A posteriori superlinear convergence bounds for block conjugate gradient

Christian E. Schaerer, Daniel B. Szyld,
and Pedro J. Torres

Report 21-07-11
July 2021. Revised May 2022 and September 2022.

Department of Mathematics
Temple University
Philadelphia, PA 19122

This report is available in the World Wide Web at
`http://www.math.temple.edu/~szyld`

A POSTERIORI SUPERLINEAR CONVERGENCE BOUNDS FOR BLOCK CONJUGATE GRADIENT *

CHRISTIAN E. SCHAERER†, DANIEL B. SZYLD‡, AND PEDRO J. TORRES†

Abstract. In this paper, we extend to the block case, the *a posteriori* bound showing superlinear convergence of Conjugate Gradients developed in [J. Comput. Applied Math., 48 (1993), pp. 327-341]; that is, we obtain similar bounds, but now for block Conjugate Gradients. We also present a series of computational experiments, illustrating the validity of the bound developed here, as well as the bound from [SIAM Review, 47 (2005), pp. 247-272] using angles between subspaces. Using these bounds, we make some observations on the onset of superlinearity, and how this onset depends on the eigenvalue distribution and the block size.

Key words. superlinear convergence, block conjugate gradient method, *a posteriori* analysis.

AMS subject classifications. 65F10, 65B99, 65F30

1. Introduction. When solving numerically linear systems of the form $Ax = b$, for a given right hand side b, and when A is a large $n \times n$ sparse symmetric positive definite (s.p.d.) matrix, the method of choice is conjugate gradient (CG) [13]. It is well known that CG exhibits superlinear convergence. In this context the term *superlinear* is understood to mean that the A-norm of the error is monotonically decreasing at first in a linear manner, but at some point the slope of this graph may change and become faster. In other words, the method is faster than linear; see, e.g., Figure 4.2 in Section 4. Several authors studied this phenomenon; see, e.g., [2, 24, 28, 29]. In particular, they studied in an *a posteriori* manner when one would expect the change from a linear regime, to a superlinear regime (i.e., a steeper slope), when this occurs.** See also [1, 3, 5, 8, 26, 31] for different analyses including studies on the effect of round-off errors.

When one has $s > 1$ right hand sides, i.e., when one wishes to solve a block system of the form

$$A\mathbf{X} = \mathbf{B}, \tag{1.1}$$

where $\mathbf{X}$ and $\mathbf{B}$ are $n \times s$ skinny-tall matrices (block vectors) with $s \ll n$, then, the block conjugate gradient (block CG) method can be considered for its solution. This block method was introduced by O'Leary in 1980 [17], and can be used either when one has multiple right-hand sides, or when one wishes to accelerate the convergence of CG by using a richer space (usually by adding random vectors to the original right-hand side b to produce $\mathbf{B}$). We have this latter situation very much in mind in our investigations.

*Christian E. Schaerer: his research is partially supported by PRONII and by CONACYT-PY under project 14-INV186. Daniel B. Szyld: his research is supported in part by the U.S. National Science Foundation under grant DMS-1418882 and U.S. Department of Energy under grant DE-SC-0016578. Pedro Torres acknowledges the financial support of Scholarship-FEEI-CONACYT-PROCIENCIA.

†Polytechnic School, National University of Asuncion, San Lorenzo, Central, Paraguay. P.O.Box: 2111 SL. (Email: cschaer@pol.una.py, pjtorres@pol.una.py). Corresponding author: P. J. Torres.

‡Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, PA 19122-6094, USA. (Email: szyld@temple.edu).

**We note that in some cases,the superlinear regime may not start before convergence is reached. In those cases, the convergence is just linear

In this paper, we extend the theory developed by van der Sluis and van der Vorst [28, 30] in order to explain the superlinear convergence of CG in the block case. Their analysis uses spectral information (and Ritz values) to bound the convergence of CG by that of a comparison method which commences with an initial vector projected onto a subspace where certain components have been deflated. This projection is based on polynomial expressions describing the CG method. We call these new *a posteriori* convergence bounds, *spectral-based bounds* for short, and they are discussed in Section 3.2.

There is another set of convergence bounds developed by Simoncini and Szyld [21, 23] where angles between subspaces are considered. We call these bounds, *subspace-based bounds* for short, and they are reviewed in Section 3.1.

We illustrate the effectiveness of these bounds with a series of numerical experiments in Section 4. It can be observed that both bounds capture well the slope of the block error or residual norm (in the appropriate block norm). As one would expect intuitively, the larger the size of the block s, the faster the convergence, and in particular, the onset of superlinearity (the point where the slope changes in the convergence curve) occurs earlier; cf. [16, 21]. The *a posteriori* bounds described in this paper reproduce properly this onset of superlinarity, both in cases of single eigenvalues or with multiple eigenvalues (or clusters). These bounds explain the observed superlinear convergence behavior of block CG: the onset of superlinarity occurs when the block Krylov subspace is close to an invariant subspace of the matrix A (subspace bound), or when the Ritz values (or latent roots of the residual polynomial) are close to the eigenvalues of the matrix A (spectral bound).

Throughout this paper, calligraphic letters $\mathcal{A}$, $\mathcal{H}$ and $\mathcal{Z}$ denote square $s \times s$ matrices and upper case letters A, V, D and H denote square $n \times n$ matrices and $\mathbf{R}$, $\mathbf{X}$, $\mathbf{U}$, $\mathbf{W}$ and $\mathbf{B}$ denote rectangular $n \times s$ matrices, matrix polynomials are denoted with upper case Greek letters (e.g., Φ, Ω, Ξ). Scalars are denoted with lower case Greek letters (e.g., λ, γ, α) and integers are normally denoted as i, j, k, m and n. A subscript on a matrix, vector, or scalar denotes an iteration number. For simplicity, we consider only real matrices, although the generalization to complex matrices is direct.

2. Block CG. In this section we review the CG method and its block counterpart. We present their formulation as minimization problems. We discuss the block Lanczos method, and its formulation using a matrix-polynomial approach.

At each step m, CG searches for the approximate solution x_m in the Krylov subspace $\mathcal{K}_m(A, r_0) = \{\sum_{k=0}^{m-1} c_k A^k r_0 : c_k \in \mathbb{R}\}$ shifted by the initial vector x_0, and such that the A-norm of the error is minimized, i.e., x_m is such that

$$\|x_* - x_m\|_A = \min_{x \in x_0 + \mathcal{K}_m} \|x_* - x\|_A = \min_{d \in A\mathcal{K}_m} \|r_0 - d\|_{A^{-1}} = \|r_m\|_{A^{-1}}, \tag{2.1}$$

where $r_m = b - Ax_m$ is the mth residual vector, x_* is the solution of $Ax = b$, and the A-norm is defined as usual as $\|v\|_A = \langle v, Av\rangle^{1/2}$ with the latter being the Euclidean inner product; see, e.g., [9, 10, 19, 22]. The latter two equalities in (2.1) indicate that minimizing the A-norm of the error is mathematically equivalent to minimizing the A^{-1}-norm of the residual. Of course the latter is never computed explicitly, but as we shall see, it is a useful way of looking at the method.

Similarly, one can construct the block Krylov subspace as

$$\mathbb{K}_m(A, \mathbf{R}_0) := \left\{ \sum_{k=0}^{m-1} A^k \mathbf{R}_0 \mathcal{C}_k : \mathcal{C}_k \in \mathbb{S} \right\}, \tag{2.2}$$

where the initial block residual is $\mathbf{R}_0 := \mathbf{B} - A\mathbf{X}_0$, with $\mathbf{X}_0$ being the initial block vector and $\mathbb{S} \subseteq \mathbb{R}^{s\times s}$ is a subspace containing the identity $\mathcal{I}_s$ and closed under matrix multiplication and transposition; see, e.g., [6, 7, 10, 11, 19].*

At each step m, block CG searches for the approximate solution on the mth block Krylov subspace (2.2), shifted by $\mathbf{X}_0$, such that it minimizes the A-norm of the (block) error. Let $\mathbf{X}_*$ be the solution of (1.1), then the approximation $\mathbf{X}_m$ and the corresponding residual $\mathbf{R}_m := \mathbf{B} - A\mathbf{X}_m$ satisfy

$$\min_{\mathbf{X}\in\mathbf{X}_0+\mathbb{K}_m} \|\mathbf{X}_* - \mathbf{X}\|_A = \min_{\mathbf{D}\in A\mathbb{K}_m} \|\mathbf{R}_0 - \mathbf{D}\|^2_{A^{-1}} = \|\mathbf{R}_m\|^2_{A^{-1}}, \tag{2.3}$$

where as usual the norm of block vectors is the Frobenious norm, so that the Z-norm (for $Z = A$ or $Z = A^{-1}$) here is defined with the inner product which induces the Frobenious norm, that is,

$$\langle \mathbf{V}, \mathbf{W}\rangle_Z := \mathrm{tr}(\mathbf{V}^T Z\mathbf{W}) = \sum_{i=1}^{s} \langle v_i, w_i\rangle_Z ,$$

where $\mathbf{V} = [v_1, \ldots, v_s]$ and $\mathbf{W} = [w_1, \ldots, w_s]$. We denote the solution of the minimization problem on the right in (2.3) by $\mathbf{D}_m$.

2.1. Block Lanczos. The block Lanczos procedure is the block version of the Lanczos method. It produces a block basis of the block Krylov susbpace $\mathbb{K}_m(A, \mathbf{R}_0)$ which we collect in the matrix $W_m = [\mathbf{U}_0, \mathbf{U}_1, \ldots, \mathbf{U}_{m-1}]$. It proceeds with a three-term recurrence as follows. Let $\mathbf{R}_0 = \mathbf{U}_0\mathcal{B}_0$ be the QR-factorization of $\mathbf{R}_0$, where $\mathbf{U}_0$ and $\mathcal{B}_0$ are real $n \times s$ and $s \times s$ matrices, respectively. Then, the sequence of block vectors $\{\mathbf{U}_i\}, i = 0, 1, 2 \ldots, m$, which have orthonormal columns and are mutually orthogonal are constructed by the three-term recurrence

$$\mathbf{U}_{i+1}\mathcal{B}_{i+1} = \mathbf{M}_i = A\mathbf{U}_i - \mathbf{U}_i\mathcal{A}_i - \mathbf{U}_{i-1}\mathcal{B}_i^T, \tag{2.4}$$

where $\mathbf{U}_{-1} = \mathbf{0}$, $\mathcal{A}_i = \mathbf{U}_i^* A\mathbf{U}_i$ and $\mathbf{U}_{i+1}\mathcal{B}_{i+1}$ is the QR-factorization of $\mathbf{M}_i$ with $\mathbf{U}_{i+1}$ orthogonal and $\mathcal{B}_{i+1}$ upper triangular; see, e.g., [9]. Thus, at each iteration step m, the following matrix relation, called the block Lanczos relation, holds

$$A[\mathbf{U}_0, \mathbf{U}_1, \ldots, \mathbf{U}_{m-1}] = [\mathbf{U}_0, \mathbf{U}_1, \ldots, \mathbf{U}_{m-1}]T_m + \mathbf{U}_m\mathcal{B}_m E_m^T, \tag{2.5}$$

where $E_m = [0_s, 0_s, \ldots, \mathcal{I}_s]^T \in \mathbb{R}^{sm\times s}$ and T_m is the block tridiagonal matrix of dimension $ms \times ms$

$$T_m = \begin{pmatrix} \mathcal{A}_0 & \mathcal{B}_1^T & & & \\ \mathcal{B}_1 & \mathcal{A}_1 & \ddots & & \\ & \mathcal{B}_2 & \ddots & & \\ & & \ddots & \mathcal{A}_{m-2} & \mathcal{B}_{m-1}^T \\ & & & \mathcal{B}_{m-1} & \mathcal{A}_{m-1} \end{pmatrix}. \tag{2.6}$$

The expression (2.5) can be rewritten as

$$A\,W_m = W_{m+1}\,\tilde{T}_m. \tag{2.7}$$

*We point out that other versions of block Krylov spaces have been used in the literature, e.g., in [4], but in this paper we will use the above definition, known as *classical*. See also [12].

where $\tilde{T}_m$ is a real block tridiagonal matrix of dimension $(m+1)s \times ms$, and $T_m = [I_{ms}, 0]\tilde{T}_m$.

