycle of IB-BGMRES within m matrix-vector products.
  $X_m = X_0 + \mathscr{V}_m Y_m$
  $R_m = \left[\mathscr{V}_m, P_{m-1}, \widetilde{W}_m\right] \left(\tilde{\Lambda}_m - \underline{\tilde{\mathscr{F}}}_m Y_m\right)$. Check convergence, and proceed if not satisfied.
  Compute k harmonic Ritz vectors of A with respect to $\mathrm{Range}(\mathscr{V}_m)$ and store them in $\widetilde{Y}_k$. That is, compute k eigenvectors g_j of $\mathscr{L}_m + \mathscr{L}_m^{-H}\mathbb{H}_m^H\mathbb{H}_m$ associated with the smallest magnitude eigenvalues θ_j and store in G_k, such that $\widetilde{Y}_k = \mathscr{V}_m G_k$.
  Let $[Q, R]$ be the reduced QR-factorization of $\underline{\tilde{\mathscr{F}}}_m G_k$.
  $C_k = \left[\mathscr{V}_m, P_{m-1}, \widetilde{W}_m\right] Q, \quad U_k = \mathscr{V}_m G_k R^{-1}$
  Set $X_1 = X_m$ and $R_1 = R_m$.
end if
if $n_{cycle} = 1$ **then**
  if an $(\ell-1)$st family of linear systems has been solved **then**
    Form the initial unitary matrix V_1 from the initial block residual $R_1 = V_1\Lambda_1$ with reduced QR-factorization.
  else
    Let $[Q, R]$ be the reduced QR-factorization of $\tilde{\Lambda}_m - \underline{\tilde{\mathscr{F}}}_m Y_m$. Set $V_1 = \left[\mathscr{V}_m, P_{m-1}, \widetilde{W}_m\right] Q$ and $\Lambda_1 = R$.
  end if
else
  Let $[Q, R]$ be the reduced QR-factorization of $\Lambda_m - \underline{\mathscr{F}}_m Y_m$. Set $V_1 = \widehat{\mathscr{V}}_{m+1} Q$ and $\Lambda_1 = R$.
end if
$n_{cycle} = n_{cycle} + 1$
To reduce unnecessary ill-conditioning of the rightmost matrix in (14) or (19), let $\mathcal{D}_k$ be a diagonal scaling matrix such that $\widehat{U}_k = U_k\mathcal{D}_k$, where the columns of $\widehat{U}_k$ have unit norm [27, Section 2.4].
Perform the block Arnoldi with inexact breakdowns as in Algorithm 3 within $m_d - k$ matrix-vector products with the linear operator $(I - C_k C_k^H)A$ and solve the least-squares problem as $\min \|\Lambda_m - \underline{\mathscr{F}}_m Y\|_F$ for $Y_m \in \mathbb{C}^{(k+n_m)\times p}$ with right-hand sides $\Lambda_m = \begin{bmatrix} 0_{k\times p} \\ \Lambda_1 \\ 0_{n_m\times p} \end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times p}$, generating Y_m, R_m^{LS}, $\mathscr{V}_m$, $\left[\mathscr{V}_m, P_{m-1}, \widetilde{W}_m\right]$, $\underline{\tilde{\mathscr{F}}}_m$, and $\mathcal{B}_m = C_k^H A\mathscr{V}_m$.
$\widehat{\mathscr{W}}_m = [\widehat{U}_k \quad \mathscr{V}_m], \quad \widehat{\mathscr{V}}_{m+1} = [C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m]$
$\underline{\mathscr{F}}_m = \begin{bmatrix} \mathscr{F}_m \\ \mathbb{H}_m \end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times(k+n_m)}$, $\mathscr{F}_m = \begin{bmatrix} \mathcal{D}_k & \mathcal{B}_m \\ 0_{(n_m+p)\times k} & \mathscr{L}_m \end{bmatrix} \in \mathbb{C}^{(k+n_m)\times(k+n_m)}$ and $\mathbb{H}_m = \begin{bmatrix} 0_{p\times k} & G_{m-1} & E_m \\ & 0 & D_m \end{bmatrix} \in \mathbb{C}^{p\times(k+n_m)}$
$X_m = X_1 + \widehat{\mathscr{W}}_m Y_m$
$R_m = \widehat{\mathscr{V}}_{m+1} R_m^{LS}$. Check convergence, and go to 36 if satisfied or else proceed.
Compute k harmonic Ritz vectors of A with respect to $\mathrm{Range}(\widehat{\mathscr{W}}_m) = [\widehat{U}_k, \mathscr{V}_m]$ and store them in $\widetilde{Y}_k$. That is, keep k eigenvectors g_j from solving the generalized eigenvalue problem $\underline{\mathscr{F}}_m^H\underline{\mathscr{F}}_m g_j = \theta_j \underline{\mathscr{F}}_m^H \widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{W}}_m g_j$ associated with the smallest magnitude eigenvalues θ_j and store in G_k, such that $\widetilde{Y}_k = \widehat{\mathscr{W}}_m G_k$.
Let $[Q, R]$ be the reduced QR-factorization of $\underline{\mathscr{F}}_m G_k$.
$C_k = \widehat{\mathscr{V}}_{m+1} Q, \quad U_k = \widehat{\mathscr{W}}_m G_k R^{-1}$
Restart with $X_1 = X_m$ and $R_1 = R_m$, i.e., go to 16.
Retain $\widetilde{Y}_k = U_k$ for the next (i.e., $(\ell+1)$st) family of linear systems.

