



**HAL**  
open science

# A block minimum residual norm subspace solver with partial convergence management for sequences of linear systems

Luc Giraud, Yan-Fei Jing, Yanfei Xiang

► **To cite this version:**

Luc Giraud, Yan-Fei Jing, Yanfei Xiang. A block minimum residual norm subspace solver with partial convergence management for sequences of linear systems. *SIAM Journal on Matrix Analysis and Applications*, 2022, 43 (2), pp.710-739. 10.1137/21m1401127 . hal-03546496v2

**HAL Id: hal-03546496**

**<https://inria.hal.science/hal-03546496v2>**

Submitted on 1 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **A BLOCK MINIMUM RESIDUAL NORM SUBSPACE SOLVER WITH PARTIAL**  
2 **CONVERGENCE MANAGEMENT FOR SEQUENCES OF LINEAR SYSTEMS**

3 L. GIRAUD\*, Y.-F. JING†, AND Y.-F. XIANG\*‡

4 **Abstract.** We are concerned with the iterative solution of linear systems with multiple right-hand sides available one group after  
5 another with possibly slowly-varying left-hand sides. For such sequences of linear systems, we first develop a new block minimum  
6 norm residual approach that combines two main ingredients. The first component exploits ideas from GCRO-DR [SIAM J. Sci.  
7 Comput., 28(5) (2006), pp. 1651–1674], enabling to recycle information from one solve to the next. The second component is the  
8 numerical mechanism to manage the partial convergence of the right-hand sides, referred to as inexact breakdown detection in IB-  
9 BGMRES [Linear Algebra Appl., 419 (2006), pp. 265–285], that enables the monitoring of the rank deficiency in the residual space  
10 basis expanded block-wise.

11 Secondly, for the class of block minimum norm residual approaches, that relies on a block Arnoldi-like equality between the  
12 search space and the residual space (e.g., any block GMRES or block GCRO variants), we introduce new search space expansion  
13 policies defined on novel criteria to detect the partial convergence. These novel detection criteria are tuned to the selected stopping  
14 criterion and targeted convergence threshold to best cope with the selected normwise backward error stopping criterion, enabling to  
15 monitor the computational effort while ensuring the final accuracy of each individual solution. Numerical experiments are reported to  
16 illustrate the numerical and computational features of both the new block Krylov solvers and the new search space block expansion  
17 policies.

18 **Key words.** Block subspace methods, augmentation, deflation, subspace recycling, partial convergence, inexact block rank  
19 deficiency, backward error stopping criterion.

20 **AMS subject classifications.** 65F10, 65N22, 15A06

21 **1. Introduction.** Many scientific and industrial simulations require the solution of a sequence of linear  
22 systems with multiple right-hand sides and possibly slowly-changing left-hand sides. In that context, one  
23 has to solve a series of linear systems of the form

$$A^{(\ell)} X^{(\ell)} = B^{(\ell)}, \quad \ell = 1, 2, \dots, \quad (1.1)$$

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

29 When solving sequences of linear systems as Equation (1.1), attractive approaches are those that can  
30 exploit information generated during the solution of a given system to accelerate the convergence for the  
31 next ones. Deflated restarting implements a similar idea between the cycles in the generalized minimum  
32 residual norm method (GMRES) [19, 21, 27]; it is realized by using a deflation subspace containing a few  
33 approximate eigenvectors deemed to hamper the convergence of the Krylov subspace methods [11–13].  
34 Another alternative technique is the subspace recycling strategy proposed in the generalized conjugate  
35 residual method with inner orthogonalization (GCRO) and deflated restarting (GCRO-DR) method [16].  
36 This latter method can reuse information accumulated in previous cycles as well as that accumulated  
37 during the solution of the previous families. Because the multiple right-hand sides of Equation (1.1) are  
38 simultaneously available, block Krylov subspace methods are often considered as the suitable candidates  
39 for their capability of sharing search subspace that can be generated using basic linear algebra subprograms,  
40 level 3 BLAS-like implementation [10]. A common issue in block Krylov subspace methods is the rank  
41 deficiency that might appear when expanding the residual spaces, which is caused by the convergence  
42 of some individual solution or a linear combination of solution vectors. Such rank deficiency problem  
43 could lead the block Arnoldi process to break down before the solutions for all the right-hand sides are  
44 found. For the sake of balancing robustness and convergence rate, Robbé and Sadkane proposed an inexact  
45 breakdown detection for the block GMRES algorithm (denoted by IB-BGMRES) [20], which could keep  
46 and reintroduce directions associated with the almost converged parts in next iteration if necessary. We refer

\*Inria, France

†School of Mathematical Sciences/Institute of Computational Science, University of Electronic Science and Technology of China,  
Chengdu, Sichuan, 611731, P. R. China

‡Cerfacs, France

47 to [1, 2, 20], for relevant works on inexact breakdown detection, as well as to [23–26, 28], for related variants  
 48 of block Krylov subspace methods for solving linear systems with multiple right-hand sides.

49 The contribution of this paper is twofold. We first show how to combine subspace recycling techniques  
 50 of GCRO-DR [16], for recycling spectral information at a new cycle/family, with the inexact breakdown  
 51 detection introduced by Robbé and Sadkane in IB-BGMRES [20], for handling the issue of almost rank  
 52 deficient block generated by the block Arnoldi procedure, to develop the IB-BGCRO-DR algorithm, a new  
 53 recycling block GCRO-DR variant with partial convergence detection. This is a natural extension of our  
 54 previous work IB-BGMRES-DR [1], that enables the deflated restarting strategy proposed by Morgan [13]  
 55 to be applied not only at restart but also when solving a sequence of linear systems. The IB-BGCRO-DR  
 56 method can reuse spectral information both from the solutions in the previous cycles and families thus  
 57 showing obvious advantages when solving sequences of linear systems like Equation (1.1). In addition,  
 58 we propose a flexible counterpart of the new algorithm, which allows the use of a mixed arithmetic  
 59 computation where all steps are computed with a selected working precision except for the preconditioner  
 60 which is performed with a reduced precision. The second contribution is related to the block search space  
 61 expansion policies that can be further developed based on the partial convergence detection. In particular,  
 62 for stopping criteria based on normwise backward error we introduce new strategies enabling to focus on  
 63 the computational effort while ensuring the final accuracy of each individual solution.

64 The remainder of this paper is organized as follows. Section 2 is devoted to the development of the new  
 65 algorithm, containing some background parts that enable us to introduce the various numerical ingredients  
 66 and notations required to design our algorithm. In Section 2.1 we first recall the governing ideas of the  
 67 minimum norm residual Krylov method GCRO in a single right-hand side setting and briefly present its  
 68 block variant in Section 2.2. Next in Section 2.3 we present how the original inexact breakdown detection  
 69 mechanism [20] introduced for block GMRES can be applied to block GCRO as well. These two main  
 70 ingredients are combined to develop the new algorithm IB-BGCRO-DR in Section 2.4 and its flexible  
 71 preconditioning variant referred to as IB-BFGCRO-DR discussed in Section 2.5. In Section 3, we describe  
 72 how to extend the original inexact breakdown detection mechanism to best adapt the computational effort  
 73 and reach the targeted accuracy prescribed by the stopping criterion defined in terms of normwise backward  
 74 errors for the individual solutions. In particular, we derive strategies to manage the situation where the  
 75 different right-hand sides need to be solved with different convergence thresholds. We also present policies  
 76 adapted to a stopping criterion based on normwise backward error on the right-hand side only (i.e. classical  
 77 residual norm scaled by the norm of the right-hand side) or the more general one used to establish the  
 78 backward stability of GMRES [14]. In Section 4, some remarks on computational and algorithmic aspects  
 79 are detailed; the associated pseudocode of the IB-BGCRO-DR algorithm is presented as well. In Section 5  
 80 we present numerical experiments that illustrate the benefits of the new algorithm with both constant and  
 81 slowly varying successive linear systems with multiple right-hand sides as well as the numerical capabilities  
 82 of the novel search space expansion policies. Finally some concluding remarks are detailed in Section 6.

83 The symbol  $\|\cdot\|$  denotes the Euclidean norm defaultly for both vectors and matrices, and the Frobenius  
 84 norm is denoted with the subscript  $F$ . The superscript  $H$  denotes the transpose conjugate and  $T$  for transpose.  
 85 Because many notations are involved, we make choices to help the readability of the paper. The vectors are  
 86 described by lowercase letter, matrices with multiple columns described by uppercase letter, the calligraphy  
 87 uppercase letters like  $\mathcal{V}$  represent the matrices whose columns are enlarged by multiple columns at each  
 88 iteration as commonly appearing in the block Krylov context, and the uppercase letter with blackboard bold  
 89 form like  $\mathbb{V}$  refers to the block Krylov basis generated at each iteration. The superscript  $\dagger$  refers to the  
 90 Moore-Penrose inverse. For convenience of the algorithm illustration and presentation, some MATLAB  
 91 notations are used. Without special note, a subscript  $j$  for a vector (in single right-hand case) or a matrix  
 92 (in block case) is used to indicate that the vector or matrix is obtained at iteration  $j$ , and a positive subscript  
 93 integer  $m$  represents the maximal iteration number of each (block) Krylov cycle. All the involved recycling  
 94 subspaces of dimension  $k$  are described as a matrix with the subscript  $k$  whose columns form a basis. A  
 95 matrix  $C \in \mathbb{C}^{m \times \ell}$  consisting of  $m$  rows and  $\ell$  columns sometimes is denoted as  $C_{m \times \ell}$  explicitly. The  
 96 identity and null matrices of dimension  $m$  are denoted respectively by  $I_m$  and  $0_m$  or just  $I$  and  $0$  when the  
 97 dimension is evident from the context. For a matrix  $C \in \mathbb{C}^{m \times \ell}$ , the singular values of  $C$  are denoted by  
 98  $\sigma_1(C) \geq \dots \geq \sigma_{\min(m, \ell)}(C)$  in descending order; furthermore we denote  $\text{span}(C)$  the space spanned by  
 99 the columns of  $C$ .

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

103 current  $\ell^{\text{th}}$  family of linear systems to be solved is

$$AX = B, \quad (1.2)$$

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

107 **2. Block GCRO-DR with partial convergence detection.** For the sake of completeness of exposure,  
 108 this section contains some possibly well-known background which enables us to introduce the numerous  
 109 notations required to describe the new algorithm and detail its properties. In that respect, we first recall the  
 110 main ingredients of the subspace recycling techniques existing in the minimum residual Krylov methods  
 111 GCRO [7] and GCRO-DR [16] that are presented in the single right-hand side context. The straightforward  
 112 extension to the multiple right-hand sides framework, that is the block formulation of GCRO-DR (BGCRO-  
 113 DR) [15, 17, 18] is next introduced. Then the driving ideas of partial convergence detection [20] as well  
 114 as the corresponding block Arnoldi-like recurrence equation are derived in the block GCRO-DR context  
 115 leading to the new IB-BGCRO-DR algorithm.

116 **2.1. GCRO.** The background of GCRO [7] is briefly reviewed first in the case of a single right-hand  
 117 side and then extended to the block case. The GCRO method relies on a given full-rank matrix  $U_k \in \mathbb{C}^{n \times k}$ ,  
 118 and a matrix  $C_k$  as the image of  $U_k$  by  $A$  satisfying the relations

$$AU_k = C_k, \quad (2.1)$$

$$C_k^H C_k = I_k. \quad (2.2)$$

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  $x_1 \in x_0 + \text{Range}(U_k)$  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

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

is defined by

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

119 Starting from the unit vector  $v_1 = r_1 / \|r_1\|$ , the Arnoldi procedure enables us to form an orthonormal basis  
 120  $V_m = [v_1, \dots, v_m]$  of the Krylov space  $\mathcal{K}_m((I - C_k C_k^H)A, v_1) = \text{span}(v_1, (I - C_k C_k^H)A v_1, \dots, ((I -$   
 121  $C_k C_k^H)A)^{m-1} v_1)$  yielding an Arnoldi-like relation in the matrix form as

$$(I - C_k C_k^H)A V_m = V_{m+1} \underline{H}_m, \quad (2.3)$$

where the top square part of  $\underline{H}_m \in \mathbb{C}^{(m+1) \times m}$  is upper Hessenberg, and only the last entry of its last row is  
 nonzero. Combining Equation (2.1) and (2.3) in one matrix form allows us to write a relation quite similar  
 to an Arnoldi equality that reads

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

where the columns of  $\widehat{W}_m = [U_k, V_m]$  defines a basis of the search space, columns of  $\widehat{V}_{m+1} = [C_k, V_{m+1}]$   
 are an orthonormal basis of the residual space and  $\underline{G}_m = \begin{bmatrix} I_k & B_m \\ 0_{(m+1) \times k} & \underline{H}_m \end{bmatrix} \in \mathbb{C}^{(k+m+1) \times (k+m)}$ ,  
 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  
 $x_1 + \text{Range}(\widehat{W}_m)$  can be written as  $x_m = x_1 + \widehat{W}_m y_m$  where

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

122 and  $c = \widehat{V}_{m+1}^H r_1 = (0_k, \|r_1\|, 0_m)^T \in \mathbb{C}^{k+m+1}$  are the components of the residual associated with  $x_1$  in  
 123 the residual space spanned by the columns of  $\widehat{V}_{m+1}$ .

GCRO and GMRES [21], both belong to the family of residual norm minimization approaches and rely on an orthonormal basis of the residual space. In addition to sharing the Arnoldi procedure to form part of 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 [20] in the context of block GMRES can be extended to block GCRO as discussed in the sequel. The purpose of the partial convergence detection is to prevent 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 expansion according to the final target accuracy.