The approximate solution of the linear systems produced by block CG at the mth iteration is given by

$$\mathbf{X}_m = \mathbf{X}_0 + W_m Y_m, \tag{2.8}$$

where $Y_m = [y_1^T, \ldots, y_m^T]^T$ with $y_i \in \mathbb{R}^{s\times s}$ is obtained by solving the equation

$$T_m Y_m = \mathcal{B}_0 E_1 \tag{2.9}$$

with $E_1 = [\mathcal{I}_s, 0_s, \ldots, 0_s]^T \in \mathbb{R}^{sm\times s}$. Since $\mathbf{R}_0 = \mathbf{U}_0\mathcal{B}_0$, then equation (2.9) is equivalent to $W_m T_m Y_m = \mathbf{R}_0$. In addition, at each block iteration it is possible to compute the residual of the solution without explicitly computing the solution, using the expression
$\mathbf{R}_m = -\mathbf{U}_m\mathcal{B}_m E_m Y_m$ which is obtained as follows (see, e.g., [6], [19])

$$\begin{aligned}
\mathbf{R}_m &= \mathbf{B} - A\mathbf{X}_m && \text{by definition of } \mathbf{R}_m \\
&= \mathbf{B} - A(\mathbf{X}_0 + W_m Y_m) && \text{by expression (2.8)} \\
&= \mathbf{R}_0 - AW_m Y_m && \text{by definition of } \mathbf{R}_0 \\
&= \mathbf{R}_0 - W_m T_m Y_m - \mathbf{U}_m\mathcal{B}_m E_m Y_m && \text{by identity(2.5)} \\
&= -\mathbf{U}_m\mathcal{B}_m E_m Y_m && \text{by identity (2.9),}
\end{aligned} \tag{2.10}$$

and notice that $\mathbf{R}_m = \mathbf{R}_0 - \mathbf{D}_m$ and $\mathbf{D}_m = AW_m Y_m$ in (2.3).

An implementation of the block Conjugate Gradient method is described in Algorithm 1 below. We note that in this algorithm, assuming exact arithmentic, one has the orthogonality properties $\mathbf{R}_k^T\mathbf{R}_j = 0$ and $\mathbf{P}_k^T A\mathbf{P}_j = 0$ with $j \neq k$, and as long as the matrices $\mathbf{P}_k$ and $\mathbf{R}_k$ retain full rank, the algorithm is well defined.

Algorithm 1 Block Conjugate Gradient - block CG[17] .

1: Given an initial approximation $\mathbf{X_0}$ to the solution matrix $\mathbf{X}^*$;
2: Set $\mathbf{R}_0 = \mathbf{B} - A\mathbf{X}_0$, $\mathbf{P}_0 = \mathbf{R}_0$;
3: **for** $k = 0, 1, 2 \ldots$ **do**
4: $\quad\mathcal{D}_k = (\mathbf{P}_k^T A\mathbf{P}_k)^{-1}\mathbf{R}_k^T\mathbf{R}_k$
5: $\quad\mathbf{X}_{k+1} = \mathbf{X}_k + \mathbf{P}_k\mathcal{D}_k$
6: $\quad\mathbf{R}_{k+1} = \mathbf{R}_k - A\mathbf{P}_k\mathcal{D}_k$
7: $\quad$**if** convergence conditions are satisfied **then**
8: $\quad\quad$break;
9: $\quad$**end if**
10: $\quad\mathcal{P}_k = (\mathbf{R}_k^T\mathbf{R}_k)^{-1}\mathbf{R}_{k+1}^T\mathbf{R}_{k+1}$
11: $\quad\mathbf{P}_{k+1} = (\mathbf{R}_{k+1} + \mathbf{P}_k\mathcal{P}_k)$
12: **end for**

2.2. Block residual polynomial. The spectral bounds are based on the analysis of the block residual polynomials. In this section we bring forth some preliminary definitions, identities and properties of matrix polynomials, and in particular the block residual polynomial.

Let $\mathbb{P}_{m,s}$ be the space of matrix-valued polynomials with elements of the form

$$\Upsilon_m(\eta) = \sum_{i=0}^{m} \eta^i \mathcal{C}_i \tag{2.11}$$

where $\eta \in \mathbb{R}$ and $\mathcal{C}_i$ are real $s \times s$ matrices. We recall the operation introduced in [15],

$$\Upsilon_m(A) \circ \mathbf{X} = \sum_{i=0}^{m} A^i \mathbf{X} \mathcal{C}_i, \tag{2.12}$$

where A is any $n \times n$ matrix, $\mathbf{X}$ is a block $n \times s$ vector and '$\circ$' is called the Gragg operator.

Denote by $\mathbb{G}_{m,s} \subset \mathbb{P}_{m,s}$ the subspace of matrix-valued polynomials with elements of the form $\Upsilon_m(\eta) = \mathcal{I}_s - \sum_{i=0}^{m-1} \eta^{i+1} \mathcal{C}_i$, i.e., the polynomias such that $\Upsilon_m(0) = \mathcal{I}_s$. Hence, using the nomenclature (2.12) and the subspace $\mathbb{G}_{m,s}$, the residual $\mathbf{R}_m$ of block CG at the mth iteration can be expressed in terms of matrix polynomials as

$$\mathbf{R}_m = \Upsilon_m(A) \circ \mathbf{R}_0 = \mathbf{R}_0 - \sum_{i=0}^{m-1} A^{i+1} \mathbf{R}_0 \mathcal{G}_i, \tag{2.13}$$

where $\Upsilon_m \in \mathbb{G}_{m,s}$, and $\mathcal{G}_i$ are $s \times s$ matrices, $i = 1, \ldots, m-1$.

Consequently, the variational formulation of block CG (2.3) can be expressed as follows using matrix-value polynomials with A and $\mathbf{R}_0$ as arguments

$$\|\mathbf{R}_m\|_{A^{-1}} = \min_{\Upsilon_m \in \mathbb{G}_{m,s}} \|\Upsilon_m(A) \circ \mathbf{R}_0\|_{A^{-1}} = \|\Phi_m(A) \circ \mathbf{R}_0\|_{A^{-1}}, \tag{2.14}$$

where $\Phi_m(\eta) \in \mathbb{G}_{m,s}$ is the solution of the minimization problem.

The three-term recurrence (2.5) of Block Lanczos can also be written in matrix polynomial form. Each matrix $\mathbf{U}_i$ in the recurrence (2.5) is a linear combinations of matrices $A^i \mathbf{R}_0 \in \mathbb{K}_m(A, \mathbf{R}_0)$ for $i = 0, 1, \ldots, m-1$. Therefore, we can set $\mathbf{U}_i = \Gamma_i(A) \circ \mathbf{R}_0$, and thus, (2.5) can be rewritten in matrix polynomial form as

$$\eta P_{m-1}(\eta) = P_{m-1}(\eta) T_m + \Gamma_m(\eta) \mathcal{B}_m E_m, \tag{2.15}$$

where $P_{m-1}(\eta) := [\Gamma_0(\eta), \Gamma_1(\eta), \ldots, \Gamma_{m-1}(\eta)]$ with $\Gamma_i \in \mathbb{P}_{i,s}$; see, e.g., [15, 20].

It can be observed from (2.15) that $\det(\lambda I - T_m) = 0$ if and only if $\det(\Gamma_m(\lambda)) = 0$. Therefore, the eigenvalues of T_m are the latent roots of $\Gamma_m(\lambda)$, hence coinciding with the Ritz values of A associated with $\mathbb{K}_m(A, \mathbf{R}_0)$. In addition, the matrix polynomials $\Gamma_m(\lambda)$ and $\Phi_m(\lambda)$ represent the block CG process dynamics but from different perspective [6, 20]. In the following proposition we show that the latent roots of these two matrix polynomials are the same.

PROPOSITION 1. *The latent roots of the block CG polynomial $\Phi_m(\eta)$ coincide with the latent roots of $\Gamma_m(\eta)$. Hence they are also the eigenvalues of T_m and the Ritz values of the matrix A associated with $\mathbb{K}_m(A, \mathbf{R}_0)$.*

Proof. From (2.10) and using matrix value representation, the block residual can be expressed as

$$\mathbf{R}_m = -(\Gamma_m(A) \circ \mathbf{U}_0) \mathcal{B}_m E_m Y_m.$$

Using (2.13) we can arrive to the following equality in matrix polynomial form

$$\Phi_m(\eta) = -\Gamma_m(\eta) \mathcal{B}_m E_m Y_m,$$

then the latents roots of $\Phi_m(\eta)$ and $\Gamma_m(\eta)$ are the same. □

3. *A posteriori* models for block CG. We are ready to present the two *a posteriori* models which explain the superlinear behavior of block CG. In the subspace-based model, the bound is based on the angle (or gap) between the block Krylov subspace and an invariant subspace of A. The A-norm of the error, or equivalently, the A^{-1}-norm of the (block) residual is bounded by the residual norm of another CG process in which the components of the corresponding eigenvectors have been deflated. In the spectral bound, we have a similar comparison CG process, and the bound is based on the difference between the eigenvalues and the Ritz values.

Let $\lambda_1, \ldots, \lambda_n$ and $v_1, \ldots, v_n$ be the eigenvalues and eigenvectors associated with the matrix A, chosen such that they form an orthonormal basis. Then $A = V\Lambda V^T$ is a spectral decomposition of matrix A with $V = [v_1, \ldots, v_n]$ and $\Lambda = \operatorname{diag}[\lambda_1, \ldots, \lambda_n]$.

The comparison CG process is a residual sequence, in which the components on the chosen invariant subspace have been deflated [21, 29]. To define this more precisely let Π_Q be a spectral projector onto an invariant subspace $\mathbb{R}(Q)$ of the matrix A with dimension k[†], where Q is an $n \times k$ matrix whose columns are k eigenvectors of A. The spectral projector is constructed as $\Pi_Q = QQ^T$ and in this case it is also an orthogonal projector with respect to the A^{-1}-inner product $\langle \cdot, \cdot \rangle_{A^{-1}}$ since the matrix A^{-1} commutes with its spectral projector, i.e., $\Pi_Q A^{-1} = A^{-1}\Pi_Q$.