I IB-BGCRO-DR with inexact breakdown detection in the initial residuals R_1 for constant left-hand side and massive number of right-hand sides

Algorithm 7 IB-BGCRO-DR WITH INEXACT BREAKDOWN DETECTION IN INITIAL RESIDUALS FOR CONSTANT LEFT-HAND SIDE AND MASSIVE NUMBER OF RIGHT-HAND SIDES.

Suppose the ℓth family of linear systems is currently to be solved. Let the current p linearly independent right-hand sides be $B = [b^{(1)}, b^{(2)}, \ldots, b^{(p)}]$ given simultaneously. Choose the maximal dimension m of the underlying block approximation subspace in each cycle, k the desired number of approximate targeted eigenvectors, ε the targeted backward error, $X_0 = [x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(p)}]$ the initial block guess. Let $r_0^{(i)} = b^{(i)} - Ax_0^{(i)},\ i = 1, \ldots, p$. Denote $R_0 = [r_0^{(1)}, r_0^{(2)}, \ldots, r_0^{(p)}]$ the initial full-rank block residual. The recast problems are $A(x^{(i)} - x_0^{(i)}) = r_0^{(i)},\ i = 1, \ldots, p$. Set $n_{cycle} = 1$.
Following step 2-12 of Algorithm 5 in Appendix G
if $n_{cycle} = 1$ **then**
  if an $(\ell - 1)$st family of linear systems has been solved **then**
    Carry out the economic singular value decomposition algorithm to detect inexact breakdown in R_1

$$R_1 = \mathbb{U}_{1,L}\Sigma_1\mathbb{V}_{1,R}^H + \mathbb{U}_{2,L}\Sigma_2\mathbb{V}_{2,R}^H = [\mathbb{V}_1, P_0]\begin{bmatrix}\Sigma_{p_1} & \\ & \Sigma_{q_1}\end{bmatrix}\mathbb{V}_{R_1}^H = [\mathbb{V}_1, P_0]\hat{\Lambda}_1, \tag{91}$$

