Electronic Transactions on Numerical Analysis.
Volume 46, pp. 424–446, 2017.

ISSN 1068–9613.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

SHARP RITZ VALUE ESTIMATES FOR RESTARTED KRYLOV SUBSPACE ITERATIONS*

MING ZHOU† AND KLAUS NEYMEYR†

Abstract. Gradient iterations for the Rayleigh quotient are elementary methods for computing the smallest eigenvalues of a pair of symmetric and positive definite matrices. A considerable convergence acceleration can be achieved by preconditioning and by computing Rayleigh-Ritz approximations from subspaces of increasing dimensions. An example of the resulting Krylov subspace eigensolvers is the generalized Davidson method. Krylov subspace iterations can be restarted in order to limit their computer storage requirements. For the restarted Krylov subspace eigensolvers, a Chebyshev-type convergence estimate was presented by Knyazev in [Soviet J. Numer. Anal. Math. Modelling, 2 (1987), pp. 371–396]. This estimate has been generalized to arbitrary eigenvalue intervals in [SIAM J. Matrix Anal. Appl., 37 (2016), pp. 955–975]. The generalized Ritz value estimate is not sharp as it depends only on three eigenvalues. In the present paper we extend the latter analysis by generalizing the geometric approach from [SIAM J. Matrix Anal. Appl., 32 (2011), pp. 443–456] in order to derive a sharp Ritz value estimate for restarted Krylov subspace iterations.

Key words. Krylov subspace, Rayleigh quotient, Rayleigh-Ritz procedure, polynomial interpolation, multigrid, elliptic eigenvalue problem.

AMS subject classifications. 65F15, 65N25

1. Introduction.

Gradient iterations for the Rayleigh quotient

$$\rho : \mathbb{R}^n \backslash \{0\} \to \mathbb{R}, \quad \rho(x) = (x, Ax)/(x, Mx) \tag{1.1}$$

serve for the computation of a moderate number of the extreme eigenvalues and the associated eigenspaces of a pair of symmetric and positive definite matrices $A, M \in \mathbb{R}^{n\times n}$. In typical applications A and M are finite element discretization matrices of self-adjoint and elliptic partial differential operators. For a sufficiently fine discretization, these matrices are large and sparse. Thus, the smallest eigenvalues of (A, M) and the associated eigenvectors should not be computed by (classical) matrix transformations [2, 6, 20, 23]. Instead, one prefers a gradient iteration for (1.1) of the basic form $x^{(\ell+1)} = x^{(\ell)} - \omega\nabla\rho(x^{(\ell)})$ since the minimizers of the Rayleigh quotient are the eigenvectors associated with the smallest eigenvalue of (A, M).

Replacing the Euclidean gradient $\nabla\rho(\cdot)$ by the A-gradient $\nabla_A\rho(\cdot) := A^{-1}\nabla\rho(\cdot)$ can result in a considerable convergence acceleration [4, 12, 16]. Practically, the A-gradient iteration can be implemented by the Rayleigh-Ritz procedure applied to the two-dimensional subspace spanned by x and $A^{-1}\nabla\rho(x)$. A further acceleration can be achieved by extending this subspace with previous iterates. The generalized Davidson method belongs to this class of A-gradient iterations if it uses the exact inverse A^{-1} for preconditioning. We call this case “exact-inverse preconditioning”. The iterative scheme, which starts with an iterate $x^{(\ell)}$ and results in a Ritz vector $x^{(\ell+1)}$ from an at most k-dimensional subspace $\mathcal{V}$, reads

$$\begin{cases} v^{(1)} = x^{(\ell)}, \quad \mathcal{V} = \mathrm{span}\{v^{(1)}\}, \\ \mathcal{V} \leftarrow \mathrm{span}\{\mathcal{V},\, A^{-1}\nabla\rho(v^{(i)})\}, \quad v^{(i+1)} \leftarrow \mathrm{RR}_{\min}(\mathcal{V}), \quad \text{for} \quad i = 1, \dots, k-1, \\ x^{(\ell+1)} = v^{(k)}. \end{cases} \tag{1.2}$$

Therein, the Rayleigh-Ritz procedure $\mathrm{RR}_{\min}(\mathcal{V})$ returns a Ritz vector associated with the smallest Ritz value of (A, M) in $\mathcal{V}$. Knyazev [8] analyzed the convergence behavior of (1.2) by estimating $\rho(x^{(\ell+1)})$. A more general estimate, which improves a classical result for Krylov subspaces by Kaniel, Saad, and Parlett [7, 20, 21], is presented in the recent work [19]. Numerical examples

*Received May 15, 2017. Accepted September 22, 2017. Published online on November 10, 2017. Recommended by D. Szyld.
†Universität Rostock, Institut für Mathematik, Ulmenstraße 69, 18055 Rostock, Germany ({ming.zhou,klaus.neymeyr}@uni-rostock.de).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

show that even the improved estimate is only sharp for $k = 2$. In fact, the special form of the estimate for the two-dimensional case coincides with the estimate for the A-gradient iteration analyzed in [16]. There, the analytic theory is mainly based on geometric arguments. A partial generalization of this geometric analysis is presented in [19] in order to prove an estimate for the angle enclosed by $x^{(\ell)}$ and the eigenspace associated with the smallest eigenvalue. This estimate is sharp for Krylov subspaces of arbitrary dimensions $k \geq 2$. In the present paper we complete this generalization in order to derive a sharp Ritz value estimate for the general cases $k > 2$. The new result is a first step toward a convergence analysis of (1.2) for the case that A^{-1} is substituted by general preconditioners $B^{-1} \approx A^{-1}$; cf. the analysis in [16] which has been generalized in [15] to preconditioned gradient iterations.