The block comparison process is defined as a CG process which commences with $\bar{\mathbf{R}}_0 = (I - \Pi_Q)\mathbf{R}_m$, i.e., with the initial residual being the mth block residual of the original CG process, but having its components in $\mathbb{R}(Q)$ deflated, i.e., such that

$$\|\bar{\mathbf{R}}_j\|_{A^{-1}} = \min_{\mathbf{D} \in A\mathbb{K}_j(A, \bar{\mathbf{R}}_0)} \|\bar{\mathbf{R}}_0 - \mathbf{D}\|_{A^{-1}}. \tag{3.1}$$

3.1. Subspace-based bound for block CG. We present an *a posteriori bound* which is a special case of those presented in [21]. As indicated above, the bound is obtained by considering a comparison CG process defined by (3.1). It is an *a posteriori bound* since it assumed that the approximate solution at the mth step $\mathbf{X}_m$ is known, and consequently, the corresponding residual $\mathbf{R}_m$. We summarize the subspace-based bound in the following theorem, where the norm of the residual j steps after the mth step is bounded by the j-th residual norm of the comparison process. The factor in the bound depends on the angle between the Krylov subspace, and the invariant subspace $\mathbb{R}(Q)$, the same subspace used for the deflation.

THEOREM 3.1 (Subspace-based bounds for block Conjugate Gradient [21]). *Consider an $n \times k$ real matrix Y, whose columns are a basis of a k-dimensional subspace of $A\mathbb{K}_m(A, \mathbf{R}_0)$. Let Q be an $n \times k$ matrix whose columns are k eigenvectors of A and Π_Q a spectral projector onto the invariant subspace $\mathbb{R}(Q)$. Let Π_Y be the A^{-1}-orthogonal projector onto $\mathbb{R}(Y)$, and let $\gamma_m = \|(I - \Pi_Y)\Pi_Q\|_{A^{-1}}$. Then*

$$\begin{aligned} \|\mathbf{R}_{m+j}\|_{A^{-1}} \leq & \min_{\mathbf{D} \in A\mathbb{K}_j(A, \mathbf{R}_m)} \{\|(I - \Pi_Q)(\mathbf{R}_m - \mathbf{D})\|_{A^{-1}} \\ & \qquad + \gamma_m \|\Pi_Q(\mathbf{R}_m - \mathbf{D})\|_{A^{-1}}\} \end{aligned} \tag{3.2}$$

$$\leq \sqrt{2} \min_{\mathbf{D} \in A\mathbb{K}_j(A, \mathbf{R}_m)} \left\| \begin{bmatrix} (I - \Pi_Q) \\ \gamma_m \Pi_Q \end{bmatrix} (\mathbf{R}_m - \mathbf{D}) \right\|_\star, \tag{3.3}$$

[†]Here and elsewhere in the paper $\mathbb{R}(Q)$ denotes the range of the matrix Q. We mention that in[21] the invariant subspace must be simple, i.e., such that there is a complementary subspace which is also invariant. Here, this condition is fulfilled automatically since the invariant subspaces consist of linear combinations of orthogonal eigenvectors.

where $\|\cdot\|_\star$ is an induced vector norm from the following inner product. Let $u_i, v_i \in \mathbb{R}^n$, $i = 1, 2$; then, if $u^T = \left[u_1^T, u_2^T\right]$, $v^T = \left[v_1^T, v_2^T\right]$, $\langle u, v\rangle_\star = \langle u_1, v_1\rangle_{A^{-1}} + \langle u_2, v_2\rangle_{A^{-1}}$.

The quantity γ_m corresponds to the angle between a subspace of $A\mathbb{K}_m(A, \mathbf{R}_0)$ and the invariant subspace since for symmetric A,

$$\gamma_m = \|(I - \Pi_Y)\Pi_Q\|_{A^{-1}} = \|\Pi_Y - \Pi_Q\|_{A^{-1}} = \sin\varphi \leq 1, \tag{3.4}$$

where φ is the maximum canonical angle between $\mathbb{R}(Y)$ and $\mathbb{R}(Q)$; see, e.g., [14, p. 56] and [25, p. 92] for details.

The computation of the upper bound is possible using the expression (3.3), which is a reformulation of (3.2) as a least squares problem of dimension $2n$. The least-squares problem (3.3) is well posed as long as $\mathbb{R}(Q) \cap A\mathbb{K}_j(A, \mathbf{R}_m) = \{0\}$.

We note that this theorem is very general and applies to *any* space $\mathbb{R}(Y)$ which is a subspace of the Krylov subspace, and *any* invariant subspace $\mathbb{R}(Q)$. Of course for our bound to be meaninful, one would take an appropriate invariant subspace close to the Krylov subspace. It is well known that the eigenvalues which are captured first as the iterations k progress, are those at the end of the spectrum; see, e.g., [28]. Thus, in our numerical experiments we choose eigenvectors corresponding to k_1 eigenvalues in the lower part of the spectrum and/or k_2 eigenvalues in the upper part of the spectrum.

3.2. Spectral-based bound for block CG. In this section, we present a convergence bound for the block CG algorithm based on spectral information. The aim is to generalize to the block case, the bounds developed in [29]. In this process, we benefited from the background material provided in [15]. The bounds are obtained by considering the optimality property (2.3) of block CG and constructing a comparison process using an auxiliary matrix polynomial. The bound uses the Ritz values, which, as we have shown, coincide with the latent roots of the residual polynomial $\Phi_m(\lambda)$, and they converge to the eigenvalues of matrix A for a sufficiently large m [20].

We begin by recalling the following result.

LEMMA 3.2. [15] *Let $\Upsilon_l(\eta) = \sum_{i=0}^{l} \eta^i \mathcal{C}_i$ be any matrix polynomial where $\mathcal{C}_i$ are $s \times s$ matrices, and let A be a $n \times n$ matrix, $\mathbf{Z}$ a $n \times s$ matrix and S any invertible matrix of order n. Then the following result holds,*

$$\Upsilon_l(A) \circ \mathbf{Z} = S\Upsilon_l(S^{-1}AS) \circ (S^{-1}\mathbf{Z}). \tag{3.5}$$

Using Lemma 3.2, we can write the norm of the block residual (2.14) in terms of polynomials evaluated at the eigenvalues of the matrix, as shown in the following lemma (adapted from [15, Ch. 4, expression (4.7)]), and it will be used in the theorems that follow.

LEMMA 3.3. *Let $A = V\Lambda V^T$ be a spectral decomposition. Let $[w_1, w_2, \ldots, w_n]^T = V^T R_0$ be the components of the block initial residual on the eigenbasis, where w_i's $(i = 1, \ldots, n)$ are $s \times 1$ matrices. Consider $\mathbf{R}_m = \Phi_m(A) \circ \mathbf{R}_0$ the residual of block CG at iteration m. Then,*

$$\|\mathbf{R}_m\|^2_{A^{-1}} = \operatorname{tr}\left(\mathbf{R}_m^T A^{-1} \mathbf{R}_m\right) = \operatorname{tr}\left[\sum_{i=1}^{n} \frac{1}{\lambda_i} \Phi_m^T(\lambda_i) w_i w_i^T \Phi_m(\lambda_i)\right]. \tag{3.6}$$

Proof. Using Lemma 3.2 and noting that $VV^T = I_n$ we can write

$$\begin{aligned}\mathbf{R}_m &= \Phi_m(A) \circ \mathbf{R}_0 \\ &= V\Phi_m(V^TAV) \circ V^T\mathbf{R}_0 = V\Phi_m(\Lambda) \circ \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_n^T \end{bmatrix} = V \begin{bmatrix} w_1^T\Phi_m(\lambda_1) \\ w_2^T\Phi_m(\lambda_2) \\ \vdots \\ w_n^T\Phi_m(\lambda_n) \end{bmatrix}.\end{aligned}$$

The results follows taking the norm $\|\cdot\|_{A^{-1}}$. ∎

Lemma 3.3 shows that the block residual can be expressed as the trace of a sum of $s \times s$ matrices with the matrix polynomial evaluated on each eigenvalue, and the weights $w_i w_i^T$ with $i = 1, \ldots, n$ are $s \times s$ symmetric positive semidefinite matrices of rank-one. Note that for $s = 1$, Lemma 3.3 reduces to the CG case. Evaluating the polynomial on each of the eigenvalues and the elimination of the operator '$\circ$' simplify the development of a superlinear bound in the block case.

As stated in the previous section, k_1 denotes the number of eigenvalues taken in the lowest part and k_2 the number of eigenvalues taken in the upper part of the spectrum to perform our analysis.

For the sake of simplicity, we begin by stating and proving the following theorem for the superlinear bound in the case of $k_1 = 1$ and $k_2 = 0$. That is, we consider the eigenpairs corresponding to the lowest part of the spectrum (λ_1, v_1) and its corresponding lowest Ritz value at the iteration m denoted $\theta_1^{(m)}$. Let $\bar{\mathbf{R}}_0$ be the block residual where the eigenvector v_1 has been deflated, and $\bar{\mathbf{R}}_j$ is the residual of the block CG process (comparison process) starting with $\bar{\mathbf{R}}_0$. We show a bound of the form $\|\mathbf{R}_{m+j}\|_{A^{-1}} \leq \alpha_m \|\bar{\mathbf{R}}_j\|_{A^{-1}}$, where the factor α_m depends on the spectral information, as shown below. In other words, we bound the norm of block residual at the $(m+j)$th iteration by the norm of the residual at the jth iteration of the comparison process. Thus, when the Ritz value approximates well the eigenvalue, the slope of the graph of the residual norm of the comparison process should mimic the slope of the residual norm we are trying to bound.

THEOREM 3.4. *Let $\mathbf{R}_{m+j}$ be the block CG residual at the $(m+j)$th iteration and $\bar{\mathbf{R}}_j$ be the residual after j iterations of block CG applied to $\bar{\mathbf{R}}_0 = (I - \Pi_Q)\mathbf{R}_m$, with $\Pi_Q = v_1 v_1^T$ and $\mathbf{R}_m = \Phi_m(A) \circ \mathbf{R}_0$, i.e.,*

$$\bar{\mathbf{R}}_j = \Psi_j(A) \circ \bar{\mathbf{R}}_0, \tag{3.7}$$

where $\Psi_j(\lambda) \in \mathbb{G}_{m,s}$ is the corresponding matrix polynomial of degree j for the new block CG residual. Then,

$$\|\mathbf{R}_{m+j}\|_{A^{-1}} \leq \alpha_{m,1,0} \|\bar{\mathbf{R}}_j\|_{A^{-1}}, \tag{3.8}$$

where $\alpha_{m,1,0} = \frac{\theta_1^{(m)}}{\lambda_1} \max_{\lambda_i \neq \lambda_1} \frac{|\lambda_i - \lambda_1|}{|\lambda_i - \theta_1^{(m)}|}$.