**2.2. Block GCRO.** The straightforward extension of the GCRO method in the block context is briefly described below. To facilitate reading, we change the calligraphy of the notations but keep the same letters to denote the block counterparts of the quantities involved in the method. Starting from the block initial guess  $X_0 = [x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(p)}] \in \mathbb{C}^{n \times p}$  and associated initial residual block  $R_0 = B - AX_0$ , one can define

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

given by

$$X_1 = X_0 + U_k C_k^H R_0, \text{ and } R_1 = (I - C_k C_k^H) R_0 \text{ such that } R_1 \in C_k^\perp. \quad (2.4)$$

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  $\mathcal{V}_m = [\mathbb{V}_1, \dots, \mathbb{V}_m]$ , 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

---

**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 matrix  $\mathbb{V}_1 \in \mathbb{C}^{n \times p}$  with orthonormal columns
  - 2: **for**  $j = 1, 2, \dots, m$  **do**
  - 3:   Compute  $W_j = (I - C_k C_k^H)A \mathbb{V}_j$
  - 4:   **for**  $i = 1, 2, \dots, j$  **do**
  - 5:      $H_{i,j} = \mathbb{V}_i^H W_j$
  - 6:      $W_j = W_j - \mathbb{V}_i H_{i,j}$
  - 7:   **end for**
  - 8:    $W_j = \mathbb{V}_{j+1} H_{j+1,j}$  (reduced  $QR$ -factorization of  $W_j$ )
  - 9: **end for**
- 

procedure leads to the matrix equality

$$(I - C_k C_k^H)A \mathcal{V}_m = \mathcal{V}_{m+1} \underline{\mathcal{H}}_m, \quad (2.5)$$

where  $\underline{\mathcal{H}}_m$  is a block Hessenberg matrix with  $(i, j)$  block defined by  $H_{i,j}$ . Similarly to the single right-hand side case, Equation (2.1) and (2.5) can be gathered in a matrix form

$$A \widehat{\mathcal{W}}_m = \widehat{\mathcal{V}}_{m+1} \underline{\mathcal{G}}_m, \quad (2.6)$$

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

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

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

144 **2.3. Block GCRO with partial convergence detection.** When one solution or a linear combination  
 145 of the solutions has converged, the block Arnoldi procedure implemented to build an orthonormal basis  
 146 of  $\mathcal{K}_j((I - C_k C_k^H)A, \mathbb{V}_1)$  needs to be modified to account for this partial convergence. This partial  
 147 convergence is characterized by a numerical rank deficiency in the new  $p$  directions that are usually  
 148 introduced for enlarging the search space at the next iteration. In [20], the authors present an elegant  
 149 numerical variant that enables the detection of what is referred to as inexact breakdowns. In that approach  
 150 the directions that have a low contribution to the residual block are discarded from the candidate set of  
 151 vectors used to expand the search space at the next iteration, but these abandoned directions are kept and  
 152 reintroduced in iterations afterwards if necessary. In this section, we try to give an insight and the main  
 153 equality required to derive the IB-BGCRO-DR algorithm. We refer the reader to the original paper [20]  
 154 for a detailed and complete description. For the sake of simplicity of exposure and easy cross-reading, we  
 155 adopt most of the notations from [1, 20].

156 Because when a partial convergence occurs, not all the space spanned by  $W_j$  is considered to build  
 157  $\mathbb{V}_{j+1}$  in order to expand the search space. For the sake of simplicity, we assume that  $p_1 = p$  and we denote  
 158 by  $p_{j+1}$  the number of columns of the block orthonormal basis vector  $\mathbb{V}_{j+1}$ . Then  $\mathbb{V}_{j+1} \in \mathbb{C}^{n \times p_{j+1}}$ ,  $W_j \in$   
 159  $\mathbb{C}^{n \times p_j}$  and  $H_{j+1,j} \in \mathbb{C}^{p_{j+1} \times p_j}$ . As a consequence the dimension of the search space  $\mathcal{K}_j((I - C_k C_k^H)A, \mathbb{V}_1)$   
 160 considered at the  $j^{\text{th}}$  iteration is no longer necessarily equal to  $j \times p$  but is equal to  $n_j = \sum_{i=1}^j p_i$ ; that is,  
 161 the sum of the column rank of  $\mathbb{V}_i$ 's ( $i = 1, \dots, j$ ).

162 When no partial convergence has occurred  $p_{j+1} = p_j = \dots = p_1 = p$ , the range of  $W_j$  has always  
 163 been used to enlarge the search space and we obtain the block relation given by Equation (2.6). To account  
 164 for a numerical deficiency in the residual block  $R_j = B - AX_j$  in a way that is described later, Robbé and  
 165 Sadkane [20] proposed to split

$$W_j = \mathbb{V}_{j+1} H_{j+1,j} + Q_j \quad (2.7)$$

166 such 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  $\mathcal{V}_j$  to  
 167 form  $\mathcal{V}_{j+1}$ . We can then extend Equation (2.6) into

$$A\widehat{\mathcal{W}}_j = \widehat{\mathcal{V}}_j \mathcal{G}_j + [0_{n \times k}, Q_{j-1}, W_j], \quad (2.8)$$

168 where  $\mathcal{G}_j \in \mathbb{C}^{(k+n_j) \times (k+n_j)}$  is the first  $k+n_j$  rows of  $\underline{\mathcal{G}}_j \in \mathbb{C}^{(k+n_j+p) \times (k+n_j)}$ ,  $Q_{j-1} = [Q_1, \dots, Q_{j-1}] \in$   
 169  $\mathbb{C}^{n \times n_{j-1}}$  accounts for all the abandoned directions. The matrix  $Q_{j-1}$  is rank deficient, and it reduces to the  
 170 zero matrix of  $\mathbb{C}^{n \times n_{j-1}}$  as long as no partial convergence has occurred.

171 In order to characterize a minimum norm solution in the space spanned by  $\widehat{\mathcal{W}}_j$  using Equation (2.8)  
 172 we need to form an orthonormal basis of the space spanned by  $[\widehat{\mathcal{V}}_j, Q_{j-1}, W_j]$ . This is performed by first  
 173 orthogonalizing  $Q_{j-1}$  against  $\widehat{\mathcal{V}}_j$ , that is  $\widetilde{Q}_{j-1} = (I - \widehat{\mathcal{V}}_j \widehat{\mathcal{V}}_j^H) Q_{j-1}$ . Because  $Q_{j-1}$  is of rank deficiency  
 174 so is  $\widetilde{Q}_{j-1}$  that can be written

$$\widetilde{Q}_{j-1} = P_{j-1} \mathbb{G}_{j-1} \text{ with } \begin{cases} P_{j-1} \in \mathbb{C}^{n \times q_j} \text{ has orthonormal columns with } \widehat{\mathcal{V}}_j^H P_{j-1} = 0, \\ \mathbb{G}_{j-1} \in \mathbb{C}^{q_j \times n_{j-1}} \text{ is of full rank with } q_j = p - p_j. \end{cases} \quad (2.9)$$

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

179 With this new basis, Equation (2.8) writes

$$\begin{aligned} A[U_k, \mathcal{V}_j] &= [C_k, \mathcal{V}_j] \begin{bmatrix} I & \mathcal{B}_j \\ 0 & \mathcal{L}_j \end{bmatrix} + \begin{bmatrix} 0_k, P_{j-1} \mathbb{G}_{j-1}, [P_{j-1}, \widetilde{W}_j] \end{bmatrix} \begin{bmatrix} E_j \\ D_j \end{bmatrix} \\ &= \begin{bmatrix} C_k, \mathcal{V}_j, [P_{j-1}, \widetilde{W}_j] \end{bmatrix} \begin{bmatrix} I_k & \mathcal{B}_j \\ 0_{(n_j+p) \times k} & \begin{matrix} \mathbb{G}_{j-1} & E_j \\ 0 & D_j \end{matrix} \end{bmatrix}, \end{aligned} \quad (2.10)$$



180 where  $\mathcal{L}_j = \begin{bmatrix} H_{1,1} & H_{1,2} & H_{1,3} & \cdots & H_{1,j} \\ H_{2,1} & H_{2,2} & H_{2,3} & \cdots & H_{2,j} \\ \mathbb{V}_3^H Q_1 & H_{3,2} & H_{3,3} & \cdots & H_{3,j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbb{V}_j^H Q_1 & \cdots & \mathbb{V}_j^H Q_{j-2} & H_{j,j-1} & H_{j,j} \end{bmatrix} \in \mathbb{C}^{n_j \times n_j}$  is no longer upper Hessenberg as  
 181 soon as one partial convergence occurs, i.e.,  $\exists \ell$ , s.t.,  $Q_\ell \neq 0$ .

Equation (2.10) can be rewritten in a more compact form as

$$A[U_k, \mathcal{V}_j] = [C_k, \mathcal{V}_j, [P_{j-1}, \widetilde{W}_j]] \mathcal{F}_j,$$

182 so that the least squares problem to be solved to compute the minimum residual norm solution associated  
 183 with the generalized Arnoldi relation (2.10) becomes

$$Y_j = \underset{Y \in \mathbb{C}^{(k+n_j) \times p}}{\operatorname{argmin}} \|\Lambda_j - \mathcal{F}_j Y\|_F, \quad (2.11)$$

184 with

$$\mathcal{F}_j = \begin{bmatrix} I_k & \mathcal{B}_j \\ 0_{(n_j+p) \times k} & \mathbb{G}_{j-1} \quad \mathcal{L}_j \quad E_j \\ & 0 & D_j \end{bmatrix} = \begin{bmatrix} \mathcal{F}_j \\ \mathbb{H}_j \end{bmatrix} \in \mathbb{C}^{(k+n_j+p) \times (k+n_j)} \quad (2.12)$$

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

187 The numerical mechanism to select  $\mathbb{V}_{j+1}$  out of  $[P_{j-1}, \widetilde{W}_j]$  follows the same ideas as discussed in [1,  
 188 20] in the context of block GMRES. The governing idea consists in building an orthonormal basis for the  
 189 directions that contribute the most to the individual residual norms and make them larger than a prescribed  
 190 threshold  $\tau$ . Specifically, the singular value decomposition (SVD) is applied to the least squares residuals

$$\Lambda_j - \mathcal{F}_j Y_j = \mathbb{U}_{1,L} \Sigma_1 \mathbb{U}_{1,R}^H + \mathbb{U}_{2,L} \Sigma_2 \mathbb{U}_{2,R}^H, \quad (2.13)$$

191 where  $\Sigma_1$  contains the  $p_{j+1}$  singular values larger than or equal to the prescribed threshold  $\tau$ . Then we  
 192 decompose  $\mathbb{U}_{1,L} = \begin{pmatrix} \mathbb{U}_1^{(1)} \\ \mathbb{U}_1^{(2)} \end{pmatrix}$  in accordance with  $[C_k, \mathcal{V}_j], [P_{j-1}, \widetilde{W}_j]$ , that is  $\mathbb{U}_1^{(1)} \in \mathbb{C}^{(k+n_j) \times p_{j+1}}$  and  
 193  $\mathbb{U}_1^{(2)} \in \mathbb{C}^{p \times p_{j+1}}$ . Because the objective is to construct an orthonormal basis, we consider a unitary matrix  
 194  $[\mathbb{W}_1, \mathbb{W}_2]$  such that  $\operatorname{Range}(\mathbb{W}_1) = \operatorname{Range}(\mathbb{U}_1^{(2)})$ . The new set of orthonormal candidate vectors used to  
 195 expand the search space

$$\mathbb{V}_{j+1} = [P_{j-1}, \widetilde{W}_j] \mathbb{W}_1 \quad (2.14)$$

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

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

196 is the new set of abandoned directions with orthonormal columns. Through this mechanism, directions that  
 197 have been abandoned at a given iteration can be reintroduced, if the residual block has a large component  
 198 along them. Furthermore, this selection strategy ensures that all the solutions have converged when  $p$   
 199 partial convergence have been detected. We do not give the details of the calculation and refer to Section 3  
 200 of [20] for a complete description, but only state that via this decomposition, the main terms that appear in  
 201 Equation (2.10) can be computed incrementally.

202 **2.4. Subspace recycling policies along with partial convergence detection.** So far, we have not  
 203 made any specific assumption on the definition of the recycling space  $U_k$  except that it has full column rank.  
 204 In the context of subspace recycling, one key point is to specify what subspace is to be recycled at restart.  
 205 At the cost of the extra storage of  $k$  vectors, block GCRO offers more flexibility than block GMRES in the  
 206 choice of the recycling space. This extra storage, that enables us to remove the constraints to have the search  
 207 space included in the residual space, allows us to consider any subspace to be deflated at restart. In particular  
 208 any of the two classical alternatives, that are Rayleigh-Ritz procedure and harmonic-Ritz procedure, can be  
 209 considered to compute the targeted approximated eigenvectors to define  $U_k$  and  $C_k$  at restart. Considering  
 210 a reasonable length of the current manuscript, we solely present the details of building a recycling subspace  
 211 based on harmonic-Ritz projection here. We refer the reader to our technical report [9, Section 2.4 and 5.1]  
 212 for the corresponding discussions on the implementation based on Rayleigh-Ritz procedure.

DEFINITION 1. *harmonic-Ritz projection.*

Consider a subspace  $\mathcal{W}$  of  $\mathbb{C}^n$ . Given a general nonsingular matrix  $A \in \mathbb{C}^{n \times n}$ ,  $\lambda \in \mathbb{C}$  and  $g \in \mathcal{W}$ ,  $(\lambda, g)$  is a harmonic-Ritz pair of  $A$  with respect to the space  $\mathcal{W}$  if and only if

$$Ag - \lambda g \perp A\mathcal{W}$$

or equivalently,

$$\forall w \in \text{Range}(A\mathcal{W}), \quad w^H (Ag - \lambda g) = 0.$$

213 The vector  $g$  is a harmonic-Ritz vector associated with the harmonic-Ritz value  $\lambda$ .

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

$$A\widehat{\mathcal{W}}_m = \widehat{\mathcal{V}}_{m+1}\widehat{\mathcal{F}}_m = \begin{bmatrix} C_k, \mathcal{V}_m, [P_{m-1}, \widetilde{W}_m] \end{bmatrix} \widehat{\mathcal{F}}_m, \quad (2.15)$$

$$X_m = X_1 + \widehat{\mathcal{W}}_m Y_m, \quad (2.16)$$

$$R_m = B - AX_m = \begin{bmatrix} C_k, \mathcal{V}_m, [P_{m-1}, \widetilde{W}_m] \end{bmatrix} (\Lambda_m - \widehat{\mathcal{F}}_m Y_m), \quad (2.17)$$

$$Y_m = \underset{Y \in \mathbb{C}^{(k+n_m) \times p}}{\text{argmin}} \|\Lambda_m - \widehat{\mathcal{F}}_m Y\|_F, \quad \Lambda_m = [0_{p \times k}, \Lambda_1^T, 0_{p \times n_m}]^T. \quad (2.18)$$

215 Then, a restart procedure has to be implemented to possibly refine the spectral information to be recycled  
 216 during the next cycle. Based on these equalities we will compute the approximated eigen-information as  
 217 shown in Proposition 1 and then use it to define the new deflation basis  $U_k^{new}$  and its orthonormal image  
 218  $C_k^{new}$  by  $A$  as described in Theorem 1.

