ANGLE-FREE CLUSTER ROBUST RITZ VALUE BOUNDS FOR RESTARTED BLOCK EIGENSOLVERS

MING ZHOU, ANDREW V. KNYAZEV, AND KLAUS NEYMEYR

Abstract. Convergence rates of block iterations for solving eigenvalue problems typically measure errors of Ritz values approximating eigenvalues. The errors of the Ritz values are commonly bounded in terms of principal angles between the initial or iterative subspace and the invariant subspace associated with the target eigenvalues. Such bounds thus cannot be applied repeatedly as needed for restarted block eigensolvers, since the left- and right-hand sides of the bounds use different terms. They must be combined with additional bounds which could cause an overestimation. Alternative repeatable bounds that are angle-free and depend only on the errors of the Ritz values have been pioneered for Hermitian eigenvalue problems in doi:10.1515/rnam.1987.2.5.371 but only for a single extreme Ritz value. We extend this result to all Ritz values and achieve robustness for clustered eigenvalues by utilizing nonconsecutive eigenvalues. Our new bounds cover the restarted block Lanczos method and its modifications with shift-and-invert and deflation, and are numerically advantageous.

1. Introduction

Fundamental vector iterations for solving matrix eigenvalue problems, e.g., the power method and the Rayleigh quotient gradient method, need to be implemented in a block form or combined with deflation in order to compute several eigenvalues and the associated eigenvectors. The block implementation is particularly crucial for multiple or clustered eigenvalues since the convergence speed of a vector iteration usually depends on consecutive eigenvalues. For instance, the power method applied to a normal matrix $A \in \mathbb{C}^{n\times n}$ with the eigenvalue arrangement $|\lambda_1| \geq \cdots \geq |\lambda_n|$ has the convergence factor $|\lambda_2/\lambda_1|$ for approximating λ_1, or generally $|\lambda_{i+1}/\lambda_i|$ for approximating λ_i where deflation is required for $i > 1$. An acceleration up to $|\lambda_{i+p}/\lambda_i|$ can be achieved by the block power method using p-dimensional subspaces instead of vectors as iterates.

Many block eigensolvers can be formulated as accelerated versions of the block power method applied to a matrix function $f(A)$. Their convergence factors thus depend on certain eigenvalues of $f(A)$. In particular, the traditional analysis of the block Lanczos method, e.g., by Saad [14, (3.10)] utilizes shifted Chebyshev polynomials as $f(\cdot)$. However, the resulting bounds could cause an overestimation for restarted iterations with low-dimensional subspaces, and combining them with additional bounds is necessary for investigating multiple outer steps. In contrast, bounds of the concise form $\varepsilon_{\text{next}} \leq \varphi\, \varepsilon_{\text{current}}$, where ε and $\varphi < 1$ denote errors and convergence factors, can be applied repeatedly giving direct bounds like $\varepsilon^{(\ell)} \leq \varphi^\ell\, \varepsilon^{(0)}$ showing the cumulative decrease factor φ^ℓ of the initial error after ℓ outer steps of a restarted block iteration.

We discover such repeatable bounds for Ritz value errors by extending convergence analysis of an abstract block iteration by Knyazev [3, 4] for Hermitian eigenvalue problems. Therein the error factor φ of the eigenvalue approximation by Ritz values is defined by distance ratios with respect to relevant eigenvalues, and the central part of the derivation utilizes monotonicity of the Rayleigh quotient rather than angle-based inequalities. Applying these bounds to the

2010 *Mathematics Subject Classification.* Primary 65F15, 65N12, 65N25.
Key words and phrases. Ritz values, subspace eigensolvers, cluster robustness. November 2, 2022.

restarted block Lanczos method [1] overcomes limitations of traditional angle-dependent bounds from [4, 9, 14] and enhances results [16] concerning cluster robustness. Further applications are considered for shift-and-invert and deflation.

We finally mention that multiplying a subspace by the matrix function $f(A)$ is interpreted in [17] as applying a graph-based filter to a block of vectors. For image denoising, this approach can extend the single-vector filtering proposed in [7] to multiple, e.g., multi-spectral, images.

1.1. Ritz value bounds for an abstract block iteration

Our investigation of restarted block eigensolvers focuses on the abstract block iteration

$$\mathcal{Y}' = f(A)\mathcal{Y} \tag{1.1}$$

with successive iterative subspaces $\mathcal{Y}$ and $\mathcal{Y}'$. The function $f(\cdot)$ can be specified afterwards for concrete eigensolvers. The original introduction of (1.1) in [3, 4] begins with a generalized eigenvalue problem. Therein a pair (M, L) of Hermitian matrices with positive definite L is considered, and (1.1) can be reformulated for the operator $L^{-1}M$ that is Hermitian with respect to the inner product induced by L. To simplify notation, we use without loss of generality the theoretically equivalent setup $A = L^{-1/2}ML^{-1/2}$ with respect to the Euclidean inner product so that A is a Hermitian matrix. The eigenvalue problems of A and (M, L) can easily be converted into each other:

$$AX = X\Lambda \quad \Leftrightarrow \quad MV = LV\Lambda \quad \text{for} \quad X = L^{1/2}V,$$

where Λ is a diagonal matrix containing eigenvalues, and the columns of X or V are the associated eigenvectors of A or (M, L).

We arrange the eigenvalues of A as $\lambda_1 \geq \cdots \geq \lambda_n$, and denote by p the block size of (1.1), i.e., $p = \dim\mathcal{Y}$. As a single-step iteration, (1.1) describes an outer step of a restarted block iteration which aims at the p largest eigenvalues of A, i.e., the subspace $\mathcal{Y}$ is the current approximation and the subspace $\mathcal{Y}'$ is an improved approximation to the invariant subspace $\mathcal{X}$ associated with $\lambda_1, \ldots, \lambda_p$. This consideration can be extended to other target eigenvalues after elementary reformulations, as introduced below in Section 4 concerning applications of our analysis and concrete eigensolvers.

Under natural assumptions (cf. Theorem 2.4), the subspace $\mathcal{Y}'$ also has dimension p, and provides more accurate Ritz values as approximate eigenvalues in comparison to those by the subspace $\mathcal{Y}$.

We first review an angle-dependent Ritz value bound based on [4, (2.20)]. By using arbitrary orthonormal basis matrices X and Y of the subspaces $\mathcal{X}$ and $\mathcal{Y}$ mentioned above, the Euclidean angle $\angle(\mathcal{X}, \mathcal{Y})$ is given by $\arccos\tau$ with the smallest singular value τ of X^HY. Then the ith largest Ritz value η_i' in $\mathcal{Y}'$ for an arbitrary index $i \in \{1, \ldots, p\}$ is shown in [4, (2.20)] to fulfill

$$(\lambda_i - \eta_i')/(\eta_i' - \lambda_n) \leq \varphi_i^2 \tan^2\angle(\mathcal{X}, \mathcal{Y}) \tag{1.2}$$

with the convergence factor

$$\varphi_i = \big(\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|\big)/\big(\min_{j\in\{1,\ldots,i\}}|f(\lambda_j)|\big). \tag{1.3}$$

The derivation of (1.2) uses the intermediate angle bound

$$\tan\angle\big(\mathcal{X}_i, f(A)\mathcal{Y}_i\big) \leq \varphi_i \tan\angle(\mathcal{X}_i, \mathcal{Y}_i) \tag{1.4}$$

for an invariant subspace $\mathcal{X}_i$ associated with $\lambda_1, \ldots, \lambda_i$ and an auxiliary subspace $\mathcal{Y}_i \subseteq \mathcal{Y}$ whose orthogonal projection on $\mathcal{X}$ coincides with $\mathcal{X}_i$. Bound (1.4) implies (1.2) by using the perturbation bound

$$(\lambda_i - \widetilde{\eta}_i)/(\widetilde{\eta}_i - \lambda_n) \leq \tan^2\angle\big(\mathcal{X}_i, f(A)\mathcal{Y}_i\big)$$

for the ith largest Ritz value $\widetilde{\eta}_i$ in $f(A)\mathcal{Y}_i$ together with the inequalities $\lambda_i \geq \eta'_i \geq \widetilde{\eta}_i$ and $\angle(\mathcal{X}_i, \mathcal{Y}_i) = \angle(\mathcal{Y}_i, \mathcal{X}) \leq \angle(\mathcal{X}, \mathcal{Y})$.

Our recent paper [17] upgrades (1.2) by majorization-based techniques from [6]. The resulting bounds are concerned with partial sums of principal angles and Ritz value errors. Therein tuplewise convergence factors enables improving comparable bounds from [9] which use scalar convergence factors.

Bound (1.2) and its upgrades cannot be applied repeatedly as needed for restarted block eigensolvers, since the left- and right-hand sides of the bounds use different terms. Alternative repeatable bounds that are angle-free and depend only on the errors of the Ritz values have been pioneered for Hermitian eigenvalue problems in [4, (2.22)] but only for a single extreme Ritz value,

$$(\lambda_p - \eta'_p)/(\eta'_p - \lambda_{p+1}) \leq \varphi_p^2\,(\lambda_p - \eta_p)/(\eta_p - \lambda_{p+1}) \tag{1.5}$$

using the smallest (pth largest) Ritz value η_p in $\mathcal{Y}$ rather than the angle $\angle(\mathcal{X}, \mathcal{Y})$. We note that (1.5) additionally requires the assumptions $\eta_p > \lambda_{p+1}$ and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_p)|$. Correspondingly, the convergence factor φ_p turns into $\big(\max_{j\in\{p+1,\ldots,n\}} |f(\lambda_j)|\big)/|f(\lambda_p)|$.

Similar bounds for other Ritz values can be derived by adapting bound (1.5) to the partial iteration $\widetilde{\mathcal{Y}}'_i = f(A)\widetilde{\mathcal{Y}}_i$ of (1.1) where the subspace $\widetilde{\mathcal{Y}}_i$ is spanned by Ritz vectors associated with the i largest Ritz values in $\mathcal{Y}$. This yields

$$(\lambda_i - \eta'_i)/(\eta'_i - \lambda_{i+1}) \leq \big((\max_{j\in\{i+1,\ldots,n\}}|f(\lambda_j)|)/|f(\lambda_i)|\big)^2\,(\lambda_i - \eta_i)/(\eta_i - \lambda_{i+1}) \tag{1.6}$$

with an alternative convergence factor. A specified form of (1.6) for the block steepest ascent method for Hermitian eigenvalue problems appears in [3, (2.5.8)]. However, the convergence factor in (1.6) could be close to 1 for clustered eigenvalues and in the case $i = 1$ simply turns into a convergence factor for vector iterations. Thus (1.6) is unable to reflect cluster robustness of block eigensolvers.

Repeatable angle-free bounds involving exclusively errors in Ritz values, such as (1.5) and (1.6), are naturally used to analyze solvers that optimize the Ritz values, e.g., the (block) Lanczos method. We aim to upgrade these bounds in the present work. Repeatable angle-based bounds for Ritz vectors in such solvers are apparently only known for single-vector versions; see [8, (2.2)] for the steepest ascent method and [11, (3.3)] for the restarted Lanczos method. Deriving sharp bounds for Ritz vectors in block solvers is one of our future topics.

1.2. Block Lanczos method

One can interpret iteration (1.1) as an A-dependent filter on the subspace $\mathcal{Y}$ aiming in our case to move $\mathcal{Y}$ closer to the A-invariant subspace $\mathcal{X}$, as can be seen, e.g., from bound (1.4). If the function $f(\cdot)$ is explicitly given and does not depend on $\mathcal{Y}$, the corresponding convergence factors can be easily explicitly derived. For example, assuming that A is positive semi-definite, the filter $f(A) = A^k$ by the k-step power method trivially gives $\varphi_i = \lambda_{p+1}^k/\lambda_i^k$ in (1.3); cf. [13, Section 2]. Similar results for indefinite A are available in [15, Section V] and [2, Subsection 5.1]. A basic proof technique therein is to skip several eigenvalues by using the orthogonal complement of the associated invariant subspace.

Advanced eigensolvers, such as the block Lanczos method, perform an extra step of determining the output subspace $\mathcal{Y}'$ via Rayleigh-Ritz optimization, in our case maximizing the Ritz values on $f(A)\mathcal{Y}$ with respect to arbitrary $f(\cdot)$. While the combined procedure may still be technically expressed by formula (1.1), the underlying optimal $f(\cdot)$ now depends on the input subspace $\mathcal{Y}$, i.e., can be viewed as a nonlinear implicitly defined filter, and a sharp convergence

factor φ_i may be difficult to determine. Thus, instead of specifying bounds like (1.2) and (1.5) directly with an unknown optimal $f(\cdot)$, one first uses a surrogate function as $f(\cdot)$, commonly based on Chebyshev polynomials, with explicitly known convergence factors φ_i in (1.2) and (1.5). Then final bounds follow from the optimality of the Ritz values.

In the block Lanczos method, with any polynomial of degree k as $f(\cdot)$, the subspace $\mathcal{Y}'$ from (1.1) is a subset of the block Krylov subspace $\mathcal{K} = \mathcal{Y} + A\mathcal{Y} + \cdots + A^k\mathcal{Y}$. Then the ith largest Ritz value ψ_i in $\mathcal{K}$ is not smaller than η_i' according to the Courant-Fischer principles. A typical choice of the surrogate function to be used for determination of the convergence factors is

$$f(\alpha) = T_k\left(1 + 2\,\frac{\alpha - \lambda_{p+1}}{\lambda_{p+1} - \lambda_n}\right) \tag{1.7}$$

with the Chebyshev polynomial T_k of the first kind. Then the convergence factor φ_i in (1.2) and (1.5) becomes

$$\sigma_i = \left[T_k\left(1 + 2\,\frac{\lambda_i - \lambda_{p+1}}{\lambda_{p+1} - \lambda_n}\right)\right]^{-1}. \tag{1.8}$$

This leads to the following bounds proposed in [3, Section 2.6]:

$$(\lambda_i - \psi_i)/(\psi_i - \lambda_n) \le \sigma_i^2 \tan^2 \angle(\mathcal{X}, \mathcal{Y}), \tag{1.9}$$

$$(\lambda_p - \psi_p)/(\psi_p - \lambda_{p+1}) \le \sigma_p^2\,(\lambda_p - \eta_p)/(\eta_p - \lambda_{p+1}). \tag{1.10}$$

The majorization-type bounds from [17] can be specified in a similar way.

The convergence factor σ_i defined in (1.8) decreases asymptotically with $k \to \infty$ as a geometric progression with the reduction rate

$$\frac{\sqrt{\kappa_i} - 1}{\sqrt{\kappa_i} + 1} \quad \text{for} \quad \kappa_i = \frac{\lambda_i - \lambda_n}{\lambda_i - \lambda_{p+1}}.$$

Therefore (1.9) or (1.10) can predict a rapid convergence of $\psi_i \to \lambda_i$ or $\psi_p \to \lambda_p$ provided that the gap between λ_p and λ_{p+1} is sufficiently large.

In the particular case $k = 2$, the block Lanczos method is reduced to one step of the block steepest ascent method. The convergence factor σ_i becomes (see [3, Section 2.5])

$$\frac{\kappa_i - 1}{\kappa_i + 1} \quad \text{or} \quad \frac{1 - \xi_i}{1 + \xi_i} \quad \text{for} \quad \xi_i = \frac{\lambda_i - \lambda_{p+1}}{\lambda_i - \lambda_n}.$$

Bound (1.9) improves the previously known result [14, Theorem 6] (with different notations of indices) in two regards:

(i) The left-hand side is not smaller than the measure $(\lambda_i - \psi_i)/(\lambda_i - \lambda_n)$ used in [14] (which is more suitable for sine-based bounds rather than tangent-based bounds).
(ii) The right-hand side does not contain critical terms which depend on Ritz values in $\mathcal{K}$ and could deteriorate the bound for clustered eigenvalues.

Nevertheless, bound (1.9) is less suitable for restarted eigensolvers in contrast to angle-free bounds like (1.10) pioneered in [3, 4] – the focus of the present investigation.