Proof. Let $\Omega_m(\lambda)$ be a matrix-valued polynomial constructed as follows,

$$\Omega_m(\lambda) = \frac{\theta_1^{(m)}}{\lambda_1}(\lambda - \lambda_1)(\lambda - \theta_1^{(m)})^{-1}\Phi_m(\lambda) \in \mathbb{G}_{m,s}. \tag{3.9}$$

By the optimality property (2.14) of block CG, we have the following relation,

$$\|\mathbf{R}_{m+j}\|_{A^{-1}} \leq \|\Psi_j(A) \circ [\Omega_m(A) \circ \mathbf{R}_0]\|_{A^{-1}}.$$

Following a procedure similar to that used in Lemma 3.3, we have

$$\|\mathbf{R}_{m+j}\|_{A^{-1}}^2 \leq \left\| V \begin{bmatrix} w_1^T \Omega_m(\lambda_1)\Psi_j(\lambda_1) \\ w_2^T \Omega_m(\lambda_2)\Psi_j(\lambda_2) \\ \vdots \\ w_n^T \Omega_m(\lambda_n)\Psi_j(\lambda_n) \end{bmatrix} \right\|_{A^{-1}}^2 . \tag{3.10}$$

Using expression (3.9), the right hand side of inequality (3.10) is equal to

$$\left\| V \begin{bmatrix} w_1^T \frac{\theta_1^{(m)}}{\lambda_1}(\lambda_1 - \lambda_1)(\lambda_1 - \theta_1^{(m)})^{-1}\Phi_m(\lambda_1)\Psi_j(\lambda_1) \\ w_2^T \frac{\theta_1^{(m)}}{\lambda_1}(\lambda_2 - \lambda_1)(\lambda_2 - \theta_1^{(m)})^{-1}\Phi_m(\lambda_2)\Psi_j(\lambda_2) \\ \vdots \\ w_n^T \frac{\theta_1^{(m)}}{\lambda_1}(\lambda_n - \lambda_1)(\lambda_n - \theta_1^{(m)})^{-1}\Phi_m(\lambda_n)\Psi_j(\lambda_n) \end{bmatrix} \right\|_{A^{-1}}^2 . \tag{3.11}$$

By construction, the first row of the matrix on the right hand side of (3.10), (or equivalently in expression (3.11)) is zero. It follows then that by evaluating the matrix polynomial $\Omega(\lambda)$ on each eigenvalue and taking into account that the matrices $w_i w_i^T$ are positive semidefinitive, we have that (3.11) is equal to

$$\begin{aligned} &\left\| V \begin{bmatrix} 0 \\ w_2^T \frac{\theta_1^{(m)}}{\lambda_1}(\lambda_2 - \lambda_1)(\lambda_2 - \theta_1^{(m)})^{-1}\Phi_m(\lambda_2)\Psi_j(\lambda_2) \\ \vdots \\ w_n^T \frac{\theta_1^{(m)}}{\lambda_1}(\lambda_n - \lambda_1)(\lambda_n - \theta_1^{(m)})^{-1}\Phi_m(\lambda_n)\Psi_j(\lambda_n) \end{bmatrix} \right\|_{A^{-1}}^2 \\ &= \left[\frac{\theta_1^{(m)}}{\lambda_1}\right]^2 \operatorname{tr}\left[\sum_{i=2}^{n}(\lambda_i - \lambda_1)(\lambda_i - \theta_1^{(m)})^{-1}\frac{1}{\lambda_i}(\Phi_m(\lambda_i)\Psi_j(\lambda_i))^T w_i w_i^T \Phi_m(\lambda_i)\Psi_j(\lambda_i)\right] . \end{aligned} \tag{3.12}$$

In the last expression we take the maximum over the scalar factors, and obtain the following bound for (3.10).

$$\leq \left[\frac{\theta_1^{(m)}}{\lambda_1}\max_{\lambda_i \neq \lambda_1}\frac{|\lambda_i - \lambda_1|}{|\lambda_i - \theta_1^{(m)}|}\right]^2 \operatorname{tr}\left[\sum_{i=2}^{n}\frac{1}{\lambda_i}(\Phi_m(\lambda_i)\Psi_j(\lambda_i))^T w_i w_i^T \Phi_m(\lambda_i)\Psi_j(\lambda_i)\right] .$$

Using Lemma 3.3 and since in the sum starts with $i = 2$, we obtain the following expression to the block CG residual.

$$\begin{aligned} \|\mathbf{R}_{m+j}\|_{A^{-1}}^2 &\leq \left[\frac{\theta_1^{(m)}}{\lambda_1}\max_{\lambda_i \neq \lambda_1}\frac{|\lambda_i - \lambda_1|}{|\lambda_i - \theta_1^{(m)}|}\right]^2 \|\Psi_j(A) \circ \Phi_m(A) \circ (I - \Pi_Q)\mathbf{R}_0\|_{A^{-1}}^2 \\ &= \left[\frac{\theta_1^{(m)}}{\lambda_1}\max_{\lambda_i \neq \lambda_1}\frac{|\lambda_i - \lambda_1|}{|\lambda_i - \theta_1^{(m)}|}\right]^2 \|\bar{\mathbf{R}}_j\|_{A^{-1}}^2 . \end{aligned} \tag{3.13}$$

The result follows by taking the square root. □

Observe that the expression in (3.9) is indeed a matrix polynomial. The Ritz value $\theta_1^{(m)}$ is a latent root of Φ_m, so that in (3.9) this root is removed, and λ_1 is

added as a root. Furthermore, this auxiliary matrix polynomial converges to the residual polynomial $\Phi_m(\lambda)$ as m increases, since he Ritz value $\theta_1^{(m)}$ converges to λ_1 and $(\lambda-\lambda_1)(\lambda-\theta_1^{(m)})^{-1} \approx \mathcal{I}_s$.

We state now the general case, that is, a comparison with a process with multiple eigenvalues deflated and at both ends of the spectrum. The proof follows in the same manner as the proof of Theorem 3.4, by appropiately constructing the auxiliary matrix polynomial $\Omega_m(\lambda)$.

THEOREM 3.5. *Let $\mathbf{R}_{m+j}$ be the block CG residual at the $(m+j)$th step and $\bar{\mathbf{R}}_j$ be the residual after j steps of block CG applied to $\bar{\mathbf{R}}_0 = (I-\Pi_Q)\mathbf{R}_m$, with $\Pi_Q = QQ^T$ and $\mathbf{R}_m = \Phi_m(A)\circ\mathbf{R}_0$, where $Q \in \mathbb{R}^{n\times k}$ with $k = k_1 + k_2$, and*

$$\bar{\mathbf{R}}_j = \Psi_j(A)\circ\bar{\mathbf{R}}_0, \tag{3.14}$$

where $\Psi_j(\lambda) \in \mathbb{G}_{m,s}$ is the corresponding block CG matrix polynomial of degree j. Then,

$$\|\mathbf{R}_{m+j}\|_{A^{-1}} \leq \alpha_{m,k_1,k_2} \left\|\bar{\mathbf{R}}_j\right\|_{A^{-1}}, \tag{3.15}$$

where α_{m,k_1,k_2} is given by

$$\alpha_{m,k_1,k_2} = \max_{j_1>k_1,\, j_2\geq k_2} \prod_{j=1}^{k_1} \frac{\theta_j^{(m)}}{\lambda_j} \left|\frac{\lambda_{j_1}-\lambda_j}{\lambda_{j_1}-\theta_j^{(i)}}\right| \prod_{j=1}^{k_2} \frac{\theta_{n+1-j}^{(m)}}{\lambda_{n+1-j}} \left|\frac{\lambda_{n+1-j}-\lambda_{n-j_2}}{\theta_{m+1-j}^{(m)}-\lambda_{n-j_2}}\right|. \tag{3.16}$$

Proof. Let $\Omega_m(\lambda)$ be a matrix polynomial constructed as follows,

$$\Omega_m(\lambda) = \prod_{j=1}^{k_1} \frac{\theta_j^{(m)}}{\lambda_j}(\lambda_{j_1}-\lambda_j)(\lambda_{j_1}-\theta_j^{(m)})^{-1}\times \prod_{j=1}^{k_2} \frac{\theta_{n+1-j}^{(m)}}{\lambda_{n+1-j}}(\lambda_{n+1-j}-\lambda_{n-j_2})(\theta_{m+1-j}^{(m)}-\lambda_{n-j_2})^{-1}\Phi_m(\lambda).$$

for $j_1 > k_1$ and $j_2 \geq k_2$, and since $\Omega_m(0) = \mathcal{I}_s$, $\Omega_m(\eta) \in \mathbb{G}_{m,s}$. The result follows using a similar procedure employed in Theorem 3.4. ∎

We note that theorems 3.4 and 3.5 reduce to the results obtained in [29] when studying superlinear bound in CG, when we consider the special case $s = 1$.

We end this section by remarking that Theorem 3.5 holds in particular when an eigenvalue has multiplicity $\kappa > 1$. In this case, in the expression for α_{m,k_1,k_2} some factors are repeated κ times.

4. Numerical Experiments. We present a series of numerical experiments illustrating the quality of both the subspace-based bounds (3.2) and the spectral-based bounds (3.8), and we compare them. We discuss different cases, namely different eigenvalue distributions of the coefficient matrix A, as well as taking different dimensions of the invariant subspaces in the comparison process. We also consider the effect of taking increasing numbers of right hand sides.

Without loss of generality, we consider diagonal matrices in our first five examples. In Example 4.6 we consider a preconditioned model problem. Recall that the coefficient α_{m,k_1,k_2} in (3.8) depends on k_1 eigenvalues in the lower part of the spectrum, and k_2 eigenvalues in the upper part of the spectrum. Different cases of k_1

and k_2 are considered, both when the eigenvalues are simple, and when they have multiplicity $\kappa > 1$. Recall also that in the bound (3.2) we have an invariant subpace whose basis are the columns of the matrix Y. In our experiments, these columns are $y_i = Az_i$, $z_i \in \mathbb{K}_m(A, \mathbf{R}_0)$, and z_i are taken to be Ritz vectors.

For each example, two sets of complementary results are presented. In one set, we follow the evolution of the Ritz values as they converge to the corresponding eigenvalues, and look at the behaviour of the two constants used in our bounds, α_{m,k_1,k_2} in (3.8) and γ_m in (3.2). In this case, we use the bounds (3.8) and (3.2) for $j = 0$ as m increases. We can observe than when the Ritz values are close enough to the eigenvalues, or when γ_m is small enough, the convergence curve changes slope; though this point is hard to pinpoit exactly. The second set of experiments show that the bounds obtained follow closely the slope of the convergence curves, especially after the change of slope, sometimes called the onset of superlinearity, has taken hold.