219 PROPOSITION 1. *At restart of IB-BGCRO-DR, the update of the recycling subspace for the next  
 220 cycle relies on the computation of harmonic-Ritz vectors  $\widehat{\mathcal{W}}_m g_i \in \text{span}(\widehat{\mathcal{W}}_m)$  of  $A$  with respect to  
 221  $\widehat{\mathcal{W}}_m = [U_k, \mathcal{V}_m] \in \mathbb{C}^{n \times (k+n_m)}$ .*

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

$$\widehat{\mathcal{F}}_m^H \widehat{\mathcal{F}}_m g_i = \theta_j \widehat{\mathcal{F}}_m^H \widehat{\mathcal{V}}_{m+1}^H \widehat{\mathcal{W}}_m g_i, \quad \text{for } 1 \leq i \leq k + n_m, \quad (2.19)$$

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

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

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

$$\forall w \in \text{Range}(A\widehat{\mathcal{W}}_m) \quad w^H (A\widehat{\mathcal{W}}_m g_i - \theta_i \widehat{\mathcal{W}}_m g_i) = 0,$$

which is equivalent to

$$(A\widehat{\mathcal{W}}_m)^H (A\widehat{\mathcal{W}}_m g_i - \theta_i \widehat{\mathcal{W}}_m g_i) = 0.$$

225 Substituting Equation (2.15) into the above one leads to

$$\left( \widehat{\mathcal{V}}_{m+1} \widehat{\mathcal{F}}_m \right)^H \left( \widehat{\mathcal{V}}_{m+1} \widehat{\mathcal{F}}_m g_i - \theta_i \widehat{\mathcal{W}}_m g_i \right) = 0. \quad (2.20)$$



Because the columns of  $\widehat{\mathcal{V}}_{m+1} = [C_k, \mathcal{V}_m, [P_{m-1}, \widetilde{W}_m]]$  generated at the end of each cycle are orthonormal, Equation (2.20) becomes

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

226 which gives the formulation (2.19).  $\square$

227

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

232 **THEOREM 1.** *At restart of the IB-BGCRO-DR, if we intend to deflate the space  $\text{span}([U_k, \mathcal{V}_m]G_k)$   
233 where  $G_k = [g_1, \dots, g_k]$  is the set of vectors associated with the targeted eigenvalues, the matrices  $U_k^{new}$   
234 and  $C_k^{new}$  to be used for the next cycle are defined by*

$$U_k^{new} = \widehat{\mathcal{W}}_m G_k R^{-1} = [U_k, \mathcal{V}_m] G_k R^{-1}, \quad (2.21)$$

$$C_k^{new} = \widehat{\mathcal{V}}_{m+1} Q = [C_k, \mathcal{V}_m, [P_{m-1}, \widetilde{W}_m]] Q, \quad (2.22)$$

235 where  $Q$  and  $R$  are the factors of the reduced QR-factorization of the tall and skinny matrix  $\underline{\mathcal{F}}_m G_k$ , which  
236 ensure that  $AU_k^{new} = C_k^{new}$  and  $(C_k^{new})^H C_k^{new} = I_k$ .

237 *Proof.* Let  $Q$  and  $R$  be the factors of the reduced QR-factorization of the tall and skinny matrix  $\underline{\mathcal{F}}_m G_k$ .  
238 And right multiplying  $G_k$  on both sides of Equation (2.15) leads to  $A\widehat{\mathcal{W}}_m G_k = \widehat{\mathcal{V}}_{m+1} \underline{\mathcal{F}}_m G_k = \widehat{\mathcal{V}}_{m+1} QR$ ,  
239 that is equivalent to  $A\widehat{\mathcal{W}}_m G_k R^{-1} = \widehat{\mathcal{V}}_{m+1} \underline{\mathcal{F}}_m G_k R^{-1} = \widehat{\mathcal{V}}_{m+1} Q$ , concluding the proof as  
240  $\text{span}(\widehat{\mathcal{W}}_m G_k R^{-1}) = \text{span}(\widehat{\mathcal{W}}_m G_k)$  and  $\widehat{\mathcal{V}}_{m+1} Q$  is the product of two matrices with orthonormal  
241 columns so are its columns.  $\square$

242

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

245 *Proof.*  $X_m^{old} = X_1 + \widehat{\mathcal{W}}_m Y_m$  where  $Y_m$  solves the least squares problem (2.18) so that  $(\Lambda_m - \underline{\mathcal{F}}_m Y_m) \in$   
246  $(\text{Range}(\underline{\mathcal{F}}_m))^\perp = \text{Null}(\underline{\mathcal{F}}_m^H)$ . We also have  $R_m^{old} = \widehat{\mathcal{V}}_{m+1} (\Lambda_m - \underline{\mathcal{F}}_m Y_m)$ , consequently

$$\begin{aligned} (C_k^{new})^H R_m^{old} &= (\widehat{\mathcal{V}}_{m+1} Q)^H (\widehat{\mathcal{V}}_{m+1} (\Lambda_m - \underline{\mathcal{F}}_m Y_m)) \\ &= (\widehat{\mathcal{V}}_{m+1} \underline{\mathcal{F}}_m G_k R^{-1})^H (\widehat{\mathcal{V}}_{m+1} (\Lambda_m - \underline{\mathcal{F}}_m Y_m)) \\ &= R^{-H} G_k^H \underbrace{\underline{\mathcal{F}}_m^H (\Lambda_m - \underline{\mathcal{F}}_m Y_m)}_{= 0 \text{ because of (2.18)}} = 0. \end{aligned}$$

247  $\square$

248

249 **2.5. A variant suited for flexible preconditioning.** All what have been described in the previous  
250 sections are naturally extended to the right preconditioning case with a fixed preconditioner  $M$ , and the  
251 central equality writes as

$$A[U_k, M\mathcal{V}_m] = [C_k, \mathcal{V}_m, [P_{m-1}, \widetilde{W}_m]] \underline{\mathcal{F}}_m. \quad (2.23)$$

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

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

and the solution is

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

If we denote  $\mathcal{M}_j$  a (possibly nonlinear) nonsingular preconditioning operator at iteration  $j$  and  $\mathcal{M}_j(\mathbb{V}_j)$  denotes the action of  $\mathcal{M}_j$  on a block vector  $\mathbb{V}_j$ , Equation (2.23) translates into

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

252 which writes in a more compact form as

$$A\widehat{\mathcal{L}}_m = \widehat{\mathcal{V}}_{m+1}\mathcal{F}_m \text{ with } \widehat{\mathcal{L}}_m = [U_k, \mathcal{L}_m] \text{ and } \widehat{\mathcal{V}}_{m+1} = \left[ C_k, \mathcal{V}_m, [P_{m-1}, \widetilde{W}_m] \right]. \quad (2.24)$$

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

256 In the context of flexible preconditioning many strategies for defining harmonic-Ritz vectors can be  
257 envisioned for GCRO-DR. Among those considered in [4], we follow the one with a lower computational  
258 cost required in solving the generalized eigenvalue problem, referred to as Strategy C in [4]. Furthermore, it  
259 also allows us to obtain counterpart properties in the flexible preconditioning case that are quite similar to the  
260 ones we have exposed in the non-preconditioned case as shown in Section 2.4. We refer to [9, Appendix A]  
261 for another two strategies for approximating targeted eigen-information. Proposition 2 indicates that with  
262 an appropriate definition of the harmonic-Ritz vectors, all the properties of IB-BGCRO-DR extend to the  
263 flexible preconditioning variant denoted as IB-BFGCRO-DR.

264 **PROPOSITION 2.** *At the end of a cycle of the IB-BFGCRO-DR algorithm, if the deflation space is*  
265 *built on the harmonic-Ritz vectors  $\mathcal{W}_m g_i \in \text{span}(\mathcal{W}_m)$  of  $A\widehat{\mathcal{L}}_m \mathcal{W}_m^\dagger$  with respect to  $\mathcal{W}_m = [\mathcal{W}_k, \mathcal{V}_m] \in$*   
266  *$\mathbb{C}^{n \times (k+n_m)}$  :*

267 1. *The harmonic-Ritz pairs  $(\theta_i, \mathcal{W}_m g_i)$  for all restarts satisfy*

$$\mathcal{F}_m^H \mathcal{F}_m g_i = \theta_j \mathcal{F}_m^H \widehat{\mathcal{V}}_{m+1}^H \mathcal{W}_m g_i, \quad \text{for } 1 \leq i \leq k + n_m, \quad (2.25)$$

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

269 2. *At restart, if  $G_k = [g_1, \dots, g_k]$  is associated with the  $k$  targeted eigenvalues, the matrices  $\mathcal{W}_k^{\text{new}}$ ,*  
270  *$U_k^{\text{new}}$  and  $C_k^{\text{new}}$  to be used for the next cycle are updated by*

$$\mathcal{W}_k^{\text{new}} = \mathcal{W}_m G_k R^{-1} = [\mathcal{W}_k, \mathcal{V}_m] G_k R^{-1}, \quad (2.26)$$

$$U_k^{\text{new}} = \widehat{\mathcal{L}}_m G_k R^{-1} = [U_k, \mathcal{L}_m] G_k R^{-1}, \quad (2.27)$$

$$C_k^{\text{new}} = \widehat{\mathcal{V}}_{m+1} Q = \left[ C_k, \mathcal{V}_m, [P_{m-1}, \widetilde{W}_m] \right] Q,$$

271 where  $Q$  and  $R$  are the factors of the reduced QR-factorization of the tall and skinny matrix  
272  $\mathcal{F}_m G_k$ , ensuring  $A U_k^{\text{new}} = C_k^{\text{new}}$  with  $(C_k^{\text{new}})^H C_k^{\text{new}} = I_k$ .

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

274 *Proof.* The proof essentially follows the same arguments as the ones developed for IB-BGCRO-DR  
275 described in Section 2.4, and we refer the reader to the [9, Appendix B] for the details.  $\square$

276  
277 We also mention that a closely related numerical technique that extends IB-BGMRES-DR in the flexible  
278 preconditioning context can be derived similarly. We refer to [9, Appendix C] where the resulting new  
279 algorithm named IB-BFGMRES-DR is detailed and its properties are described.

280 **3. Search space expansion policies governed by the stopping criterion.** In this section we  
281 describe a few novel policies to expand the search space that generalize the original one considered for  
282 inexact breakdown detection [20]. In particular we first show how numerical criteria to detect the partial  
283 convergence and expand the search space expansion can be tuned to ensure that a targeted threshold for a  
284 prescribed stopping criterion based on the individual backward error solution will be eventually satisfied.  
285 Secondly, we present how computational constraints can be taken into account, and combined with any of  
286 the previous numerical criteria, to best cope with the performance of the underlying computer architecture.

287 The partial convergence detection shortly described in Section 2.3 ensures that if all the singular values  
 288 of the least squares residual are smaller than the threshold  $\tau$ , then all the linear system residual norms are  
 289 also smaller than  $\tau$  (i.e.,  $p$  partial convergences have occurred). This is due to the following inequality

$$\forall i \quad \|b^{(i)} - Ax_j^{(i)}\| \leq \|B - AX_j\| = \|\Lambda_j - \mathcal{F}_j Y_j\| = \sigma_{\max}(\Lambda_j - \mathcal{F}_j Y_j) < \tau, \quad (3.1)$$

290 which follows from the fact that the 2-norm of a matrix is an upper bound of the 2-norm of its individual  
 291 columns and  $\widehat{\mathcal{V}}_{j+1}$  has orthonormal columns.

292 **3.1. Search space expansion policy governed by  $\eta_b$ .** A classical stopping criterion for the solution  
 293 of a linear system  $Ax = b$  is based on backward error analysis and consists in stopping the iteration when

$$\eta_b(x_j) = \frac{\|b - Ax_j\|}{\|b\|} \leq \varepsilon. \quad (3.2)$$

294 This criterion was considered in [1] where it was consequently proposed to define  $\tau = \varepsilon \min_{i=1, \dots, p} \|b^{(i)}\|$ .

295 With this choice, when the iteration complies with Equation (3.1), we have

$$\eta_b(x_j^{(i)}) \leq \frac{\|b - Ax_j^{(i)}\|}{\min_{i=1, \dots, p} \|b^{(i)}\|} \leq \varepsilon. \quad (3.3)$$

296 When the different right-hand sides have very different norms in magnitude, the subspace expansion  
 297 associated with this criterion might not be effective as the upper bound in Equation (3.3) will not be  
 298 tight. This leads to enlarging the search space with directions that are not relevant (generating useless  
 299 computation). In that context a better choice would be to better focus on the space expansion to reduce  
 300 more the residual associated with right-hand side of large norm. For that purpose, the idea is to perform the  
 301 SVD not directly on the least squares residual but on its scaled least squares residual.

302 **PROPOSITION 3.1.** *Performing the SVD of the scaled least squares residuals  $(\Lambda_j - \mathcal{F}_j Y_j)D_{b, \varepsilon}$  with*  
 303 *threshold  $\tau = 1$  and  $D_{b, \varepsilon} = \varepsilon^{-1} \text{diag}(\|b^{(1)}\|^{-1}, \dots, \|b^{(p)}\|^{-1})$  ensures that when  $p$  partial convergences*  
 304 *have occurred, so that the search space cannot be enlarged, the current individual iterates comply with the*  
 305 *stopping criterion (3.2).*

*Proof.* This is a direct consequence of the following inequalities

$$\max_{i=1, \dots, p} \frac{\|b^{(i)} - Ax_j^{(i)}\|}{\varepsilon \|b^{(i)}\|} \leq \|(B - AX_j)D_{b, \varepsilon}\| = \|(\Lambda_j - \mathcal{F}_j Y_j)D_{b, \varepsilon}\| \leq 1$$

306 that implies  $\forall i \quad \eta_b(x_j^{(i)}) \leq \varepsilon$ .  $\square$

307

308 In some applications all the solutions associated with a block of right-hand sides do not need to be  
 309 solved with the same accuracy. That is, we may have to solve a family of right-hand sides  $B = [b^{(1)}, \dots, b^{(p)}]$   
 310 with individual convergence thresholds  $\varepsilon^{(i)}$  for the solution associated with each right-hand side  $b^{(i)}$  ( $i =$   
 311  $1, \dots, p$ ), thus we have a more general version of Equation (3.2) as

$$\eta_{b^{(i)}}(x_j^{(i)}) = \frac{\|b^{(i)} - Ax_j^{(i)}\|}{\|b^{(i)}\|} \leq \varepsilon^{(i)}. \quad (3.4)$$

312 In that context, the subspace expansion policy can be easily adapted to ensure the convergence for each  
 313 individual accuracy.