1.3. Restarted block Lanczos method

In practice, the block Lanczos method needs to be restarted for ensuring numerical stability and reasonable storage requirements. The corresponding block Krylov subspace is extended up to a fixed index k, and then reduced to a subspace spanned by selected Ritz vectors, e.g.,

$$\mathcal{Y}^{(\ell+1)} \quad \xleftarrow{\mathrm{RR}[p]} \quad \mathcal{Y}^{(\ell)} + A\mathcal{Y}^{(\ell)} + \cdots + A^k\mathcal{Y}^{(\ell)} \tag{1.11}$$

where the Rayleigh-Ritz procedure RR[p] extracts Ritz vectors associated with the p largest Ritz values. Bounds (1.9) and (1.10) can easily be adapted to (1.11) as single-step Ritz value bounds

$$(\lambda_i - \psi_i^{(\ell+1)})/(\psi_i^{(\ell+1)} - \lambda_n) \le \sigma_i^2 \tan^2 \angle(\mathcal{X}, \mathcal{Y}^{(\ell)}) \tag{1.12}$$

and

$$(\lambda_p - \psi_p^{(\ell+1)})/(\psi_p^{(\ell+1)} - \lambda_{p+1}) \le \sigma_p^2\, (\lambda_p - \psi_p^{(\ell)})/(\psi_p^{(\ell)} - \lambda_{p+1}). \tag{1.13}$$

We note that (1.12) cannot directly describe the convergence behavior of the Ritz value sequence $(\psi_i^{(\ell)})_{\ell\in\mathbb{N}}$. A possible way out is extending (1.12) in multiple steps. This requires additional bounds for $\angle(\mathcal{X}, \mathcal{Y}^{(\ell+1)})$ in terms of $\psi_i^{(\ell+1)}$. One such bound following [4, pp. 382] is

$$\sin^2 \angle(\mathcal{X}, \mathcal{Y}^{(\ell+1)}) \le (\lambda_1 - \psi_p^{(\ell+1)})/(\lambda_1 - \lambda_{p+1})$$

which however could cause an overestimation, especially in the case $\psi_p^{(\ell+1)} \le \lambda_{p+1}$. Similarly, a more accurate approach based on the intermediate bound (1.4) requires a reasonable (and rather challenging) comparison of the angles $\angle\big(\mathcal{X}_i, f(A)\mathcal{Y}_i^{(\ell)}\big)$ and $\angle\big(\mathcal{X}_i, \mathcal{Y}_i^{(\ell+1)}\big)$. Ritz value bounds from [9, 14] and majorization-type bounds from [17] also depend on angles and cannot easily be adapted to multiple steps. The angle-dependence is thus an obstacle for deriving direct Ritz value bounds for the restarted block Lanczos method or other similar methods.

In contrast, angle-free bound (1.13) clearly demonstrates that $(\psi_p^{(\ell)})_{\ell\in\mathbb{N}}$ converges with respect to the measure $(\lambda_p - \psi)/(\psi - \lambda_{p+1})$. A repeated application of (1.13) leads to the a priori bound

$$(\lambda_p - \psi_p^{(\ell)})/(\psi_p^{(\ell)} - \lambda_{p+1}) \le \sigma_p^{2\ell}\, (\lambda_p - \psi_p^{(0)})/(\psi_p^{(0)} - \lambda_{p+1}).$$

Moreover, (1.13) can be generalized to $(\psi_i^{(\ell)})_{\ell\in\mathbb{N}}$ for $i \in \{1, \dots, p\}$ as in [16, Section 3], namely,

$$(\lambda_t - \psi_i^{(\ell+1)})/(\psi_i^{(\ell+1)} - \lambda_{t+1}) \le \widetilde{\sigma}_t^2\, (\lambda_t - \psi_i^{(\ell)})/(\psi_i^{(\ell)} - \lambda_{t+1}). \tag{1.14}$$

Therein $\psi_i^{(\ell)}$ is assumed to be located in an arbitrary eigenvalue interval $(\lambda_{t+1}, \lambda_t)$ for $t \ge i$, and the convergence factor is defined by

$$\widetilde{\sigma}_t = \left[T_k\left(1 + 2\,\frac{\lambda_t - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-1}.$$

Bound (1.14) for $t = i$ simply follows from (1.6). Nevertheless, (1.14) is rather appropriate for well-separated eigenvalues on account of the gap between the consecutive eigenvalues λ_t and λ_{t+1}. Generalizing (1.13) with convergence factors like (1.8) is desirable for interpreting the practically observed cluster robustness of the restarted block Lanczos method.

1.4. Aim and outline

In this paper, we generally consider the abstract block iteration (1.1) and extend the angle-free Ritz value bound (1.5) to arbitrarily located Ritz values. The error of a Ritz value ψ is measured by $(\lambda_t - \psi)/(\psi - \lambda_{t+1})$ or $(\lambda_{t-p+i} - \psi)/(\psi - \lambda_{t+1})$. The former is related to bounds (1.6) and (1.14) mentioned above, whereas the latter is suitable for deriving convergence factors which are independent of the eigenvalues $\lambda_{t-p+i+1}, \dots, \lambda_t$. In the special case $t = p$, these irrelevant eigenvalues read $\lambda_{i+1}, \dots, \lambda_p$, i.e., those skipped in convergence factors (1.3) and (1.8).

Angle-free Ritz value bounds using such measures are also available in our investigation of block preconditioned gradient-type eigensolvers; cf. [19, 20]. These methods can be regarded as perturbed versions of the block power method or a restarted block Krylov subspace iteration of degree 2. Some proof techniques in [19, 20] are motivated by analyzing the case of exact shift-inverse preconditioning in connection with special forms of (1.1). The present paper aims at generalizing this approach to (1.1) as a basis for upgrading the analysis of more efficient block preconditioned eigensolvers including LOBPCG [5] and the block Davidson method [10].

Preparing for our main analysis, we formulate in Section 2 two important arguments introducing decisive auxiliary vectors and intermediate bounds. By combining these arguments in Section 3, we get desirable extensions of (1.5). Therein a partial iteration of (1.1) is constructed with certain subspaces orthogonal to the invariant subspace associated with $\lambda_{i+1}, \ldots, \lambda_p$ for arbitrary $i \in \{1, \ldots, p\}$. The approximation of λ_i by Ritz values from this partial iteration is analyzed analogously to (1.5) with a novel accuracy measure for block eigensolvers that allows deriving repeatable cluster-friendly bounds extending (1.5) to all Ritz values. Specifically, under the same assumptions as for (1.5) concerning the final phase of a restarted block eigensolver, we derive in Theorem 3.2 the repeatable bound

$$(\lambda_i - \widetilde{\eta}_i')/(\widetilde{\eta}_i' - \lambda_{p+1}) \leq \frac{\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|^2}{|f(\lambda_i)|^2}\ (\lambda_i - \widetilde{\eta}_i)/(\widetilde{\eta}_i - \lambda_{p+1})$$

for the ith largest Ritz values $\widetilde{\eta}_i$ and $\widetilde{\eta}_i'$ generated by the partial iteration. Theorem 3.4 utilizes a similar partial iteration within the orthogonal complement of the invariant subspace associated with $\lambda_{t-p+i+1}, \ldots, \lambda_t$, and generalizes the above bound to an arbitrary outer step of a restarted block eigensolver,

$$(\lambda_{t-p+i} - \widetilde{\eta}_i')/(\widetilde{\eta}_i' - \lambda_{t+1}) \leq \frac{\max_{j\in\{t+1,\ldots,n\}}|f(\lambda_j)|^2}{|f(\lambda_{t-p+i})|^2}\ (\lambda_{t-p+i} - \widetilde{\eta}_i)/(\widetilde{\eta}_i - \lambda_{t+1}).$$

These repeatable bounds enable explicit multi-step error estimation for various restarted block eigensolvers. Therein convergence factors with nonconsecutive eigenvalues are preserved and improve bounds like (1.6) in the case of clustered eigenvalues.

Section 4 introduces applications to the restarted block Lanczos method and its modifications concerning shift-and-invert and deflation. Section 5 presents numerical examples for demonstrating our new results. Some detailed proofs are given in Section 6 (Appendix).

Frequently used symbols and settings are collected in Lemma 2.1.

2. Preliminaries

A remarkable argument for deriving the angle-free Ritz value bound (1.5) is concerned with inequalities of the Rayleigh quotient where two vectors are compared with respect to their orthogonal projections on eigenspaces [3, Lemma 2.3.2] by Knyazev. Therewith certain auxiliary vectors can be constructed in low-dimensional subspaces and produce an appropriate intermediate bound. In Subsection 2.1, we introduce this argument within a modified derivation of (1.5) for preparing extensions of (1.5). A further argument introduced in Subsection 2.2 is related to biorthogonal vectors originally used in the analysis of the block power method by Rutishauser [13]. We construct similar vectors for skipping irrelevant eigenvalues mentioned in Subsection 1.4.

2.1. Derivation of a basic angle-free bound

As the starting point of our analysis, we modify the derivation of the known bound (1.5) in a style that facilitates extensions. For the reader's convenience, we recall some basic settings.

Lemma 2.1. *Consider the abstract block iteration* (1.1)*, i.e.,* $\mathcal{Y}' = f(A)\mathcal{Y}$ *for a Hermitian matrix* $A \in \mathbb{C}^{n\times n}$*. The eigenvalues of* A *are arranged as* $\lambda_1 \geq \cdots \geq \lambda_n$*, and* $x_1, \ldots, x_n$ *are associated orthonormal eigenvectors. Let* $Y \in \mathbb{C}^{n\times p}$ *be an arbitrary basis matrix of the subspace* $\mathcal{Y}$*. The Ritz values of* A *in* $\mathcal{Y}$ *are arranged as* $\eta_1 \geq \cdots \geq \eta_p$*. If* $\eta_p > \lambda_{p+1}$*, and* $f(\lambda_j) \neq 0$ *for each* $j \in \{1, \ldots, p\}$*, then* $f(A)Y$ *is a basis matrix of* $\mathcal{Y}'$*, and* $\dim\mathcal{Y}' = p$*.*

Proof. → Subsection 6.1.

Subsequently, we reformulate [3, Lemma 2.3.2] for deriving intermediate bounds.

Lemma 2.2. *With the settings from Lemma 2.1, denote by $\rho(\cdot)$ the Rayleigh quotient with respect to A. Then the following statements hold for arbitrary vectors $u, v \in \mathbb{C}^n \backslash \{0\}$ where u satisfies $\lambda_l \geq \rho(u) \geq \lambda_{l+1}$.*

(a) If $|x_j^H v| \geq |x_j^H u|$ $\forall$ $j \in \{1, \ldots, l\}$ and $|x_j^H v| \leq |x_j^H u|$ $\forall$ $j \in \{l+1, \ldots, n\}$, then $\rho(v) \geq \rho(u)$.

(b) If $|x_j^H v| \leq |x_j^H u|$ $\forall$ $j \in \{1, \ldots, l\}$ and $|x_j^H v| \geq |x_j^H u|$ $\forall$ $j \in \{l+1, \ldots, n\}$, then $\rho(u) \geq \rho(v)$.

We omit the proof of Lemma 2.2 and refer the reader to [11, Lemma 3.2]. Although the eigenvalues are assumed to be simple in [3, 11], the proof therein does not utilize strict inequalities in $\lambda_l \geq \rho(u) \geq \lambda_{l+1}$ and is thus compatible with multiple eigenvalues.

Lemma 2.2 motivates some auxiliary vectors by reweighting projections of a Ritz vector with respect to eigenvectors.

Lemma 2.3. *Following Lemma 2.1, an arbitrary Ritz vector y' associated with the smallest (pth largest) Ritz value η_p' in $\mathcal{Y}'$ can be represented by $y' = f(A)Yg$ with a certain $g \in \mathbb{C}^p \backslash \{0\}$. Assume in addition $\lambda_p > \eta_p'$ and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_p)| > \nu_p$ for $\nu_p = \max_{j \in \{p+1, \ldots, n\}} |f(\lambda_j)|$. Denote by $\rho(\cdot)$ the Rayleigh quotient with respect to A. Then*

$$\lambda_p > \eta_p' = \rho(y') \geq \rho(y^\circ) \geq \rho(y) \geq \eta_p > \lambda_{p+1} \tag{2.1}$$

holds for $y = Yg$ and $y^\circ = f(\lambda_p) \sum_{j=1}^p x_j x_j^H y + \nu_p \sum_{j=p+1}^n x_j x_j^H y$. Moreover,

$$\frac{\rho(y_p) - \rho(y^\circ)}{\rho(y^\circ) - \rho(y - y_p)} = \frac{\nu_p^2}{|f(\lambda_p)|^2} \frac{\rho(y_p) - \rho(y)}{\rho(y) - \rho(y - y_p)} \tag{2.2}$$

holds for $y_p = \sum_{j=1}^p x_j x_j^H y$.

Proof. → Subsection 6.2.

Now the derivation of (1.5) can be completed by combining intermediate bounds.

Theorem 2.4. *Consider the abstract block iteration* (1.1)*. The eigenvalues of A are arranged as $\lambda_1 \geq \cdots \geq \lambda_n$, and the Ritz values of A in $\mathcal{Y}$ as $\eta_1 \geq \cdots \geq \eta_p$ for $p = \dim \mathcal{Y}$. If $\eta_p > \lambda_{p+1}$, and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_p)| > \max_{j \in \{p+1, \ldots, n\}} |f(\lambda_j)|$, then $\dim \mathcal{Y}' = p$, and the smallest (pth largest) Ritz value η_p' in $\mathcal{Y}'$ fulfills* (1.5)*, i.e.,*

$$\frac{\lambda_p - \eta_p'}{\eta_p' - \lambda_{p+1}} \leq \frac{\max_{j \in \{p+1, \ldots, n\}} |f(\lambda_j)|^2}{|f(\lambda_p)|^2} \frac{\lambda_p - \eta_p}{\eta_p - \lambda_{p+1}}.$$

Proof. → Subsection 6.3.

Remark 2.1. *The above derivation of* (1.5) *is based on [3, pp. 79–83], but provides more details which are crucial for extending* (1.5) *and deriving more flexible angle-free Ritz value bounds. In our main analysis, this approach is adapted to invariant subspaces which enclose appropriate subsets of the iterative subspaces from* (1.1)*. The orthogonal complements of such invariant subspaces are associated with irrelevant eigenvalues mentioned in Subsection 1.4.*

2.2. Skipping eigenvalues

The auxiliary vectors introduced in Lemma 2.3 are related to the convergence measure $(\lambda_p - \psi)/(\psi - \lambda_{p+1})$ with two consecutive eigenvalues. This fact inspires an analogous approach within the invariant subspace $\operatorname{span}\{x_{i+1}, \ldots, x_p\}^\perp$. Therein λ_i and λ_{p+1} become neighboring eigenvalues so that intermediate bounds like (2.2) enable bounds for the ith largest Ritz value

with respect to $(\lambda_i - \psi)/(\psi - \lambda_{p+1})$. For this purpose, we introduce an appropriate subset of the subspace $\mathcal{Y}$ from (1.1).

Lemma 2.5. *With the settings from Lemma 2.1, the intersection $\mathcal{Y}_{[1,i]\cup(p,n]} = \mathcal{X}_{[1,i]\cup(p,n]} \cap \mathcal{Y}$ of the invariant subspace $\mathcal{X}_{[1,i]\cup(p,n]} = \operatorname{span}\{x_{i+1}, \ldots, x_p\}^\perp$ and $\mathcal{Y}$ has at least dimension i. If $\eta_p > \lambda_{p+1}$, then $\dim\mathcal{Y}_{[1,i]\cup(p,n]} = i$, and the orthogonal projection of $\mathcal{Y}_{[1,i]\cup(p,n]}$ on the invariant subspace $\mathcal{X} = \operatorname{span}\{x_1, \ldots, x_p\}$ coincides with $\operatorname{span}\{x_1, \ldots, x_i\}$. In addition, if $f(\lambda_j) \neq 0$ for each $j \in \{1, \ldots, i\}$, then $\mathcal{Y}'_{[1,i]\cup(p,n]} = f(A)\mathcal{Y}_{[1,i]\cup(p,n]}$ also has dimension i.*

Proof. → Subsection 6.4.

Remark 2.2. *Lemma 2.5 suggests a partial iteration $\mathcal{Y}'_{[1,i]\cup(p,n]} = f(A)\mathcal{Y}_{[1,i]\cup(p,n]}$ of (1.1) where the block size is i and the target invariant subspace is $\operatorname{span}\{x_1, \ldots, x_i\}$. By adapting the derivation of (1.5) to this partial iteration, we get a bound for the smallest (ith largest) Ritz value in $\mathcal{Y}'_{[1,i]\cup(p,n]}$. Subsequent simple transformations lead to a cluster robust convergence factor for (1.1) where the eigenvalues $\lambda_{i+1}, \ldots, \lambda_p$ are skipped.*

The property $\dim\mathcal{Y}_{[1,i]\cup(p,n]} = i$ can also be ensured by assuming $\angle(\mathcal{X}, \mathcal{Y}) < \pi/2$ as in derivations of angle-dependent bounds [3, 9] and majorization-type bounds [6, 17]. This assumption is evidently equivalent to the invertibility of $X^H Y$ in the proof of Lemma 2.5. Moreover, the columns of $Y(X^H Y)^{-1}$ correspond to auxiliary vectors utilized in [13] for analyzing the block power method.