EXAMPLE 4.1. Our first example is for the case $s = 1$, and thus, the block method reduces to the standard case, and our bounds reduce to that presented in [29]. We consider $k_1 = 1$, $k_2 = 0$, that is, only one Ritz value (and one Ritz vector), at the lowest part of the spectrum. We do so as to examine three situations in which the Ritz value $\theta_1^{(m)}$ is either very close to the first eigenvalue λ_1 (Case 4.1.a), $\theta_1^{(m)}$ is between λ_1 and λ_2 (Case 4.1.b), or $\theta_1^{(m)} > \lambda_2$ (Case 4.1.c), as was done in [30]. We thus study a 100×100 diagonal matrix with eigenvalues 0.1, 0.2, 0.3, 0.4, 5, ... , 100. The right-hand side is a vector with all unit entries, the initial approximation $\mathbf{X}_0 = x_0$ is the zero vector, and the initial residual is $\mathbf{R}_0 = b$; *i.e.*, the residual has equal components in all eigenvectors that form an eigenbasis so that no one eigendirection is favored.

Recall that in the expressions (3.2) and (3.15), $\mathbf{R}_{m+j}$ is the block CG residual at the $(m + j)$th step, and $\bar{\mathbf{R}}_j$ is the residual after j steps of block CG applied to $\bar{\mathbf{R}}_0 = (I - \Pi_Q)\mathbf{R}_m$. The counter m specifies the status of iteration advance of block CG when the bound starts to be employed. Hence, in the first set of experiments, as m increases, θ_1 convergences towards λ_1, as shown, e.g., in Table 4.1 below, and for fixed m, as j increases, we look at the behavior of the bounds, as shown, e.g., in Table 4.2.

Case 4.1.a. The first Ritz value θ_1 is very close to λ_1. Our first set of experiments are reported in Table 4.1, where we show the behavior of $\alpha_{m,1,0}$ and γ_m at each iteration m. Bounds (3.2) and (3.8) are computed with $j = 0$ and are presented for several values of m. In other words, the bounds reduced to

$$b_1 := \|\bar{\mathbf{R}}_0\|_{A^{-1}} + \gamma_m \|\Pi_Q \mathbf{R}_m\|_{A^{-1}}, \quad \text{and} \quad b_2 := \alpha_{m,1,0} \|\bar{\mathbf{R}}_0\|_{A^{-1}}. \tag{4.1}$$

The comparison residual at iteration m is $\bar{\mathbf{R}}_0 = \|(I - \Pi_Q)\mathbf{R}_m\|_{A^{-1}}$. Observe that despite of the fact that the convergence of the Ritz vectors is slow and non-monotone [18], which has influence on γ_m, the subspace bound b_1 gives sharper estimates than the spectral bound b_2; see Table 4.1.

Observe that as m increases, then $\alpha_{m,1,0}$ decreases (ideally it tends to one) and γ_m also decreases (ideally tends to zero). Recall that when $\alpha_{m,1,0} = 1$ and $\gamma_m = 0$, then both bounds coincide. However, also note that the spectral bound b_2 is more sensitive to changes $\alpha_{m,1,0}$ than the bound b_1 to changes in γ_m. This means that γ_m does not need to be close to zero in order for the subspace bound b_2 obtain a better approximation than the spectral bound b_1.

In the second set of experiments, we examine our two bounds in the same stage of convergence of the Ritz values to their corresponding eigenvalues. To this end, we fix $m = 33$ and take iterations j from $j = 1$ to $j = 10$. For easier reading we denote

Table 4.1: Case 4.1.a. The first Ritz value $\theta_1^{(m)}$ is very close to smallest eigenvalue λ_1. Parameters $k_1 = 1$, $k_2 = 0$, $s = 1$ and $j = 0$. Factors γ_m and $\alpha_{m,1,0}$ are computed using (3.4) and (3.8), respectively. The subspace bound b_1 and the spectral bound b_2 correspond to (4.1).

m	θ_1	θ_1/λ_1	γ_m	$\alpha_{m,1,0}$	b_1	b_2	$\|\bar{\mathbf{R}}_0\|_{A^{-1}}$	$\|\mathbf{R}_m\|_{A^{-1}}$
30	0.12659	1.26597	0.61127	1.72469	0.77367	0.92530	0.53650	0.66210
31	0.12280	1.22800	0.56286	1.59067	0.67380	0.79217	0.49801	0.58784
32	0.11708	1.17082	0.48434	1.41203	0.53402	0.60523	0.42862	0.48069
33	0.11138	1.11386	0.39936	1.25698	0.39652	0.42922	0.34147	0.36825
34	0.10771	1.07712	0.33933	1.16715	0.29890	0.31279	0.26799	0.28305

Table 4.2: Case 4.1.a. The first Ritz value θ_1 very close to smallest eigenvalue λ_1. Parameters are $k_1 = 1$, $k_2 = 0$, $s = 1$, and iteration $m = 33$. Factors $\alpha_{m,k_1,k_2} = \alpha_{33,1,0}$ and $\gamma_m = \gamma_{33}$ are computed using (3.8) and (3.4), respectively. The subspace bound $b_{1,j}$ and the spectral bound $b_{2,j}$ are given by (4.2).

θ_1	$\alpha_{33,1,0}$	γ_{33}	$\|\mathbf{R}_{33}\|_{A^{-1}}$
0.11138	1.25698	0.39936	0.36825

j	$b_{1,j}$	$b_{2,j}$	$\|\bar{\mathbf{R}}_j\|_{A^{-1}}$	$\|\mathbf{R}_{33+j}\|_{A^{-1}}$
1	0.34058	0.35903	0.28563	0.28305
2	0.29627	0.30365	0.24157	0.23350
3	0.27034	0.27142	0.21594	0.20939
4	0.25733	0.25542	0.20321	0.19836
5	0.25064	0.24739	0.19684	0.19279
6	0.24635	0.24246	0.19293	0.18896
7	0.24228	0.23803	0.18944	0.18476
8	0.23651	0.2320	0.18474	0.17775
9	0.22602	0.22135	0.17654	0.16362
10	0.20638	0.20139	0.16165	0.13714

our bounds as follows,

$$\begin{cases} b_{1,j} &= \min_{\mathbf{D}\in A\mathbb{K}_j(A,\mathbf{R}_m)} \{\|(I-\Pi_Q)(\mathbf{R}_m-\mathbf{D})\|_{A^{-1}} \\ & \qquad\qquad +\gamma_m\|\Pi_Q(\mathbf{R}_m-\mathbf{D})\|_{A^{-1}}\} \qquad \text{and} \\ b_{2,j} &= \alpha_{m,k_1,k_2}\|\bar{\mathbf{R}}_j\|_{A^{-1}}. \end{cases} \tag{4.2}$$

Table 4.2 shows the behavior of expressions the subspace bound $b_{1,j}$ and the spectral bound $b_{2,j}$ Note that for this example, both bounds behave similarly.

Case 4.1.b. The Ritz value θ_1 is between λ_1 and λ_2. The first set of experiments, for $m = 21, \ldots, 25$, are presented in Table 4.3. In this interval, the factor $\alpha_{m,1,0}$ varies considerable with m. This is in contrast with γ_m which does not vary as much.. The variation of $\alpha_{m,1,0}$ is due to the proximity of $\theta_1^{(m)}$ to λ_2, which yields a very rough estimate by the spectral bound b_2 of the block CG behavior. The estimate is improved either once $\theta_1^{(m)}$ moves away from λ_2 or by replacing the factor α_m by $\theta_1^{(m)}/\lambda_1$ as suggested in [28]. The latter results in a bound ($\frac{\theta_1}{\lambda_1}\|\bar{\mathbf{R}}_0\|_{A^{-1}}$) which is sharper than either of our bounds in this case.

For the second set of experiments, we consider $m = 23$, when $\theta_1^{(m)}$ fall almost exactly in the middle of $[\lambda_1, \lambda_2]$. Hence, the factor $\alpha_{m,1,0}$ is large and the spectral bound $b_{2,j}$ overestimates the residual norm, as shown in Table 4.4. Note that for this case, the subspace bound $b_{1,j}$ is sharper than the spectral bound $b_{2,j}$ for all j.

Case 4.1.c. When $\theta_1^{(m)}$ is outside the interval $[\lambda_1, \lambda_2]$. This occurs when

Table 4.3: Case 4.1.b. Ritz value θ_1 in $[\lambda_1, \lambda_2]$. Parameters $k_1 = 1$, $k_2 = 0$, $s = 1$, and $j = 0$. Factors $\alpha_{m,1,0}$ and γ_m are computed using (3.8) and (3.4), respectively. The subspace bound b_1 and the spectral bound b_2 correspond to (4.1).

m	$\theta_1^{(m)}$	θ_1/λ_1	γ_m	$\alpha_{m,1,0}$	b_1	b_2	$\|\bar{\mathbf{R}}_0\|_{A^{-1}}$	$\|\mathbf{R}_m\|_{A^{-1}}$
21	0.20181	2.0181	0.86898	110.93237	2.11922	113.62360	1.02426	1.62383
22	0.17786	1.7786	0.80929	8.0359	1.78204	6.28417	0.964314	1.39672
23	0.15595	1.5595	0.74794	3.5404	1.43954	2.47657	0.85200	1.15887
24	0.14271	1.4271	0.70553	2.4913	1.18655	1.59363	0.73686	0.97427
25	0.13622	1.3622	0.68186	2.1359	1.03745	1.29275	0.65704	0.86194

Table 4.4: Case 4.1.b. Residual bounds at $m = 23$, when θ_1 is in the middle of $[\lambda_1, \lambda_2]$. Parameters $k_1 = 1$, $k_2 = 0$ and $s = 1$. Factors $\alpha_{23,1,0}$ corresponds to (3.8) and γ_m to (3.4). The subspace bound $b_{1,j}$ and the spectral bound $b_{2,j}$ correspond to (4.2).

	θ_1	$\alpha_{23,1,0}$	γ_{23}	$\|\mathbf{R}_{23}\|_{A^{-1}}$
	0.15595	3.54044	0.74794	1.15887
j	$b_{1,j}$	$b_{2,j}$	$\|\bar{\mathbf{R}}_j\|_{A^{-1}}$	$\|\mathbf{R}_{23+j}\|_{A^{-1}}$
1	1.28592	2.47657	0.69950	0.97427
2	1.15295	2.01543	0.56925	0.86194
3	1.06385	1.71144	0.48339	0.80285
4	1.01284	1.54247	0.43567	0.77255
5	0.98455	1.45406	0.41070	0.75507
6	0.96687	1.40444	0.39668	0.74175
7	0.95238	1.36986	0.38691	0.72694
8	0.93562	1.33628	0.37743	0.70399
9	0.90988	1.29108	0.36466	0.66210
10	0.86496	1.21714	0.34378	0.58784

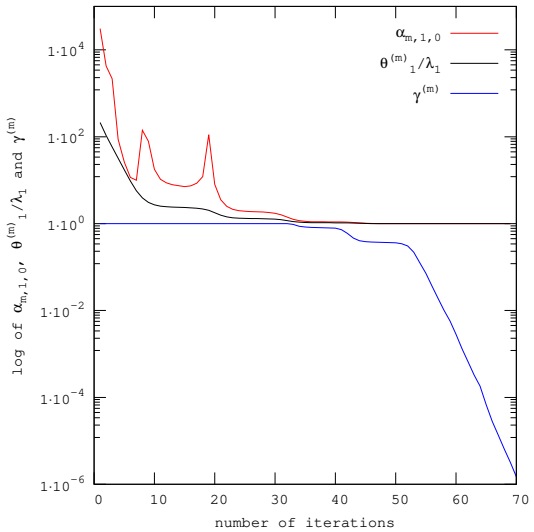


Fig. 4.1: Example 4.1.c. Behavior of the factors $\alpha_{m,1,0}$ and γ_m of our bounds given in (4.1) for each m, the number of iterations. The smallest Ritz value $\theta_1^{(m)}$ at the mth iteration and smallest eigenvalue λ_1 of the matrix A.