314 **COROLLARY 2.** *Performing the SVD of the scaled least squares residuals  $(\Lambda_j - \mathcal{F}_j Y_j)D_{b, \varepsilon_i}$*   
 315 *with threshold  $\tau = 1$  and  $D_{b, \varepsilon_i} = \text{diag}((\varepsilon_1 \|b^{(1)}\|)^{-1}, \dots, (\varepsilon_p \|b^{(p)}\|)^{-1})$  ensures that when  $p$  partial*  
 316 *convergences have occurred the current individual iterates comply with the stopping criterion (3.4).*

317 **3.2. Search space expansion policy governed by  $\eta_{A, b}$ .** One can also adapt the expansion policy  
 318 described in the previous section to the situation where the stopping criterion is based on the normwise  
 319 backward error on  $A$  and  $b$ , defined by

$$\eta_{A, b}(x_j) = \frac{\|b - Ax_j\|}{\|b\| + \|A\| \|x_j\|} \leq \varepsilon. \quad (3.5)$$

320 It suffices to define accordingly the scaled least squares residuals in the SVD that is involved in the search  
 321 space expansion. We notice that this type of stopping criterion will have a computational penalty as the  
 322 iterates of all individual iterations have to be computed to calculate their norm.

323 **COROLLARY 3.** *Performing the SVD of the scaled least squares residual  $(\Lambda_j - \mathcal{F}_j Y_j) D_{A,b,\varepsilon}$  with  
 324 threshold  $\tau = 1$  and  $D_{A,b,\varepsilon} = \varepsilon^{-1} \text{diag}((\|A\| \|x_j^{(1)}\| + \|b^{(1)}\|)^{-1}, \dots, (\|A\| \|x_j^{(p)}\| + \|b^{(p)}\|)^{-1})$  ensures  
 325 that when  $p$  partial convergences have occurred, the current individual iterates comply with the stopping  
 326 criterion (3.5).*

327 We do not develop further these ideas but similarly we could define expansion policies where for each  
 328 solution we can select either  $\eta_b$  or  $\eta_{A,b}$  as stopping criterion with individual threshold setting.

329 The occurrence of  $p$  partial convergences is a sufficient condition that ensures the convergence of the  
 330  $p$  solution vectors, but the convergence might happen before and a more classic stopping criterion can be  
 331 accommodated at a low computational cost. Given the norms of true residuals are very close to those of the  
 332 least squares residuals when the loss of orthogonality of the generated block Krylov basis is not too serious,  
 333 one can also check the convergence by looking at the norm of the least squares residual, which is easy to  
 334 compute. Let  $Q_j^{LS} R_j^{LS}$  be a full QR-factorization of  $\mathcal{F}_j$  (i.e.,  $Q_j^{LS}$  is unitary), then

$$\Lambda_j - \mathcal{F}_j Y_j = Q_j^{LS} \begin{pmatrix} 0_{(n_j+k) \times p} \\ R_j^{LS} \end{pmatrix}, \quad (3.6)$$

335 where  $R_j^{LS} \in \mathbb{C}^{p \times p}$  are the last  $p$  rows of  $(Q_j^{LS})^H \Lambda_j$  so that  $\|b^{(i)} - Ax_j^{(i)}\| = \|R_j^{LS}(:, i)\|$ . Those residual  
 336 norm calculations are part of the stopping criterion based on  $\eta_b$  or  $\eta_{A,b}$

337 **3.3. Search space expansion policy governed by computational performance.** Based on any of  
 338 these expansion policies, the abandoned directions at a given iteration might be reintroduced in a subsequent  
 339 one, thereby we can trade on the considered numerical policy and select for the subspace expansion only  
 340 a subset of those eligible. In particular, it might be relevant to choose a prescribed block size  $p^{CB}$  (here  
 341 the superscript  $^{CB}$  stands for Computational Blocking) that is suited to best cope with the computational  
 342 features on a given platform rather than selecting the numerical block size  $p_{j+1}$  defined as the number  
 343 of singular values larger than or equal to the prescribed threshold  $\tau = 1$ . In that respect, we consider a  
 344 subspace expansion policy so that the block size at the end of step  $j$  is defined as  $p_{j+1}^{CB} = \min(p^{CB}, p_{j+1})$ .  
 345 We refer this variant as Inexact Breakdown Block GCRO-DR with computational blocking (denoted by  
 346 IB-BGCRO-DR-CB).

347 Note that all the subspace expansion policies discussed in Section 3 could be applied to any other  
 348 block minimum residual norm methods equipped with the partial convergence detection such as the IB-  
 349 BGMRES [20] and IB-BGMRES-DR [1] algorithms.

350 **4. Remarks on some computational and algorithmic aspects.** The mathematical description made  
 351 in the previous section assumes exact calculation. In practice, the numerical behavior of the algorithms does  
 352 depend on the numerical algorithms selected to perform the computation in finite precision arithmetic. In  
 353 particular, all the above descriptions assume the orthonormality of the residual basis; it ensures the norm  
 354 equality of the true linear system residual and their least squares counterpart which governs the numerical  
 355 search space expansion policies described in the previous section. In our implementation, for the block  
 356 Arnoldi procedure (See Algorithm 1), we consider the block Modified Gram-Schmidt (BMGS) algorithm  
 357 with reduced QR factorization based on Householder reflections of the final tall and skinny block (referred  
 358 to as (BMGS  $\circ$  HouseQR) in [3]). In addition, at restart the re-orthogonalization of the recycling space  $C_k$   
 359 and initial block residual vector  $[\mathbb{V}_1, P_0]$  in Equation (4.2) is performed a vector at a time using Modified  
 360 Gram-Schmidt. For the sake of conciseness, we do not necessarily give the full technical details of what we  
 361 briefly expose in the core of the paper but sometimes refer to a particular part in the appendix.

362 **4.1. Inexact breakdown and re-orthogonalization at restart.** For the sake of simplicity of  
 363 exposure, in the previous sections we made the assumption that the initial residual block was of full rank. In  
 364 practice, this constraint can be removed by applying the partial convergence detection to the initial residual  
 365 block. In that case, only a subspace of the space spanned by the columns of the initial residual block will  
 366 be selected to define the first search space and the abandoned directions are kept in the basis of the residual  
 367 space. This has two main consequences:

- 368 1. The first iteration needs some extra attention to set up the initial basis  $\mathbb{V}_1$  and abandoned directions  
 369  $P_0$  defined in Equation (2.9).

2. A consequence of having abandoned directions in the first search space is that the projection of the initial residual block in the residual space, that defines the right-hand side of the least squares residual solved at each block iteration, will no longer have the nested block structure that is expanded by a  $p \times p$  zero block at each block iteration as presented in Equation (2.18).

Without loss of generality, let us present the partial convergence detection and re-orthogonalization at restart where the recycling subspace  $U_k^{new}$  and  $C_k^{new}$  are defined by Equation (2.21) and (2.22), so that mathematically  $AU_k^{new} = C_k^{new}$  and  $(C_k^{new})^H C_k^{new} = I_k$  and the initial residual block  $R_1^{new} = R_1$  in Corollary 1 is orthogonal to  $C_k^{new}$ . For a prescribed stopping criterion and convergence threshold, let us denote  $D_\varepsilon$  the diagonal matrix used to select the space expansion described in the Section 3. Let

$$R_1 D_\varepsilon = [\mathbb{V}_1^{new}, P_0^{new}] \begin{bmatrix} \Sigma_{p_1} & \\ & \Sigma_{q_1} \end{bmatrix} \mathbb{V}_{R_1}^H = [\mathbb{V}_1^{new}, P_0^{new}] \hat{\Lambda}_1', \quad (4.1)$$

where  $\mathbb{V}_1^{new} \in \mathbb{C}^{n \times p_1}$ ,  $P_0^{new} \in \mathbb{C}^{n \times q_1}$  with  $p_1 + q_1 = p$ , and  $\Sigma_{p_1}$  contains the  $p_1$  singular values of  $R_1 D_\varepsilon$  larger than or equal to the prescribed  $\tau$ , and  $\Sigma_{q_1}$  the ones smaller than  $\tau$ .

We first perform an MGS re-orthogonalization of the columns of  $[C_k^{new}, [\mathbb{V}_1^{new}, P_0^{new}]]$  that writes

$$[C_k^{new}, [\mathbb{V}_1^{new}, P_0^{new}]] = [C_k, [\mathbb{V}_1, P_0]] \begin{bmatrix} R_{11} & R_{12} \\ & R_{22} \end{bmatrix}, \quad (4.2)$$

where all the columns of  $[C_k, [\mathbb{V}_1, P_0]]$  are orthogonal to each other,  $\begin{bmatrix} R_{11} & R_{12} \\ & R_{22} \end{bmatrix} \in \mathbb{C}^{(k+p) \times (k+p)}$  is an upper triangular matrix with  $R_{11} \in \mathbb{C}^{k \times k}$  and  $R_{22} \in \mathbb{C}^{p \times p}$ . Next, we update  $U_k = U_k^{new} R_{11}^{-1}$  to satisfy Equation (2.1), and  $\mathcal{V}_1 = \mathbb{V}_1$  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. Form  $W_1 = A\mathbb{V}_1$  and orthogonalize it (using BMGS  $\circ$  HouseQR) against the set of orthonormal vectors that are part of the residual space  $[C_k, \mathbb{V}_1, P_0]$  which enables the computation of the entries of  $\mathcal{B}_1 = C_k^H W_1$ ,  $\mathcal{L}_{1,1} = \mathbb{V}_1^H W_1$  and  $E_1 = P_0^H W_1$ .
2. The resulting block  $\bar{W}_1$  formally writes  $\bar{W}_1 = W_1 - C_k \mathcal{B}_1 - \mathbb{V}_1 \mathcal{L}_{1,1} - P_0 E_1$  with  $\bar{W}_1 = \tilde{W}_1 D_1$  being its reduced  $QR$ -factorization.
3. In matrix form the above relations also writes

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

So that we have the first Arnoldi-like relation

$$A[U_k, \mathbb{V}_1] = [C_k, \mathbb{V}_1, [P_0, \tilde{W}_1]] \underline{\mathcal{F}}_1 \quad (4.3)$$

with

$$\underline{\mathcal{F}}_1 = \begin{bmatrix} I_k & \mathcal{B}_1 \\ 0_{(p_1+p) \times k} & \begin{bmatrix} \mathcal{L}_{1,1} \\ \tilde{\mathbb{H}}_1 \end{bmatrix} \end{bmatrix} \in \mathbb{C}^{(k+p_1+p) \times (k+p_1)} \text{ and } \tilde{\mathbb{H}}_1 = \begin{bmatrix} E_1 \\ D_1 \end{bmatrix} \in \mathbb{C}^{p \times p_1}.$$

4. Next, define the minimum norm solution  $X_2 = X_1 + [U_k, \mathbb{V}_1]Y$  and notice that  $R_1$  belongs to the space  $[C_k, \mathbb{V}_1, P_0, \tilde{W}_1]$  where its components in this orthogonal basis are given by  $[C_k, \mathbb{V}_1, P_0, \tilde{W}_1]^H R_1$ . From Equation (4.3) we have

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

and then from Equation (4.1), we have

$$R_1 = [\mathbb{V}_1^{new}, P_0^{new}] \hat{\Lambda}_1' D_\varepsilon^{-1} = [\mathbb{V}_1^{new}, P_0^{new}] \hat{\Lambda}_1 \text{ with } \hat{\Lambda}_1 = \hat{\Lambda}_1' D_\varepsilon^{-1}. \quad (4.4)$$

398 So that from (4.2), the right-hand side of the above least squares residual reads

$$\begin{aligned}
\Lambda_1 &= [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [\mathbb{V}_1^{new}, P_0^{new}] \hat{\Lambda}_1 = [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [C_k R_{12} + [\mathbb{V}_1, P_0] R_{22}] \hat{\Lambda}_1 \\
&= \left( [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H C_k R_{12} + [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [\mathbb{V}_1, P_0] R_{22} \right) \hat{\Lambda}_1 \\
&= \begin{bmatrix} R_{12} \\ 0_{(p_1+p) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} 0_{k \times p_1} & 0_{k \times q_1} \\ I_{p_1} & 0_{p_1 \times q_1} \\ 0_{q_1 \times p_1} & I_{q_1} \\ 0_{p_1 \times p_1} & 0_{p_1 \times q_1} \end{bmatrix} R_{22} \hat{\Lambda}_1 \in \mathbb{C}^{(k+p_1+p) \times p}. \tag{4.5}
\end{aligned}$$

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

$$Y_1 = \operatorname{argmin}_{Y \in \mathbb{C}^{(k+p_1) \times p}} \|\Lambda_1 - \mathcal{F}_1 Y\|_F.$$

400 6. Execute the search space expansion policy following the IB principles

401 (a) compute the SVD of the scaled least squares residual

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

402 (b) Compute  $\mathbb{W}_1$  and  $\mathbb{W}_2$  such that  $\operatorname{Range}(\mathbb{W}_1) = \operatorname{Range}(\mathbb{U}_1^{(2)}) \in \mathbb{C}^{p \times p_2}$  with  $\mathbb{U}_{1,L} =$   
403  $\begin{pmatrix} \mathbb{U}_1^{(1)} \\ \mathbb{U}_1^{(2)} \end{pmatrix} \in \mathbb{C}^{(k+p_1+p) \times p_2}$ ,  $[\mathbb{W}_1, \mathbb{W}_2]$  is unitary and  $\mathbb{W}_2 \in \mathbb{C}^{p \times q_2}$  with  $p_2 + q_2 = p$ .  
404

(c) Compute the new orthonormal matrices  $\mathbb{V}_2$  and  $P_1$  as

$$\mathbb{V}_2 = [P_0, \widetilde{W}_1] \mathbb{W}_1 \in \mathbb{C}^{n \times p_2}, \quad P_1 = [P_0, \widetilde{W}_1] \mathbb{W}_2 \in \mathbb{C}^{n \times q_2},$$

as well as the last block row matrix  $\mathcal{L}_{2,:}$  of  $\mathcal{L}_1$  and  $\mathbb{G}_1$  as

$$\mathcal{L}_{2,:} = \mathbb{W}_1^H \widetilde{\mathbb{H}}_1 \in \mathbb{C}^{p_2 \times p_1}, \quad \mathbb{G}_1 = \mathbb{W}_2^H \widetilde{\mathbb{H}}_1 \in \mathbb{C}^{q_2 \times p_1}.$$

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

406 Whenever a partial convergence is detected in  $R_1$ , some of its components (along  $P_0^{new}$ ) are firstly  
407 abandoned but could be reintroduced in some subsequent iterations. One of the consequences, is that the last  
408  $q_1$  columns of the least squares right-hand side problem will evolve from one iteration to the next, depending  
409 on how some of the  $P_0^{new}$  directions will be re-introduced in the search space along the iterations. There is  
410 a way to incrementally update the least squares right-hand side to be discussed in the next proposition.