In an analogous analysis concerning the location $\eta_p > \lambda_{t+1}$ for a certain $t \geq p$, we use $\widetilde{\mathcal{X}} = \operatorname{span}\{x_{t-p+i+1}, \ldots, x_t\}^\perp$ instead of $\mathcal{X}_{[1,i]\cup(p,n]}$. Then $\dim(\widetilde{\mathcal{X}} \cap \mathcal{Y}) \geq i$ also holds, but the case $\dim(\widetilde{\mathcal{X}} \cap \mathcal{Y}') > i$ cannot easily be excluded by assumptions on Ritz values or angles. Correspondingly, we consider an arbitrary i-dimensional subset of $\widetilde{\mathcal{X}} \cap \mathcal{Y}$ in the further analysis.

3. Main results

The angle-free Ritz value bound (1.5) (with the settings from Theorem 2.4) deals with the pth largest Ritz value η'_p in $\mathcal{Y}'$ located in the eigenvalue interval $(\lambda_{p+1}, \lambda_p)$. Our main results are three extensions of (1.5) concerning the ith largest Ritz value η'_i for an arbitrary $i \leq p$. The corresponding eigenvalue intervals are $(\lambda_{p+1}, \lambda_i)$, $(\lambda_{t+1}, \lambda_{t-p+i})$ for $t \geq p$, and $(\lambda_{t+1}, \lambda_t)$ for $t \geq i$, respectively.

3.1. Bound depending on nonconsecutive eigenvalues in the final phase

The first extension of (1.5) is directly based on Lemma 2.5. Therein the assumption $\eta_p > \lambda_{p+1}$ is related to the final phase of a restarted block eigensolver. We adapt the derivation of (1.5) to the partial iteration $\mathcal{Y}'_{[1,i]\cup(p,n]} = f(A)\mathcal{Y}_{[1,i]\cup(p,n]}$ within the invariant subspace $\operatorname{span}\{x_{i+1}, \ldots, x_p\}^\perp$.

Similarly to Lemma 2.3, we begin with a Ritz vector in $\mathcal{Y}'_{[1,i]\cup(p,n]}$ and construct auxiliary vectors for producing intermediate bounds.

Lemma 3.1. *Following Lemma 2.5, an arbitrary Ritz vector $\widetilde{y}'$ associated with the smallest (ith largest) Ritz value $\widetilde{\eta}'_i$ in $\mathcal{Y}'_{[1,i]\cup(p,n]}$ can be represented by $\widetilde{y}' = f(A)\widetilde{y}$ with a certain $\widetilde{y} \in \mathcal{Y}_{[1,i]\cup(p,n]} \backslash \{0\}$. Assume in addition $\lambda_i > \widetilde{\eta}'_i$ and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_i)| > \nu_p$ for $\nu_p = \max_{j\in\{p+1,\ldots,n\}} |f(\lambda_j)|$. Denote by $\rho(\cdot)$ the Rayleigh quotient with respect to A, and define*

$$\widetilde{y}_i = \textstyle\sum_{j=1}^{i} x_j x_j^H \widetilde{y}, \qquad \widetilde{y}^\circ = f(\lambda_i)\widetilde{y}_i + \nu_p(\widetilde{y} - \widetilde{y}_i).$$

Then it holds that

$$\lambda_i > \widetilde{\eta}_i' = \rho(\widetilde{y}') \geq \rho(\widetilde{y}^\circ) \geq \rho(\widetilde{y}) \geq \widetilde{\eta}_i > \lambda_{p+1} \tag{3.1}$$

where $\widetilde{\eta}_i$ is the smallest (ith largest) Ritz value in $\mathcal{Y}_{[1,i]\cup(p,n]}$. Moreover,

$$\frac{\rho(\widetilde{y}_i) - \rho(\widetilde{y}^\circ)}{\rho(\widetilde{y}^\circ) - \rho(\widetilde{y} - \widetilde{y}_i)} = \frac{\nu_p^2}{|f(\lambda_i)|^2} \frac{\rho(\widetilde{y}_i) - \rho(\widetilde{y})}{\rho(\widetilde{y}) - \rho(\widetilde{y} - \widetilde{y}_i)}. \tag{3.2}$$

Proof. → Subsection 6.5.

Lemma 3.1 enables a generalization of Theorem 2.4 which provides angle-free bounds for the smallest Ritz value in $\mathcal{Y}'_{[1,i]\cup(p,n]}$ and the ith largest Ritz value in $\mathcal{Y}'$.

Theorem 3.2. *Consider the abstract block iteration* (1.1)*. The eigenvalues of A are arranged as $\lambda_1 \geq \cdots \geq \lambda_n$, and $x_1, \ldots, x_n$ are associated orthonormal eigenvectors. Define $\mathcal{X}_{[1,i]\cup(p,n]} = \operatorname{span}\{x_{i+1}, \ldots, x_p\}^\perp$ for $p = \dim\mathcal{Y}$ and $i \in \{1, \ldots, p\}$ (including $\mathcal{X}_{[1,i]\cup(p,n]} = \mathbb{C}^n$ for $i = p$). Let $\mathcal{Y}_{[1,i]\cup(p,n]} = \mathcal{X}_{[1,i]\cup(p,n]} \cap \mathcal{Y}$, and $\mathcal{Y}'_{[1,i]\cup(p,n]} = f(A)\mathcal{Y}_{[1,i]\cup(p,n]}$.*

Assume that the smallest (pth largest) Ritz value η_p in $\mathcal{Y}$ is larger than λ_{p+1}, and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_p)| > \max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|$. Then $\dim\mathcal{Y}' = p$, and $\dim\mathcal{Y}_{[1,i]\cup(p,n]} = \dim\mathcal{Y}'_{[1,i]\cup(p,n]} = i$. Moreover, it holds that

$$\frac{\lambda_i - \widetilde{\eta}_i'}{\widetilde{\eta}_i' - \lambda_{p+1}} \leq \frac{\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|^2}{|f(\lambda_i)|^2} \frac{\lambda_i - \widetilde{\eta}_i}{\widetilde{\eta}_i - \lambda_{p+1}} \tag{3.3}$$

for the smallest (ith largest) Ritz values $\widetilde{\eta}_i$ and $\widetilde{\eta}_i'$ in $\mathcal{Y}_{[1,i]\cup(p,n]}$ and $\mathcal{Y}'_{[1,i]\cup(p,n]}$. Consequently, the ith largest Ritz value η_i' in $\mathcal{Y}'_{[1,i]\cup(p,n]}$ fulfills

$$\frac{\lambda_i - \eta_i'}{\eta_i' - \lambda_{p+1}} \leq \frac{\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|^2}{|f(\lambda_i)|^2} \frac{\lambda_i - \eta_p}{\eta_p - \lambda_{p+1}}. \tag{3.4}$$

Proof. The assumptions $\eta_p > \lambda_{p+1}$ and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_p)| > \max_{j\in\{p+1,\ldots,n\}} |f(\lambda_j)|$ ensure the dimension statements according to Theorem 2.4 and Lemma 2.5.

Bound (3.3) is trivial for $\lambda_i = \widetilde{\eta}_i'$. In the nontrivial case $\lambda_i > \widetilde{\eta}_i'$, Lemma 3.1 is applicable. Then (3.1) leads to

$$\rho(\widetilde{w}) \geq \lambda_i > \widetilde{\eta}_i' = \rho(\widetilde{y}') \geq \rho(\widetilde{y}^\circ) \geq \rho(\widetilde{y}) \geq \widetilde{\eta}_i > \lambda_{p+1} \geq \rho(\widetilde{z})$$

for $\widetilde{w} = \widetilde{y}_i$, $\widetilde{z} = \widetilde{y} - \widetilde{y}_i$. Combining this with (3.2) implies (3.3) analogously to (6.2).

Finally, (3.3) is extended as (3.4) by using $\lambda_i \geq \eta_i' \geq \widetilde{\eta}_i' \geq \widetilde{\eta}_i \geq \eta_p > \lambda_{p+1}$. □

The abstract block iteration (1.1) and the partial iteration $\mathcal{Y}'_{[1,i]\cup(p,n]} = f(A)\ \mathcal{Y}_{[1,i]\cup(p,n]}$ both produce subsets of the trial subspace of a restarted block eigensolver so that Theorem 3.2 gives upper bounds for the convergence rates of Ritz values therein. In particular, Theorem 3.2 indicates that the convergence rate of the ith largest Ritz value in the final phase does not depend on the possibly clustered eigenvalues $\lambda_{i+1}, \ldots, \lambda_p$.

Bounds (3.3) and (3.4) coincide with the known bound (1.5) in the special case $i = p$, and are thus extensions of (1.5) to $i \leq p$. They supplement (1.5) especially for a block size p with $\lambda_p \approx \lambda_{p+1}$ but $\lambda_i \gg \lambda_{p+1}$. Therein the convergence factor $\big(\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|^2\big)/|f(\lambda_i)|^2$ can be bounded away from 1. Correspondingly, it is meaningful to observe individual residual vectors instead of an entire residual matrix in the stopping criterion of a restarted block eigensolver.

3.2. Bound depending on nonconsecutive eigenvalues in general

In the second extension of (1.5), the assumption $\eta_p > \lambda_{p+1}$ is to be relaxed as $\eta_p > \lambda_{t+1}$ for $t \geq p$ so that the resulting bound is applicable to each outer step of a restarted block eigensolver. We first generalize Lemma 2.5 and Lemma 3.1 where the skipped eigenvalues $\lambda_{i+1}, \ldots, \lambda_p$ become $\lambda_{t-p+i+1}, \ldots, \lambda_t$, i.e., the indices are shifted by $t-p$.

Lemma 3.3. *With the settings from Lemma 2.1, denote for $t \in \{p, \ldots, n-1\}$ and $i \in \{1, \ldots, p\}$ the invariant subspace* $\operatorname{span}\{x_{t-p+i+1}, \ldots, x_t\}^\perp$ *by $\widetilde{\mathcal{X}}$. Then the intersection $\widetilde{\mathcal{Y}} = \widetilde{\mathcal{X}} \cap \mathcal{Y}$ has at least dimension i. Moreover, if $\eta_p > \lambda_{t+1}$, and $f(\lambda_j) \neq 0$ for each $j \in \{1, \ldots, t-p+i\}$, then $\widetilde{\mathcal{Y}}_i' = f(A)\widetilde{\mathcal{Y}}_i$ has dimension i for an arbitrary i-dimensional subspace $\widetilde{\mathcal{Y}}_i \subseteq \widetilde{\mathcal{Y}}$.*

Consequently, an arbitrary Ritz vector $\widetilde{y}'$ associated with the smallest (ith largest) Ritz value $\widetilde{\eta}_i'$ in $\widetilde{\mathcal{Y}}_i'$ can be represented by $\widetilde{y}' = f(A)\widetilde{y}$ with a certain $\widetilde{y} \in \widetilde{\mathcal{Y}}_i \backslash \{0\}$. Assume in addition $\lambda_{t-p+i} > \widetilde{\eta}_i'$ and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_{t-p+i})| > \nu_t$ for $\nu_t = \max_{j\in\{t+1,\ldots,n\}} |f(\lambda_j)|$. Denote by $\rho(\cdot)$ the Rayleigh quotient with respect to A, and define

$$\widetilde{w} = \textstyle\sum_{j=1}^{t-p+i} x_j x_j^H \widetilde{y}, \qquad \widetilde{z} = \sum_{j=t+1}^{n} x_j x_j^H \widetilde{y}, \qquad \widetilde{y}^\circ = f(\lambda_{t-p+i})\widetilde{w} + \nu_t \widetilde{z}.$$

Then it holds that

$$\lambda_{t-p+i} > \widetilde{\eta}_i' = \rho(\widetilde{y}') \geq \rho(\widetilde{y}^\circ) \geq \rho(\widetilde{y}) \geq \widetilde{\eta}_i > \lambda_{t+1} \tag{3.5}$$

where $\widetilde{\eta}_i$ is the smallest (ith largest) Ritz value in $\widetilde{\mathcal{Y}}_i$. Moreover,

$$\frac{\rho(\widetilde{w}) - \rho(\widetilde{y}^\circ)}{\rho(\widetilde{y}^\circ) - \rho(\widetilde{z})} = \frac{\nu_t^2}{|f(\lambda_{t-p+i})|^2} \frac{\rho(\widetilde{w}) - \rho(\widetilde{y})}{\rho(\widetilde{y}) - \rho(\widetilde{z})}. \tag{3.6}$$

Proof. → Subsection 6.6.

The subspaces $\widetilde{\mathcal{Y}}_i$ and $\widetilde{\mathcal{Y}}_i'$ from Lemma 3.3 motivate a partial iteration of (1.1) within $\widetilde{\mathcal{X}}$ due to $\widetilde{\mathcal{Y}}_i \subseteq \widetilde{\mathcal{X}} \cap \mathcal{Y}$ and $\widetilde{\mathcal{Y}}_i' \subseteq \widetilde{\mathcal{X}} \cap \mathcal{Y}'$. A corresponding upgrade of Theorem 3.2 provides angle-free bounds for the smallest Ritz value in $\widetilde{\mathcal{Y}}_i'$ and the ith largest Ritz value in $\mathcal{Y}'$ concerning the eigenvalue interval $(\lambda_{t+1}, \lambda_{t-p+i})$.

Theorem 3.4. *Consider the abstract block iteration* (1.1)*. The eigenvalues of A are arranged as $\lambda_1 \geq \cdots \geq \lambda_n$, and $x_1, \ldots, x_n$ are associated orthonormal eigenvectors. Define $\widetilde{\mathcal{Y}} = \widetilde{\mathcal{X}} \cap \mathcal{Y}$ with the invariant subspace $\widetilde{\mathcal{X}} = \operatorname{span}\{x_{t-p+i+1}, \ldots, x_t\}^\perp$ for $p = \dim\mathcal{Y}$, $t \in \{p, \ldots, n-1\}$ and $i \in \{1, \ldots, p\}$ (including $\widetilde{\mathcal{X}} = \mathbb{C}^n$ for $i = p$).*

Assume that the smallest (pth largest) Ritz value η_p in $\mathcal{Y}$ is larger than λ_{t+1}, and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_t)| > \max_{j\in\{t+1,\ldots,n\}} |f(\lambda_j)|$. Then $\dim\widetilde{\mathcal{Y}} \geq i$, and $\dim\widetilde{\mathcal{Y}}_i' = i$ holds for $\widetilde{\mathcal{Y}}_i' = f(A)\widetilde{\mathcal{Y}}_i$ with an arbitrary i-dimensional subspace $\widetilde{\mathcal{Y}}_i \subseteq \widetilde{\mathcal{Y}}$. Moreover, denote by $\widetilde{\eta}_i$ and $\widetilde{\eta}_i'$ the smallest (ith largest) Ritz values in $\widetilde{\mathcal{Y}}_i$ and $\widetilde{\mathcal{Y}}_i'$. Then either $\widetilde{\eta}_i' \geq \lambda_{t-p+i}$, or

$$0 < \frac{\lambda_{t-p+i} - \widetilde{\eta}_i'}{\widetilde{\eta}_i' - \lambda_{t+1}} \leq \frac{\max_{j\in\{t+1,\ldots,n\}} |f(\lambda_j)|^2}{|f(\lambda_{t-p+i})|^2} \frac{\lambda_{t-p+i} - \widetilde{\eta}_i}{\widetilde{\eta}_i - \lambda_{t+1}}. \tag{3.7}$$

Consequently, $\dim\mathcal{Y}' = p$, and the ith largest Ritz value η_i' in $\mathcal{Y}'$ fulfills either $\eta_i' \geq \lambda_{t-p+i}$ or

$$0 < \frac{\lambda_{t-p+i} - \eta_i'}{\eta_i' - \lambda_{t+1}} \leq \frac{\max_{j\in\{t+1,\ldots,n\}} |f(\lambda_j)|^2}{|f(\lambda_{t-p+i})|^2} \frac{\lambda_{t-p+i} - \eta_p}{\eta_p - \lambda_{t+1}}. \tag{3.8}$$

Proof. The dimension statements $\dim\widetilde{\mathcal{Y}} \geq i$ and $\dim\widetilde{\mathcal{Y}}_i' = i$ directly follow from Lemma 3.3.

Bound (3.7) is stated in the nontrivial case $\lambda_{t-p+i} > \widetilde{\eta}_i'$ which is also considered in Lemma 3.3. The intermediate results (3.5) and (3.6) imply (3.7) analogously to (6.2).

Furthermore, $\dim\mathcal{Y}' = p$ is indeed the special form of the statement $\dim\widetilde{\mathcal{Y}}_i' = i$ for $i = p$. Bound (3.8) follows from (3.7) according to $\lambda_{t-p+i} > \eta_i' \geq \widetilde{\eta}_i' \geq \widetilde{\eta}_i \geq \eta_p > \lambda_{t+1}$. □

Theorem 3.4 serves to discuss the global convergence behavior of a restarted block eigensolver. Provided that the block size p exceeds the size of each eigenvalue cluster, i.e., λ_{t-p+i} is not close to λ_{t+1}, the convergence factor $\left(\max_{j\in\{t+1,\dots,n\}}|f(\lambda_j)|^2\right)/|f(\lambda_{t-p+i})|^2$ with a properly defined $f(\cdot)$, e.g., a shifted Chebyshev polynomial, can well reflect cluster robustness.