$m \leq 20$. The factors $\alpha_{m,1,0}$ and γ_m have very different behavior as reported in Figure 4.1. Observe that $\alpha_{m,1,0}$ has large variations when $\theta_1^{(m)}$ approaches a wrong eigenvalue, and consequently the spectral bound computed with these factors do not provide any useful information. On the other hand, for $m \leq 20$ the value of γ_m remains almost equal to 1, that means that the subspace bound $b_{1,j}$ in (4.2) can be computed as $\|(I - \Pi_Q)\bar{\mathbf{R}}_j\|_{A^{-1}} + \|\Pi_Q \bar{\mathbf{R}}_j\|_{A^{-1}}$.

In summary, Example 4.1 illustrates the behavior of the bounds in several stages of the convergence of the Ritz value $\theta_1^{(m)}$ towards the eigenvalue λ_1. In **Case 4.1.a**, both bounds behave similarly when the Ritz value has sufficiently converged to its corresponding eigenvalue. In the other two cases, the subspace bound is better than the spectral bound when the Ritz value $\theta_1^{(m)}$ is far away of the corresponding eigenvalue λ_1.

EXAMPLE 4.2. Here, we consider the four smallest eigenvalues in the construction of the bounds, i.e., $k_1 = 4$ and $k_2 = 0$. In this experiment, the matrix A and the initial iterate $\mathbf{X}_0$ are the same as in Example 4.1.

In the first set of experiments, we present in Table 4.5, the convergence of the Ritz values $\theta_i^{(m)}$, $i = 1, \ldots, 4$, to the four lowest eigenvalues, and the behavior of $\alpha_{m,k_1,k_2} = \alpha_{m,4,0}$ and γ_m. In addition, we present the subspace bound b_1 and the spectral bound b_2 given in (4.1), and the norms of the comparison residual $\bar{\mathbf{R}}_0$ and the block CG residual $\mathbf{R}_m$. Note that since $\theta_4^{(m)}$ is close to an eigenvalue different than λ_4, namely λ_5, the factor $\alpha_{m,4,0}$ varies widely.

Table 4.5: Example 4.2. Ritz values $\theta_i^{(m)}$, $i = 1, \cdots, 4$. The bounds area computed with $j = 0$, and parameters $k_1 = 4$ and $k_2 = 0$. Factors $\alpha_{m,4,0}$ corresponds to (3.16) and γ_m to (3.4). The subspace bound b_1 and the spectral bound b_2 correspond to (4.1). The four smallest eigenvalues are $\lambda_1 = 0.1, \lambda = 0.2, \lambda = 0.3$ and $\lambda = 0.4$.

m	$\theta_1^{(m)}$	$\theta_2^{(m)}$	$\theta_3^{(m)}$	$\theta_4^{(m)}$	$\alpha_{m,4,0}$	γ_m	b_1	b_2	$\|\bar{\mathbf{R}}_0\|_{A^{-1}}$	$\|\mathbf{R}_m\|_{A^{-1}}$
39	0.10485	0.24301	0.39061	5.00336	29249.6	0.99998	0.23883	1340.984	0.04584	0.19836
40	0.10469	0.24235	0.39033	4.99812	52236.3	0.99996	0.22012	1547.637	0.02962	0.19279
41	0.10458	0.24189	0.39013	4.98636	7124.5	0.99983	0.21305	184.532	0.02590	0.18896
42	0.10446	0.24138	0.38991	4.35890	131.885	0.98594	0.21336	4.53408	0.03437	0.18476
43	0.10426	0.24049	0.38949	2.33890	16.9294	0.95700	0.21476	0.88256	0.05213	0.17775
44	0.10384	0.23842	0.38838	1.10663	5.3865	0.88939	0.20429	0.40350	0.07491	0.16362
45	0.10302	0.23329	0.38414	0.58421	2.40210	0.75256	0.16799	0.16799	0.09033	0.13714
46	0.10185	0.22274	0.35799	0.42276	1.46327	0.55404	0.11397	0.12184	0.08327	0.10002

The second set of experiments are presented in Figure 4.2, where we see the behavior of the residual norm $\|\mathbf{R}_m\|_{A^{-1}}$ of the block CG, the subspace bound $b_{1,j}$ and spectral bound $b_{2,j}$ as expressed in (4.2). Note that at iteration $m = 45$ the spectral bound $b_{2,j}$ is sharper than the subspace bound $b_{1,j}$. This is because the subspace bound $b_{1,j}$ with a moderate γ_m ($\gamma_m = 0.75256$) amplifies significantly the residual component on the selected eigenspace $\mathbb{R}(Q)$. At iteration $m = 50$ the selected Ritz values are very close to its respective eigenvalues and the resulting bounds are almost the same.

EXAMPLE 4.3. This example is designed to analyze the effect of the block size s on the bounds and on the convergence of the block CG in the presence of a single eigenvalue near the origin, i.e., $k_1 = 1$ and $k_2 = 0$. This experiment considers a diagonal matrix of dimension 404×404 with eigenvalues on the diagonal with values $0.0005, 0.08, \ldots, 2.42$, that is, the matrix has one isolated eigenvalue near to zero and the other 403 eigenvalues are equally distributed between 0.08 and 2.42. It is expected

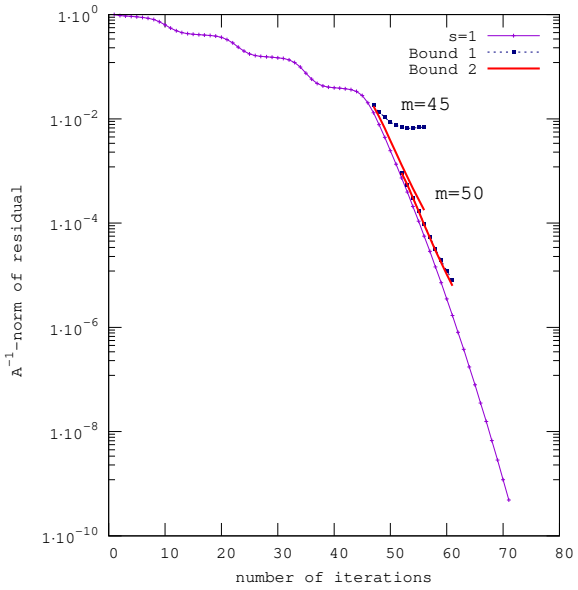


Fig. 4.2: Example 4.2. Block CG residual $\|\mathbf{R}_m\|_{A^{-1}}$, subspace bound 1 and spectral bound 2 from (4.2) computed with $m = 45$ and $m = 50$. Parameters $k_1 = 4$ and $k_2 = 0$.

to observe superlinear convergence once the method captures the lowest eigenvalue 0.005.

Figure 4.3 shows the convergence history for block sizes $s = 1$, 4 and 8. In this example, and in those in the rest of the section, when $s > 1$, the matrix $\mathbf{B}$ has repeated copies of b, but the initial set of vectors in $\mathbf{X}_0$ are nonzero and randomly generated, implying that $\mathbf{R}_0$ has all distinct columns. On each residual history, we plot the subspace bound $b_{1,j}$ and the spectral bound $b_{2,j}$ at two different stages of convergence: soon after the onset of superlinearity occurs, and a numbero of iterations later. Observe that for values of m relatively large, both bounds capture well the slope of the superlinear regime. In addition, observe that as the block size is increased, block CG hastens the onset of superlinear convergence, but going from $s = 4$ to $s = 8$ the difference in the onset is moderate.

Observe also that near the onset of superlinearity (for earlier m) the spectral bound $b_{2,j}$ is sharper than subspace bound $b_{1,j}$. This is due to the fact that factor γ_m has moderate values ($\gamma_m \in [0.5, 1]$)), introducing in the subspace bound $b_{1,j}$ a contribution from the invariant subspace $\mathbb{R}(Q)$. It is important to remark that moderate values of γ_m are in connection with the rate of convergence of Ritz vectors, which is smaller than that of the convergence of the Ritz values. This can be appreciated by comparing Figures 4.4(a) and 4.4(b), that the convergence of $\alpha_{m,1,0}$ (to one) is faster than the convergence of γ_m (to zero).

In addition, in Figure 4.4(a), it can be observed the erratic convergence behavior of $\alpha_{m,1,0}$. In fact, it takes large oscillatory values before approaching the final interval of convergence of the Ritz values. On the other hand, the behavior of γ_m is smooth and well behaved; see Figure 4.4(b). This indicates that the subspace bound $b_{1,j}$ is less sensitive to the stage of convergence of the Ritz value, hence it can be used safely to describe the behavior in this region.

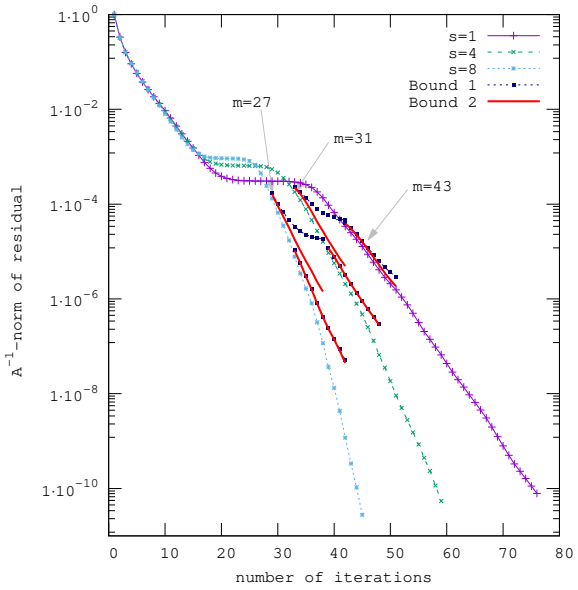


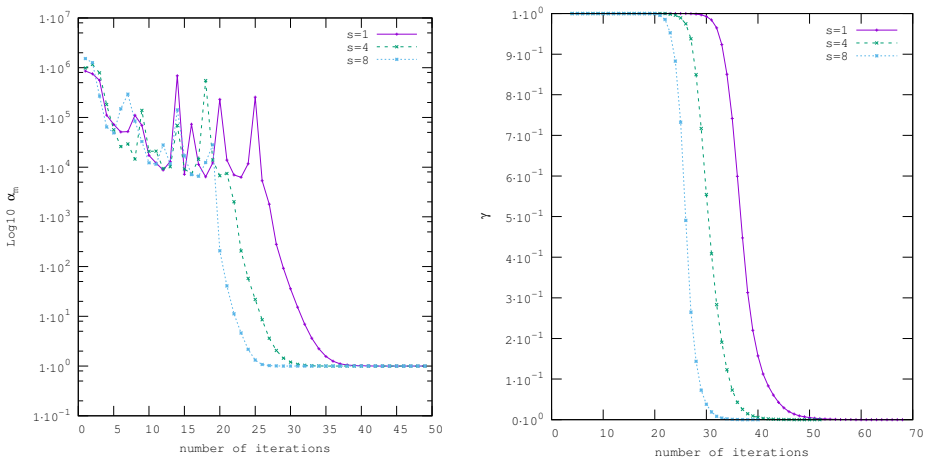
Fig. 4.3: Example 4.3. Residual convergence behavior for the block CG $\|\mathbf{R}_{m+j}\|_{A^{-1}}$ for $s = 1, 4, 8$, subspace and spectral bounds $b_{1,j}$ and $b_{2,j}$ correspond to expressions (4.2). Parameters $k_1 = 1$ and $k_2 = 0$. Note: $m = 43$ correspond to ($\alpha = 1.0254$, $\gamma = 0.1572$), $m = 31$ correspond to ($\alpha = 1.201$, $\gamma = 0.4083$), and $m = 27$ correspond to ($\alpha = 1.0758$, $\gamma = 0.2648$).