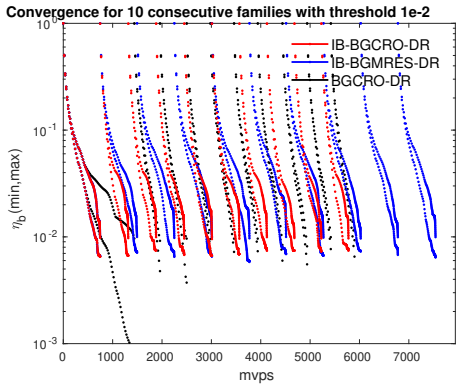
    where $\sigma_{\min}(\Sigma_1) \geq \epsilon_j^{(R)} > \sigma_{\max}(\Sigma_2)$.
    Set $\mathbb{V}_1 = \mathbb{U}_{1,L} \in \mathbb{C}^{n\times p_1}$, $P_0 = \mathbb{U}_{2,L} \in \mathbb{C}^{n\times q_1}$ with $p_1 + q_1 = p$ and $\hat{\Lambda}_1 = \begin{bmatrix}\Sigma_{p_1} & \\ & \Sigma_{q_1}\end{bmatrix}\mathbb{V}_{R_1}^H \in \mathbb{C}^{p\times p}$.
  else
    Compute $R_m^{LS} = \tilde{\Lambda}_m - \underline{\tilde{\mathscr{F}}}_m Y_m = \mathscr{V}_{m+1}^H R_m = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]^H R_m \in \mathbb{C}^{(n_m+p)\times p}$ the least-squares residuals at the end of of one cycle of Algorithm 3 with operator A. Carry out the economic singular value decomposition algorithm to detect inexact breakdown in R_m^{LS} as shown in equation (91) where R_1 changed into R_m^{LS}.
    Set $\mathbb{V}_1 = \mathscr{V}_{m+1}\mathbb{U}_{1,L} \in \mathbb{C}^{n\times p_1}$, $P_0 = \mathscr{V}_{m+1}\mathbb{U}_{2,L} \in \mathbb{C}^{n\times q_1}$ with $p_1 + q_1 = p$ and $\hat{\Lambda}_1 \in \mathbb{C}^{p\times p}$.
  end if
else
  Compute $R_m^{LS} = \Lambda_m - \underline{\mathscr{F}}_m Y_m = \widehat{\mathscr{V}}_{m+1}^H R_m = \left[C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]^H R_m \in \mathbb{C}^{(k+n_m+p)\times p}$ the least-squares residuals at the end of of one cycle of Algorithm 3 with operator $(I - C_kC_k^H)A$. Carry out the economic singular value decomposition algorithm to detect inexact breakdown in R_m^{LS} as shown in equation (91) where R_1 changed into R_m^{LS}.
  Set $\mathbb{V}_1 = \widehat{\mathscr{V}}_{m+1}\mathbb{U}_{1,L} \in \mathbb{C}^{n\times p_1}$, $P_0 = \widehat{\mathscr{V}}_{m+1}\mathbb{U}_{2,L} \in \mathbb{C}^{n\times q_1}$ with $p_1 + q_1 = p$ and $\hat{\Lambda}_1 \in \mathbb{C}^{p\times p}$.
end if
$n_{cycle} = n_{cycle} + 1$. Perform Algorithm 4 to fulfill the initialization of Algorithm 3.
if the current family is the first one being solved **then**
  To reduce unnecessary ill-conditioning of the rightmost matrix in (14) or (19), let $\mathcal{D}_k$ be a diagonal scaling matrix such that $\widehat{U}_k = U_k\mathcal{D}_k$, where the columns of $\widehat{U}_k$ have unit norm [27, Section 2.4].