3.3. Bound depending on consecutive eigenvalues

The above extensions of (1.5) give bounds (3.4) and (3.8) for the ith largest Ritz value η_i' in $\mathcal{Y}'$ in terms of the pth largest Ritz value η_p in $\mathcal{Y}$. The next extension deals with the relation between η_i' and η_i with respect to the eigenvalue interval $(\lambda_{t+1},\lambda_t)$ for $t\geq i$. The resulting bound includes (1.5) by setting $t=i=p$. We begin with the following auxiliary terms.

Lemma 3.5. *With the settings from Lemma 2.1, let $y_1,\dots,y_p$ be orthonormal Ritz vectors associated with the Ritz values $\eta_1\geq\cdots\geq\eta_p$ in $\mathcal{Y}$. Define $\widetilde{\mathcal{Y}}_i=\operatorname{span}\{y_1,\dots,y_i\}$ for $i\in\{1,\dots,p\}$. If $\eta_i>\lambda_{t+1}$ for a certain $t\in\{i,\dots,n-p+i-1\}$, and $f(\lambda_j)\neq 0$ for each $j\in\{1,\dots,t\}$, then $\widetilde{\mathcal{Y}}_i'=f(A)\widetilde{\mathcal{Y}}_i$ has dimension i.*

Consequently, an arbitrary Ritz vector $\widetilde{y}'$ associated with the smallest (ith largest) Ritz value $\widetilde{\eta}_i'$ in $\widetilde{\mathcal{Y}}_i'$ can be represented by $\widetilde{y}'=f(A)\widetilde{y}$ with a certain $\widetilde{y}\in\widetilde{\mathcal{Y}}_i\backslash\{0\}$. Assume in addition $\lambda_t>\widetilde{\eta}_i'$ and $|f(\lambda_1)|\geq\cdots\geq|f(\lambda_t)|>\nu_t$ for $\nu_t=\max_{j\in\{t+1,\dots,n\}}|f(\lambda_j)|$. Denote by $\rho(\cdot)$ the Rayleigh quotient with respect to A, and define

$$\widetilde{w}=\textstyle\sum_{j=1}^{t}x_jx_j^H\widetilde{y},\qquad \widetilde{z}=\sum_{j=t+1}^{n}x_jx_j^H\widetilde{y},\qquad \widetilde{y}^\circ=f(\lambda_t)\widetilde{w}+\nu_t\widetilde{z}.$$

Then it holds that

$$\lambda_t>\widetilde{\eta}_i'=\rho(\widetilde{y}')\geq\rho(\widetilde{y}^\circ)\geq\rho(\widetilde{y})\geq\eta_i>\lambda_{t+1}, \tag{3.9}$$

$$\frac{\rho(\widetilde{w})-\rho(\widetilde{y}^\circ)}{\rho(\widetilde{y}^\circ)-\rho(\widetilde{z})}=\frac{\nu_t^2}{|f(\lambda_t)|^2}\,\frac{\rho(\widetilde{w})-\rho(\widetilde{y})}{\rho(\widetilde{y})-\rho(\widetilde{z})}. \tag{3.10}$$

Proof. Despite the different definition of $\widetilde{\mathcal{Y}}_i$, we can formally reuse the proof of Lemma 3.3 from Subsection 6.6. A slight modification with the substitution $t-p+i\to t$ proves Lemma 3.5. □

In contrast to Lemma 3.3, the auxiliary subspaces $\widetilde{\mathcal{Y}}_i$ and $\widetilde{\mathcal{Y}}_i'$ from Lemma 3.5 do not build a meaningful partial iteration since $\widetilde{\mathcal{Y}}_i$ is spanned by Ritz vectors in $\mathcal{Y}$, but $\widetilde{\mathcal{Y}}_i'$ is not necessarily spanned by Ritz vectors in $\mathcal{Y}'$. For this reason, we omit in the following theorem a Ritz value bound like (3.7) concerning auxiliary subspaces.

Theorem 3.6. *Consider the abstract block iteration* (1.1)*. The eigenvalues of A are arranged as $\lambda_1\geq\cdots\geq\lambda_n$. Assume that the ith largest Ritz value η_i in $\mathcal{Y}$ is larger than λ_{t+1} for a certain $t\in\{i,\dots,n-p+i-1\}$, and $|f(\lambda_1)|\geq\cdots\geq|f(\lambda_t)|>\max_{j\in\{t+1,\dots,n\}}|f(\lambda_j)|$. Then $\dim\mathcal{Y}'\geq i$, and the ith largest Ritz value η_i' in $\mathcal{Y}'$ fulfills either $\eta_i'\geq\lambda_t$ or*

$$0<\frac{\lambda_t-\eta_i'}{\eta_i'-\lambda_{t+1}}\leq\frac{\max_{j\in\{t+1,\dots,n\}}|f(\lambda_j)|^2}{|f(\lambda_t)|^2}\,\frac{\lambda_t-\eta_i}{\eta_i-\lambda_{t+1}}. \tag{3.11}$$

Proof. The subspace $\widetilde{\mathcal{Y}}_i'$ defined in Lemma 3.5 is a subset of $\mathcal{Y}'$. Thus $\dim\mathcal{Y}'\geq\dim\widetilde{\mathcal{Y}}_i'=i$.

Bound (3.11) is stated in the nontrivial case $\lambda_t>\eta_i'$. Then $\lambda_t>\eta_i'\geq\widetilde{\eta}_i'$ allows applying (3.9) and (3.10) so that a bound for $\widetilde{\eta}_i'$ is obtained analogously to (6.2). This implies (3.11) according to $\lambda_t>\eta_i'\geq\widetilde{\eta}_i'>\lambda_{t+1}$. □

The convergence factor in (3.11) uses consecutive eigenvalues and is thus less appropriate for describing cluster robustness in comparison to the convergence factors in (3.4) and (3.8).

However, the term $(\lambda_t - \eta_i)/(\eta_i - \lambda_{t+1})$ can lead to a better bound in the first phase of a restarted block eigensolver, especially if the relevant eigenvalues are not tightly clustered.

4. Applications to restarted block eigensolvers

The angle-free Ritz value bounds from Section 3 can typically be applied to the restarted block Lanczos method by utilizing shifted Chebyshev polynomials as $f(\cdot)$ in the abstract block iteration (1.1). Results by this approach are presented in Subsection 4.1, beginning with bounds for one outer step which are comparable with some known angle-dependent bounds from [4, 14]. A decisive advantage of our angle-free bounds is that applying them to multiple outer steps does not require additional bounds for connecting successive steps, and thus avoids certain overestimations. Subsection 4.2 deals with the convergence analysis of restarted block eigensolvers with shift-and-invert, e.g., for computing eigenvalues of a self-adjoint elliptic partial differential operator. Therein our angle-free bounds can easily be adapted to a corresponding generalized eigenvalue problem. A related discussion on deflation is given in Subsection 4.3.

4.1. Application to the restarted block Lanczos method

Following Subsection 1.3, we represent the restarted block Lanczos method by (1.11), and observe the Ritz value sequence $(\psi_i^{(\ell)})_{\ell\in\mathbb{N}}$ where $\psi_i^{(\ell)}$ is the ith largest Ritz value in the ℓth iterative subspace $\mathcal{Y}^{(\ell)}$.

By regarding $\mathcal{Y}^{(\ell)}$ as $\mathcal{Y}$ in the abstract block iteration (1.1), one can select a real polynomial of degree k as $f(\cdot)$ so that $\mathcal{Y}'$ is a subset of $\mathcal{Y}^{(\ell+1)}$, and the ith largest Ritz value η_i' in $\mathcal{Y}'$ is a lower bound for $\psi_i^{(\ell+1)}$. Constructing a "sharp" $f(\cdot)$ that enables $\eta_i' = \psi_i^{(\ell+1)}$ involves certain interpolating polynomials which cannot easily be represented in an explicit form; cf. [18, Subsection 3.3]. In contrast to this, the standard approach from [3, 12, 14] for investigating the Lanczos method weakly minimizes the convergence factor concerning an eigenvalue interval. For instance, minimizing $\varphi_p = \big(\max_{j\in\{p+1,\dots,n\}}|f(\lambda_j)|\big)/|f(\lambda_p)|$ in (1.5) can be weakened as minimizing $\big(\max_{\lambda\in[\lambda_n,\lambda_{p+1}]}|f(\lambda)|\big)/|f(\lambda_p)|$. This results in the shifted Chebyshev polynomial (1.7); cf. [17, Lemma 4.1]. We reuse (1.7) together with its alternative

$$f(\alpha) = T_k\left(1 + 2\,\frac{\alpha - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right) \tag{4.1}$$

with $t \geq p$ or $t \geq i$ for specifying the new bounds from Section 3.

Theorem 4.1. *Consider the restarted block Lanczos method* (1.11) *with the block size p. The eigenvalues of A are arranged as $\lambda_1 \geq \cdots \geq \lambda_n$, and the Ritz values of A in $\mathcal{Y}^{(\ell)}$ as $\psi_1^{(\ell)} \geq \cdots \geq \psi_p^{(\ell)}$. Let T_k be the Chebyshev polynomial (of the first kind) of degree k, and $i \in \{1,\dots,p\}$.*

(a) If $\psi_p^{(\ell)} > \lambda_{t+1}$ for a certain $t \in \{p,\dots,n-1\}$, then either $\psi_i^{(\ell+1)} \geq \lambda_{t-p+i}$, or

$$0 < \frac{\lambda_{t-p+i} - \psi_i^{(\ell+1)}}{\psi_i^{(\ell+1)} - \lambda_{t+1}} \leq \left[T_k\left(1 + 2\,\frac{\lambda_{t-p+i} - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-2} \frac{\lambda_{t-p+i} - \psi_p^{(\ell)}}{\psi_p^{(\ell)} - \lambda_{t+1}}. \tag{4.2}$$

In the latter case, consider orthonormal eigenvectors $x_1,\dots,x_n$ associated with $\lambda_1,\dots,\lambda_n$. Then the subspace $\widetilde{\mathcal{Y}} = \operatorname{span}\{x_{t-p+i+1},\dots,x_t\}^{\perp} \cap \mathcal{Y}^{(\ell)}$ has at least dimension i. By using the smallest (ith largest) Ritz value $\widetilde{\eta}_i$ in an arbitrary i-dimensional subspace $\widetilde{\mathcal{Y}}_i \subseteq \widetilde{\mathcal{Y}}$, it holds that

$$0 < \frac{\lambda_{t-p+i} - \psi_i^{(\ell+1)}}{\psi_i^{(\ell+1)} - \lambda_{t+1}} \leq \left[T_k\left(1 + 2\,\frac{\lambda_{t-p+i} - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-2} \frac{\lambda_{t-p+i} - \widetilde{\eta}_i}{\widetilde{\eta}_i - \lambda_{t+1}}. \tag{4.3}$$

(b) If $\psi_i^{(\ell)} > \lambda_{t+1}$ for a certain $t \in \{i, \dots, n-p+i-1\}$, then either $\psi_i^{(\ell+1)} \geq \lambda_t$, or

$$0 < \frac{\lambda_t - \psi_i^{(\ell+1)}}{\psi_i^{(\ell+1)} - \lambda_{t+1}} \leq \left[T_k\left(1 + 2\,\frac{\lambda_t - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-2} \frac{\lambda_t - \psi_i^{(\ell)}}{\psi_i^{(\ell)} - \lambda_{t+1}}. \tag{4.4}$$

Proof. The statement (a) follows from Theorem 3.4. We set $\mathcal{Y} = \mathcal{Y}^{(\ell)}$ and define $f(\cdot)$ by (4.1). Then $\mathcal{Y}' \subseteq \mathcal{Y}^{(\ell+1)}$ so that $\eta_i' \leq \psi_i^{(\ell+1)}$. Moreover, the assumptions in Theorem 3.4 are fulfilled: $\eta_p = \psi_p^{(\ell)} > \lambda_{t+1}$, and $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_t)| > 1 = \max_{j\in\{t+1,\dots,n\}}|f(\lambda_j)|$. Specifying (3.8) implies (4.2) according to $\lambda_{t-p+i} > \psi_i^{(\ell+1)} \geq \eta_i' \geq \eta_p = \psi_p^{(\ell)} > \lambda_{t+1}$. An analogous specification of (3.7) leads to (4.3).

Similarly, the statement (b) is proved by Theorem 3.6 where bound (3.11) is specified. □

Theorem 4.1 extends the angle-free Ritz value bound (1.13) which focuses on the smallest Ritz value and specifies [4, (2.22)]. The statements deal with all Ritz values. In particular, (4.2) is comparable with existing angle-dependent bounds from [14, Theorem 6] and [4, (2.20)]. The angle-independence enables more accurate predictions in the case of large angle-dependent factors. Furthermore, the limitation of angle-dependent bounds for investigating multiple outer steps mentioned in Subsection 1.3 can be overcome by a direct generalization of Theorem 4.1.

Theorem 4.2. *With the settings from Theorem 4.1, the following statements hold for m outer steps of the restarted block Lanczos method* (1.11)*.*

(a) If $\psi_p^{(\ell)} > \lambda_{t+1}$ for a certain $t \in \{p, \dots, n-1\}$, then either $\psi_i^{(\ell+m)} \geq \lambda_{t-p+i}$, or

$$0 < \frac{\lambda_{t-p+i} - \psi_i^{(\ell+m)}}{\psi_i^{(\ell+m)} - \lambda_{t+1}} \leq \left[T_k\left(1 + 2\,\frac{\lambda_{t-p+i} - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-2m} \frac{\lambda_{t-p+i} - \psi_p^{(\ell)}}{\psi_p^{(\ell)} - \lambda_{t+1}}. \tag{4.5}$$

In the latter case, consider orthonormal eigenvectors $x_1, \dots, x_n$ associated with $\lambda_1, \dots, \lambda_n$. Then the subspace $\widetilde{\mathcal{Y}} = \operatorname{span}\{x_{t-p+i+1}, \dots, x_t\}^\perp \cap \mathcal{Y}^{(\ell)}$ has at least dimension i. By using the smallest (ith largest) Ritz value $\widetilde{\eta}_i$ in an arbitrary i-dimensional subspace $\widetilde{\mathcal{Y}}_i \subseteq \widetilde{\mathcal{Y}}$, it holds that

$$0 < \frac{\lambda_{t-p+i} - \psi_i^{(\ell+m)}}{\psi_i^{(\ell+m)} - \lambda_{t+1}} \leq \left[T_k\left(1 + 2\,\frac{\lambda_{t-p+i} - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-2m} \frac{\lambda_{t-p+i} - \widetilde{\eta}_i}{\widetilde{\eta}_i - \lambda_{t+1}}. \tag{4.6}$$

(b) If $\psi_i^{(\ell)} > \lambda_{t+1}$ for a certain $t \in \{i, \dots, n-p+i-1\}$, then either $\psi_i^{(\ell+m)} \geq \lambda_t$, or

$$0 < \frac{\lambda_t - \psi_i^{(\ell+m)}}{\psi_i^{(\ell+m)} - \lambda_{t+1}} \leq \left[T_k\left(1 + 2\,\frac{\lambda_t - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-2m} \frac{\lambda_t - \psi_i^{(\ell)}}{\psi_i^{(\ell)} - \lambda_{t+1}}. \tag{4.7}$$