EXAMPLE 4.4. The objective of this example is to analyze the effect of the block size s on the convergence of block CG and the behavior of the bounds to capture the superlinearity in presence of a cluster of eigenvalues near the origin. To this end, we consider a diagonal matrix of dimension 404×404 with eigenvalues on the diagonal with values 0.0005, 0.0015, 0.0025, 0.0035, 0.0045, 0.0055, $0.08, \ldots, 2.42$. The matrix has six clustered eigenvalues near zero and the rest distributed uniformly between 0.08 and 2.42.

Figure 4.5 shows the convergence history of block CG, considering the invariant subspace associated to the six lowest eigenvalues, i.e., both bounds are computed using $k_1 = 6$ and $k_2 = 0$ for each block size ($s = 1$, 2, 4 and 8). Our observations here are very similar to those in the previous example, where a single eigenvalue was considered. In particular, note that that when the block size is increased, the superlinear behavior starts earlier. This shows the dependency of the onset of the superlinear behavior on the block size. Moreover, as in the previous example, this example suggests that block CG (with $s > 1$) hastens the onset of the superlinearity in the presence of clustered eigenvalues.

EXAMPLE 4.5. In this example we analyze effect of an eigenvalue with algebraic multiplicity $\kappa > 1$ on the two bounds we study. To this end, we consider a 384×384 matrix with an eigenvalue $\lambda = 0.0005$ with algebraic multiplicity 5 in the lowest part of its spectrum. The rest of eigenvalues are uniformly distributed between 0.065 and 5.42.

Figure 4.6 shows the block CG residual, the subspace bound $b_{1,j}$, and the residual bound $b_{2,j}$. When the parameter $k_1 = s$, the block size (see Figures 4.6(a) and 4.6(c)), the spectral bound $b_{2,j}$ (defined in (4.2) with α_{m,k_1,k_2} approximates adequately the behavior of the residual. On the other hand, if $k_1 > s$ (see Figures 4.6(b) and 4.6(d)),



(a) Behavior of $\alpha_{m,1,0}$

(b) Behavior of γ_m

Fig. 4.4: Example 4.3. Behavior of $\alpha_{m,1,0}$ and γ_m correspond to the number of iterations m. $\alpha_{m,1,0}$ is computed with expression (3.16) and γ_m is computed with expression (3.4). Parameters $k_1 = 1$, $k_2 = 0$ and $s = 1, 4, 8$.

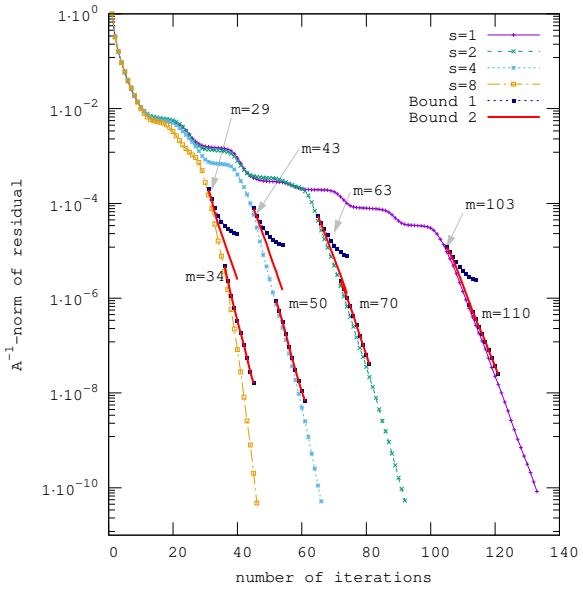
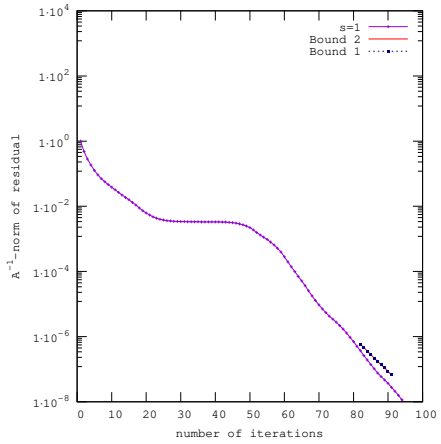


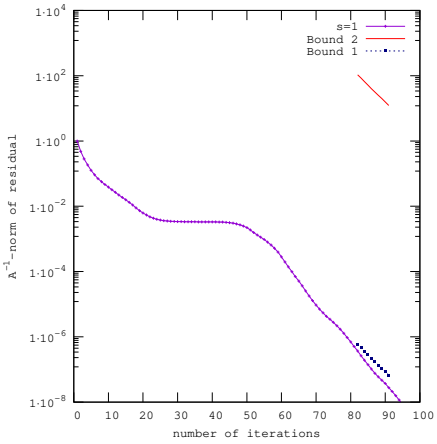
Fig. 4.5: Example 4.4. Residual convergence behavior for the block CG $\|\mathbf{R}_{m+j}\|_{A^{-1}}$ for $s = 1, 2, 4, 8$, The subspace bound $b_{1,j}$ and the spectral bound $b_{2,j}$ are given by (4.2). The parameters are $k_1 = 6$ and $k_2 = 0$. Note: $m = 29$ correspond to ($\alpha = 1.080$, $\gamma = 0.2702$), $m = 43$ correspond to ($\alpha = 1.1630$, $\gamma = 0.3674$), $m = 63$ correspond to ($\alpha = 1.1072$, $\gamma = 0.3097$) and $m = 103$ correspond to ($\alpha = 1.2057$, $\gamma = 0.4102$).

then the bound captures the slope but it is far from sharp. In all cases the subspace bound $b_{1,j}$ approximates adequately the behavior of the residual.

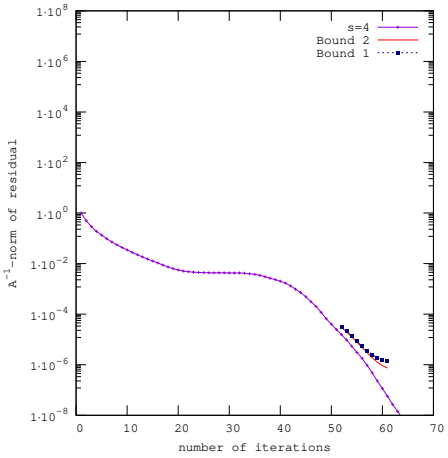
At this point, it is important to remark that the CG polynomial only captures one copy of the eigenvalue with multiplicity, but block CG can find up to s copies in the



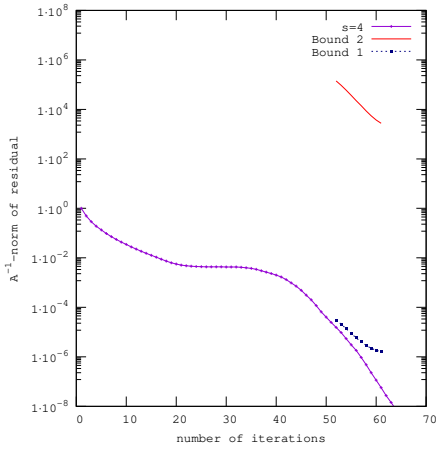
(a) $k_1 = 1$, $k_2 = 0$ and $s = 1$.



(b) $k_1 = 2$, $k_2 = 0$ and $s = 1$.



(c) $k_1 = 4$, $k_2 = 0$ and $s = 4$.



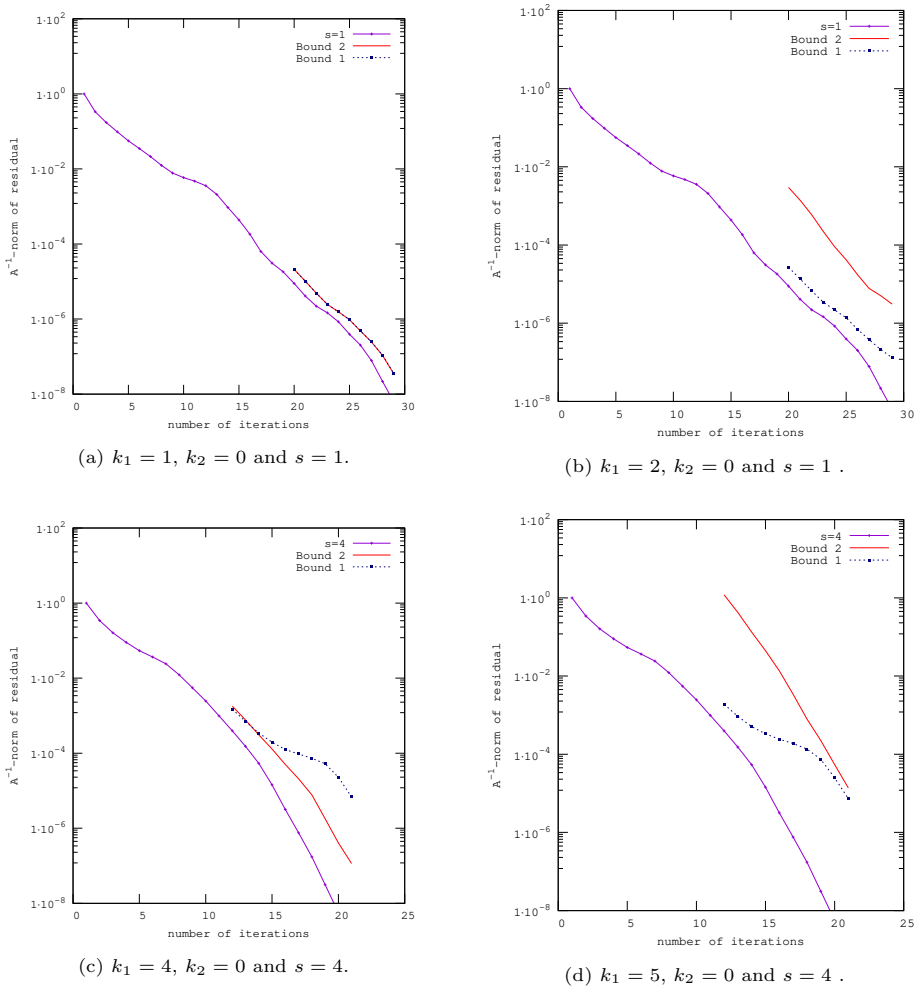
(d) $k_1 = 5$, $k_2 = 0$ and $s = 4$.