1.1. The A-gradient and restarted Krylov subspace iterations. The A-gradient of (1.1) reads

$$A^{-1}\nabla\rho(x) = \big(2/(x^T Mx)\big)\big(x - \rho(x)A^{-1}Mx\big) \in \operatorname{span}\{x, A^{-1}Mx\}.$$

Hence the restarted iteration (1.2) works in subspaces $\mathcal{V}$ of the Krylov subspace

$$\mathcal{K}^k(x^{(\ell)}) \ = \operatorname{span}\{x^{(\ell)}, A^{-1}Mx^{(\ell)}, \ldots, (A^{-1}M)^{k-1}x^{(\ell)}\}.$$

We consider the nontrivial case that the dimension of $\mathcal{V}$ increases by 1 in each step so that $\dim \mathcal{K}^k(x^{(\ell)}) = k$. All this allows us to consider (1.2) as a restarted Krylov subspace iteration

$$x^{(\ell+1)} \quad \leftarrow \quad \mathrm{RR}_{\min}\big(\mathcal{K}^k(x^{(\ell)})\big). \tag{1.3}$$

Alternatively, the iteration can be interpreted as an Invert-Lanczos process [14]. The required matrix-vector products $A^{-1}w$ are computed by solving linear systems in A; see [9, 10, 11]. An approximate solution of these linear systems amounts to an inexact-inverse preconditioning. Block versions of (1.3) serve for the simultaneous computation of several eigenvalues; cf. the block Lanczos algorithm [3, 5].

Classical convergence estimates for eigenvalue and eigenvector approximations in Krylov subspaces have been presented by Kaniel, Saad, and Parlett [7, 20, 21]. A generalization of Theorem 12.4.1 in [20] to the generalized eigenvalue problem for (A, M) provides an upper bound for the approximation error of the smallest Ritz value in $\mathcal{K}^k(x^{(\ell)})$ which depends on differences of eigenvalues and angles between $x^{(\ell)}$ and certain invariant subspaces. This generalization also provides upper bounds for angles between eigenvectors and $\mathcal{K}^k(x^{(\ell)})$. However, these estimates cannot be applied recursively in order to formulate a priori estimates for multiple steps of the restarted Krylov subspace iteration (1.3). For instance, in order to estimate $\rho(x^{(\ell+2)})$ in terms of $x^{(\ell)}$, the classical results cannot provide suitable estimates of $x^{(\ell+1)}$ and $x^{(\ell)}$; cf. [19, Section 2].

For the derivation of an a priori estimate for (1.3), we choose the convergence measure

$$\Delta_{i,i+1}(\theta) = \big(\theta - \lambda_i\big)/\big(\lambda_{i+1} - \theta\big), \tag{1.4}$$

which describes the relative position of an eigenvalue approximation θ with respect to two neighboring eigenvalues $\lambda_i < \lambda_{i+1}$ of (A, M). The ratio (1.4) has been used in various papers on sharp convergence estimates for gradient-type eigensolvers; see [1, 8, 15, 16, 17, 18]. In particular, [8] provides for (1.3) the Ritz value estimate

$$\Delta_{1,2}\big(\rho(x^{(\ell+1)})\big) \leq \big[\,T_{k-1}(1 + 2\gamma_1)\,\big]^{-2}\,\Delta_{1,2}\big(\rho(x^{(\ell)})\big). \tag{1.5}$$

The convergence factor contains a Chebyshev polynomial of degree $k - 1$ and the gap ratio $\gamma_1 := (\lambda_1^{-1} - \lambda_2^{-1})/(\lambda_2^{-1} - \lambda_{\max}^{-1})$ of reciprocal eigenvalues, and it does not depend on the iterate $x^{(\ell)}$. Thus, the a priori estimate $\Delta_{1,2}\big(\rho(x^{(\ell)})\big) \leq \big[\,T_{k-1}(1+2\gamma_1)\,\big]^{-2\ell}\,\Delta_{1,2}\big(\rho(x^{(0)})\big)$ holds. Furthermore, (1.5) gives a tighter bound in comparison to the classical Ritz value estimate from [20]. This fact has been illustrated in [19, Section 2.1] by a numerical example. A generalization concerning arbitrary eigenvalue intervals is given by [19, Theorem 3.5]. There, the generalized estimate concerns the case that $\rho(x^{(\ell)})$ is contained in an arbitrary interval $(\lambda_i, \lambda_{i+1})$ for $i \geq 1$ and has the form

$$\Delta_{i,i+1}\big(\rho(x^{(\ell+1)})\big) \leq [\,T_{k-1}(1+2\gamma_i)\,]^{-2}\,\Delta_{i,i+1}\big(\rho(x^{(\ell)})\big) \tag{1.6}$$

with $\gamma_i := (\lambda_i^{-1} - \lambda_{i+1}^{-1})/(\lambda_{i+1}^{-1} - \lambda_{\max}^{-1})$. The convergence factor $[\,T_{k-1}(1+2\gamma_i)\,]^{-2}$ shows that a larger dimension k or a larger distance between λ_i and λ_{i+1} yields faster convergence. However, the numerical examples in [19] show that the Ritz value estimate (1.6) is not sharp for $k > 2$. The simple explanation is that the convergence factor $[\,T_{k-1}(1+2\gamma_i)\,]^{-2}$ depends only on three eigenvalues λ_i, λ_{i+1}, and λ_m for arbitrary $k > 2$. In this paper our focus is on deriving a sharp Ritz value estimate which depends for increasing k on an increasing number of eigenvalues of (A, M). Our convergence analysis can be simplified considerably by using an A-geometry representation as in [19], i.e., the generalized eigenvalue problem $Ax = \lambda Mx$ is represented by the standard eigenvalue problem $Hy = \lambda^{-1}y$ by considering $y = A^{1/2}x$ and $H = A^{-1/2}MA^{-1/2}$. Thus, we can start with the setting of the following theorem which contains the main results of [19].