Proof. For proving (a), we use the indexed form

$$\mathcal{Y}^{(\ell+1)} = f(A)\mathcal{Y}^{(\ell)} \tag{4.8}$$

of the abstract block iteration (1.1), and define $f(\cdot)$ by (4.1). According to the Courant-Fischer principles, (1.11) is a stepwisely accelerated version of (4.8) with respect to Ritz values. By setting the ℓth iterative subspace of (1.11) as $\mathcal{Y}^{(\ell)}$ in (4.8), we only need to verify (4.5) for (4.8). Therein we define $\widetilde{\mathcal{X}} = \operatorname{span}\{x_{t-p+i+1}, \dots, x_t\}^\perp$ and $\widetilde{\mathcal{Y}}^{(\ell)} = \widetilde{\mathcal{X}} \cap \mathcal{Y}^{(\ell)}$. The assumption $\psi_p^{(\ell)} > \lambda_{t+1}$ and the definition of $f(\cdot)$ allow applying Theorem 3.4 to $\mathcal{Y} = \mathcal{Y}^{(\ell)}$ which first shows $\dim\big(f(A)\widetilde{\mathcal{Y}}_i^{(\ell)}\big) = i$ for an arbitrary i-dimensional subspace $\widetilde{\mathcal{Y}}_i^{(\ell)} \subseteq \widetilde{\mathcal{Y}}^{(\ell)}$. Combining this with

$$f(A)\widetilde{\mathcal{Y}}_i^{(\ell)} \subseteq \big(f(A)\widetilde{\mathcal{X}}\big) \cap \big(f(A)\mathcal{Y}^{(\ell)}\big) \subseteq \widetilde{\mathcal{X}} \cap \mathcal{Y}^{(\ell+1)} = \widetilde{\mathcal{Y}}^{(\ell+1)}$$

ensures that $\widetilde{\mathcal{Y}}_i^{(\ell+1)} = f(A)\widetilde{\mathcal{Y}}_i^{(\ell)}$ is an i-dimensional subspace within $\widetilde{\mathcal{Y}}^{(\ell+1)}$ and defines a partial iteration of (4.8). In the nontrivial case $\lambda_{t-p+i} > \psi_i^{(\ell+m)}$ for the restarted block Lanczos method (1.11), we get $\lambda_{t-p+i} > \psi_i^{(\ell+m)} \geq \cdots \geq \psi_i^{(\ell+1)}$ so that λ_{t-p+i} is also larger than the

corresponding ith Ritz values produced by (4.8) and the partial iteration $\widetilde{\mathcal{Y}}_i^{(\ell+1)} = f(A)\widetilde{\mathcal{Y}}_i^{(\ell)}$. Then adapting (3.7) to the ith largest Ritz values $\widetilde{\eta}_i^{(\ell)}$ and $\widetilde{\eta}_i^{(\ell+1)}$ in $\widetilde{\mathcal{Y}}_i^{(\ell)}$ and $\widetilde{\mathcal{Y}}_i^{(\ell+1)}$ yields

$$0 < \frac{\lambda_{t-p+i} - \widetilde{\eta}_i^{(\ell+1)}}{\widetilde{\eta}_i^{(\ell+1)} - \lambda_{t+1}} \leq \left[T_k\left(1 + 2\,\frac{\lambda_{t-p+i} - \lambda_{t+1}}{\lambda_{t+1} - \lambda_n}\right)\right]^{-2} \frac{\lambda_{t-p+i} - \widetilde{\eta}_i^{(\ell)}}{\widetilde{\eta}_i^{(\ell)} - \lambda_{t+1}}.$$

A repeated application thereof leads to a multi-step bound concerning $\widetilde{\eta}_i^{(\ell)}$ and $\widetilde{\eta}_i^{(\ell+m)}$. Then (4.5) and (4.6) with $\widetilde{\eta}_i = \widetilde{\eta}_i^{(\ell)}$ are verified by considering $\lambda_{t-p+i} > \psi_i^{(\ell+m)} \geq \widetilde{\eta}_i^{(\ell+m)} \geq \widetilde{\eta}_i^{(\ell)} \geq \psi_p^{(\ell)} > \lambda_{t+1}$.

The statement (b) simply follows from Theorem 4.1 (b) by repeatedly applying (4.4) in the nontrivial case $\lambda_t > \psi_i^{(\ell+m)}$. □

Theorem 4.2 indicates that the Ritz value sequence $(\psi_i^{(\ell)})_{\ell\in\mathbb{N}}$ approaches or exceeds the right end of the considered eigenvalue interval after sufficiently many outer steps. Therein the convergence toward an interior eigenvalue λ_s with $s > i$ is not excluded, although rare in practice. In general, a strict increase $\psi_i^{(\ell+1)} > \psi_i^{(\ell)}$ occurs if the current iterative subspace $\mathcal{Y}^{(\ell)}$ contains no eigenvectors; cf. [16, Corollary 2] or a self-contained explanation in Subsection 6.7. Based on this fact, the case $\psi_p^{(\ell)} = \lambda_{t+1}$ or $\psi_i^{(\ell)} = \lambda_{t+1}$ which is not included in the assumption of Theorem 4.2 can be discussed as follows: If some Ritz vectors in $\mathcal{Y}^{(\ell)}$ are already eigenvectors, we can adapt Theorem 4.2 to a reduced iterative subspace spanned by other Ritz vectors. Otherwise $\psi_p^{(\ell+1)} > \psi_p^{(\ell)} = \lambda_{t+1}$ or $\psi_i^{(\ell+1)} > \psi_i^{(\ell)} = \lambda_{t+1}$ holds so that Theorem 4.2 is applicable after updating the index t.

4.2. Shift-and-invert

The analysis in Subsection 4.1 is concerned with a Hermitian matrix A and the computation of its largest eigenvalues. A trivial extension to computing the smallest eigenvalues can be made by the substitution $A \to -A$. Now we formulate further extensions starting with a generalized eigenvalue problem $Lv = \alpha Sv$ for Hermitian matrices $L, S \in \mathbb{C}^{n\times n}$ where S is positive definite.

A basic eigensolver for computing eigenvalues of (L, S) close to a noneigenvalue shift $\beta \in \mathbb{R}$ is the block shift-and-invert iteration

$$\mathcal{Z}^{(\ell+1)} = L_\beta^{-1} S \mathcal{Z}^{(\ell)} \quad \text{with} \quad L_\beta = L - \beta S \tag{4.9}$$

which is implemented by solving linear systems for L_β.

In particular, if β is smaller than the smallest eigenvalue of (L, S), then the shifted matrix L_β is positive definite so that (4.9) can be reformulated as

$$L_\beta^{1/2}\mathcal{Z}^{(\ell+1)} = \big(L_\beta^{-1/2} S L_\beta^{-1/2}\big) L_\beta^{1/2}\mathcal{Z}^{(\ell)}$$

which corresponds to the block power method for $A = L_\beta^{-1/2} S L_\beta^{-1/2}$. Moreover, the eigenvalues of (L, S) and A, arranged as $\alpha_1 \leq \cdots \leq \alpha_n$ and $\lambda_1 \geq \cdots \geq \lambda_n$, can be converted into each other by $\lambda_i = (\alpha_i - \beta)^{-1}$. The conversion between Ritz values is similar: the ith smallest Ritz value $\vartheta_i^{(\ell)}$ of (L, S) in $\mathcal{Z}^{(\ell)}$ and the ith largest Ritz value $\psi_i^{(\ell)}$ of A in $\mathcal{Y}^{(\ell)} = L_\beta^{1/2}\mathcal{Z}^{(\ell)}$ fulfill $\psi_i^{(\ell)} = (\vartheta_i^{(\ell)} - \beta)^{-1}$. Therewith bounds from Section 3 can be specified by $f(\alpha) = \alpha$ and then transformed for (4.9). For instance, (3.3) leads to a counterpart of (4.5) with the convergence factor $(\lambda_{p+1}/\lambda_i)^{2m}$ which is equivalent to

$$\frac{\vartheta_i^{(\ell+m)} - \alpha_i}{\alpha_{p+1} - \vartheta_i^{(\ell+m)}} \leq \left(\frac{\alpha_i - \beta}{\alpha_{p+1} - \beta}\right)^{2m} \frac{\vartheta_p^{(\ell)} - \alpha_i}{\alpha_{p+1} - \vartheta_p^{(\ell)}}. \tag{4.10}$$

Furthermore, a restarted version of the block Davidson method with L_β^{-1} as preconditioner can be represented by

$$\mathcal{Z}^{(\ell+1)} \quad \xleftarrow{\mathrm{RR}[L,S,p]} \quad \mathcal{Z}^{(\ell)} + L_\beta^{-1} S \mathcal{Z}^{(\ell)} + \cdots + (L_\beta^{-1} S)^k \mathcal{Z}^{(\ell)} \tag{4.11}$$

where the Rayleigh-Ritz procedure $\mathrm{RR}[L,S,p]$ extracts Ritz vectors associated with the p smallest Ritz values of (L,S). As (4.11) is equivalent to the restarted block Lanczos method (1.11) for $A = L_\beta^{-1/2} S L_\beta^{-1/2}$ and $\mathcal{Y}^{(\ell)} = L_\beta^{1/2} \mathcal{Z}^{(\ell)}$, the specified bounds from Subsection 4.1 can be adapted to (4.11) analogously to (4.10). For instance, (4.5) with $t = p$ corresponds to

$$\frac{\vartheta_i^{(\ell+m)} - \alpha_i}{\alpha_{p+1} - \vartheta_i^{(\ell+m)}} \le \left[T_k \left(1 + 2 \, \frac{(\alpha_i - \beta)^{-1} - (\alpha_{p+1} - \beta)^{-1}}{(\alpha_{p+1} - \beta)^{-1} - (\alpha_n - \beta)^{-1}} \right) \right]^{-2m} \frac{\vartheta_p^{(\ell)} - \alpha_i}{\alpha_{p+1} - \vartheta_p^{(\ell)}}. \tag{4.12}$$

The above extension concerning the case $\beta < \alpha_1$ can easily be modified for the case $\beta > \alpha_n$ where $-L_\beta$ is positive definite.

In the remaining case $\alpha_1 < \beta < \alpha_n$, computing the largest/smallest eigenvalues smaller/larger than β is equivalent to computing the largest eigenvalues of $(-L_\beta, M)$ or (L_β, M) for $M = L_\beta S^{-1} L_\beta$ which is positive definite. This allows again applying bounds from Section 3 where A is defined by $M^{-1/2}(\pm L_\beta) M^{-1/2}$, i.e., a counterpart of $L_\beta^{-1/2} S L_\beta^{-1/2}$ from the above extension.

4.3. Deflation

Our angle-free Ritz value bounds reflect different convergence rates of individual Ritz values. In practice, the convergence is typically checked by easily computable residual norms of Ritz vectors. Sufficiently accurate Ritz vectors form a matrix against which further iterative subspaces are orthogonalized explicitly or implicitly within a Rayleigh-Ritz procedure. Moreover, one can add random vectors to the next iterative subspace for keeping the block size unchanged. By ignoring errors of accepted Ritz vectors, we can assume that they span an invariant subspace $\mathcal{V}$, and restrict the investigation of further steps to the orthogonal complement of $\mathcal{V}$.

As an example, we consider again the generalized eigenvalue problem from Subsection 4.2, and denote by $v_1, \ldots, v_n$ S-orthonormal eigenvectors associated with the eigenvalues $\alpha_1 \le \cdots \le \alpha_n$ of (L,S). We observe the restarted block eigensolvers (4.9) and (4.11) in the case that the invariant subspace $\mathcal{V} = \mathrm{span}\{v_1, \ldots, v_c\}$ is known. By deflation, further iterative subspaces are contained in $\widetilde{\mathcal{V}} = \mathrm{span}\{v_{c+1}, \ldots, v_n\}$.

We generally discuss Ritz values in an arbitrary subspace $\mathcal{Z} \subseteq \widetilde{\mathcal{V}}$ with an S-orthonormal basis matrix Z of $\mathcal{Z}$. The Ritz values of (L,S) in $\mathcal{Z}$ are thus given by the eigenvalues of $Z^H L Z$. Moreover, by using the matrix $\widetilde{V} = [v_{c+1}, \ldots, v_n]$, we get the S-orthogonal projector $P = \widetilde{V}\widetilde{V}^H S$ on $\widetilde{\mathcal{V}}$ so that $Z = PZ$, and

$$Z^H L Z = Z^H P^H L P Z = Z^H S \widetilde{V} \widetilde{V}^H L \widetilde{V} \widetilde{V}^H S Z = \widetilde{Z}^H D \widetilde{Z}$$

for $D = \widetilde{V}^H L \widetilde{V} = \mathrm{diag}(\alpha_{c+1}, \ldots, \alpha_n)$ and $\widetilde{Z} = \widetilde{V}^H S Z$. Therein $\widetilde{Z}$ is Euclidean orthonormal since $\widetilde{Z}^H \widetilde{Z} = Z^H S \widetilde{V} \widetilde{V}^H S Z = Z^H S P Z = Z^H S Z$ and Z is S-orthonormal. Thus the eigenvalues of $\widetilde{Z}^H D \widetilde{Z}$ coincide with the Ritz values of D in $\mathrm{span}\{\widetilde{Z}\}$. In summary, the Ritz values of (L,S) in $\mathcal{Z}$ are just those of D in the subspace $\mathrm{span}\{\widetilde{Z}\} = \widetilde{V}^H S \mathcal{Z}$.

Consequently, an eigensolver for (L,S) after deflation can be represented by an eigensolver for D. We first observe (4.9). The above paragraph shows that the iterate $\mathcal{Z}^{(\ell)} \subseteq \widetilde{\mathcal{V}}$ shares Ritz values with $\widetilde{\mathcal{Z}}^{(\ell)} = \widetilde{V}^H S \mathcal{Z}^{(\ell)}$. A corresponding representation of $\mathcal{Z}^{(\ell)}$ is $\mathcal{Z}^{(\ell)} = P\mathcal{Z}^{(\ell)} = \widetilde{V}\widetilde{V}^H S \mathcal{Z}^{(\ell)} = \widetilde{V}\widetilde{\mathcal{Z}}^{(\ell)}$. The next iterate $\mathcal{Z}^{(\ell+1)} = L_\beta^{-1} S \mathcal{Z}^{(\ell)}$ is also a subset of $\widetilde{\mathcal{V}}$ due to

$L_\beta^{-1}S\mathcal{Z}^{(\ell)} \subseteq L_\beta^{-1}S\widetilde{\mathcal{V}} \subseteq \widetilde{\mathcal{V}}$. In addition, it holds that

$$L_\beta^{-1}S\mathcal{Z}^{(\ell)} = PL_\beta^{-1}S\mathcal{Z}^{(\ell)} = PL_\beta^{-1}SP\mathcal{Z}^{(\ell)} = \widetilde{V}\widetilde{V}^H SL_\beta^{-1}S\widetilde{V}\widetilde{V}^H S\mathcal{Z}^{(\ell)}$$
$$= \widetilde{V}\widetilde{V}^H S\widetilde{V}D_\beta^{-1}\widetilde{V}^H S\mathcal{Z}^{(\ell)} = \widetilde{V}D_\beta^{-1}\widetilde{\mathcal{Z}}^{(\ell)}$$

with $D_\beta = \operatorname{diag}(\alpha_{c+1}-\beta,\dots,\alpha_n-\beta)$. Thus the subspace $\widetilde{\mathcal{Z}}^{(\ell+1)} = \widetilde{V}^H S\mathcal{Z}^{(\ell+1)}$ representing the next iterate $\mathcal{Z}^{(\ell+1)}$ fulfills

$$\widetilde{\mathcal{Z}}^{(\ell+1)} = \widetilde{V}^H S(L_\beta^{-1}S\mathcal{Z}^{(\ell)}) = \widetilde{V}^H S(\widetilde{V}D_\beta^{-1}\widetilde{\mathcal{Z}}^{(\ell)}) = D_\beta^{-1}\widetilde{\mathcal{Z}}^{(\ell)},$$

i.e., (4.9) is represented by $\widetilde{\mathcal{Z}}^{(\ell+1)} = D_\beta^{-1}\widetilde{\mathcal{Z}}^{(\ell)}$.

Similarly, $(L_\beta^{-1}S)^j\mathcal{Z}^{(\ell)} = \widetilde{V}D_\beta^{-j}\widetilde{\mathcal{Z}}^{(\ell)}$ holds for $j \geq 1$ and gives the representation

$$\widetilde{\mathcal{Z}}^{(\ell+1)} \quad \xleftarrow{\mathrm{RR}[D,p]} \quad \widetilde{\mathcal{Z}}^{(\ell)} + D_\beta^{-1}\widetilde{\mathcal{Z}}^{(\ell)} + \cdots + D_\beta^{-k}\widetilde{\mathcal{Z}}^{(\ell)}$$

of (4.11). Specifying bounds from Section 3 for these eigensolvers for D produces explicit bounds analogously to Subsection 4.2. The relevant eigenvalues are from the set $\{\alpha_{c+1},\dots,\alpha_n\}$, e.g., a counterpart of (4.12) reads

$$\frac{\vartheta_i^{(\ell+m)} - \alpha_{c+i}}{\alpha_{c+p+1} - \vartheta_i^{(\ell+m)}} \leq \left[T_k\left(1 + 2\,\frac{(\alpha_{c+i}-\beta)^{-1} - (\alpha_{c+p+1}-\beta)^{-1}}{(\alpha_{c+p+1}-\beta)^{-1} - (\alpha_n-\beta)^{-1}}\right)\right]^{-2m} \frac{\vartheta_p^{(\ell)} - \alpha_{c+i}}{\alpha_{c+p+1} - \vartheta_p^{(\ell)}}.$$

Therein the convergence factor can be refined by enlarging the shift β up to α_c. This enables an acceleration with respect to the number of steps, but solving linear systems for L_β with enlarged β could be more costly so that the total computational time is not necessarily reduced.