411 PROPOSITION 3. At each iteration of IB-BGCRO-DR, the new least squares problem reads

$$Y_{j+1} = \operatorname{argmin}_{Y \in \mathbb{C}^{(k+n_{j+1}) \times p}} \|\Lambda_{j+1} - \mathcal{F}_{j+1} Y\|_F, \quad \Lambda_{j+1} \in \mathbb{C}^{(k+n_{j+1}+p) \times p}, \quad j = 0, 1, 2, \dots \tag{4.6}$$

412 with the updated right-hand sides being

$$\Lambda_{j+1} = \begin{bmatrix} R_{12} \\ 0_{(n_j+p+p_{j+1}) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} 0_{k \times p_1} & 0_{k \times q_1} \\ I_{p_1} & \Phi_{j+1} \\ 0_{(n_j+p-p_1) \times p_1} & \\ 0_{p_{j+1} \times p_1} & 0_{p_{j+1} \times q_1} \end{bmatrix} R_{22} \hat{\Lambda}_1, \tag{4.7}$$

413 where  $\Phi_{j+1} = \begin{bmatrix} \Phi_j(1:n_j,:) \\ [\mathbb{W}_1, \mathbb{W}_2]^H \begin{bmatrix} \Phi_j(n_j+1:n_j+q_j,:) \\ 0_{p_j \times q_1} \end{bmatrix} \end{bmatrix} \in \mathbb{C}^{(n_j+p) \times q_1}$  for  $j = 0, 1, 2, \dots$ , with  
414  $\Phi_1 = \begin{bmatrix} 0_{p_1 \times q_1} \\ I_{q_1} \end{bmatrix} \in \mathbb{C}^{p \times q_1}$ ,  $q_j = p - p_j$  ( $j > 0$ ) and  $[\mathbb{W}_1, \mathbb{W}_2]$  is unitary as defined in the search space  
415 expansion algorithm based on IB principles,  $R_{12} \in \mathbb{C}^{k \times p}$  and  $R_{22} \in \mathbb{C}^{p \times p}$  are two block components of  
416 the upper triangular matrix as shown in the right-hand side of Equation (4.2).

417 Proof. We refer the reader to Appendix A for the proving details.  $\square$

418  
419 Based on the above discussions, the IB-BGCRO-DR algorithm with partial convergence detection in the  
420 initial residual block and updated right-hand sides of the least squares residual is presented in Algorithm 2  
421 for solving a series of linear systems with slowly-changing left-hand sides.



---

**Algorithm 2** IB-BGCRO-DR for slowly-changing left-hand sides and massive number of right-hand sides
 

---

**Require:**  $A \in \mathbb{C}^{n \times n}$  left-hand side of current family (supposed not vary much compared to previous one)

**Require:**  $B \in \mathbb{C}^{n \times p}$  the block of right-hand-sides and  $X_0 \in \mathbb{C}^{n \times p}$  the block initial guess

**Require:**  $m$  maximum number of Arnoldi step within a cycle

**Require:**  $p^{CB}$  a given constant number satisfying  $1 \leq p^{CB} \leq p$  for computational blocking

**Require:**  $D_\varepsilon \in \mathbb{C}^{p \times p}$  a diagonal matrix used to select the space expansion described in the Section 3

**Require:**  $U_k, C_k \in \mathbb{C}^{n \times k}$  the recycling subspaces supposed be empty for the first family and obtained after solving previous slow-changing family

- 1: Compute  $R_0 = B - AX_0$   
*/\* Some families have already been solved ? \*/*
  - 2: **if** the recycling space is not empty,  $U_k \neq 0$  **then**
  - 3: Apply the reduced  $QR$ -factorization to  $AU_k$  for updating  $U_k$  and  $C_k$  for the current family such that the  $U_k$  and  $C_k$  satisfy Equation (2.1) and (2.2). Compute  $R_1$  and  $X_1$  as described in Equation (2.4)
  - 4: **else**
  - 5: Set  $R_1 = R_0, X_1 = X_0, U_k = 0, C_k = 0$
  - 6: **end if**  
*/\* Loop over the restarts \*/*
  - 7: **while** the stopping criterion based on Section 3.1 or 3.2 is not met **do**
  - 8: Apply partial convergence detection in the scaled (least squares) residual block following Section 4.1  
*/\* Arnoldi loop \*/*
  - 9: **for**  $j = 2, 3, \dots, m$  **do**
  - 10: Orthogonalize  $AV_j$  against  $C_k$  as  $W_j = (I - C_k C_k^H)AV_j$ . Then orthogonalize  $W_j$  against previous block orthonormal vector  $\mathcal{V}_j = [\mathbb{V}_1, \dots, \mathbb{V}_j]$  as  

$$W_j = AV_j - C_k C_k^H AV_j - \mathcal{V}_j \mathcal{L}_{1,1:j}, \text{ where } \mathcal{L}_{1,1:j} = \mathcal{V}_j^H (W_j) = \mathcal{V}_j^H (AV_j) \text{ is a block column matrix}$$
  - 11: Set  $\mathcal{L}_j = [\mathcal{L}_{j-1}, \mathcal{L}_{1,1:j}] \in \mathbb{C}^{n_j \times n_j}, \mathcal{B}_j = [\mathcal{B}_{j-1}, C_k^H AV_j] \in \mathbb{C}^{k \times n_j}$
  - 12: Orthogonalize  $W_j$  against  $P_{j-1}$  and carry out its reduced  $QR$ -factorization as  

$$\widetilde{W}_j D_j = W_j - P_{j-1} E_j, \text{ where } E_j = P_{j-1}^H W_j$$
  - 13: Compute  $Y_j$  by solving the least squares problem described in Equation (2.11) (or (4.6)) with  $\mathcal{F}_j$  shown in Equation (2.12) composed by  $\mathcal{F}_j$  and  $\mathbb{H}_j$  but with the updated right-hand side  $\Lambda_j$  as shown in Equation (4.7) instead
  - 14: **if** the stopping criterion is met **then**
  - 15: **return**  $X_j = X_1 + [U_k, \mathcal{V}_j] Y_j, U_k$  and  $C_k$
  - 16: **end if**
  - 17: Singular value decomposition of the residuals scaled by  $D_\varepsilon$   

$$(\Lambda_j - \mathcal{F}_j Y) D_\varepsilon = \mathbb{U}_{1,L} \Sigma_1 \mathbb{V}_{1,R}^H + \mathbb{U}_{2,L} \Sigma_2 \mathbb{V}_{2,R}^H \text{ with } \sigma_{\min}(\Sigma_1) \geq 1 > \sigma_{\max}(\Sigma_2)$$
  - 18: **if** Computational blocking of Section 3.3 is activated **then**
  - 19:  $\mathbb{U}_{1,L} = \mathbb{U}_{1,L}(:, 1 : p_j^{CB})$  with  $p_j^{CB} = \min(p^{CB}, nl_{\Sigma_1}), nl_{\Sigma_1}$  refers to column number of  $\Sigma_1$
  - 20: **end if**
  - 21: Following item 6 described in Section 4.1 for computing  $\mathbb{W}_1$  and  $\mathbb{W}_2$
  - 22: Compute orthonormal matrices  $\mathbb{V}_{j+1}$  and  $P_j$ , the last block row matrix  $\mathcal{L}_{j+1,:}$  of  $\mathcal{L}_j$ , and  $G_j$  as  

$$\mathbb{V}_{j+1} = [P_{j-1}, \widetilde{W}_j] \mathbb{W}_1, P_j = [P_{j-1}, \widetilde{W}_j] \mathbb{W}_2, \mathcal{L}_{j+1,:} = \mathbb{W}_1^H \mathbb{H}_j, \mathbb{G}_j = \mathbb{W}_2^H \mathbb{H}_j, \underline{\mathcal{L}}_j = \begin{pmatrix} \mathcal{L}_j \\ \mathcal{L}_{j+1,:} \end{pmatrix}$$
  - 23: **end for**  
*/\* Restart procedure \*/*
  - 24: Compute the solution  $X_m$  as described in Equation (2.16) and residual  $R_m$  according to (2.17)
  - 25: Compute the targeted harmonic-Ritz vectors  $G_k = [g_1, \dots, g_k]$  by solving the generalized eigenvalue problem (2.19) described in Proposition 1
  - 26: Update the values of  $U_k$  and  $C_k$  respectively by Equation (2.21) and (2.22) described in Theorem 1
  - 27: Restart with  $X_1 = X_m, \mathcal{V}_{m+1}, R_1^{LS} = \Lambda_m - \mathcal{F}_m Y_m (R_1 = R_m = \mathcal{V}_{m+1} R_1^{LS})$
  - 28: **end while**
  - 29: **return**  $X_j$  for approximation of the current family;  $U_k, C_k$  for the next family to be solved
-

422 **4.2. Solution of the least squares problem and cheap SVD calculation of the scaled least squares**  
 423 **residual.** Computing the full  $QR$ -factorization of the matrices involved in the least squares problems allows  
 424 us to reuse its  $Q$  factor to compute the SVD of the least squares residual using a QR-SVD algorithm such  
 425 that the actual SVD decomposition is performed on a  $p \times p$  block  $R_j^{\ell s} D_\varepsilon$ , where  $R_j^{\ell s}$  appeared in the right-  
 426 hand side of Equation (3.6), at each iteration (we refer to Appendix B for the details of this calculation).  
 427 Note that this observation applies naturally to the IB-BGMRES [20] and IB-BGMRES-DR [1] algorithms  
 428 as well.

429 **5. Numerical experiments.** In the following sections we illustrate different numerical features of the  
 430 novel algorithm introduced above. For the sake of comparison, in some of the experiments we also display  
 431 results of closely related block methods such as BGCRO-DR [17, 18, 22, 29] or IB-BGMRES-DR [1]. All the  
 432 numerical experiments have been run using a MATLAB prototype, so that the respective performances of the  
 433 algorithms are evaluated in term of number of matrix-vector products, denoted as  $mvps$  (and preconditioner  
 434 applications in the preconditioned case) required to converge.

435 For each set of block of right-hand sides, referred to as a family, the block initial guess is equal to  
 436  $0 \in \mathbb{C}^{n \times p}$ , where  $p$  is the number of right-hand sides. The block right-hand side  $B = [b^{(1)}, b^{(2)}, \dots, b^{(p)}] \in$   
 437  $\mathbb{C}^{n \times p}$  is composed of  $p$  linearly independent vectors generated randomly (using the same seed when block  
 438 methods are compared). While any part of the spectrum could be considered to define the recycling space we  
 439 consider for all the experiments the approximated eigenvectors associated with the  $k$  smallest approximated  
 440 eigenvalues in magnitude. The maximum dimension of the search space in each cycle is set to be  $m_d =$   
 441  $15 \times p$ . To illustrate the potential benefit of IB-BGCRO-DR when compared to another block solver, we  
 442 consider the overall potential gain when solving a sequence of  $\ell$  families defined as

$$\text{Gain}(\ell) = \frac{\sum_{s=1}^{\ell} \#mvps(\text{method})^{(s)}}{\sum_{s=1}^{\ell} \#mvps(\text{IB-BGCRO-DR})^{(s)}}. \quad (5.1)$$

443 **5.1. Benefits of recycling between the families.** To illustrate the benefits of recycling spectral  
 444 information from one family to the next as well as the computational saving due to the partial convergence  
 445 detection mechanism, we first report on experiments with BGCRO-DR, IB-BGCRO-DR and IB-BGMRES-  
 446 DR on a series of linear systems with constant left-hand side. Following the spirit of the test examples  
 447 considered in [12], we consider a bidiagonal matrix of size 5000 with upper diagonal unity so that its  
 448 spectrum is defined by the diagonal entries:  $0.1, 1, 2, 3, \dots, 4999$ , which is denoted as Matrix 1. We  
 449 consider experiments with a family size  $p = 20$ , the size of the recycled space  $k = 30$  and the maximal  
 dimension of the search space  $m_d = 300$ .

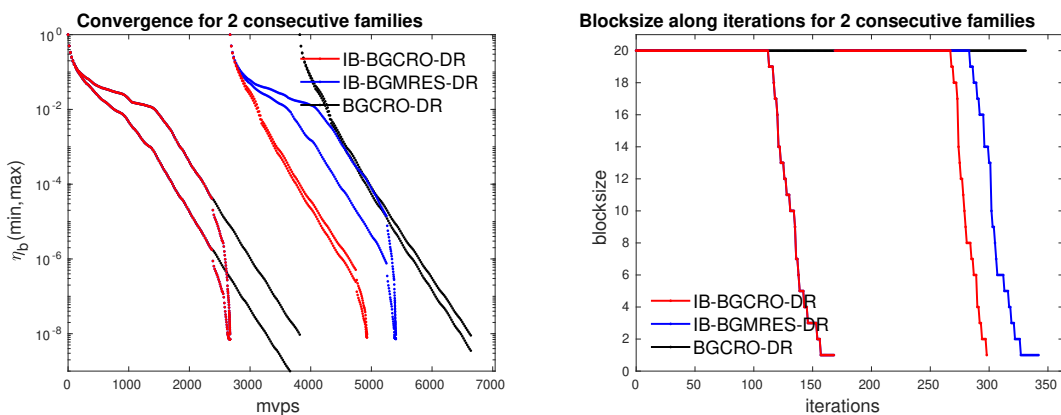


FIG. 5.1. Comparison history for Section 5.1. IB-BGCRO-DR with BGCRO-DR and IB-BGMRES-DR by solving Matrix 1 ( $p = 20$ ,  $m_d = 300$  and  $k = 30$ ). Left: convergence histories of the largest/smallest backward errors  $\eta_b(i)$  at each  $mvps$  for 2 consecutive families. Right: varying blocksize (i.e.  $p_j$ ) along the iterations.