THEOREM 1.1. *Let $\mu_1 > \mu_2 > \cdots > \mu_m$ be the distinct eigenvalues of the symmetric and positive definite matrix $H \in \mathbb{R}^{n\times n}$, and let $\mu(\cdot)$ be the Rayleigh quotient with respect to H. Consider a Ritz vector y' associated with the largest Ritz value of H in the Krylov subspace $\mathcal{K} := \mathcal{K}_H^k(y) = \operatorname{span}\{y, Hy, \ldots, H^{k-1}y\}$ with $y \in \mathbb{R}^n\backslash\{0\}$ and $k \geq 2$. If $\mu(y) \in (\mu_{i+1}, \mu_i)$, then it holds that*

$$\frac{\mu_i - \mu(y')}{\mu(y') - \mu_{i+1}} \leq [\,T_{k-1}(1+2\gamma_i)\,]^{-2}\,\frac{\mu_i - \mu(y)}{\mu(y) - \mu_{i+1}} \tag{1.7}$$

with the Chebyshev polynomial T_{k-1} and the gap ratio $\gamma_i := (\mu_i - \mu_{i+1})/(\mu_{i+1} - \mu_m)$. If y is not H-orthogonal to the eigenspace $\mathcal{W}_1$ associated with μ_1, then

$$\tan\angle_H(y', \mathcal{W}_1) \leq \prod_{j=1}^{k-1} \frac{\mu_2 - \mu_{m+1-j}}{\mu_1 - \mu_{m+1-j}}\,\tan\angle_H(y, \mathcal{W}_1). \tag{1.8}$$

Here $\angle_H$ denotes an angle with respect to the inner product induced by H.

1.2. Relation to the analyses by Saad and Knyazev. This paper aims at improving the Ritz value estimate (1.7) for the restarted Krylov subspace iteration (1.3). We extend the ellipsoid analysis from [19] in order to generalize the two-dimensional ellipse analysis from [16]. The new estimate depends on an interpolating polynomial generated by k eigenvalues; see Lemma 3.2. In comparison to the Chebyshev polynomials used in the classical theory on Krylov subspaces, this polynomial is not derived from an optimization with respect to the representation of a Krylov subspace by polynomials, but it is derived from geometry-based arguments. Therefore, our analysis can hardly be compared to the Chebyshev-type analysis by Saad [22, eq. (6.45), (6.53)]. However, a proof technique suggested by Knyazev for deriving (1.5) (cf. [8, eq. (1.9), (1.18)]) is used in the proof of Theorem 3.6, but the argument is applied to an interpolating polynomial instead of a Chebyshev polynomial. Among the k supporting eigenvalues, the smallest and the largest eigenvalue are given explicitly, and the monotonicity of our polynomial is described by the interpolation conditions. Thus, the speed of convergence depends on the distance between two neighboring eigenvalues μ_i and μ_{i+1}. The improved quality of the new estimate in comparison to the Chebyshev-type estimates (1.6) and (1.7) is demonstrated in Figure 3.2 for a model problem and in Tables 4.2 and 4.3 for a high-dimensional PDE eigenproblem.

1.3. Overview of the paper. The remaining part of the paper is structured as follows. The drawback of the dimension reduction technique from [16] applied to Krylov subspaces of dimension $k > 2$ is discussed in Section 2. Then a short review is given of the analytic arguments from [19] which are required here to extend the ellipsoid analysis. The main convergence analysis in

Section 3 provides first an implicit sharp Ritz value estimate in terms of the settings of Theorem 1.1. Monotonicity arguments show that the slowest convergence is attained in a $(k+1)$-dimensional invariant subspace. This allows us to derive explicit convergence estimates. Especially for $k = 3$ an explicit sharp estimate is compared with the Chebyshev-type estimate (1.7) for some numerical examples. Finally, in Section 4 the iteration (1.3) is numerically tested by an adaptive multigrid eigensolver for elliptic operator eigenvalue problems. The improvement of the estimates is illustrated by numerical tests.

2. Auxiliary arguments on Krylov subspaces. In the case $k = 2$ the estimates (1.6) and (1.7) are sharp and coincide with the estimates for the steepest descent/ascent iteration as presented in [16]. A natural way to gain a sharp estimate also for $k > 2$ is to use the same framework and to generalize the arguments to higher-dimensional subspaces. In [16] the analysis starts by deriving necessary conditions on the slowest convergence by differentiating smooth curves on a level set of the Rayleigh quotient. The generalization of this analysis to restarted Krylov subspace iterations shows that the slowest convergence is attained in invariant subspaces which are associated with $2k-1$ distinct eigenvalues. In contrast to this, the sharp Ritz vector estimate (1.8) in Theorem 1.1 depends on only $k+1$ eigenvalues and suggests to use a $(k+1)$-dimensional auxiliary subspace. For this reason we prefer the analysis framework which has been used in [19] for proving the estimate (1.8).

We begin with the setting of Theorem 1.1. If $\mathcal{K}$ is H-invariant, then $\mu(y')$ is an eigenvalue of H and fulfills $\mu(y') \geq \mu(y) > \mu_{i+1}$. This means $\mu(y') \geq \mu_i$ so that the left-hand side of (1.7) is non-positive; the estimate holds trivially. If $\mathcal{K}$ is not H-invariant, then we first consider the case $\mu(y) \in (\mu_2, \mu_1)$ since the auxiliary arguments for (1.8) are related to the extremal eigenvalues. Let $y = \sum_{j=1}^{m} w_j$ be the eigenspace expansion of y, i.e., w_j is the orthogonal projection of y onto the eigenspace associated with μ_j for each j with respect to the Euclidean inner product and also with respect to the inner product induced by H. Here w_1 is nonzero since otherwise $\mu(y) \leq \mu_2$. Then the auxiliary subspace

$$\mathcal{U} := \operatorname{span}\{w_1, \mathcal{K}\} = \operatorname{span}\{w_1, y, Hy, \ldots, H^{k-1}y\} \tag{2.1}$$

can be used for a mini-dimensional analysis. Next we collect some basic properties of $\mathcal{U}$.

LEMMA 2.1. *In the setting of Theorem 1.1 and with $\mu(y) \in (\mu_2, \mu_1)$, let U be an orthonormal matrix whose column space equals $\mathcal{U}$ defined by* (2.1)*. We further define the $\mathcal{U}$-representations*

$$\widehat{H} := U^T H U, \qquad \widehat{y} := U^T y.$$

If $\mathcal{K}$ is not H-invariant, then the following assertions hold.