end if
Then perform the block Arnoldi with inexact breakdowns as in Algorithm 3 for $j = 2, \ldots, m$ with the linear operator $(I - C_kC_k^H)A$ and update the least-squares right-hand sides (90) as description in Proposition 7, generating $\mathscr{V}_m$, $\left[\mathscr{V}_m, P_{m-1}, \widetilde{W}_m\right]$, $\underline{\tilde{\mathscr{F}}}_m$, Y_m, $R_m^{LS} = \Lambda_m - \underline{\mathscr{F}}_m Y_m$, and $\mathcal{B}_m = C_k^H A\mathscr{V}_m$.
$\widehat{\mathscr{W}}_m = [\widehat{U}_k \quad \mathscr{V}_m]$, $\widehat{\mathscr{V}}_{m+1} = [C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m]$,
$\underline{\mathscr{F}}_m = \begin{bmatrix}\mathscr{F}_m \\ \mathbb{H}_m\end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times(k+n_m)}$, $\mathscr{F}_m = \begin{bmatrix}\mathcal{D}_k & \mathcal{B}_m \\ 0_{(n_m+p)\times k} & \mathscr{L}_m\end{bmatrix} \in \mathbb{C}^{(k+n_m)\times(k+n_m)}$, $\mathbb{H}_m = \begin{bmatrix}0_{p\times k} & G_{m-1} & E_m \\ & 0 & D_m\end{bmatrix} \in \mathbb{C}^{p\times(k+n_m)}$
$X_m = X_1 + \widehat{\mathscr{W}}_m Y_m$
$R_m = \widehat{\mathscr{V}}_{m+1}R_m^{LS}$. Check convergence, and go to 30 if satisfied or else proceed.
if the current family is the first one being solved **then**
  Compute k harmonic Ritz vectors of A with respect to $\mathrm{Range}(\widehat{\mathscr{W}}_m) = [\widehat{U}_k, \mathscr{V}_m]$ and store them in $\widetilde{Y}_k$. That is, keep k eigenvectors g_j from solving the generalized eigenvalue problem $\underline{\mathscr{F}}_m^H\underline{\mathscr{F}}_m g_j = \theta_j\underline{\mathscr{F}}_m^H\widehat{\mathscr{V}}_{m+1}^H\widehat{\mathscr{W}}_m g_j$ associated with the smallest magnitude eigenvalues θ_j and store in G_k, such that $\widetilde{Y}_k = \widehat{\mathscr{W}}_m G_k$.
  Let $[Q, R]$ be the reduced QR-factorization of $\underline{\mathscr{F}}_m G_k$.
  $C_k = \widehat{\mathscr{V}}_{m+1}Q$, $U_k = \widehat{\mathscr{W}}_m G_k R^{-1}$
end if
Restart with $X_1 = X_m$ and $R_1 = R_m$, i.e., go to 3.
Retain $C_k, U_k, \widehat{U}_k$ and $\mathcal{D}_k$ for the next (i.e., $(\ell + 1)$st) family of linear systems.