450 In the left plot of Figure 5.1 we display the convergence histories for solving two consecutive families  
 451 with the  $\eta_b$ -based stopping criterion. Several observations can be made. Because IB-BGMRES-DR, IB-  
 452 BGCRO-DR and BGCRO-DR do not have a deflation space to start with for the first family, the convergence  
 453 histories of these three solvers overlap as long as no partial convergence is detected. After this first partial  
 454

455 convergence, the convergence rate of IB-BGCRO-DR and IB-BGMRES-DR becomes faster (in terms of  
 456 *mvps*) than that of BGCRO-DR, and the former two convergence histories mostly overlap as the two  
 457 IB solvers remain mathematically equivalent. For the second and subsequent families, the capability  
 458 to start with a deflation space shows its benefit for BGCRO-DR and IB-BGCRO-DR. It is because IB-  
 459 BGMRES-DR needs a few restarts to capture this spectral information again and refines it in its subsequent  
 460 search spaces construction process; eventually it exhibits a convergence rate similar to the BGCRO-DR  
 461 counterpart. For the sake of comparison and to illustrate the benefit of the partial convergence detection we  
 462 also display the convergence histories of BGCRO-DR which always requires more *mvps* compared to its  
 463 IB counterpart. Those extra *mvps* mostly concur to improve the solution quality for some right-hand sides  
 464 beyond the targeted accuracy.

465 To visualize the effect of the partial convergence detection, we also report in the right plot of Figure 5.1  
 466 the size of search space expansion  $p_j$  as a function of the iterations. Because BGCRO-DR does not  
 467 implement the partial convergence detection, its search space is increased by  $p = 20$  at each iteration.  
 468 For the other two block IB-solvers, the block size monotonically decreases down to 1. Note that the partial  
 469 convergence detection is implemented in the initial (least squares) residual block in IB-BGCRO-DR, thus  
 470 its block size does not jump back to the original block size  $p$  at restart. By construction, IB-BGMRES-DR  
 471 implements the partial convergence detection at restart so that the same observation applies.

Number of families	Method	<i>mvps</i>	<i>its</i>
2	BGCRO-DR	6640	332
	IB-BGMRES-DR	5404	343
	IB-BGCRO-DR	<b>4928</b>	<b>299</b>
20	BGCRO-DR	56940	2847
	IB-BGMRES-DR	53772	3454
	IB-BGCRO-DR	<b>45652</b>	<b>2637</b>

TABLE 5.1

Numerical results in both terms of *mvps* and *its* for Section 5.1 with Matrix 1 ( $p = 20$ ,  $m_d = 300$  and  $k = 30$ ).

472 A summary of the *mvps* and the number of block iterations (referred to as *its*) is given in Table 5.1  
 473 that shows the benefit of using IB-BGCRO-DR.

474 In the rest of this paper, the Matrix 1 is chosen as the constant left-hand side in the following  
 475 Section 5.2- 5.4, in which the related parameters are likewise set to be  $p = 20$ ,  $k = 30$  and  $m_d = 300$   
 476 defaultly.

477 **5.2. Subspace expansion governed by the convergence criterion  $\eta_{A,b}$ .** In this section we show  
 478 the capability of the novel subspace expansion policy to drive the individual backward errors  $\eta_{A,b}$  down  
 479 to different accuracies and its benefit with respect to the original BGCRO-DR method. In Figure 5.2, we  
 480 display the convergence histories of the IB and IB-free method for three different convergence thresholds,  
 481 from the less stringent on the left to the most stringent on the right. We can firstly observe that the first  
 482 iteration, where the partial convergence detection starts to act, depends on the targeted accuracy as it can  
 have been expected from the associated threshold on the singular values of the least squares residual. The

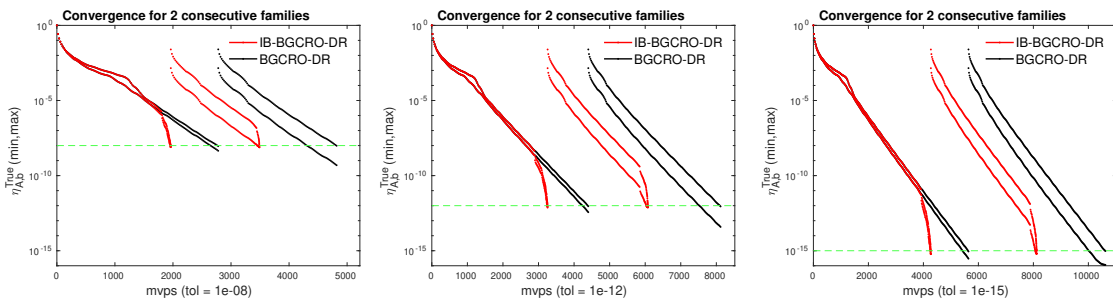


FIG. 5.2. Convergence histories of the largest/smallest  $\eta_{A,b(i)}(x_j^{(i)})$  at each *mvps* for 2 consecutive families for Section 5.2 with different convergence thresholds. Comparison of IB-BGCRO-DR with BGCRO-DR by solving Matrix 1 ( $p = 20$ ,  $m_d = 300$  and  $k = 30$ ).

483  
484  
485  
486  
487  
488  
489  
490  
491  
492

second interesting observation is that IB-BGCRO-DR is able to decrease  $\eta_{A,b}$  down to a very low value close to the machine epsilon, that is  $\mathcal{O}(10^{-16})$ . This latter result mostly reveals the orthogonality quality of the residual space basis computed by (BMGS  $\circ$  HouseQR) in the block Arnoldi implementation and the re-orthogonalization using MGS between all the columns of the recycling subspace  $C_k$  and the initial block Arnoldi basis at restart. This ensures that the least squares residual norms to be quite close to the linear system residual ones. This latter fact ensures the relevance of the space expansion policy, that monitors the linear system residual norms through the least squares residual ones. To illustrate the orthonormal quality of the basis  $\widehat{\mathcal{V}}_{j+1} = [C_k, \mathcal{V}_j, [P_{j-1}, \widetilde{W}_j]]$ , we display in Figure 5.3 the loss of orthogonality along *mvps* that is defined by

$$\text{Loss-Orth} = \left\| \widehat{\mathcal{V}}_{j+1}^H \widehat{\mathcal{V}}_{j+1} - I_{j+1} \right\|. \quad (5.2)$$

493  
494  
495  
496  
497

In a quite similar manner to MGS-GMRES, that is backward-stable [14], it can be observed that the loss of orthogonality mostly appears when the solutions of the linear systems converge. Note that without the re-orthogonalization at restart, the loss of orthogonality tends to be accumulated along restart which prevents the value of Loss-Orth to be close to the machine epsilon. Refer to [9, Figure 5.7] for the corresponding results without applying re-orthogonalization to all the columns of  $[C_k, [V_1, P_0]]$  at restart.

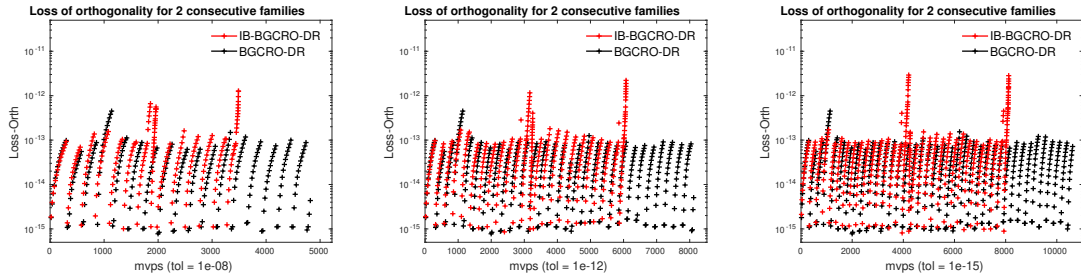


FIG. 5.3. Loss-Orth defined in Equation (5.2) of GCRO-variants with stopping criterion based on  $\eta_{A,b^{(i)}}(x_j^{(i)})$  at each *mvps* for 2 consecutive families for Section 5.2 with different convergence thresholds. Comparison of IB-BGCRO-DR with BGCRO-DR for solving Matrix 1 ( $p = 20$ ,  $m_d = 300$  and  $k = 30$ ).

498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514

**5.3. Subspace expansion policy for individual convergence thresholds for  $\eta_b$ .** To illustrate this feature, we consider a family of  $p$  right-hand sides and a convergence threshold  $10^{-4}$  for the first  $p/2$  right-hand sides and  $10^{-8}$  for the last  $p/2$  ones. As an estimate of the computational benefit of this feature, we also compare with calculations where all the right-hand sides are solved with the most stringent threshold, that is  $10^{-8}$ . We display in the left part of Figure 5.4, 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 one (associated with the first  $p/2$  solutions) only goes down to  $10^{-4}$ . If we compare the convergence histories 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; after this point IB-BGCRO-DR-VA somehow focuses on the new directions (produced by *mvps* given for the  $x$ -axis) to reduce the residual norms of the remaining  $p/2$  solutions that have not yet converged. The right plot of Figure 5.4 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 to the most stringent one if this feature had not been designed. In this case the individual monitoring of the convergence saves around 45 % of *mvps* on this example. Those results are summarized in Table 5.2.

515  
516

We refer to [9, Figure F.1 and Table F.1 of Appendix F] for an illustration of extending such individual control to the block solver IB-BGMRES-DR that can also accommodate this feature.

517  
518  
519

**5.4. Expansion policy governed by computational performance.** As discussed in Section 3.3, only a subset of the candidate directions exhibited by the partial convergence detection mechanism can be eventually selected to expand the search space at the next block iteration; we denote this maximum size

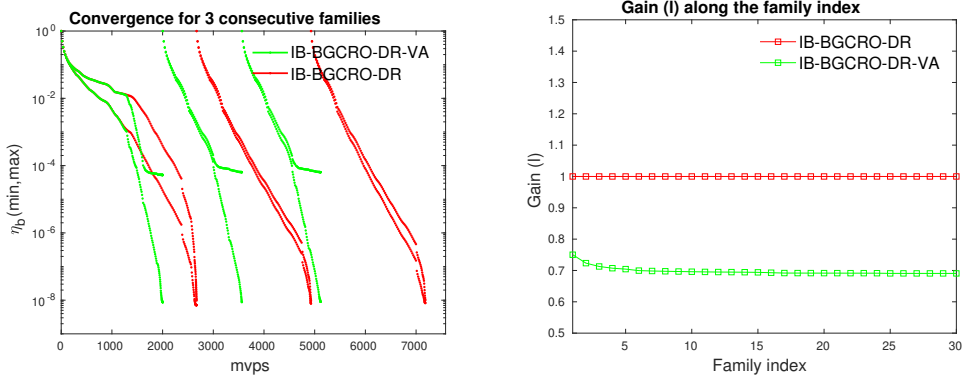


FIG. 5.4. Comparison of IB-BGCRO-DR to IB-BGCRO-DR-VA for Section 5.3 with Matrix 1 ( $p = 20$ ,  $m_d = 300$  and  $k = 30$ ). Left: convergence histories of the largest/smallest backward errors  $\eta_{b(i)}$  at each mvps for 3 consecutive families. Right: Gain ( $l$ ) defined in Equation (5.1) of IB-BGCRO-DR-VA to IB-BGCRO-DR versus family index.

Number of families	Method	<i>mvps</i>	<i>its</i>
3	IB-BGCRO-DR	7182	428
	IB-BGCRO-DR-VA	<b>5119</b>	<b>395</b>
30	IB-BGCRO-DR	68263	3932
	IB-BGCRO-DR-VA	<b>47143</b>	<b>3566</b>

TABLE 5.2

Numerical results of IB-BGCRO-DR with fixed/varying accuracy for each right-hand side in terms of *mvps* and *its* for Section 5.3, where the coefficient matrix is Matrix 1 with  $p = 20$ ,  $m_d = 300$  and  $k = 30$ .

520 as  $p^{CB}$  and refer to this variant as IB-BGCRO-DR-CB where the CB stands for Computational Blocking.  
 521 In Table 5.3 we show the effect of this algorithmic parameter on *mvps* and *its* for the solutions of 3 and 30  
 522 families with Matrix 1 when  $p^{CB}$  varies from 1 to 15 for a number of right-hand sides  $p = 20$ . Generally,  
 523 the smaller  $p^{CB}$  is, the smaller *mvps*, but the larger *its*. Although reported only on one example this trend  
 524 has been observed in all our numerical experiments. Depending on the computational efficiency or cost  
 525 of the *mvps* with respect to the computational weight of the least squares problem and SVD of the scaled  
 526 least squares residual, this gives opportunities to monitor the overall computational effort to the complete  
 527 solution.

Number of families	Method	<i>mvps</i>	<i>its</i>
3	IB-BGCRO-DR	7182	<b>428</b>
	IB-BGCRO-DR-CB ( $p^{CB} = 15$ )	<b>6934</b>	467
	IB-BGCRO-DR-CB ( $p^{CB} = 10$ )	6941	668
	IB-BGCRO-DR-CB ( $p^{CB} = 5$ )	6968	1312
	IB-BGCRO-DR-CB ( $p^{CB} = 1$ )	6966	6444
30	IB-BGCRO-DR	68262	<b>3932</b>
	IB-BGCRO-DR-CB ( $p^{CB} = 15$ )	<b>65364</b>	4303
	IB-BGCRO-DR-CB ( $p^{CB} = 1$ )	65823	60836

TABLE 5.3

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 5.4, where the coefficient matrix is Matrix 1 with  $p = 20$ ,  $m_d = 300$  and  $k = 30$ .

528 Similar to previous subsections, we notice that this subspace expansion policy is also applicable to  
 529 IB-BGMRES-DR and we refer to [9, Figure G.1 and Table G.1 of Appendix G] for an illustration.