5. Numerical examples

We compare our angle-free Ritz value bounds with their angle-dependent counterparts which are related to [4, (2.20)] by Knyazev and more accurate than similar traditional bounds from [14, Theorem 6] by Saad especially for clustered eigenvalues; cf. the comparison in [16, Example 3]. The angle-dependent Ritz value bounds in unitarily invariant norms from [9, Theorem 8.2] are not included as they deal with a tuple of Ritz value errors and cannot individually be applied to the ith Ritz value unless $i = 1$.

5.1. Example 1

We reuse the test matrix from [17, Example 1], i.e., the diagonal matrix $A = \operatorname{diag}(\lambda_1,\dots,\lambda_n)$ with $n = 900$ and

$$\lambda_1 = 2, \quad \lambda_2 = 1.6, \quad \lambda_3 = 1.4, \quad \lambda_j = 1 - (j-3)/n \ \text{ for } \ j = 4,\dots,n.$$

following [14, Subsection 4.2] and [9, Example 7.3].

In Figure 1, we demonstrate bounds from Theorem 4.1 concerning one outer step of the restarted block Lanczos method (1.11). Therein (1.11) is implemented with 15 inner steps and the block size 3 in each of 1000 runs with randomly constructed $\mathcal{Y}^{(\ell)}$ and full orthogonalization. We document the Ritz value error $\lambda_i - \psi_i^{(\ell+1)}$ for $i \in \{1, 2, 3\}$ and each inner step. The associated mean values among 1000 samples are displayed by "Lanczos" curves. In addition, mean values of Ritz value errors for $f(A)\mathcal{Y}^{(\ell)}$ with the shifted Chebyshev polynomial (1.7) are drawn as "Chebyshev" circles. We note that the differences between these two types of data concerning individual Ritz values are less obvious than comparing Ritz value sums as in [17, Example 1].

Subsequently, we apply bounds (4.3) and (4.4) to the first inner steps by determining the index t for the assumption $\psi_p^{(\ell)} > \lambda_{t+1}$ or $\psi_i^{(\ell)} > \lambda_{t+1}$. Once $\psi_p^{(\ell+1)}$ or $\psi_i^{(\ell+1)}$ in a $(c+1)$th inner

step exceeds an eigenvalue larger than λ_{t+1}, we update the index t and observe a subspace $\widehat{\mathcal{Y}}$ spanned by p orthonormal Ritz vectors associated with the p largest Ritz values in the current block Krylov subspace $\mathcal{Y}^{(\ell)} + A\mathcal{Y}^{(\ell)} + \cdots + A^c\mathcal{Y}^{(\ell)}$. Then the block Krylov subspace $\widehat{\mathcal{K}} = \widehat{\mathcal{Y}} + A\widehat{\mathcal{Y}} + \cdots + A^{k-c}\widehat{\mathcal{Y}}$ is a subset of $\mathcal{K} = \mathcal{Y}^{(\ell)} + A\mathcal{Y}^{(\ell)} + \cdots + A^k\mathcal{Y}^{(\ell)}$. Thus adapting (4.3) and (4.4) to $\widehat{\mathcal{K}}$ provides appropriate bounds for $\mathcal{K}$, i.e., for further inner steps up to the next update. We convert the evaluated bounds into upper bounds of $\lambda_i - \psi_i^{(\ell+1)}$. The associated mean values are displayed in Figure 1 by "Bound$_1$" and "Bound$_2$" corresponding to (4.3) and (4.4). Furthermore, by using orthonormal eigenvectors $x_1, \ldots, x_p$ associated with $\lambda_1, \ldots, \lambda_p$, we evaluate the angle-dependent bound

$$\frac{\lambda_i - \psi_i^{(\ell+1)}}{\psi_i^{(\ell+1)} - \lambda_n} \le \left[T_k\left(1 + 2\,\frac{\lambda_i - \lambda_{p+1}}{\lambda_{p+1} - \lambda_n}\right)\right]^{-2} \tan^2 \angle(\mathcal{X}_i, \mathcal{Y}_i^{(\ell)}) \tag{5.1}$$

in terms of $\mathcal{X}_i = \operatorname{span}\{x_1, \ldots, x_i\}$ and $\mathcal{Y}_i^{(\ell)} = \operatorname{span}\{x_{i+1}, \ldots, x_p\}^\perp \cap \mathcal{Y}^{(\ell)}$ which can be derived based on (1.4) and improves (1.12) (specification of [4, (2.20)]) for $i < p$. Upper bounds of $\lambda_i - \psi_i^{(\ell+1)}$ generated by (5.1) are displayed by "Bound$_3$".

The comparison in Figure 1 indicates that Bound$_1$ is generally more advantageous than the other two bounds. The overestimation by Bound$_1$ in several of the first inner steps for $i=1$ is related to the assumption $\psi_p^{(\ell)} > \lambda_{t+1}$ of (4.3) and the considerably different convergence rates of $\psi_1^{(\ell)}$ and $\psi_p^{(\ell)}$. In contrast, Bound$_2$ using the assumption $\psi_i^{(\ell)} > \lambda_{t+1}$ provides better alternatives for these steps, but is less accurate in further steps due to the convergence factor in (4.4) with consecutive eigenvalues. The benefit of Bound$_3$ is visible in two inner steps for $i=1$. Afterwards the updated Bound$_1$ shares the convergence factor with Bound$_3$, cf. (4.3) for $t=p$ and (5.1), and becomes more accurate thanks to the angle-independence. Moreover, Bound$_2$ coincides with Bound$_1$ for $i=3$.

In Figure 2, we implement 6 outer steps of the restarted block Lanczos method (1.11) with the block size 3. Each outer step contains 4 inner steps. The comparison again utilizes mean values among 1000 samples with random initial subspaces. Mean values of Ritz value errors from (1.11) and its stepwise modification by the shifted Chebyshev polynomial (1.7) are contained in "Lanczos" curves and "Chebyshev" circles. We count $4 + (4-1) \times 5 = 19$ inner steps. Restart occurs in the iteration indices 4, 7, 10, 13, 16 where we use Theorem 4.2 for determining several nodes in "Bound$_1$" and "Bound$_2$". The other nodes concerning the rest inner steps are generated by Theorem 4.1 as above for Figure 1. Thus Bound$_1$ illustrates a combination "(4.6)+(4.3)", and Bound$_2$ corresponds to "(4.7)+(4.4)". In addition, based on (5.1), we evaluate

$$\left[T_c\left(1 + 2\,\frac{\lambda_i - \lambda_{p+1}}{\lambda_{p+1} - \lambda_n}\right)\right]^{-2} \left[T_k\left(1 + 2\,\frac{\lambda_i - \lambda_{p+1}}{\lambda_{p+1} - \lambda_n}\right)\right]^{-2(s-1)} \tan^2 \angle(\mathcal{X}_i, \mathcal{Y}_i^{(\ell)}) \tag{5.2}$$

in the $(c+1)$th inner step of the sth outer step for generating "Bound$_3$". These three bounds behave similar to their counterparts in Figure 1. Bound$_1$ generally improves Bound$_2$ and Bound$_3$ by using nonconsecutive eigenvalues in convergence factors and angle-free constant terms.

5.2. Example 2

We reuse the test matrix from [17, Example 2], i.e., $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ with $n = 3600$ and

$$\lambda_1 = 2.05, \quad \lambda_2 = 2, \quad \lambda_3 = 1.95, \quad \lambda_4 = 1.65, \quad \lambda_5 = 1.6, \quad \lambda_6 = 1.55,$$
$$\lambda_7 = 1.45, \quad \lambda_8 = 1.4, \quad \lambda_9 = 1.35, \quad \lambda_j = 1 - (j-9)/n \ \text{ for } \ j = 10, \ldots, n.$$

Therein $\lambda_1, \ldots, \lambda_9$ are considered as target eigenvalues and build three clusters. Correspondingly, we implement the restarted block Lanczos method (1.11) for the block size $p = 9$. Further settings and bound evaluations are the same as those used in Example 1.

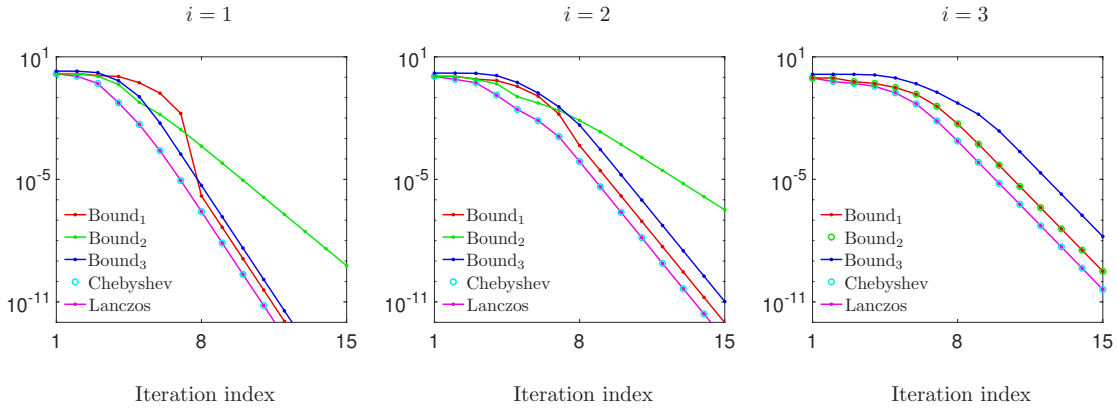

FIGURE 1. Numerical comparison between several Ritz value bounds concerning one outer step of the restarted block Lanczos method (1.11) in Example 1. The three bound curves are determined by bounds (4.3), (4.4) and (5.1), respectively.

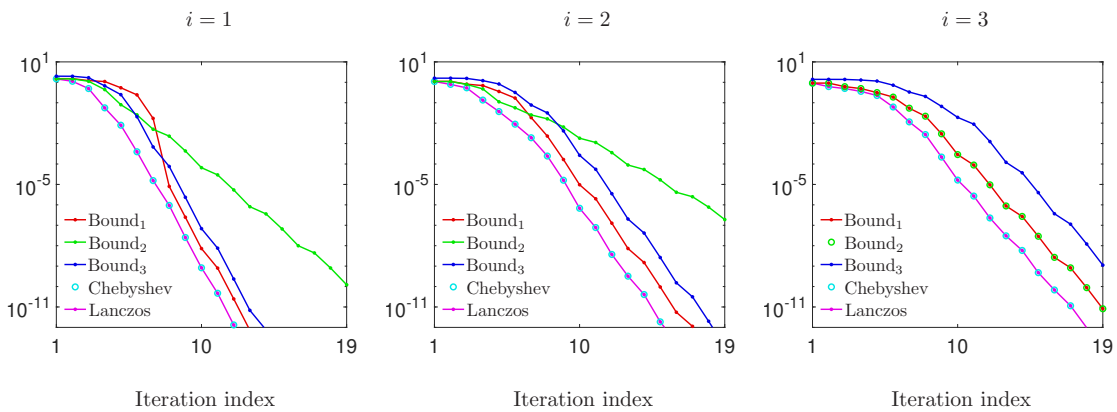

FIGURE 2. Numerical comparison between several Ritz value bounds concerning 6 outer steps (19 inner steps) of the restarted block Lanczos method (1.11) in Example 1. The three bound curves are determined by bound combinations "(4.6)+(4.3)", "(4.7)+(4.4)" and (5.2), respectively.

Figure 3 and Figure 4 illustrate one outer step (15 inner steps) and 6 outer steps (19 inner steps) of (1.11), respectively. We document Ritz value errors for the indices 2, 5, 8 regarding the eigenvalue clusters. The cluster robustness of (1.11) is clearly reflected by Bound$_1$ and Bound$_3$, whereas Bound$_2$ depending on consecutive eigenvalues gives a more substantial overestimation in the final phase in comparison to Example 1. Moreover, the distances between Bound$_1$ and Bound$_3$ become more evident due to relatively larger angle terms.

Conclusions

The convergence theory of block eigensolvers [4, 9, 14] is dominated by angle-dependent bounds where angles build constant terms beside central convergence factors. For analyzing restarted versions, it is necessary to adjust these terms for associated convergence measures. The consequent overestimation is expected to be avoided by direct bounds which can be applied repeatedly. We extend one such bound, the angle-free Ritz value bound [4, (2.22)], to more Ritz values and flexible locations. Therein an argument with biorthogonal vectors for investigating the block power method by Rutishauser [13] is modified by considering intersections of iterative subspaces and certain invariant subspaces. This enables generating partial iterations with

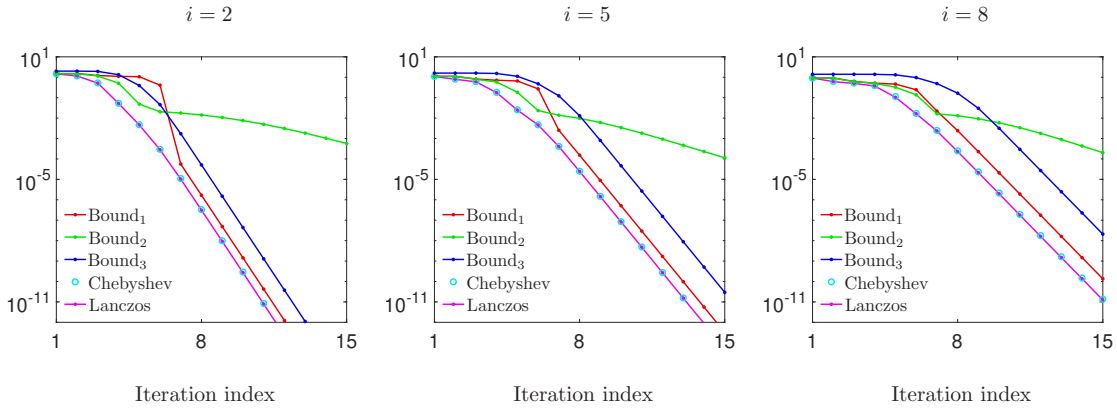

FIGURE 3. Numerical comparison between several Ritz value bounds concerning one outer step of the restarted block Lanczos method (1.11) in Example 2. The three bound curves are determined by bounds (4.3), (4.4) and (5.1), respectively.

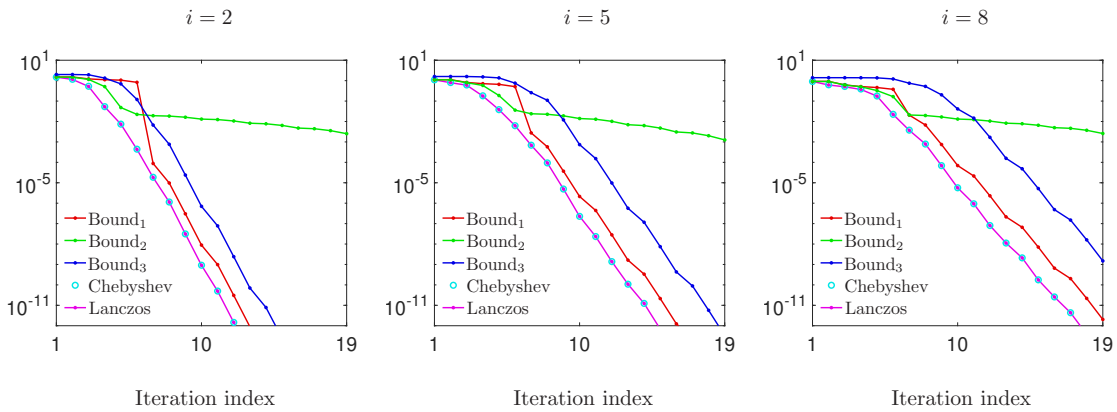

FIGURE 4. Numerical comparison between several Ritz value bounds concerning 6 outer steps (19 inner steps) of the restarted block Lanczos method (1.11) in Example 2. The three bound curves are determined by bound combinations "(4.6)+(4.3)", "(4.7)+(4.4)" and (5.2), respectively.

which the derivation of [4, (2.22)] is upgraded. The resulting bounds are particularly applicable to the restarted block Lanczos method for multiple outer steps, and improve comparable angle-dependent bounds. An open issue is whether the above approach can be generalized to sums of Ritz value errors for supplementing angle-dependent majorization bounds from [17]. Furthermore, the application to restarted block eigensolvers with shift-and-invert is potentially extendable to corresponding preconditioned versions, similarly to recent results on block preconditioned gradient-type eigensolvers in [19, 20].