(a) Left multiplication of $\mathcal{K}$ with U^T results in the Krylov subspace

$$\widehat{\mathcal{K}} := \operatorname{span}\{\widehat{y}, \widehat{H}\widehat{y}, \ldots, \widehat{H}^{k-1}\widehat{y}\}.$$

The pair (θ, v) is a Ritz pair of H in $\mathcal{K}$ if and only if $(\theta, \widehat{v})$ with $\widehat{v} := U^T v$ is a Ritz pair of $\widehat{H}$ in $\widehat{\mathcal{K}}$.

(b) $\mathcal{U}$ has the dimension $k+1$. Thus, there are $k+1$ Ritz values of H in $\mathcal{U}$. These are denoted by $\alpha_1 \geq \cdots \geq \alpha_{k+1}$. The α_i are the eigenvalues of $\widehat{H}$ and are strictly interlaced by the Ritz values $\theta_1 \geq \cdots \geq \theta_k$ of H in $\mathcal{K}$, i.e.,

$$\alpha_1 > \theta_1 > \alpha_2 > \cdots > \alpha_k > \theta_k > \alpha_{k+1}. \tag{2.2}$$

Further, the eigenspace of $\widehat{H}$ associated with the eigenvalue $\alpha_1 = \mu_1$ is the column space of $U^T w_1$.

(c) The eigenvalues $\alpha_1, \ldots, \alpha_{k+1}$ of $\widehat{H}$ are simple due to (2.2)*. Let $\widehat{u}_1, \ldots, \widehat{u}_{k+1}$ be the associated orthonormal eigenvectors and $\widehat{\mu}(\cdot)$ be the Rayleigh quotient with respect to $\widehat{H}$. Then the affine space*

$$\widehat{\mathcal{U}} := \widehat{u}_1 + \operatorname{span}\{\widehat{u}_2, \ldots, \widehat{u}_{k+1}\}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

contains a vector

$$\widetilde{y} := \widehat{u}_1 + \sum_{j=2}^{k+1} \beta_j \widehat{u}_j$$

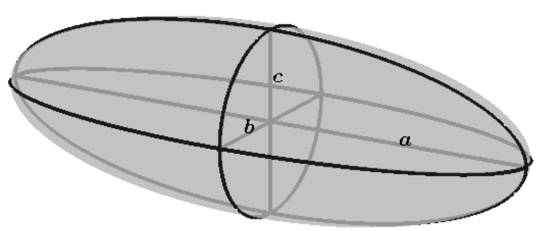
for which $U\widetilde{y}$ is collinear with y. The H-invariance of $\mathcal{K}$ guarantees that all coefficients β_j are nonzero. There exists a further vector $\widetilde{y}'$ in $\widehat{\mathcal{U}}$ for which $(\theta_1, U\widetilde{y}')$ is a Ritz pair of H in $\mathcal{K}$. The level set $\{\widehat{u} \in \widehat{\mathcal{U}}\ ;\ \widehat{\mu}(\widehat{u}) = \theta_1\}$ forms an ellipsoid. Namely, the coefficients $\widehat{\beta}_j$ in the representation $\widehat{u} = \widehat{u}_1 + \sum_{j=2}^{k+1} \widehat{\beta}_j \widehat{u}_j$ fulfill the ellipsoid equation

$$\sum_{j=2}^{k+1} \frac{\widehat{\beta}_j^2}{\dfrac{\alpha_1 - \theta_1}{\theta_1 - \alpha_j}} = 1. \tag{2.3}$$

The proof of this lemma is given in [19, Lemma 3.3 and Lemma 3.4]. In particular, the ellipsoid defined by (2.3) can be derived by the definition of the Rayleigh quotient $\widehat{\mu}(\cdot)$ since

$$\theta_1 = \widehat{\mu}(\widehat{u}) = \frac{\widehat{u}^T \widehat{H} \widehat{u}}{\widehat{u}^T \widehat{u}} = \frac{\alpha_1 + \sum_{j=2}^{k+1} \alpha_j \widehat{\beta}_j^2}{1 + \sum_{j=2}^{k+1} \widehat{\beta}_j^2} \quad \Rightarrow \quad \sum_{j=2}^{k+1} (\theta_1 - \alpha_j)\widehat{\beta}_j^2 = \alpha_1 - \theta_1.$$

The quotients $(\alpha_1 - \theta_1)/(\theta_1 - \alpha_j)$ are positive due to (2.2) and are equal to the squares of the semi-axes of the ellipsoid; see Figure 2.1 for an illustration.



$$a = \sqrt{\frac{\alpha_1 - \theta_1}{\theta_1 - \alpha_2}}$$

$$b = \sqrt{\frac{\alpha_1 - \theta_1}{\theta_1 - \alpha_3}}$$

$$c = \sqrt{\frac{\alpha_1 - \theta_1}{\theta_1 - \alpha_4}}$$

FIG. 2.1. *The ellipsoid defined by (2.3) in the case $k = 3$.*

3. Sharp Ritz value estimates. Lemma 2.1 is the basis for the following advanced ellipsoid analysis, which is inspired by [19]. We use the properties of the auxiliary subspace $\mathcal{U}$ in order to derive a sharp estimate for restarted Krylov subspace iterations. First, Theorem 3.1 provides an estimate in terms of the Ritz values of H in $\mathcal{U}$. In Section 3.2, Lemma 3.2 contains an improved estimate using eigenvalues of H and monotonicity arguments. This result shows that the slowest convergence is attained in a $(k+1)$-dimensional invariant subspace. Additionally, similar estimates concerning arbitrary eigenvalue intervals, namely $\mu(y) \in (\mu_{i+1}, \mu_i)$, for $i = 1, \ldots, m-1$, are formulated in Lemma 3.3. In Section 3.3 we improve the representation of the sharp estimate especially for three-dimensional Krylov subspaces. Moreover, the drawback of the Chebyshev-type estimate (1.7) is discussed and illustrated by a numerical comparison. In Section 3.4 the main results are summarized and restated in explicit form for the generalized eigenvalue problem $Ax = \lambda Mx$.