530 **5.5. Behavior on sequences of slowly-varying left-hand sides problems.** The example used in  
 531 this section is from a finite element fracture mechanics problem in the field of Fatigue and Fracture of  
 532 Engineering Components (denoted as *FFEC* collection), which is fully documented in [16, Section 4.1].

533 Over 2000 linear systems of size  $3988 \times 3988$  from *FFEC* collection need to be solved in order to capture  
 534 the fracture progression, and among them 151 linear systems 400 – 550 representing a typical subset of the  
 535 fracture progression in which many cohesive elements break are examined in [16]. The solutions of these  
 536 linear systems have been investigated using both GCRO-DR and GCROT (generalized conjugate residual  
 537 with inner orthogonalization and outer truncation), and we refer to [8] for a comprehensive experimental  
 538 analysis. For our numerical experiments we borrow the ten linear systems numbered 400 – 409 from this  
 539 *FFEC* collection. For each set of linear system we select the matrix and the corresponding right-hand sides  
 540 that we expand to form a block of  $p = 20$  by appending random linearly independent vectors.

541 We display the convergence histories for solving the first 3 consecutive families of such linear systems  
 542 in the left plot of Figure 5.5. For the solution of the first linear system, the observations on the IB and DR  
 543 mechanisms discussed in Section 5.1 apply. Even though the coefficient matrix has changed, the recycling  
 544 spectral information computed for the previous family still enables a faster convergence at the beginning  
 545 of the solution of the next one. Specifically, for the solution of the first family the convergence histories  
 546 of the two methods fully overlap until the first partial convergence occurs, as until this step the two methods  
 547 are identical. From the initial slope of the subsequent families, it can be seen that the sequence of matrices  
 548 are close enough to ensure that the recycled space from one system to the next still makes benefit to the  
 549 convergence. The benefit of the partial convergence detection is also illustrated on that example since IB-  
 550 BGCRO-DR still outperforms BGCRO-DR. The overall benefit in term of *mvp*s saving is illustrated in  
 551 the right plot on a sequence of 10 linear systems, where the saving is more than 65 % with respect to  
 552 BGCRO-DR. Corresponding results are summarized in Table 5.4.

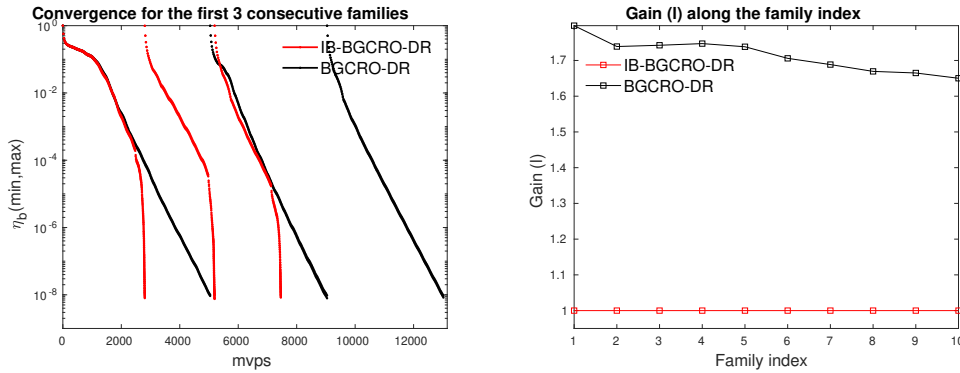


FIG. 5.5. Convergence results of IB-BGCRO-DR and BGCRO-DR on a sequence of slowly-changing left-hand sides described in Section 5.5, where the coefficient matrices are built on *FFEC* with  $p = 20$ ,  $m_d = 300$  and  $k = 15$ .

Number of families	Method	<i>mvp</i> s	<i>its</i>
3	BGCRO-DR	13050	651
	IB-BGCRO-DR	<b>7489</b>	<b>540</b>
10	BGCRO-DR	39935	1990
	IB-BGCRO-DR	<b>24200</b>	<b>1658</b>

TABLE 5.4

Numerical results in terms of *mvp*s and *its* for Section 5.5 with  $p = 20$ ,  $m_d = 300$  and  $k = 15$ .

553 **5.6. A variant suited for flexible preconditioning.** In this section, we illustrate the numerical  
 554 behavior of the flexible variant IB-BFGCRO-DR that we have derived in Section 2.5 and make comparison  
 555 with closely related variants namely BFGCRO-DR (a straightforward block extension of FGCRO-DR [5]).

556 We consider a representative quantum chromodynamics (QCD) matrix from the University of Florida  
 557 sparse matrix collection [6]. It is the conf5.4-0018x8-0500 matrix denoted as  $B_{\text{QCD}}$  of size  $49152 \times$   
 558  $49152$  with the critical parameter  $\kappa_c = 0.17865$  as a model problem. Thirty families of linear systems are  
 559 constructed that are defined as  $A^{(\ell)} = I - \kappa_c(\ell)B_{\text{QCD}}$  with  $0 \leq \kappa_c(\ell) < \kappa_c$  and  $\ell = 1, 2, \dots, 30$ . We  
 560 use the MATLAB function `linspace(0.1780, 0.1786, 30)` to generate the parameters  $\kappa_c(\ell)$  for a sequence of  
 561 matrices and observe that those matrices have the same eigenvectors associated with shifted eigenvalues. A



562 sequence of  $p = 12$  successive canonical basis vectors are chosen to be the block of right-hand sides for a  
 563 given left-hand side matrix following [16, Section 4.3] so that the complete set of the right-hand sides for  
 564 the  $\ell$  linear systems reduces to the first  $p \times \ell$  columns of the identity matrix. This choice could be supported  
 565 by the fact that the problem of numerical simulations of QCD on a four-dimensional space-time lattice for  
 566 solving QCD ab initio (cf. [16, Section 4.3]) has a  $12 \times 12$  block structure, and then a system with 12  
 567 right-hand sides related to a single lattice site is often of interest to solve.

568 The flexible preconditioner is defined by a 32-bit  $ILU(0)$  factorization of the matrix involved in the  
 569 linear system. In a 64-bit calculation framework, the preconditioning consists in casting the set of directions  
 570 to be preconditioned in 32-bit format, performing the forward/backward substitution in 32-bit calculation  
 571 and casting back the solutions in 64-bit arithmetic. The rounding applied to the vectors, cast from 64 to  
 572 32-bit format, has a nonlinear effect that makes the preconditioner nonlinear.

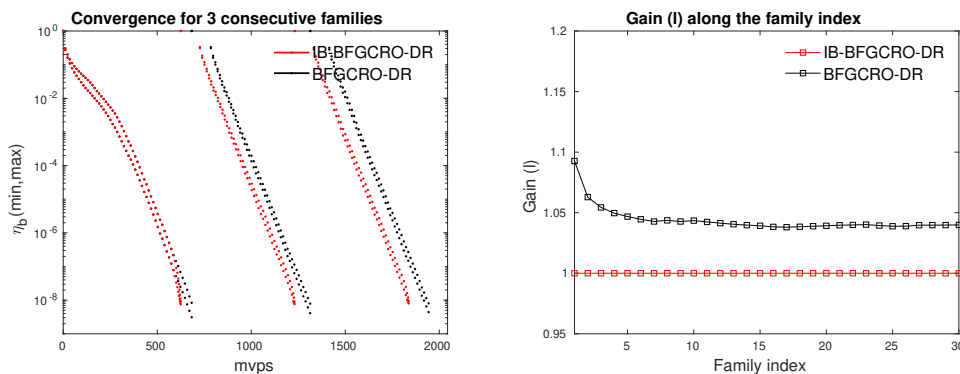


FIG. 5.6. Behavior of the BGCRO-DR-solvers with flexible preconditioner on families of QCD matrices described in Section 5.6 with  $p = 12$ ,  $m_d = 180$  and  $k = 90$ . Left: convergence histories of the largest/smallest backward errors  $\eta_{b(i)}$  at each mvps for 3 consecutive families. Right: Gain ( $l$ ) of the block methods with respect to IB-BFGCRO-DR along family index.

Number of families	Method	mvps	its
3	BFGCRO-DR	1944	<b>147</b>
	IB-BFGCRO-DR	<b>1838</b>	148
30	BFGCRO-DR	18774	<b>1347</b>
	IB-BFGCRO-DR	<b>18054</b>	1350

TABLE 5.5

Numerical results in terms of mvps and its for Section 5.6 with  $p = 12$ ,  $m_d = 15 \times p = 180$  and  $k = 90$ .

573 For those experiments, we attempt to favor the recycling of the space, because the matrices share the  
 574 same invariant space, so that we choose a relative large value for  $k$  that is  $k = m_d/2$ . We report in the  
 575 left plot of Figure 5.6, the convergence histories of the two flexible block variants. Similarly to what has  
 576 already been observed previously the convergences are very similar on the first family and only differ when  
 577 the partial convergence detection becomes active mostly in the last restart. For the second and third families,  
 578 one can see that IB-BFGCRO-DR and BFGCRO-DR have identical convergence speed. One can observe  
 579 a shift in the convergence histories between the end of the solution of one family and the beginning of  
 580 the next one for both IB-BFGCRO-DR and BFGCRO-DR. This shift is due to the extra  $k$  mvps that have to be  
 581 performed when the matrix changes in order to adapt the recycling space as follows

- 582 1. compute  $A^{(\ell+1)}U_k^{(\ell)} = \tilde{C}_k$
- 583 2. compute the reduced  $QR$ -factorization of  $\tilde{C}_k = C_k^{(\ell+1)}R$
- 584 3. update the basis of the deflation space  $U_k^{(\ell+1)} = U_k^{(\ell)}R^{-1}$  so that  $A^{(\ell+1)}U_k^{(\ell+1)} = C_k^{(\ell+1)}$ .

585 Because  $k$  is large, we can clearly see this shift in the left plot of Figure 5.6. For this parameter selection  
 586 in this section, it can be noticed that the dominating effect on the convergence improvement is due to the  
 587 space recycling and not the partial convergence detection. This observation is highlighted in the right plot  
 588 of Figure 5.6, where the benefit of using IB-BFGCRO-DR rather than BFGCRO-DR does diminish when

589 compared to previous experiments and is only about 4%. Numerical details are summarized in Table 5.5.

590 **6. Concluding remarks.** In this paper, we develop a new variant of the block GCRO-DR method  
 591 denoted as IB-BGCRO-DR that inherits the appealing genes of its two parents [16, 20]. First, it inherits the  
 592 capabilities to speed up the convergence rate when solving sequences of linear systems by recycling spectral  
 593 information from one family to the next. Second, the extended search space expansion policy enabled by the  
 594 so-called partial convergence detection allows us to focus on the convergence by considering only the most  
 595 important directions. Along this line, we introduce stopping-criterion driven search space expansion policies  
 596 that enable us to ensure that a prescribed threshold used for the partial convergence detection will eventually  
 597 lead to reach a prescribed threshold for a backward error based stopping criterion. While introduced in the  
 598 block GCRO context, those policies apply to any block minimum residual norm approach that relies on  
 599 an Arnoldi-like relation and includes both block GMRES and GCRO variants. In exact arithmetic, these  
 600 policies exploit the close link between the least squares residuals and the linear system residuals, which is  
 601 guaranteed by the orthonormal basis of the residual space. Through numerical experiments, we show that  
 602 the MGS re-orthogonalization between the columns of recycling space and initial block Arnoldi basis at  
 603 restart combined with (BMGS  $\circ$  HouseQR) in the block Arnoldi algorithm seems to generate good enough  
 604 orthonormal basis to ensure that such a property does also hold in finite precision calculation. Following  
 605 ideas from [14], it would be a future research work to theoretically establish that this class of subspace  
 606 augmentation algorithms is backward stable. To comply with mixed-precision calculation, the flexible  
 607 preconditioning variant is also proposed, which would be of interest for emerging computing platforms  
 608 where mixed-precision calculation could be a way to reduce data movement, which is foreseen as one of  
 609 the major bottleneck to reach high performance.

610 **Acknowledgments.** We would like to thank Matthieu Simonin (who developed a C++ implementation  
 611 of the solver - available on <https://gitlab.inria.fr/solverstack/fabulous/> - in the framework of the PRACE 6IP  
 612 project) for his reading and comments on earlier version of this document. The second author is supported  
 613 by NSFC (12071062, 61772003), Key Projects of Applied Basic Research in Sichuan Province (Grant No.  
 614 2020YJ0216), and Science Strength Promotion Programme of UESTC.

615 Finally, we are extremely grateful to the Associate Editor and the anonymous referees for the  
 616 stimulating and constructive exchanges during the reviewing process. Their comments and questions helped  
 617 us not only to improve the readability of the article, but also to enrich its scientific content.