6. Appendix

6.1. Proof of Lemma 2.1

The assumption $\eta_p > \lambda_{p+1}$ ensures that $X^H Y$ has full rank for $X = [x_1, \ldots, x_p]$, since otherwise there exists a vector $\widehat{g} \in \mathbb{C}^p \backslash \{0\}$ satisfying $X^H Y \widehat{g} = 0$ and causing a contradiction: The vector $Y\widehat{g} \in \mathcal{Y}$ belongs to the invariant subspace $\operatorname{span}\{x_{p+1}, \ldots, x_n\}$ so that $\lambda_{p+1} \geq \rho(Y\widehat{g}) \geq \eta_p$ holds and contradicts $\eta_p > \lambda_{p+1}$.

Moreover, the diagonal matrix $D = \operatorname{diag}\big(f(\lambda_1), \ldots, f(\lambda_p)\big)$ is invertible due to the assumption on $f(\cdot)$ so that $DX^H Y$ has full rank, and $f(A)Y$ also does since $DX^H Y = X^H f(A) Y$. Then

$f(A)Y$ is a basis matrix of $\mathcal{Y}'$ since $\mathcal{Y}' = f(A)\,\mathrm{span}\{Y\} = \mathrm{span}\{f(A)Y\}$. This trivially implies $\dim\mathcal{Y}' = p$. □

6.2. Proof of Lemma 2.3

The relations $\lambda_p > \eta_p' = \rho(y')$ and $\rho(y) \geq \eta_p > \lambda_{p+1}$ in (2.1) simply follow from the settings and the Courant-Fischer principles. The remaining relation $\rho(y') \geq \rho(y^\circ) \geq \rho(y)$ can be shown in three steps by Lemma 2.2.

(i) The relation $\rho(y) \geq \eta_p > \lambda_{p+1}$ ensures $\lambda_l \geq \rho(y) \geq \lambda_{l+1}$ for a certain index $l \in \{1, \ldots, p\}$. According to $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_p)| > \max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)| \geq 0$, it holds that

$$\frac{|f(\lambda_1)|}{|f(\lambda_l)|} \geq \cdots \geq \frac{|f(\lambda_l)|}{|f(\lambda_l)|} = 1 \geq \cdots \geq \frac{|f(\lambda_p)|}{|f(\lambda_l)|},$$

and

$$\frac{|f(\lambda_j)|}{|f(\lambda_l)|} \leq \frac{\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|}{|f(\lambda_p)|} < 1 \quad \forall\, j > p.$$

By defining $w' = y'/f(\lambda_l) = f(A)y/f(\lambda_l)$, we get $\rho(w') = \rho(y')$, and

$$|x_j^H w'| = \frac{|x_j^H f(A)y|}{|f(\lambda_l)|} = \frac{|f(\lambda_j)x_j^H y|}{|f(\lambda_l)|} = \frac{|f(\lambda_j)|}{|f(\lambda_l)|}\,|x_j^H y| \begin{cases} \geq |x_j^H y| & \forall\, j \leq l, \\ \leq |x_j^H y| & \forall\, j > l. \end{cases}$$

Then applying Lemma 2.2 (a) to $u = y$ and $v = w'$ yields $\rho(w') \geq \rho(y)$ so that

$$\lambda_p > \eta_p' = \rho(y') = \rho(w') \geq \rho(y) > \lambda_{p+1}.$$

(ii) The assumption on $f(\cdot)$ ensures $f(\lambda_p) \neq 0$. Then $y^\circ \neq 0$ holds since otherwise

$$0 = x_j^H y^\circ = f(\lambda_p)x_j^H y \quad \forall\ j \leq p \qquad \Rightarrow \qquad x_j^H y = 0 \quad \forall\ j \leq p \qquad \Rightarrow \qquad \rho(y) \leq \lambda_{p+1}$$

contradicts $\rho(y) \geq \eta_p > \lambda_{p+1}$. Thus $\mu(y^\circ)$ can be defined. According to the result $\lambda_p > \rho(y') > \lambda_{p+1}$ from (i) and

$$|x_j^H y^\circ| = \begin{cases} |f(\lambda_p)|\,|x_j^H y| \leq |f(\lambda_j)|\,|x_j^H y| = |x_j^H f(A)y| = |x_j^H y'| & \forall\, j \leq p, \\ \left(\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|\right)|x_j^H y| \geq |f(\lambda_j)|\,|x_j^H y| = |x_j^H y'| & \forall\, j > p, \end{cases}$$

we can apply Lemma 2.2 (b) to $l = p$, $u = y'$ and $v = y^\circ$. This implies $\rho(y') \geq \rho(y^\circ)$.

(iii) The vector $w^\circ = y^\circ/f(\lambda_p) \neq 0$ fulfills

$$|x_j^H w^\circ| = \begin{cases} |f(\lambda_p)|\,|x_j^H y|/|f(\lambda_p)| = |x_j^H y| & \forall\, j \leq p, \\ \left(\max_{j\in\{p+1,\ldots,n\}}|f(\lambda_j)|\right)|x_j^H y|/|f(\lambda_p)| \leq |x_j^H y| & \forall\, j > p. \end{cases}$$

Using this together with the result $\lambda_p > \rho(y) > \lambda_{p+1}$ from (i) allows applying Lemma 2.2 (a) to $l = p$, $u = y$ and $v = w^\circ$. Then we get $\rho(y^\circ) = \rho(w^\circ) \geq \rho(y)$.

Summarizing (ii) and (iii) yields $\rho(y') \geq \rho(y^\circ) \geq \rho(y)$.

For deriving (2.2), we denote y_p and $y - y_p$ by w and z so that

$$y = w + z \quad \text{and} \quad y^\circ = f(\lambda_p)w + \nu_p z.$$

Then y and y° can be analyzed within the subspace $\mathrm{span}\{w,\, z\}$.

Therein w and z are nonzero vectors since otherwise $y = z$ or $y = w$ holds so that $\rho(y) \leq \lambda_{p+1}$ or $\rho(y) \geq \lambda_p$ contradicts (2.1). Thus $\rho(w)$ and $\rho(z)$ can be defined. Applying the

orthogonality properties $w^H z = 0$ and $w^H A z = 0$ to $y = w + z$ gives

$$\begin{aligned} \rho(y) = \frac{y^H A y}{y^H y} &= \frac{w^H A w + z^H A z}{w^H w + z^H z} = \frac{\rho(w)\|w\|_2^2 + \rho(z)\|z\|_2^2}{\|w\|_2^2 + \|z\|_2^2} \\ \Rightarrow \quad \frac{\rho(w) - \rho(y)}{\rho(y) - \rho(z)} &= \frac{\|z\|_2^2}{\|w\|_2^2}. \end{aligned} \tag{6.1}$$

Combining this with an analogous result for $y^\circ = f(\lambda_p) w + \nu_p z$ yields

$$\frac{\rho(w) - \rho(y^\circ)}{\rho(y^\circ) - \rho(z)} = \frac{\|\nu_p z\|_2^2}{\|f(\lambda_p) w\|_2^2} = \frac{\nu_p^2}{|f(\lambda_p)|^2} \frac{\|z\|_2^2}{\|w\|_2^2} = \frac{\nu_p^2}{|f(\lambda_p)|^2} \frac{\rho(w) - \rho(y)}{\rho(y) - \rho(z)}$$

and implies (2.2). □

6.3. Proof of Theorem 2.4

The assumption on $f(\cdot)$ ensures $f(\lambda_j) \neq 0$ for $j \leq p$ so that Lemma 2.1 is applicable and gives $\dim \mathcal{Y}' = p$. For verifying (1.5), we can skip the trivial case $\lambda_p = \eta_p'$. Then $\lambda_p > \eta_p'$ holds so that Lemma 2.3 is applicable. Denoting y_p and $y - y_p$ by w and z, (2.1) implies

$$\rho(w) \geq \lambda_p > \eta_p' = \rho(y') \geq \rho(y^\circ) \geq \rho(y) \geq \eta_p > \lambda_{p+1} \geq \rho(z).$$

Subsequently, simple monotonicity arguments lead to

$$\begin{aligned} &\left(\frac{\lambda_p - \eta_p'}{\eta_p' - \lambda_{p+1}}\right)\left(\frac{\lambda_p - \eta_p}{\eta_p - \lambda_{p+1}}\right)^{-1} \leq \left(\frac{\lambda_p - \rho(y^\circ)}{\rho(y^\circ) - \lambda_{p+1}}\right)\left(\frac{\lambda_p - \rho(y)}{\rho(y) - \lambda_{p+1}}\right)^{-1} \\ &= \left(\frac{\lambda_p - \rho(y^\circ)}{\lambda_p - \rho(y)}\right)\left(\frac{\rho(y) - \lambda_{p+1}}{\rho(y^\circ) - \lambda_{p+1}}\right) \leq \left(\frac{\rho(w) - \rho(y^\circ)}{\rho(w) - \rho(y)}\right)\left(\frac{\rho(y) - \rho(z)}{\rho(y^\circ) - \rho(z)}\right) \\ &= \left(\frac{\rho(w) - \rho(y^\circ)}{\rho(y^\circ) - \rho(z)}\right)\left(\frac{\rho(w) - \rho(y)}{\rho(y) - \rho(z)}\right)^{-1} \overset{(2.2)}{=} \frac{\max_{j \in \{p+1,\ldots,n\}} |f(\lambda_j)|^2}{|f(\lambda_p)|^2} \end{aligned} \tag{6.2}$$

which results in (1.5). □

6.4. Proof of Lemma 2.5

We denote the three subspaces with the subscript $[1, i] \cup (p, n]$ by $\widetilde{\mathcal{X}}$, $\widetilde{\mathcal{Y}}$ and $\widetilde{\mathcal{Y}}'$, respectively. The dimension comparison

$$\dim(\widetilde{\mathcal{X}} \cap \mathcal{Y}) = \dim \widetilde{\mathcal{X}} + \dim \mathcal{Y} - \dim(\widetilde{\mathcal{X}} + \mathcal{Y}) \geq (n - p + i) + p - n = i$$

gives $\dim \widetilde{\mathcal{Y}} \geq i$. In the case $\eta_p > \lambda_{p+1}$, the strict inequality $\dim \widetilde{\mathcal{Y}} > i$ does not hold, since otherwise the smallest (jth largest with $j > i$) Ritz value $\widetilde{\eta}$ in $\widetilde{\mathcal{Y}}$ fulfills $\widetilde{\eta} \leq \lambda_{p+1}$ and $\widetilde{\eta} \geq \eta_p$ due to $\widetilde{\mathcal{Y}} \subseteq \widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}} \subseteq \mathcal{Y}$ so that $\eta_p \leq \lambda_{p+1}$.

As shown in the proof of Lemma 2.1, $X^H Y$ is an invertible $p \times p$ matrix for $X = [x_1, \ldots, x_p]$ and an arbitrary basis matrix Y of $\mathcal{Y}$. Then the $n \times i$ matrix $Y_i = Y(X^H Y)^{-1}[e_1, \ldots, e_i]$ has full rank where $e_1, \ldots, e_p$ are columns of the $p \times p$ identity matrix. Thus $\operatorname{span}\{Y_i\}$ is an i-dimensional subset of $\mathcal{Y}$. Moreover, since $[x_{i+1}, \ldots, x_p]^H Y_i$ is a zero matrix due to

$$\begin{aligned} [x_{i+1}, \ldots, x_p]^H Y_i &= [e_{i+1}, \ldots, e_p]^H X^H Y (X^H Y)^{-1} [e_1, \ldots, e_i] \\ &= [e_{i+1}, \ldots, e_p]^H [e_1, \ldots, e_i], \end{aligned}$$

it holds that $\operatorname{span}\{Y_i\} \subseteq \widetilde{\mathcal{X}}$ and consequently $\operatorname{span}\{Y_i\} = \widetilde{\mathcal{Y}}$. By using XX^H as the orthogonal projector $P_{\mathcal{X}}$ on $\mathcal{X}$, we get

$$(XX^H) Y_i = XX^H Y (X^H Y)^{-1} [e_1, \ldots, e_i] = X\,[e_1, \ldots, e_i] = [x_1, \ldots, x_i]$$

so that $P_{\mathcal{X}} \widetilde{\mathcal{Y}} = \operatorname{span}\{x_1, \ldots, x_i\}$.

In addition, for $X_i = [x_1, \ldots, x_i]$ and $D_i = \operatorname{diag}\big(f(\lambda_1), \ldots, f(\lambda_i)\big)$, it holds that $X_i^H f(A) Y_i = D_i X_i^H Y_i = D_i$, and the assumption on $f(\cdot)$ ensures that D_i is invertible. Thus $f(A)Y_i$ has full rank so that the subspace $\widetilde{\mathcal{Y}}' = f(A)\operatorname{span}\{Y_i\} = \operatorname{span}\{f(A)Y_i\}$ has dimension i. □

6.5. Proof of Lemma 3.1

Following the proof of Lemma 2.5 with simplified subspace notation, we use again the basis matrices Y_i and $f(A)Y_i$ of $\widetilde{\mathcal{Y}}$ and $\widetilde{\mathcal{Y}}'$. Then an arbitrary Ritz vector $\widetilde{y}'$ associated with $\widetilde{\eta}_i'$ can be represented by $\widetilde{y}' = f(A)Y_i g$ with a certain $g \in \mathbb{C}^i \backslash \{0\}$ so that $\widetilde{y}' = f(A)\widetilde{y}$ for $\widetilde{y} = Y_i g$.

For verifying (3.1), the relation $\lambda_i > \widetilde{\eta}_i' = \rho(\widetilde{y}')$ is trivial. Moreover, $\widetilde{y} \in \widetilde{\mathcal{Y}} \subseteq \mathcal{Y}$ implies $\rho(\widetilde{y}) \geq \widetilde{\eta}_i \geq \eta_p$. Combining this with the assumption $\eta_p > \lambda_{p+1}$ from Lemma 2.5 gives $\rho(\widetilde{y}) \geq \widetilde{\eta}_i > \lambda_{p+1}$. Subsequently, $\rho(\widetilde{y}') \geq \rho(\widetilde{y}^\circ) \geq \rho(\widetilde{y})$ follows from Lemma 2.2 in three steps, with some nontrivial detailed changes in comparison to the proof of Lemma 2.3.

(i) We have $\lambda_l \geq \rho(\widetilde{y}) \geq \lambda_{l+1}$ for a certain index $l \in \{1, \ldots, p\}$ due to $\rho(\widetilde{y}) \geq \widetilde{\eta}_i > \lambda_{p+1}$. If $l < i$, the assumption $|f(\lambda_1)| \geq \cdots \geq |f(\lambda_i)| > \max_{j\in\{p+1,\ldots,n\}} |f(\lambda_j)|$ leads to

$$\frac{|f(\lambda_1)|}{|f(\lambda_l)|} \geq \cdots \geq \frac{|f(\lambda_l)|}{|f(\lambda_l)|} = 1 \geq \cdots \geq \frac{|f(\lambda_i)|}{|f(\lambda_l)|},$$

and

$$\frac{|f(\lambda_j)|}{|f(\lambda_l)|} \leq \frac{\max_{j\in\{p+1,\ldots,n\}} |f(\lambda_j)|}{|f(\lambda_i)|} < 1 \quad \forall\, j > p.$$

Then $\widetilde{w}' = \widetilde{y}'/f(\lambda_l) = f(A)\widetilde{y}/f(\lambda_l)$ fulfills $\rho(\widetilde{w}') = \rho(\widetilde{y}')$,

$$|x_j^H \widetilde{w}'| = \frac{|x_j^H f(A)\widetilde{y}|}{|f(\lambda_l)|} = \frac{|f(\lambda_j) x_j^H \widetilde{y}|}{|f(\lambda_l)|} \begin{cases} \geq |x_j^H \widetilde{y}| & \forall\, j \leq l, \\ \leq |x_j^H \widetilde{y}| & \forall\, j \in \{l+1, \ldots, i,\, p+1, \ldots, n\}. \end{cases}$$

If $l \geq i$, we redefine $\widetilde{w}'$ by $\widetilde{w}' = \widetilde{y}'/f(\lambda_i)$ so that

$$|x_j^H \widetilde{w}'| = \frac{|x_j^H f(A)\widetilde{y}|}{|f(\lambda_i)|} = \frac{|f(\lambda_j) x_j^H \widetilde{y}|}{|f(\lambda_i)|} \begin{cases} \geq |x_j^H \widetilde{y}| & \forall\, j \leq i, \\ \leq |x_j^H \widetilde{y}| & \forall\, j > p. \end{cases}$$

In both cases, a combination with $|x_j^H \widetilde{w}'| = 0 = |x_j^H \widetilde{y}| \;\; \forall\, j \in \{i+1, \ldots, p\}$ gives

$$|x_j^H \widetilde{w}'| \geq |x_j^H \widetilde{y}| \;\; \forall\, j \leq l \quad \text{and} \quad |x_j^H \widetilde{w}'| \leq |x_j^H \widetilde{y}| \;\; \forall\, j > l.$$

Thus Lemma 2.2 (a) is applicable to $u = \widetilde{y}$ and $v = \widetilde{w}'$, and implies

$$\lambda_i > \eta_i' = \rho(\widetilde{y}') = \rho(\widetilde{w}') \geq \rho(\widetilde{y}) > \lambda_{p+1}.$$

(ii) The vector $\widetilde{y}^\circ$ is nonzero since otherwise

$$0 = x_j^H \widetilde{y}^\circ = f(\lambda_i) x_j^H \widetilde{y} \quad \forall\;\; j \leq i \quad \Rightarrow \quad x_j^H \widetilde{y} = 0 \quad \forall\;\; j \leq i \quad \Rightarrow \quad x_j^H \widetilde{y} = 0 \quad \forall\;\; j \leq p$$

so that $\rho(\widetilde{y}) \leq \lambda_{p+1}$ holds and contradicts $\rho(\widetilde{y}) \geq \widetilde{\eta}_i > \lambda_{p+1}$. Thus $\mu(y^\circ)$ can be defined. The result $\lambda_i > \rho(\widetilde{y}') > \lambda_{p+1}$ from (i) indicates $\lambda_l \geq \rho(\widetilde{y}') \geq \lambda_{l+1}$ for a certain index $l \in \{i, \ldots, p\}$. In addition, combining

$$|x_j^H \widetilde{y}^\circ| = \begin{cases} |f(\lambda_i)|\, |x_j^H \widetilde{y}| \leq |f(\lambda_j)|\, |x_j^H \widetilde{y}| = |x_j^H f(A)\widetilde{y}| = |x_j^H \widetilde{y}'| & \forall\, j \leq i, \\ \big(\max_{j\in\{p+1,\ldots,n\}} |f(\lambda_j)|\big)\, |x_j^H \widetilde{y}| \geq |f(\lambda_j)|\, |x_j^H \widetilde{y}| = |x_j^H \widetilde{y}'| & \forall\, j > p \end{cases}$$

with $|x_j^H \widetilde{y}^\circ| = 0 = |x_j^H \widetilde{y}'| \;\; \forall\, j \in \{i+1, \ldots, p\}$ ensures that Lemma 2.2 (b) is applicable to $u = \widetilde{y}'$ and $v = \widetilde{y}^\circ$. Thus $\rho(\widetilde{y}') \geq \rho(\widetilde{y}^\circ)$.