3.1. An auxiliary estimate. We start with the nontrivial case that the Krylov subspace $\mathcal{K}$ is not H-invariant. The analysis extends the geometry-based one in [19] and generalizes the derivation of a sharp bound for the steepest descent method in [16].

THEOREM 3.1. *Let $\mu_1 > \mu_2 > \cdots > \mu_m$ be the distinct eigenvalues of the symmetric and positive definite matrix $H \in \mathbb{R}^{n\times n}$. The corresponding Rayleigh quotient is $\mu(y) = (y^T Hy)/(y^T y)$. Consider for $y \in \mathbb{R}^n\backslash\{0\}$ the Krylov subspace $\mathcal{K} := \mathcal{K}_H^k(y) = \operatorname{span}\{y, Hy, \ldots, H^{k-1}y\}$, $k > 2$, and the H-projection w_1 of y to the eigenspace associated with μ_1. If $\mathcal{K}$ is not H-invariant and $\mu(y) \in (\mu_2, \mu_1)$, then the subspace $\mathcal{U} := \operatorname{span}\{w_1, \mathcal{K}\}$ is $(k+1)$-dimensional. Let $\alpha_1 \geq \cdots \geq \alpha_{k+1}$ be the Ritz values of H in $\mathcal{U}$. Then the largest Ritz value θ_1 of H in $\mathcal{K}$ satisfies*

$$\left(\frac{\alpha_1-\theta_1}{\theta_1-\alpha_2}\right)\left(\frac{\alpha_1-\mu(y)}{\mu(y)-\alpha_2}\right)^{-1} \leq [\,p(\alpha_1)\,]^{-2}. \tag{3.1}$$

Here $p(\cdot)$ is a polynomial of degree $k-1$ that interpolates the k pairs $\big(\alpha_j, (-1)^j\big)$ for $j = 2, \ldots, k+1$. Equality is attained in the limit case $\mu(y) \to \alpha_1$.

Proof. Lemma 2.1 proves $\dim\mathcal{U} = k+1$. The proof of (3.1) is given in five steps.

(i) We represent y and a Ritz vector associated with θ_1 by using the auxiliary vectors $\widetilde{y}$, $\widetilde{y}\,' \in \widehat{\mathcal{U}}$ as introduced in Lemma 2.1. Then the $\widehat{\mathcal{U}}$-representation of the Krylov subspace $\mathcal{K}$ is shown to be a hyperplane in $\mathbb{R}^k$.

(ii) A representation of the hyperplane from (i) is derived depending on the Ritz values $\alpha_1, \ldots, \alpha_{k+1}$ and the coefficients $\beta_2, \ldots, \beta_{k+1}$ of $\widetilde{y}$ from Lemma 2.1.

(iii) We represent $\widetilde{y}\,'$ by using the geometric property of the ellipsoid (2.3).

(iv) An intermediate bound for the left-hand side of (3.1) is proved concerning the $\widehat{\mathcal{U}}$-coefficients of $\widetilde{y}$ and $\widetilde{y}\,'$.

(v) An upper bound is proved for the intermediate bound by (iv). This bound is shown to be equal to $[\,p(\alpha_1)\,]^{-2}$ with the interpolating polynomial $p(\cdot)$.

We now go into detail.

(i) According to Lemma 2.1, we use the Rayleigh quotient $\widehat{\mu}(\cdot)$ with respect to the matrix $\widehat{H} := U^T HU$. It holds that

$$\widehat{\mu}(\widetilde{y}) = \widehat{\mu}(U^T y) = \frac{(U^T y)^T (U^T HU)(U^T y)}{(U^T y)^T (U^T y)} = \frac{y^T (UU^T) H (UU^T) y}{y^T (UU^T) y}.$$

Since $y, Hy \in \mathcal{U}$ and UU^T is the projection matrix onto $\mathcal{U}$, it holds that $(UU^T)y = y$ and $(UU^T)H(UU^T)y = Hy$ so that $\widehat{\mu}(\widetilde{y}) = (y^T Hy)/(y^T y) = \mu(y)$. Further, (a) in Lemma 2.1 shows that $(\theta_1, \widetilde{y}\,')$ is a Ritz pair of $\widehat{H}$ in $\widehat{\mathcal{K}}$ since $(\theta_1, U\widetilde{y}\,')$ is a Ritz pair of H in $\mathcal{K}$, and since $\widetilde{y}\,' = U^T(U\widetilde{y}\,')$. Moreover, $\theta_1 = \widehat{\mu}(\widetilde{y}\,')$ is the largest Ritz value of $\widehat{H}$ in $\widehat{\mathcal{K}}$; i.e., the maximum of $\widehat{\mu}(\cdot)$ in $\widehat{\mathcal{K}}$. Thus, the analysis can be restricted to the Krylov subspace $\widehat{\mathcal{K}}$.

Due to $\widetilde{y}$, $\widetilde{y}\,' \in \widehat{\mathcal{U}}$, we define for each $\widehat{u} \in \widehat{\mathcal{U}}$ the coefficient vector

$$P\widehat{u} := (\widehat{\beta}_2, \ldots, \widehat{\beta}_{k+1})^T \in \mathbb{R}^k$$

with respect to the expansion $\widehat{u} = \widehat{u}_1 + \sum_{j=2}^{k+1} \widehat{\beta}_j \widehat{u}_j$. For each subspace $\widehat{\mathcal{V}} \subseteq \widehat{\mathcal{U}}$ we define $P\widehat{\mathcal{V}} := \{P\widehat{u}\ ;\ \widehat{u} \in \widehat{\mathcal{V}}\}$. Then we consider the level set $\mathcal{S} := \{\widehat{u} \in \widehat{\mathcal{U}}\ ;\ \widehat{\mu}(\widehat{u}) = \theta_1\}$ and the intersection $\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}}$. Their $\widehat{\mathcal{U}}$-representations $P\mathcal{S}$ and $P(\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}})$ can be interpreted geometrically. The set $P\mathcal{S}$ is the ellipsoid given by (2.3). The convexity of the ellipsoid and the fact that $\widetilde{y}\,'$ maximizes the Rayleigh quotient $\widehat{\mu}(\cdot)$ in $\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}}$ show that $P(\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}})$ is a tangential hyperplane of $P\mathcal{S}$; see Figure 3.1. The point of tangency is $P\widetilde{y}\,'$.