## 618 REFERENCES

- 619 [1] E. Agullo, L. Giraud, and Y.-F. Jing. Block GMRES method with inexact breakdowns and deflated restarting. *SIAM J. Matrix*  
 620 *Anal. Appl.*, 35(4):1625–1651, 2014.
- 621 [2] H. Calandra, S. Gratton, R. Lago, X. Vasseur, and L. M. Carvalho. A modified block flexible GMRES method with deflation  
 622 at each iteration for the solution of non-Hermitian linear systems with multiple right-hand sides. *SIAM J. Sci. Comput.*,  
 623 35:S345–S367, 2013.
- 624 [3] E. Carson, K. Lund, M. Rozloznic, and S. Thomas. Block Gram-Schmidt algorithms and their stability properties. *Linear*  
 625 *Algebra Appl.*, 638:150–195, 2022.
- 626 [4] L. M. Carvalho, S. Gratton, R. Lago, and X. Vasseur. A flexible generalized Conjugate Residual method with inner  
 627 orthogonalization and deflated restarting. Technical Report TR/PA/10/10, CERFACS, Toulouse, France, 2010.
- 628 [5] L. M. Carvalho, S. Gratton, R. Lago, and X. Vasseur. A flexible generalized conjugate residual method with inner  
 629 orthogonalization and deflated restarting. *SIAM J. Matrix Anal. Appl.*, 32:1212–1235, 2011.
- 630 [6] T. A. Davis and Y. Hu. The University of Florida Sparse Matrix Collection. *ACM Trans. Math. Softw.*, 38:1:1–1:25, 2011.
- 631 [7] E. de Sturler. Nested Krylov methods based on GCR. *J. Comput. Appl. Math.*, 67:15–41, 1996.
- 632 [8] E. de Sturler. Truncation strategies for optimal Krylov subspace methods. *SIAM J. Numer. Anal.*, 36:864–889, 1999.
- 633 [9] L. Giraud, Y.-F. Jing, and Y.-F. Xiang. A block minimum residual norm subspace solver with partial convergence management  
 634 for sequences of linear systems. Research Report 9393, Inria, 2021.
- 635 [10] M. H. Gutknecht. Block Krylov space methods for linear systems with multiple right-hand sides: An introduction. In I. S.  
 636 Duff, A. H. Siddiqi, and O. Christensen, editors, *Modern Mathematical Models, Methods and Algorithms for Real World*  
 637 *Systems*, pages 420–447. Anamaya Publishers, New Delhi, India, 2006.
- 638 [11] R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM J. Matrix Anal. Appl.*, 16:1154–1171, 1995.
- 639 [12] R. B. Morgan. GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 24(1):20–37, 2002.
- 640 [13] R. B. Morgan. Restarted block GMRES with deflation of eigenvalues. *Appl. Numer. Math.*, 54(2):222–236, 2005.
- 641 [14] C. C. Paige, M. Rozloznic, and Z. Strakos. Modified Gram-Schmidt (MGS), least squares, and backward stability of MGS-  
 642 GMRES. *SIAM J. Matrix Anal. Appl.*, 28(1):264–284, 2006.
- 643 [15] M. L. Parks. The Iterative Solution of a Sequence of Linear Systems Arising from Nonlinear Finite Element Analysis. Ph.D.  
 644 Dissertation UIUCDCS-R-2005-2497, University of Illinois at Urbana-Champaign, 2005.
- 645 [16] M. L. Parks, E. de Sturler, G. Mackey, D.D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems.  
 646 *SIAM J. Sci. Comput.*, 28(5):1651–1674, 2006.
- 647 [17] M. L. Parks and K. M. Soodhalter. Block GCRO-DR. in Belos package of the Trilinos C++ Library, 2011.

- 648 [18] M. L. Parks, K. M. Soodhalter, and D. B. Szyld. A block recycled GMRES method with investigations into aspects of solver  
649 performance, 2016. <http://arxiv.org/abs/1604.01713>.
- 650 [19] L. G. Ramos, R. Kehl, and R. Nabben. Projections, deflation, and multigrid for nonsymmetric matrices. *SIAM J. Matrix Anal.*  
651 *Appl.*, 41(1):83–105, 2020.
- 652 [20] M. Robbé and M. Sadkane. Exact and inexact breakdowns in the block GMRES method. *Linear Algebra Appl.*, 419:265–285,  
653 2006.
- 654 [21] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM*  
655 *J. Sci. Stat. Comput.*, 7:856–869, 1986.
- 656 [22] K. Soodhalter. Krylov Subspace Methods with Fixed Memory Requirements: Nearly Hermitian Linear Systems and Subspace  
657 Recycling. Ph.D. dissertation, Temple University, 2012.
- 658 [23] D.-L. Sun, B. Carpentieri, T.-Z. Huang, and Y.-F. Jing. A spectrally preconditioned and initially deflated variant of the restarted  
659 block GMRES method for solving multiple right-hand sides linear systems. *Internat. J. Mech. Sci.*, 144:775–787, 2018.
- 660 [24] D.-L. Sun, T.-Z. Huang, B. Carpentieri, and Y.-F. Jing. Flexible and deflated variants of the block shifted GMRES method. *J.*  
661 *Comput. Appl. Math.*, 345:168–183, 2019.
- 662 [25] D.-L. Sun, T.-Z. Huang, B. Carpentieri, and Y.-F. Jing. A new shifted block GMRES method with inexact breakdowns for  
663 solving multi-shifted and multiple right-hand sides linear systems. *J. Sci. Comput.*, 78:746–769, 2019.
- 664 [26] D.-L. Sun, T.-Z. Huang, Y.-F. Jing, and B. Carpentieri. A block GMRES method with deflated restarting for solving  
665 linear systems with multiple shifts and multiple right-hand sides. *Numer. Linear Algebra Appl.*, 25, 2018. e2148.  
666 <https://doi.org/10.1002/nla.2148>.
- 667 [27] A. Tajaddini, G. Wu, F. S. Movahed, and N. Azizizadeh. Two new variants of the simpler block GMRES method with vector  
668 deflation and eigenvalue deflation for multiple linear systems. *J. Sci. Comput.*, 86, 2021.
- 669 [28] Y.-F. Xiang, Y.-F. Jing, and T.-Z. Huang. A new projected variant of the deflated block conjugate gradient method. *J. Sci.*  
670 *Comput.*, 80:1116–1138, 2019.
- 671 [29] F. Xue and H. C. Elman. Fast inexact subspace iteration for generalized eigenvalue problems with spectral transformation.  
672 *Linear Algebra Appl.*, 435:601–622, 2011.

673 **Appendix A. Proof of Proposition 3.**

674 *Proof.* From Equation (4.1), (4.2) and (4.4), the initial residual block  $R_1$  with partial convergence  
675 detection at restart could be described as

$$\begin{aligned} R_1 &= [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1][C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H R_1 = [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1][C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [\mathbb{V}_1^{new}, P_0^{new}] \hat{\Lambda}_1 \\ &= [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1] \left( [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H C_k R_{12} + [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1]^H [\mathbb{V}_1, P_0] R_{22} \right) \hat{\Lambda}_1 \\ &= [C_k, \mathbb{V}_1, P_0, \widetilde{W}_1] \Lambda_1 \text{ with } \Lambda_1 = \begin{bmatrix} R_{12} \\ 0_{(p_1+p) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} 0_{k \times p_1} & 0_{k \times q_1} \\ I_{p_1} & 0_{p_1 \times q_1} \\ 0_{q_1 \times p_1} & I_{q_1} \\ 0_{p_1 \times p_1} & 0_{p_1 \times q_1} \end{bmatrix} R_{22} \hat{\Lambda}_1, \end{aligned}$$

676 by  $[\mathbb{V}_1^{new}, P_0^{new}] = C_k R_{12} + [\mathbb{V}_1, P_0] R_{22}$  obtained from Equation (4.2). That can also be written as

$$\Lambda_1 = \begin{bmatrix} R_{12} \\ 0_{(p_1+p) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} 0_{k \times p_1} & 0_{k \times q_1} \\ I_{p_1} & \Phi_1 \\ 0_{q_1 \times p_1} & \\ 0_{p_1 \times p_1} & 0_{p_1 \times q_1} \end{bmatrix} R_{22} \hat{\Lambda}_1,$$

677 where  $\Phi_1 = \begin{bmatrix} 0_{p_1 \times q_1} \\ I_{q_1} \end{bmatrix} \in \mathbb{C}^{p \times q_1}$  and  $q_1 + p_1 = p$ .

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

$$\begin{aligned} \Lambda_{j+1} &= [C_k, \mathcal{V}_{j+1}, [P_j, \widetilde{W}_{j+1}]]^H R_1 = [C_k, \mathcal{V}_j, V_{j+1}, [P_j, \widetilde{W}_{j+1}]]^H R_1 \\ &= [C_k, \mathcal{V}_j, [P_{j-1}, \widetilde{W}_j] \mathbb{W}_1, [P_{j-1}, \widetilde{W}_j] \mathbb{W}_2, \widetilde{W}_{j+1}]^H R_1 = [C_k, \mathcal{V}_j, [P_{j-1}, \widetilde{W}_j] [\mathbb{W}_1, \mathbb{W}_2], \widetilde{W}_{j+1}]^H [\mathbb{V}_1^{new}, P_0^{new}] \hat{\Lambda}_1 \\ &= \left( [C_k, \mathcal{V}_j, [P_{j-1}, \widetilde{W}_j] [\mathbb{W}_1, \mathbb{W}_2], \widetilde{W}_{j+1}]^H C_k R_{12} + [C_k, \mathcal{V}_j, [P_{j-1}, \widetilde{W}_j] [\mathbb{W}_1, \mathbb{W}_2], \widetilde{W}_{j+1}]^H [\mathbb{V}_1, P_0] R_{22} \right) \hat{\Lambda}_1 \\ &= \begin{bmatrix} R_{12} \\ 0_{(n_j+p+p_{j+1}) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} C_k^H \mathbb{V}_1 & C_k^H P_0 \\ \mathcal{V}_j^H \mathbb{V}_1 & \mathcal{V}_j^H P_0 \\ [V_{j+1}, P_j]^H \mathbb{V}_1 & [\mathbb{W}_1, \mathbb{W}_2]^H [P_{j-1}, \widetilde{W}_j]^H P_0 \\ \widetilde{W}_{j+1}^H \mathbb{V}_1 & \widetilde{W}_{j+1}^H P_0 \end{bmatrix} R_{22} \hat{\Lambda}_1 \\ &= \begin{bmatrix} R_{12} \\ 0_{(n_j+p+p_{j+1}) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} 0_{k \times p_1} & 0_{k \times q_1} \\ I_{p_1} & \Phi_j(1:n_j, :) \\ 0_{(n_j-p_1) \times p_1} & [\mathbb{W}_1, \mathbb{W}_2]^H \begin{bmatrix} P_{j-1}^H \\ \widetilde{W}_j^H \end{bmatrix} P_0 \\ 0_{p \times p_1} & \\ 0_{p_{j+1} \times p_1} & 0_{p_{j+1} \times q_1} \end{bmatrix} R_{22} \hat{\Lambda}_1 \\ &= \begin{bmatrix} R_{12} \\ 0_{(n_j+p+p_{j+1}) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} 0_{k \times p_1} & 0_{k \times q_1} \\ I_{p_1} & \Phi_j(1:n_j, :) \\ 0_{(n_j-p_1) \times p_1} & [\mathbb{W}_1, \mathbb{W}_2]^H \begin{bmatrix} \Phi_j(n_j+1:n_j+q_j, :) \\ 0_{p_j \times q_1} \end{bmatrix} \\ 0_{p \times p_1} & \\ 0_{p_{j+1} \times p_1} & 0_{p_{j+1} \times q_1} \end{bmatrix} R_{22} \hat{\Lambda}_1 \\ &= \begin{bmatrix} R_{12} \\ 0_{(n_j+p+p_{j+1}) \times p} \end{bmatrix} \hat{\Lambda}_1 + \begin{bmatrix} 0_{k \times p_1} & 0_{k \times q_1} \\ I_{p_1} & \Phi_{j+1} \\ 0_{(n_j+p-p_1) \times p_1} & \\ 0_{p_{j+1} \times p_1} & 0_{p_{j+1} \times q_1} \end{bmatrix} R_{22} \hat{\Lambda}_1 \end{aligned}$$

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

680

681 **Appendix B. The SVD decomposition of the least squares residual and the solution of the least**  
682 **squares problem.** The partial convergence detection mechanism allows to extract from the residual spaces  
683 new directions to expand the search space at the next iteration of the block method. The selection consists in  
684 extracting the directions that contribute the most to the scaled residual block and is based on the SVD of the  
685 scaled least squares residual. In this section, we detail how the solution of the least squares problem (2.11)  
686 enables to compute easily and cheaply the SVD of the associated scaled (least squares) residual block. The  
687 least squares problem

$$Y_j = \operatorname{argmin}_{Y \in \mathbb{C}^{(k+n_j) \times p}} \|\Lambda_j - \underline{\mathcal{F}}_j Y\|_F, \text{ with } \underline{\mathcal{F}}_j \in \mathbb{C}^{(k+n_j+p) \times (k+n_j)} \quad (\text{B.1})$$

688 is solved by using a full  $QR$ -factorization of  $\underline{\mathcal{F}}_j = Q_j^{LS} R_j^{LS}$ , where the superscript  $LS$  comes from Least  
689 Squares,  $Q_j^{LS} = [Q_j^{LS(1)}, Q_j^{LS(2)}]$  with  $Q_j^{LS(1)} \in \mathbb{C}^{(k+n_j+p) \times (k+n_j)}$  and  $Q_j^{LS(2)} \in \mathbb{C}^{(k+n_j+p) \times p}$ ,  $R_j^{LS} =$   
690  $\begin{bmatrix} R_j^{LS(1)} \\ 0_{p \times (k+n_j)} \end{bmatrix} \in \mathbb{C}^{(k+n_j+p) \times (k+n_j)}$  with  $R_j^{LS(1)} \in \mathbb{C}^{(k+n_j) \times (k+n_j)}$  is an upper triangular matrix, from  
691 which the reduced  $QR$ -factorization of  $\underline{\mathcal{F}}_j$  is formulated as  $\underline{\mathcal{F}}_j = Q_j^{LS(1)} R_j^{LS(1)}$  if  $Q_j^{LS(1)}$  is considered  
692 as an orthogonal basis of  $\underline{\mathcal{F}}_j$ . Thus, we could still formulate  $Y_j$  in a relatively economic way as

$$Y_j = (R_j^{LS(1)})^{-1} ((Q_j^{LS(1)})^H \Lambda_j) \in \mathbb{C}^{(k+n_j) \times p}, \quad (\text{B.2})$$

693 from which we could deduce the residual of the least squares problem described in Equation (3.6) as follows:

$$\begin{aligned} \Lambda_j - \underline{\mathcal{F}}_j Y_j &= \Lambda_j - Q_j^{LS} R_j^{LS} Y_j = Q_j^{LS} ((Q_j^{LS})^H \Lambda_j - R_j^{LS} Y_j), \\ &= Q_j^{LS} \left( \begin{bmatrix} (Q_j^{LS(1)})^H \\ (Q_j^{LS(2)})^H \end{bmatrix} \Lambda_j - \begin{bmatrix} R_j^{LS(1)} \\ 0_{p \times (k+n_j)} \end{bmatrix} Y_j \right), \\ &= Q_j^{LS} \left( \begin{bmatrix} 0_{(k+n_j) \times (k+n_j+p)} \\ (Q_j^{LS(2)})^H \end{bmatrix} \Lambda_j \right), \\ &= Q_j^{LS} \begin{pmatrix} 0_{(k+n_j) \times p} \\ R_j^{\ell s} \end{pmatrix}, \end{aligned}$$

where  $R_j^{\ell s} = (Q_j^{LS(2)})^H \Lambda_j \in \mathbb{C}^{p \times p}$  are the last  $p$  rows of  $(Q_j^{LS})^H \Lambda_j$ . The SVD of scaled residual  $R_j^{\ell s} D_\varepsilon$  can be written as

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

so that the SVD of the scaled least squares residual is

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