J IB-BGCRO-DR with inexact breakdown detection in the initial residuals R_1 for slowly-changing left-hand side and massive number of right-hand sides

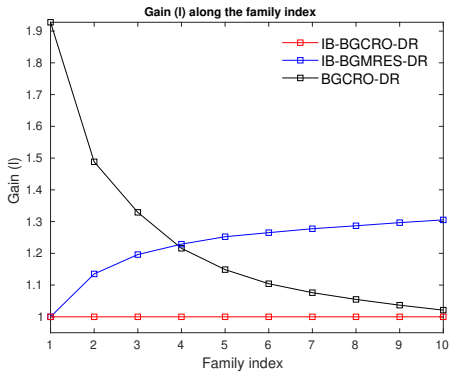
Algorithm 8 IB-BGCRO-DR WITH INEXACT BREAKDOWN DETECTION IN INITIAL RESIDUALS FOR SLOWLY-CHANGING LEFT-HAND SIDE AND MASSIVE NUMBER OF RIGHT-HAND SIDES.

Suppose the ℓth family of linear systems is currently to be solved. Let the current p linearly independent right-hand sides be $B = [b^{(1)}, b^{(2)}, \ldots, b^{(p)}]$ given simultaneously. Choose the maximal dimension m of the underlying block approximation subspace in each cycle, k the desired number of approximate targeted eigenvectors, ε the targeted backward error, $X_0 = [x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(p)}]$ the initial block guess. Let $r_0^{(i)} = b^{(i)} - Ax_0^{(i)},\ i = 1, \ldots, p$. Denote $R_0 = [r_0^{(1)}, r_0^{(2)}, \ldots, r_0^{(p)}]$ the initial full-rank block residual. The recast problems are $A(x^{(i)} - x_0^{(i)}) = r_0^{(i)},\ i = 1, \ldots, p$. Set $n_{cycle} = 1$.
Following step 2-15 of Algorithm 6 in Appendix H
if $n_{cycle} = 1$ **then**
  if an $(\ell - 1)$st family of linear systems has been solved **then**
    Carry out the economic singular value decomposition algorithm to detect inexact breakdown in R_1 as shown in (91). Set $\mathbb{V}_1 = \mathbb{U}_{1,L} \in \mathbb{C}^{n\times p_1}$, $P_0 = \mathbb{U}_{2,L} \in \mathbb{C}^{n\times q_1}$ with $p_1 + q_1 = p$ and $\hat{\Lambda}_1 = \begin{bmatrix} \Sigma_{p_1} & \\ & \Sigma_{q_1} \end{bmatrix} \mathbb{V}_{R_1}^H \in \mathbb{C}^{p\times p}$.
  else
    Compute $R_m^{LS} = \tilde{\Lambda}_m - \underline{\tilde{\mathscr{F}}}_m Y_m = \mathscr{V}_{m+1}^H R_m = \left[\mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]^H R_m \in \mathbb{C}^{(n_m+p)\times p}$ the least-squares residuals at the end of of one cycle of Algorithm 3 with operator A. Carry out the economic singular value decomposition algorithm to detect inexact breakdown in R_m^{LS} as shown in equation (91) where R_1 changed into R_m^{LS}.
    Set $\mathbb{V}_1 = \mathscr{V}_{m+1}\mathbb{U}_{1,L} \in \mathbb{C}^{n\times p_1}$, $P_0 = \mathscr{V}_{m+1}\mathbb{U}_{2,L} \in \mathbb{C}^{n\times q_1}$ with $p_1 + q_1 = p$ and $\hat{\Lambda}_1 = \begin{bmatrix} \Sigma_{p_1} & \\ & \Sigma_{q_1} \end{bmatrix} \mathbb{V}_{R_m^{LS}}^H \in \mathbb{C}^{p\times p}$.
  end if
else
  Compute $R_m^{LS} = \Lambda_m - \underline{\mathscr{F}}_m Y_m = \widehat{\mathscr{V}}_{m+1}^H R_m = \left[C_k, \mathscr{V}_m, [P_{m-1}, \widetilde{W}_m]\right]^H R_m \in \mathbb{C}^{(k+n_m+p)\times p}$ the least-squares residuals at the end of of one cycle of Algorithm 3 with operator $(I - C_k C_k^H)A$. Carry out the economic singular value decomposition algorithm to detect inexact breakdown in R_m^{LS} as shown in equation (91) where R_1 changed into R_m^{LS}.