Fig. 4.6: Example 4.5. Residual convergence behavior for block CG. The subspace bound $b_{1,j}$ and the spectral bound $b_{2,j}$ are given by (4.2). (a) Parameters values $s = k_1$ and $\alpha_{80,1,0} = 1.00002$ and $s = 1$, (b) Parameters values $k_1 = 2$ and $\alpha_{80,2,0} = 1.83221 \times 10^8$ and $s = 1$,, (c) Parameters values $s = k_1$ and $\alpha_{50,4,0} = 1.00108$ and $s = 4$, and (d) Parameters values $k_1 = 5$ and $\alpha_{50,5,0} = 4.718055 \times 10^9$ and $s = 4$.

case of repeated eigenvalues [27]. This remark also applies for clustered eigenvalues. Hence, if $s = k_1 \leq \kappa$, block CG captures s eigenvalues and the spectral bound $b_{2,j}$ which uses these $k_1 = s$ eigenvalues approximates well the residual. However, if $s < k_1 \leq \kappa$, then the bound is expected to capture more eigenvalues than the block CG is able to capture, and consequently the approximation bound is not sharp. Note the different horizontal scales in Figures 4.6(c) and (d) for $s = 4$, compared to that in Figures 4.6(a) and (b) for $s = 1$.

The analysis of this example suggest that when an eigenvalue with algebraic multiplicity κ is considered, the spectral-based bound approximates adequately the residual behavior when the number of eigenvalues taken as reference (denoted by $k_1 \leq \kappa$) is at most equal to the block size s.

In this example, it can also be observed that as the block size s is increased, the number of eigenvalues captured by block CG is larger, hence the onset of the superlinear convergence occurs earlier (compare for instance Figures 4.6(b) and 4.6(d). This is in accordance with the observation made in Example 4.4 showing a close relationship between block size s and the onset of superlinear convergence.



(a) $k_1 = 1$, $k_2 = 0$ and $s = 1$.

(b) $k_1 = 2$, $k_2 = 0$ and $s = 1$.

(c) $k_1 = 4$, $k_2 = 0$ and $s = 4$.

(d) $k_1 = 5$, $k_2 = 0$ and $s = 4$.

Fig. 4.7: Example 4.6. Poisson matrix preconditioned using an incomplete Cholesky factorization. Residual convergence behavior for the block CG, the bound $b_{2,j}$ corresponding to expression (4.2) and $\|\bar{\mathbf{R}}_j\|_{A^{-1}}$. (a) Parameters $k_1 = 1$ and $\alpha_{20,1,0} = 1.01$. (b) Parameters $k_1 = 2$ and $\alpha_{20,2,0} = 89.4478$. (c) Parameters $k_1 = 4$ and $\alpha_{10,4,0} = 1.448558$. (d) Parameters $k_1 = 7$ and $\alpha_{10,7,0} = 29.224$.

EXAMPLE 4.6. For our last example we consider a 400×400 matrix obtained with a standard discretization of the 2D Poisson equation, presonditioned with incomplete Cholesky factorization with no fill. Thus, the coefficient matrix is $\tilde{A} = L^{-1}AL^{-T}$. The maximum eigenvalue is 1.2015 and the minimum is 0.0724. The ten smallest eigenvalues are 0.0724, 0.1652, 0.1699, 0.2483, 0.2971, 0.2994, 0.3486, 0.3742, 0.4362, 0.4367, 0.4396, 0.4802. The rest of the eigenvalues are distributed between 0.5014 and

1.2015.

Figure 4.7 shows the behavior of the residual, together with the subspace and spectral bounds $b_{1,j}$ and $b_{2,j}$ as in (4.2). In one set of experiments, we consider the block size $s = 1$, and parameters $k_1 = 1$ and $k_2 = 0$, and $k_1 = 2$ and $k_2 = 0$. Figure 4.7(a) shows a good approximation of the bounds to the block CG residual. This is in accordance with the observation made that the convergence is mainly due to the convergence of the first eigenvalue. It can also be observed (similar to Example 4.5) that when k_1 is larger than s, the bounds capture the slope of the residual but the spectral bound $b_{2,j}$ is far from sharp (see Figure 4.7(b)). Similar observation can be done for a larger block sizes, for instance, Figures 4.7(c) and 4.7(d) for values of $k_1 = 4$, $k_2 = 0$ and $s = 4$, and $k_1 = 5$, $k_2 = 0$ and $s = 5$, respectively.

Finally, comparing Figures 4.7(c) and 4.7(d), since more eigenvalues are captured when the block size is increased, the onset of the superlinear convergence occurs earlier. Again, note the different horizontal scale of these two figures for $s = 4$ as compared to those with $s = 1$.

5. Conclusions. We extended the *a posteriori* spectral bound introduced by van der Sluis and van der Vorst [28] to the block CG case. We have also implemented the subspace bound introduced by Simoncini and Szyld [21]. Numerical experiments show that both bounds capture the slope of the residual after the onset of superlinearity. In addition, when there is a cluster of eigenvalues in the lower part of the matrix spectrum, the spectral bound captures better the superlinearity even at the near the onset of superlinearity, while the subspace bound is sharper in presence of repeated eigenvalues.

Analyzing the bounds and the residual behavior of the block CG method, it can be observed that the block method accelerates the convergence because it captures more eigenvalues. Therefore, in the presence of a cluster or repeated eigenvalues in the lower part of the spectrum, the larger the block size s, the earlier is the onset of the superlinear convergence. Hence in these cases we suggest the use of block CG with the block size of the order of to the size of the cluster.

REFERENCES

[1] O. Axelsson and J. Karátson, *On the rate of convergence of the conjugate gradient method for linear operators in hilbert space*, Numerical Functional Analysis and Optimization, 23 (2002), pp. 285–302.

[2] B. Beckermann and A. B. Kuijlaars, *Superlinear CG convergence for special right-hand sides*, Electronic Transactions on Numerical Analysis, 14 (2002), pp. 1–19.

[3] S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, and C. D. Meyer, *GMRES and the minimal polynomial*, BIT Numerical Mathematics, 36 (1996), pp. 664–675.

[4] A. El Guennouni, K. Jbilou, and H. Sadok, *The block Lanczos method for linear systems with multiple right-hand sides*, Applied Numerical Mathematics, 51 (2004), pp. 243 – 256.

[5] H. C. Elman, D. J. Silvester, Andrew, and A. J. Wathen, *Performance and analysis of saddle point preconditioners for the discrete steady-state navier-stokes equations*, Numer. Math, 90 (2000), pp. 665–688.

[6] A. Frommer, K. Lund, and D. B. Szyld, *Block Krylov subspace methods for functions of matrices*, Electronic Transactions on Numerical Analysis, 47 (2017), pp. 100–126.

[7] ———, *Block Krylov subspace methods for functions of matrices II: Modified block FOM*, SIAM Journal on Matrix Analysis and Applications, 41 (2020), pp. 804–837.

[8] T. Gergelits and Z. Strakos, *Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations*, Numerical Algorithms, 65 (2014), p. 759?782.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 4th ed., 2013.

[10] A. Greenbaum, *Iterative methods for solving linear systems*, SIAM, Philadelphia, 1997.
[11] M. H. Gutknecht, *Block Krylov space methods for linear systems with multiple right-hand sides: An introduction*, in Modern Mathematical Models, Methods and Algorithms for Real World Systems, A. Siddiqi, I. Duff, and O. Christensen, eds., Anamaya, New Delhi, 2007, pp. 420–447.
[12] M. H. Gutknecht and T. Schmelzer, *The block grade of a block Krylov space*, Linear Algebra and its Applications, 430 (2009), pp. 174 – 185.
[13] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, Journal of Research of the National Bureau of Standards, 49 (1952).
[14] T. Kato, *Perturbation theory for linear operators*, Springer, Berlin, 2013.
[15] M. D. Kent, *Chebyshev, Krylov, Lanczos: Matrix Relationship and Computations*, PhD thesis, Department of Computer Science - Stanford University, 1989.
[16] M. Kubínová and K. M. Soodhalter, *Admissible and attainable convergence behavior of block arnoldi and gmres*, SIAM Journl on Matrix Analysis and Applications, 41 (2020), pp. 464–486.
[17] D. P. O'Leary, *The block conjugate gradient algorithm and related methods*, Linear Algebra and its Applications, 29 (1980), pp. 293–322.
[18] Y. Saad, *On the Rates of Convergence of the Lanczos and the Block-Lanczos Methods*, SIAM Journal on Numerical Analysis, 17 (1980), pp. 687–706.
[19] Y. Saad, *Iterative methods for sparse linear systems*, vol. 82, SIAM, Philadelphia, 2003.
[20] V. Simoncini, *Ritz and Pseudo-Ritz values using matrix polynomials*, Linear Algebra and its Applications, 241 (1996), pp. 787 – 801.
[21] V. Simoncini and D. B. Szyld, *On the occurrence of superlinear convergence of exact and inexact Krylov subspace methods*, SIAM Review, 47 (2005), pp. 247–272.
[22] ———, *Recent computational developments in Krylov subspace methods for linear systems*, Numerical Linear Algebra with Applications, 14 (2007), pp. 1–59.
[23] ———, *On the superlinear convergence of MINRES*, in Numerical Mathematics and Advanced Applications 2011 - Proceedings of ENUMATH 2011, the 9th European Conference on Numerical Mathematics and Advanced Applications, Leicester, September 2011, Berlin and Heidelberg, 2013, Springer, pp. 733–740.
[24] G. Sleijpen and A. van der Sluis, *Further results on the convergence behavior of conjugate-gradients and Ritz values*, Linear Algebra and its Applications, 246 (1996), pp. 233 – 278.
[25] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, London, 1990.
[26] Z. Strakos, *On the real convergence rate of the conjugate gradient method*, Linear Algebra and its Applications, 154-156 (1991), pp. 535–549.
[27] R. Underwood, *An iterative block Lanczos method for the solution of large sparse symmetric eigenproblems*, PhD thesis, Department of Computer Science, Tech Rept. 496, Stanford University, 1975.
[28] A. van der Sluis and H. A. van der Vorst, *The rate of convergence of Conjugate Gradients*, Numerische Mathematik, 48 (1986), pp. 543–560.
[29] A. van der Sluis and H. A. van der Vorst, *The convergence behavior of Ritz values in the presence of close eigenvalues*, Linear Algebra and its Applications, 88 (1987), pp. 651–694.
[30] H. A. van der Vorst and C. Vuik, *The superlinear convergence behaviour of GMRES*, Journal of Computational and Applied Mathematics, 48 (1993), pp. 327 – 341.
[31] R. Winther, *Some superlinear convergence results for the conjugate gradient method*, SIAM Journal on Numerical Analysis, 17 (1980), pp. 14–17.