(iii) It holds that $\lambda_l \geq \rho(\widetilde{y}) \geq \lambda_{l+1}$ for a certain index $l \in \{i, \ldots, p\}$ due to $\lambda_i > \rho(\widetilde{y}) > \lambda_{p+1}$ from (i). The auxiliary vector $\widetilde{w}^\circ = \widetilde{y}^\circ / f(\lambda_i)$ is nonzero and fulfills

$$|x_j^H \widetilde{w}^\circ| = \begin{cases} |f(\lambda_i)|\, |x_j^H \widetilde{y}| / |f(\lambda_i)| = |x_j^H \widetilde{y}| & \forall\ j \leq i, \\ \big(\max_{j \in \{p+1,\ldots,n\}} |f(\lambda_j)|\big)\, |x_j^H \widetilde{y}| / |f(\lambda_i)| \leq |x_j^H \widetilde{y}| & \forall\ j > p \end{cases}$$

and $|x_j^H \widetilde{w}^\circ| = 0 = |x_j^H \widetilde{y}|\ \ \forall\ j \in \{i+1, \ldots, p\}$. This allows applying Lemma 2.2 (a) to $u = \widetilde{y}$ and $v = \widetilde{w}^\circ$ so that $\rho(\widetilde{y}^\circ) = \rho(\widetilde{w}^\circ) \geq \rho(\widetilde{y})$.

According to (ii) and (iii), we get $\rho(\widetilde{y}') \geq \rho(\widetilde{y}^\circ) \geq \rho(\widetilde{y})$.

The derivation of (3.2) is based on the representations

$$\widetilde{y} = \widetilde{w} + \widetilde{z} \quad \text{and} \quad \widetilde{y}^\circ = f(\lambda_i)\widetilde{w} + \nu_p \widetilde{z}$$

with $\widetilde{w} = \widetilde{y}_i$ and $\widetilde{z} = \widetilde{y} - \widetilde{y}_i$. The property (3.1) excludes $\widetilde{w} = 0$ or $\widetilde{z} = 0$ which would lead to $\rho(\widetilde{y}) \leq \lambda_{p+1}$ or $\rho(\widetilde{y}) \geq \lambda_i$. Thus $\rho(\widetilde{w})$ and $\rho(\widetilde{z})$ can be defined. By using $\widetilde{w}^H \widetilde{z} = 0$ and $\widetilde{w}^H A \widetilde{z} = 0$,

$$\frac{\rho(\widetilde{w}) - \rho(\widetilde{y})}{\rho(\widetilde{y}) - \rho(\widetilde{z})} = \frac{\|\widetilde{z}\|_2^2}{\|\widetilde{w}\|_2^2} \quad \text{and} \quad \frac{\rho(\widetilde{w}) - \rho(\widetilde{y}^\circ)}{\rho(\widetilde{y}^\circ) - \rho(\widetilde{z})} = \frac{\|\nu_p \widetilde{z}\|_2^2}{\|f(\lambda_i)\widetilde{w}\|_2^2}$$

can be shown analogously to (6.1), and result in (3.2). □

6.6. Proof of Lemma 3.3

The property $\dim \widetilde{\mathcal{Y}} \geq i$ follows from a dimension comparison as in the proof of Lemma 2.5.

For verifying $\dim \widetilde{\mathcal{Y}}_i' = i$, we use an arbitrary basis matrix $\widetilde{Y}_i \in \mathbb{C}^{n \times i}$ of $\widetilde{\mathcal{Y}}_i$ together with $X_{t-p+i} = [x_1, \ldots, x_{t-p+i}]$. The assumption $\eta_p > \lambda_{t+1}$ ensures that $X_{t-p+i}^T \widetilde{Y}_i$ has full rank since otherwise there exists a vector $\widehat{g} \in \mathbb{C}^i \backslash \{0\}$ with $X_{t-p+i}^T \widetilde{Y}_i \widehat{g} = 0$ so that the vector $\widehat{y} = \widetilde{Y}_i \widehat{g}$ is orthogonal to $x_1, \ldots, x_{t-p+i}$. Combining this with $\widehat{y} \in \widetilde{\mathcal{Y}}_i \subseteq \widetilde{\mathcal{Y}} \subseteq \widetilde{\mathcal{X}} = \operatorname{span}\{x_{t-p+i+1}, \ldots, x_t\}^\perp$ shows that $\widehat{y}$ belongs to $\operatorname{span}\{x_{t+1}, \ldots, x_n\}$. Then $\lambda_{t+1} \geq \rho(\widehat{y}) \geq \eta_p$ holds (due to $\widehat{y} \in \widetilde{\mathcal{Y}}_i \subseteq \widetilde{\mathcal{Y}} \subseteq \mathcal{Y}$) and contradicts $\eta_p > \lambda_{t+1}$. In addition, the assumption $f(\lambda_j) \neq 0\ \ \forall\ j \in \{1, \ldots, t-p+i\}$ suggests the invertible diagonal matrix $D_{t-p+i} = \operatorname{diag}\big(f(\lambda_1), \ldots, f(\lambda_{t-p+i})\big)$. Therewith $X_{t-p+i}^T f(A) \widetilde{Y}_i = D_{t-p+i} X_{t-p+i}^T \widetilde{Y}_i$ has full rank so that $f(A)\widetilde{Y}_i$ also does, and $\widetilde{\mathcal{Y}}_i' = f(A) \operatorname{span}\{\widetilde{Y}_i\} = \operatorname{span}\{f(A)\widetilde{Y}_i\}$ has dimension i.

Furthermore, by using $f(A)\widetilde{Y}_i$ as a basis matrix of $\widetilde{\mathcal{Y}}_i'$, we get the representation $\widetilde{y}' = f(A)\widetilde{Y}_i g$ with a certain $g \in \mathbb{C}^i \backslash \{0\}$. Thus $\widetilde{y}' = f(A)\widetilde{y}$ for $\widetilde{y} = \widetilde{Y}_i g$.

The properties (3.5) and (3.6) can be shown by modifying the derivations of (3.1) and (3.2) in Subsection 6.5 with shifted indices. □

6.7. Strictly increasing Ritz values

We explain the fact that the Ritz values strictly increase during the restarted block Lanczos method (1.11) as long as the current iterative subspace contains no eigenvectors.

With the settings from Theorem 4.1, we show $\psi_i^{(\ell+1)} > \psi_i^{(\ell)}$ for each $i \in \{1, \ldots, p\}$ provided that there are no eigenvectors in $\mathcal{Y}^{(\ell)}$. Therein we denote by $y_1, \ldots, y_p$ orthonormal Ritz vectors associated with the Ritz values $\psi_1^{(\ell)} \geq \cdots \geq \psi_p^{(\ell)}$ in $\mathcal{Y}^{(\ell)}$, and define $\widetilde{\mathcal{Y}}_i = \operatorname{span}\{y_1, \ldots, y_i\}$. By using an arbitrary $\lambda < \lambda_n$, the matrix $\widetilde{A} = A - \lambda I$ is Hermitian positive definite, and the subspace $\widetilde{\mathcal{Y}}_i' = \widetilde{A}\widetilde{\mathcal{Y}}_i$ has dimension i. Then the ith largest Ritz value $\widetilde{\eta}_i'$ in $\widetilde{\mathcal{Y}}_i'$ fulfills $\widetilde{\eta}_i' \leq \psi_i^{(\ell+1)}$ according to

$$\widetilde{A}\widetilde{\mathcal{Y}}_i \subseteq \widetilde{A}\mathcal{Y}^{(\ell)} \subseteq \mathcal{Y}^{(\ell)} + A\mathcal{Y}^{(\ell)} + \cdots + A^k \mathcal{Y}^{(\ell)}.$$

Moreover, an arbitrary Ritz vector $\widetilde{y}'$ associated with $\widetilde{\eta}_i'$ can be represented by $\widetilde{y}' = \widetilde{A}\widetilde{y}$ with a certain $\widetilde{y} \in \widetilde{\mathcal{Y}}_i \backslash \{0\}$.

We further use the Rayleigh quotients $\rho(\cdot)$ and $\widetilde{\rho}(\cdot)$ with respect to A and $\widetilde{A}$. Then

$$\widetilde{\rho}(\widetilde{y}) = \frac{\widetilde{y}^H \widetilde{A}\widetilde{y}}{\widetilde{y}^H \widetilde{y}} \leq \frac{\|\widetilde{y}\|_2 \|\widetilde{A}\widetilde{y}\|_2}{\|\widetilde{y}\|_2^2} = \frac{(\widetilde{y},\, \widetilde{A}\widetilde{y})_{\widetilde{A}}}{\|\widetilde{y}\|_2 \|\widetilde{A}\widetilde{y}\|_2} \leq \frac{\|\widetilde{y}\|_{\widetilde{A}} \|\widetilde{A}\widetilde{y}\|_{\widetilde{A}}}{\|\widetilde{y}\|_2 \|\widetilde{A}\widetilde{y}\|_2} = \big(\widetilde{\rho}(\widetilde{y})\, \widetilde{\rho}(\widetilde{y}')\big)^{1/2}$$

implies $\widetilde{\rho}(\widetilde{y}) \leq \widetilde{\rho}(\widetilde{y}')$, i.e., $\rho(\widetilde{y}) - \lambda \leq \rho(\widetilde{y}') - \lambda$ so that $\rho(\widetilde{y}) \leq \rho(\widetilde{y}')$. Therein the equality holds if and only if $\widetilde{y}$ is collinear with $\widetilde{A}\widetilde{y}$, i.e., $\widetilde{y}$ is an eigenvector of $\widetilde{A}$ as well as A. Consequently, if $\mathcal{Y}^{(\ell)}$ contains no eigenvectors, we get $\rho(\widetilde{y}) < \rho(\widetilde{y}')$ so that $\psi_i^{(\ell+1)} \geq \widetilde{\eta}_i' = \rho(\widetilde{y}') > \rho(\widetilde{y}) \geq \psi_i^{(\ell)}$.

REFERENCES

[1] J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*. SIAM, Philadelphia, PA, 2002. doi:10.1137/1.9780898719192

[2] S. M. Greene, R. J. Webber, T. C. Berkelbach, and J. Weare, *Approximating matrix eigenvalues by subspace iteration with repeated random sparsification*, SIAM J. Sci. Comput., 44 (2022), pp. A3067–A3097. doi:10.1137/21M1422513

[3] A. V. Knyazev, *Computation of Eigenvalues and Eigenvectors for Mesh Problems: Algorithms and Error Estimates*, Dept. Numerical Math. USSR Academy of Sciences, Moscow, 1986 (in Russian). `https://shorturl.at/gNQ29`

[4] A. V. Knyazev, *Convergence rate estimates for iterative methods for mesh symmetric eigenvalue problem*, Soviet J. Num. Anal. Math. Modelling, 2 (1987), pp. 371–396. doi:10.1515/rnam.1987.2.5.371

[5] A. V. Knyazev, *Toward the Optimal Preconditioned Eigensolver: Locally Optimal Block Preconditioned Conjugate Gradient Method*, SIAM J. Sci. Comput., 23 (2001), pp. 517–541. doi:10.1137/S1064827500366124

[6] A. V. Knyazev and M. E. Argentati, *Rayleigh-Ritz majorization error bounds with applications to FEM*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1521–1537. doi:10.1137/08072574X

[7] A. V. Knyazev and A. Malyshev, *Accelerated graph-based spectral polynomial filters*, IEEE 25th International Workshop on Machine Learning for Signal Processing, (2015), pp. 1–6. doi:10.1109/MLSP.2015.7324315

[8] A. V. Knyazev and A. L. Skorokhodov, *On exact estimates of the convergence rate of the steepest ascent method in the symmetric eigenvalue problem*, Linear Algebra Appl., 154–156 (1991), pp. 245–257. doi:10.1016/0024-3795(91)90379-B

[9] R. C. Li and L. H. Zhang, *Convergence of the block Lanczos method for eigenvalue clusters*, Numer. Math., 131 (2015), pp. 83–113. doi:10.1007/s00211-014-0681-6

[10] C. W. Murray, S. C. Racine, and E. R. Davidson, *Improved algorithms for the lowest few eigenvalues and associated eigenvectors of large matrices*, J. Comput. Phys., 103 (1992), pp. 382–389. doi:10.1016/0021-9991(92)90409-R

[11] K. Neymeyr and M. Zhou, *Convergence analysis of restarted Krylov subspace eigensolvers*, SIAM J. Matrix Anal. Appl. 37 (2016), pp. 955–975. doi:10.1137/16M1056481

[12] B. N. Parlett, *The Symmetric Eigenvalue Problem*. SIAM, Philadelphia, PA, 1998. doi:10.1137/1.9781611971163

[13] H. Rutishauser, *Computational aspects of F.L. Bauer's simultaneous iteration method*, Numer. Math., 13 (1969), pp. 4–13. doi:10.1007/BF02165269

[14] Y. Saad, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706. doi:10.1137/0717059

[15] G. W. Stewart, *Accelerating the orthogonal iteration for the eigenvectors of a Hermitian matrix*, Numer. Math., 13 (1969), pp. 362–376. doi:10.1007/BF02165413

[16] M. Zhou, *Convergence estimates of nonrestarted and restarted block-Lanczos methods*, Numer. Linear Algebra Appl., 25 (2018), e2182. doi:10.1002/nla.2182

[17] M. Zhou, M. E. Argentati, A. V. Knyazev, and K. Neymeyr, *Majorization-type cluster robust bounds for block filters and eigensolvers*, Technical Report, 2022. `https://arxiv.org/abs/2201.04517`

[18] M. Zhou and K. Neymeyr, *Sharp Ritz value estimates for restarted Krylov subspace iterations*, Electron. Trans. Numer. Anal., 46 (2017), pp. 424–446. `https://etna.math.kent.edu/volumes/2011-2020/vol46`

[19] M. Zhou and K. Neymeyr, *Cluster robust estimates for block gradient-type eigensolvers*, Math. Comp., 88 (2019), pp. 2737–2765. doi:10.1090/mcom/3446

[20] M. Zhou and K. Neymeyr, *Convergence rates of individual Ritz values in block preconditioned gradient-type eigensolvers*, Technical Report, 2022. `https://arxiv.org/abs/2206.00585`

Universität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18055 Rostock, Germany

E-mail address: `ming.zhou at uni-rostock (dot) de`

Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO 80217-3364, USA

E-mail address: `andrew.knyazev at ucdenver (dot) edu`

Universität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18055 Rostock, Germany

E-mail address: `klaus.neymeyr at uni-rostock (dot) de`