  Set $\mathbb{V}_1 = \widehat{\mathscr{V}}_{m+1}\mathbb{U}_{1,L} \in \mathbb{C}^{n\times p_1}$, $P_0 = \widehat{\mathscr{V}}_{m+1}\mathbb{U}_{2,L} \in \mathbb{C}^{n\times q_1}$ with $p_1 + q_1 = p$ and $\hat{\Lambda}_1 = \begin{bmatrix} \Sigma_{p_1} & \\ & \Sigma_{q_1} \end{bmatrix} \mathbb{V}_{R_m^{LS}}^H \in \mathbb{C}^{p\times p}$.
end if
$n_{cycle} = n_{cycle} + 1$. Perform Algorithm 4 to fulfill the initialization of Algorithm 3.
To reduce unnecessary ill-conditioning of the rightmost matrix in (14) or (19), let $\mathcal{D}_k$ be a diagonal scaling matrix such that $\widehat{U}_k = U_k\mathcal{D}_k$, where the columns of $\widehat{U}_k$ have unit norm [27, Section 2.4].
Then perform the block Arnoldi with inexact breakdowns as in Algorithm 3 for $j = 2, \ldots, m$ with the linear operator $(I - C_k C_k^H)A$ and update the least-squares right-hand sides (90) as description in Proposition 7, generating $\mathscr{V}_m$, $\left[\mathscr{V}_m, P_{m-1}, \widetilde{W}_m\right]$, $\underline{\tilde{\mathscr{F}}}_m$, Y_m, $R_m^{LS} = \Lambda_m - \underline{\mathscr{F}}_m Y_m$, and $\mathcal{B}_m = C_k^H A\mathscr{V}_m$.
$\widehat{\mathscr{W}}_m = [\widehat{U}_k \quad \mathscr{V}_m]$, $\widehat{\mathscr{V}}_{m+1} = [C_k, \mathscr{V}_m, P_{m-1}, \widetilde{W}_m]$
$\underline{\mathscr{F}}_m = \begin{bmatrix} \mathscr{F}_m \\ \mathbb{H}_m \end{bmatrix} \in \mathbb{C}^{(k+n_m+p)\times(k+n_m)}$, $\mathscr{F}_m = \begin{bmatrix} \mathcal{D}_k & \mathcal{B}_m \\ 0_{(n_m+p)\times k} & \mathscr{L}_m \end{bmatrix} \in \mathbb{C}^{(k+n_m)\times(k+n_m)}$ and $\mathbb{H}_m = \begin{bmatrix} 0_{p\times k} & G_{m-1} & E_m \\ & 0 & D_m \end{bmatrix} \in \mathbb{C}^{p\times(k+n_m)}$
$X_m = X_1 + \widehat{\mathscr{W}}_m Y_m$
$R_m = \widehat{\mathscr{V}}_{m+1} R_m^{LS}$. Check convergence, and go to 25 if satisfied or else proceed.
Compute k harmonic Ritz vectors of A with respect to $\mathrm{Range}(\widehat{\mathscr{W}}_m) = [\widehat{U}_k, \mathscr{V}_m]$ and store them in $\widetilde{Y}_k$. That is, keep k eigenvectors g_j from solving the generalized eigenvalue problem $\underline{\mathscr{F}}_m^H \underline{\mathscr{F}}_m g_j = \theta_j \underline{\mathscr{F}}_m^H \widehat{\mathscr{V}}_{m+1}^H \widehat{\mathscr{W}}_m g_j$ associated with the smallest magnitude eigenvalues θ_j and store in G_k, such that $\widetilde{Y}_k = \widehat{\mathscr{W}}_m G_k$.
Let $[Q, R]$ be the reduced QR-factorization of $\underline{\mathscr{F}}_m G_k$.
$C_k = \widehat{\mathscr{V}}_{m+1} Q, \quad U_k = \widehat{\mathscr{W}}_m G_k R^{-1}$
Restart with $X_1 = X_m$ and $R_1 = R_m$, i.e., go to 3.
Retain $\widetilde{Y}_k = U_k$ for the next (i.e., $(\ell + 1)$st) family of linear systems.