(ii) The hyperplane $P(\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}})$ can be represented by $P\widehat{u} + \mathcal{W}$ with an arbitrary $\widehat{u} \in \widehat{\mathcal{U}} \cap \widehat{\mathcal{K}}$ and a $(k-1)$-dimensional subspace $\mathcal{W} \subset \mathbb{R}^k$. Suitable $\widehat{u}$ and $\mathcal{W}$ can be constructed by using the intersection points v_i of $\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}}$ and the axes $\widehat{u}_1 + \operatorname{span}\{\widehat{u}_i\}$ with

$$v_i := (\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}}) \cap (\widehat{u}_1 + \operatorname{span}\{\widehat{u}_i\}), \quad i = 2, \ldots, k+1.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

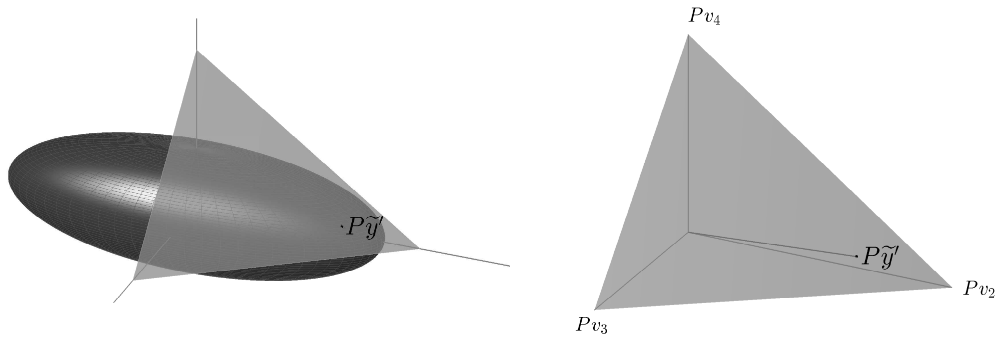


FIG. 3.1. *Geometry in the affine space $\widehat{\mathcal{U}}$ in the case $k = 3$. Left: ellipsoid $P\mathcal{S}$ and its tangential hyperplane $P(\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}})$. Right: representation of the tangential hyperplane by its intersection points with the coordinate axes.*

Due to $\widehat{u}_1 + \operatorname{span}\{\widehat{u}_i\} \subset \widehat{\mathcal{U}}$, it holds that $v_i = \widehat{\mathcal{K}} \cap (\widehat{u}_1 + \operatorname{span}\{\widehat{u}_i\})$. Since $\widetilde{y}$ is collinear with $\widehat{y}$ (as $U\widetilde{y}$ and $y = U\widehat{y}$ are collinear) and $v_i \in \widehat{\mathcal{K}} = \operatorname{span}\{\widehat{y}, \widehat{H}\widehat{y}, \ldots, \widehat{H}^{k-1}\widehat{y}\}$, the vectors v_i can be represented by $v_i = q_i(\widehat{H})\widetilde{y}$ for certain polynomials $q_i(\cdot)$ whose degrees are less than or equal to $k-1$. Next, we determine the explicit form of $q_i(\cdot)$. With the expansion $\widetilde{y} = \widehat{u}_1 + \sum_{j=2}^{k+1} \beta_j \widehat{u}_j$ from Lemma 2.1, we get

$$v_i = q_i(\widehat{H})\widetilde{y} = q_i(\widehat{H}) \left(\widehat{u}_1 + \sum_{j=2}^{k+1} \beta_j \widehat{u}_j \right) = q_i(\alpha_1)\widehat{u}_1 + \sum_{j=2}^{k+1} \beta_j q_i(\alpha_j) \widehat{u}_j.$$

Together with $v_i \in \widehat{u}_1 + \operatorname{span}\{\widehat{u}_i\}$ and $\beta_j \neq 0$, we conclude that

$$q_i(\alpha_1) = 1 \quad \text{and} \quad q_i(\alpha_j) = 0 \quad \text{for} \quad j = 2, \ldots, k+1, \quad j \neq i.$$

This interpolation problem in the k pairwise different nodes $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_{k+1}$ is uniquely solved by the Lagrange polynomial

$$q_i(\alpha) = \prod_{j=2,\ j\neq i}^{k+1} \frac{\alpha - \alpha_j}{\alpha_1 - \alpha_j}.$$

This yields

$$v_i = \widehat{u}_1 + \beta_i q_i(\alpha_i)\widehat{u}_i = \widehat{u}_1 + \beta_i \prod_{j=2,\ j\neq i}^{k+1} \frac{\alpha_i - \alpha_j}{\alpha_1 - \alpha_j}\, \widehat{u}_i.$$

The corresponding $\widehat{\mathcal{U}}$-representations are

$$Pv_i := \beta_i \kappa_i e_{i-1} \quad \text{with} \quad \kappa_i := \prod_{j=2,\ j\neq i}^{k+1} \frac{\alpha_i - \alpha_j}{\alpha_1 - \alpha_j}, \quad i = 2, \ldots, k+1, \tag{3.2}$$

where $e_1, \ldots, e_k \in \mathbb{R}^k$ are the standard basis vectors. By using these k intersection points, we obtain a representation $P\widehat{u} + \mathcal{W}$ of $P(\widehat{\mathcal{U}} \cap \widehat{\mathcal{K}})$ with $\mathcal{W} = \operatorname{span}\{Pv_3 - Pv_2, \ldots, Pv_{k+1} - Pv_2\}$ and $\widehat{u} = v_2$; see Figure 3.1. Here $Pv_i = \beta_i \kappa_i e_{i-1}$ and $\beta_i \kappa_i \neq 0$ prove that $\dim \mathcal{W} = k - 1$.

(iii) The representation $P\widehat{u} + \mathcal{W}$ from (ii) allows us to describe $\widetilde{y}\,'$ componentwise: we interpret the ellipsoid $P\mathcal{S}$ as the unit sphere in $\mathbb{R}^k$ with respect to the norm $\|\cdot\|_D$ induced by the diagonal matrix $D = \operatorname{diag}(\delta_2, \ldots, \delta_{k+1}) \in \mathbb{R}^{k\times k}$ with

$$\delta_j = \left(\frac{\alpha_1 - \theta_1}{\theta_1 - \alpha_j} \right)^{-1}, \quad j = 2, \ldots, k+1. \tag{3.3}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Since $P\widetilde{y}'$ is the point of tangency associated with $P\mathcal{S}$ and the tangential hyperplane $P\widehat{u}+\mathcal{W}$, the vector $P\widetilde{y}'$ is orthogonal to the subspace $\mathcal{W}$ with respect to the inner product $(\cdot,\cdot)_D$ induced by D. Because of $\dim\mathcal{W}=k-1$, the D-orthogonal complement of $\mathcal{W}$ in $\mathbb{R}^k$ is one-dimensional so that $P\widetilde{y}'$ is collinear with any $v\in\mathbb{R}^k\backslash\{0\}$ satisfying $v\perp_D\mathcal{W}$. Such a vector v can easily be determined by using the basis vectors Pv_i-Pv_2 of $\mathcal{W}$. We set, e.g., $v=(\gamma_1,\ldots,\gamma_k)^T$ and solve the system of equations

$$(v,\,Pv_i-Pv_2)_D=0,\quad i=2,\ldots,k+1.$$

Since $D=\operatorname{diag}(\delta_2,\ldots,\delta_{k+1})$ and $Pv_i=\beta_i\kappa_i e_{i-1}$, these equations have the detailed form

$$\gamma_{i-1}\,\delta_i\,(\beta_i\kappa_i)-\gamma_1\,\delta_2\,(\beta_2\kappa_2)=0,\quad i=2,\ldots,k+1.$$

A particular solution of this linear system with degenerate rank has, for $j=1,\ldots,k$, the components $\gamma_j=(\delta_{j+1}\beta_{j+1}\kappa_{j+1})^{-1}$ so that $P\widetilde{y}'$ is collinear to

$$v=\left(\frac{1}{\delta_2\beta_2\kappa_2},\ldots,\frac{1}{\delta_{k+1}\beta_{k+1}\kappa_{k+1}}\right)^T. \tag{3.4}$$

(iv) In order to apply the detailed representation (3.4), we derive for the left-hand side of (3.1) an intermediate bound that is related to $P\widetilde{y}$ and $P\widetilde{y}'$. First, (b) in Lemma 2.1 and the Courant-Fischer principles show that $\mu_1=\alpha_1>\mu_2\geq\alpha_2$. Then $\mu(y)\in(\mu_2,\mu_1)$ and $\widehat{\mu}(\widetilde{y})=\mu(y)$ yield $\widehat{\mu}(\widetilde{y})\in(\alpha_2,\alpha_1)$. Next, the expansion $\widetilde{y}=\widehat{u}_1+\sum_{j=2}^{k+1}\beta_j\widehat{u}_j$ gives the equation

$$\widehat{\mu}(\widetilde{y})=\frac{\alpha_1+\sum_{j=2}^{k+1}\alpha_j\beta_j^2}{1+\sum_{j=2}^{k+1}\beta_j^2}$$

so that $\alpha_1-\widehat{\mu}(\widetilde{y})=\sum_{j=2}^{k+1}\beta_j^2\big(\widehat{\mu}(\widetilde{y})-\alpha_j\big)$ and

$$\frac{\alpha_1-\mu(y)}{\mu(y)-\alpha_2}=\frac{\alpha_1-\widehat{\mu}(\widetilde{y})}{\widehat{\mu}(\widetilde{y})-\alpha_2}=\sum_{j=2}^{k+1}{\beta_j}^2\left(\frac{\widehat{\mu}(\widetilde{y})-\alpha_j}{\widehat{\mu}(\widetilde{y})-\alpha_2}\right)\geq\sum_{j=2}^{k+1}{\beta_j}^2\left(\frac{\theta_1-\alpha_j}{\theta_1-\alpha_2}\right) \tag{3.5}$$

by using $\theta_1\geq\widehat{\mu}(\widetilde{y})$ and the monotonicity of the function $f(a)=(\alpha-\alpha_j)/(\alpha-\alpha_2)$. Thus,

$$\left(\frac{\alpha_1-\theta_1}{\theta_1-\alpha_2}\right)\left(\frac{\alpha_1-\mu(y)}{\mu(y)-\alpha_2}\right)^{-1}\leq\left(\frac{\alpha_1-\theta_1}{\theta_1-\alpha_2}\right)\left(\sum_{j=2}^{k+1}{\beta_j}^2\left(\frac{\theta_1-\alpha_j}{\theta_1-\alpha_2}\right)\right)^{-1}$$

$$=\left(\sum_{j=2}^{k+1}{\beta_j}^2\left(\frac{\theta_1-\alpha_j}{\alpha_1-\theta_1}\right)\right)^{-1}\overset{(3.3)}{=}\left(\sum_{j=2}^{k+1}\delta_j{\beta_j}^2\right)^{-1}=\|P\widetilde{y}\|_D^{-2}.$$

However, the intermediate bound $\|P\widetilde{y}\|_D^{-2}$ depends on the semi-axes $\delta_2,\ldots,\delta_{k+1}$, whereas the bound in (3.1) only depends on the Ritz values of H in $\mathcal{U}$. Hence, a further reformulation is necessary. Since $P\widetilde{y},\,P\widetilde{y}'\in P(\widehat{\mathcal{U}}\cap\widehat{\mathcal{K}})=P\widehat{u}+\mathcal{W}$, it holds that $P\widetilde{y}-P\widetilde{y}'\in\mathcal{W}$. Then the orthogonality $P\widetilde{y}'\perp_D\mathcal{W}$ implies that

$$(P\widetilde{y}-P\widetilde{y}',\,P\widetilde{y}')_D=0\quad\Rightarrow\quad(P\widetilde{y},\,P\widetilde{y}')_D=\|P\widetilde{y}'\|_D^2=1.$$