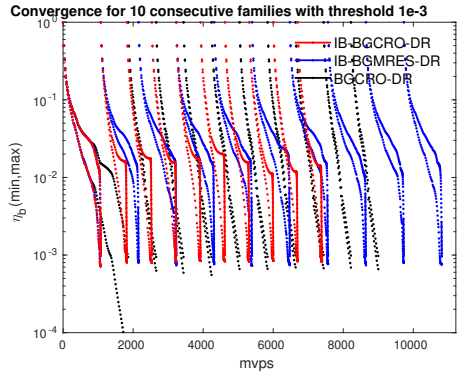
K Numerical results for various target accuracy



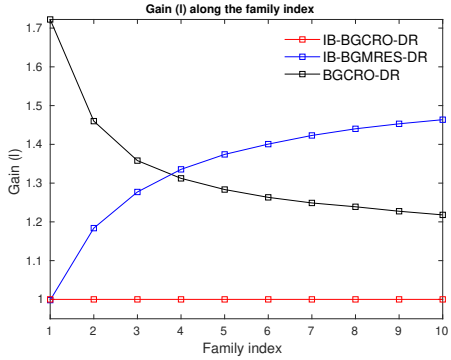
Target accuracy 10^{-2}



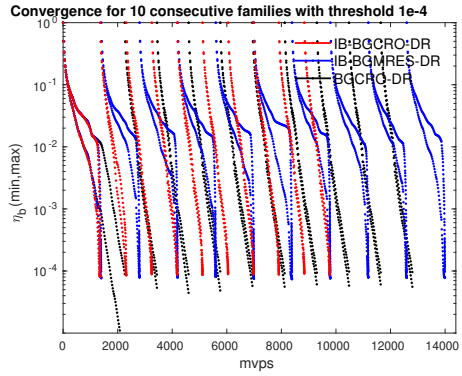
Target accuracy 10^{-2}



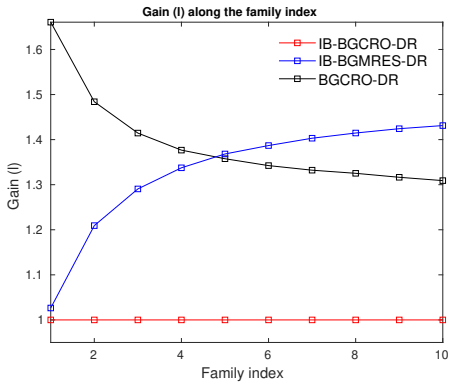
Target accuracy 10^{-3}



Target accuracy 10^{-3}



Target accuracy 10^{-4}



Target accuracy 10^{-4}

Figure 8: History of Section 4.3 for the behavior in case of different target accuracy (10^{-2}, 10^{-3} and 10^{-4}) for the families constructed by bidiagonal Matrix 1 with parameters setting as $p = 20$, $m_d = 300$ and $k = 30$. Left: convergence histories of IB-BGCRO-DR, BGCRO-DR and IB-BGMRES-DR on the largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 10 consecutive families. Right: Gain (l) of the block methods with respect to IB-BGCRO-DR along family index.

L Numerical results for IB-BGMRES-DR-CB

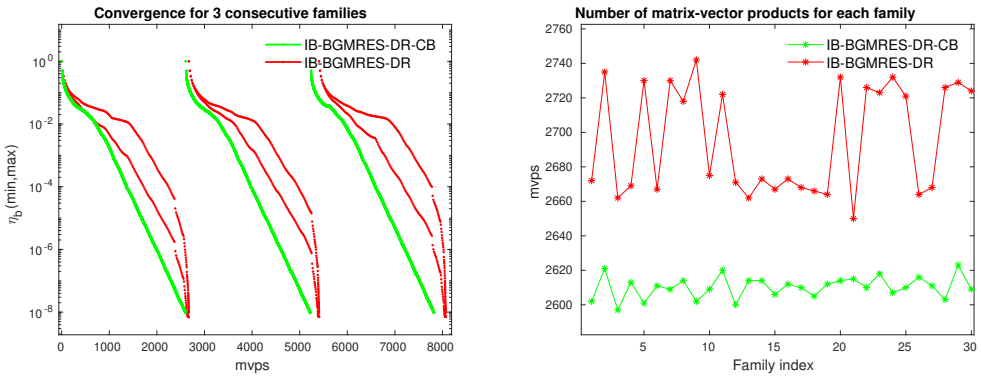


Figure 9: Comparison of IB-BGMRES-DR to IB-BGMRES-DR-CB on families constructed by bidiagonal Matrix 1 with parameters setting as $p^{CB} = 1$, $p = 20$, $m_d = 300$ and $k = 15$. Left: convergence histories of largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 3 consecutive families. Right: number of consumed $mvps$ verse family index.

Family number	Method	$mvps$	its
3	IB-BGMRES-DR	8069	**515**
	IB-BGMRES-DR-CB ($p^{CB} = 15$)	7844	561
	IB-BGMRES-DR-CB ($p^{CB} = 1$)	**7820**	7250
30	IB-BGMRES-DR	80861	**5198**
	IB-BGMRES-DR-CB ($p^{CB} = 1$)	**78308**	72608

Table 8: Numerical results of IB-BGMRES-DR, IB-BGMRES-DR-CB with parameter $p^{CB} = 1, 15$ in terms of $mvps$ and its, where the involving parameters for bidiagonal Matrix 1 are set to be $p = 20$, $m_d = 300$ and $k = 30$.

M Numerical results for IB-BGMRES-DR-VA

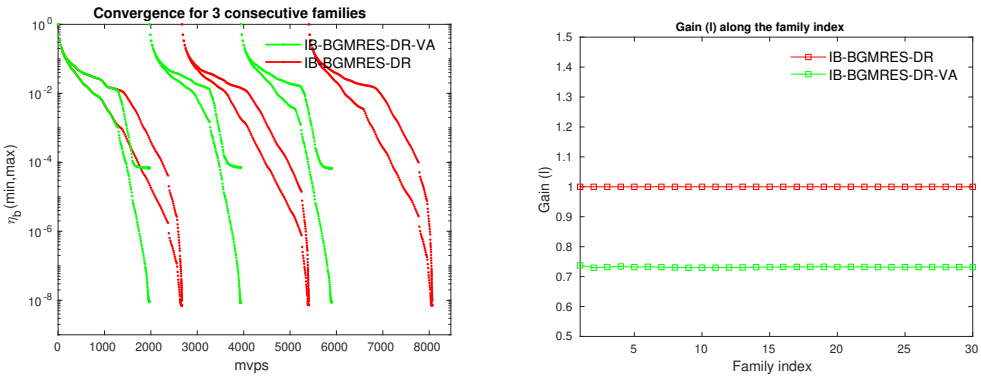
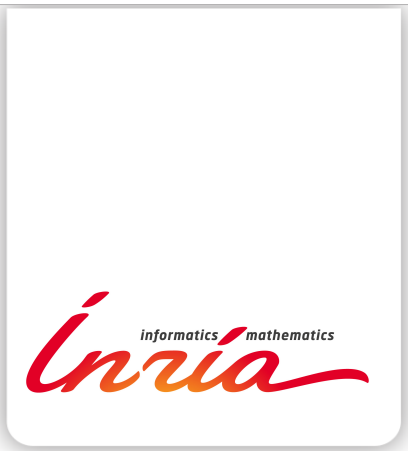


Figure 10: Comparison of IB-BGMRES-DR to IB-BGMRES-DR-VA on families built by bidiagonal Matrix 1 with parameters as $p = 20$, $m_d = 300$ and $k = 15$. Left: convergence histories of largest/smallest backward errors $\eta_{b^{(i)}}$ at each $mvps$ for 3 consecutive families. Right: Gain (ℓ) of IB-BGMRES-DR-VA to IB-BGMRES-DR verse family index.

Family number	Method	$mvps$	its
3	IB-BGMRES-DR	8066	515
	IB-BGMRES-DR-VA	**5903**	**490**
30	IB-BGMRES-DR	80717	5191
	IB-BGMRES-DR-VA	**59069**	**4957**

Table 9: Numerical results of IB-BGMRES-DR and IB-BGMRES-DR-VA in terms of $mvps$ and its, where the coefficient matrix is bidiagonal Matrix 1 with involving parameters defined as $p = 20$, $m_d = 300$ and $k = 30$.



RESEARCH CENTRE
BORDEAUX – SUD-OUEST

200 avenue de la Vieille Tour
33405 Talence Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399