Here the last equality holds since $P\widetilde{y}'$ belongs to the unit sphere with respect to $\|\cdot\|_D$; see (iii). Furthermore, we get

$$\cos\angle_D(P\widetilde{y},\,P\widetilde{y}')=\frac{(P\widetilde{y},\,P\widetilde{y}')_D}{\|P\widetilde{y}\|_D\|P\widetilde{y}'\|_D}=\frac{1}{\|P\widetilde{y}\|_D}.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Therefore, the intermediate bound $\|P\widetilde{y}\|_D^{-2}$ coincides with $\cos^2\angle_D(P\widetilde{y},\, P\widetilde{y}\,')$ and, for v from (3.4), with $\cos^2\angle_D(P\widetilde{y},\, v)$ since $P\widetilde{y}\,'$ is collinear with v.

(v) Finally, $\cos^2\angle_D(P\widetilde{y},\, v)$ needs to be compared with a further bound which only depends on $\alpha_1,\dots,\alpha_{k+1}$. We begin with the representation

$$\cos^2\angle_D(P\widetilde{y}, v) = \frac{(P\widetilde{y},\, v)_D^2}{\|P\widetilde{y}\|_D^2\|v\|_D^2} \overset{(3.4)}{=} \frac{\left(\sum_{j=2}^{k+1}\kappa_j^{-1}\right)^2}{\left(\sum_{j=2}^{k+1}\delta_j\beta_j^2\right)\left(\sum_{j=2}^{k+1}\delta_j^{-1}\beta_j^{-2}\kappa_j^{-2}\right)}.$$

The denominator can be bounded from below by using the Cauchy-Schwarz inequality,

$$\left(\sum_{j=2}^{k+1}\delta_j\beta_j^2\right)^{1/2}\left(\sum_{j=2}^{k+1}\delta_j^{-1}\beta_j^{-2}\kappa_j^{-2}\right)^{1/2} \ge \sum_{j=2}^{k+1}\left(\delta_j^{1/2}\beta_j\right)\left(\delta_j^{-1/2}\beta_j^{-1}|\kappa_j|^{-1}\right) = \sum_{j=2}^{k+1}|\kappa_j|^{-1}.$$

This yields

$$\cos^2\angle_D(P\widetilde{y},\, v) \le \frac{\left(\sum_{j=2}^{k+1}\kappa_j^{-1}\right)^2}{\left(\sum_{j=2}^{k+1}|\kappa_j|^{-1}\right)^2}.$$

According to the definition of $\kappa_2,\dots,\kappa_{k+1}$ in (3.2), we consider the polynomials

$$l_j(\alpha) = \prod_{i=2,\ i\neq j}^{k+1}\frac{\alpha-\alpha_i}{\alpha_j-\alpha_i}, \quad j = 2,\dots,k+1,$$

which are the Lagrange basis polynomials with respect to $\alpha_2,\dots,\alpha_{k+1}$. Then the sum $\sum_{j=2}^{k+1} l_j(\cdot)$ is a constant function equal to 1 so that $\sum_{j=2}^{k+1}\kappa_j^{-1} = \sum_{j=2}^{k+1} l_j(\alpha_1) = 1$. Furthermore, the relation $\alpha_1 < \alpha_2 < \cdots < \alpha_{k+1}$ shows that $|\kappa_j|^{-1} = (-1)^j\kappa_j^{-1}$. Thus,

$$\sum_{j=2}^{k+1}|\kappa_j|^{-1} = \sum_{j=2}^{k+1}(-1)^j\kappa_j^{-1} = \sum_{j=2}^{k+1}(-1)^j l_j(\alpha_1) = p(\alpha_1)$$

with the interpolating polynomial $p(\cdot)$ for the pairs $\big(\alpha_j, (-1)^j\big)$, $j = 2,\dots,k+1$, so that $\cos^2\angle_D(P\widetilde{y},\, v) \le [\,p(\alpha_1)\,]^{-2}$. Combining this with (iv) results in the estimate (3.1).

In the limit case $\mu(y)\to\alpha_1$, which means that $\widehat{\mu}(\widetilde{y})\to\alpha_1$ and $\theta_1\to\alpha_1$, the inequality in (3.5) for the intermediate bound in (iv) turns into an equality. Additionally, the Cauchy-Schwarz inequality applied in (v) turns into an equality by setting suitable coefficients β_j such that the two corresponding vectors are collinear. Thus, equality in (3.1) can be attained. □

3.2. An improved estimate in terms of the eigenvalues of H. In order to improve the usefulness and applicability of the estimate (3.1), we use the Courant-Fischer principles to bound the Ritz values $\alpha_1,\dots,\alpha_{k+1}$ in terms of the eigenvalues of H.

LEMMA 3.2. *Let $\mu_1 > \mu_2 > \cdots > \mu_m$ be the distinct eigenvalues of the symmetric and positive definite matrix $H \in \mathbb{R}^{n\times n}$, and let $\mu(\cdot)$ be the Rayleigh quotient with respect to H. Consider a Ritz vector y' associated with the largest Ritz value of H in the Krylov subspace $\mathcal{K} := \mathcal{K}_H^k(y) = \operatorname{span}\{y, Hy, \dots, H^{k-1}y\}$ with $y \in \mathbb{R}^n\backslash\{0\}$ and $k > 2$. If $\mu(y) \in (\mu_2, \mu_1)$, then it holds that*

$$\frac{\mu_1-\mu(y')}{\mu(y')-\mu_2} \le \left(\min_J\ p_J(\mu_1)\right)^{-2}\ \frac{\mu_1-\mu(y)}{\mu(y)-\mu_2}, \tag{3.6}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where J runs through all $(k-2)$-element subsets of $\{3,\dots,m-1\}$. Further, $p_J(\cdot)$ is a polynomial of degree $k-1$ through the k points $\big(\mu_2,1\big)$, $\big(\mu_m,(-1)^{k+1}\big)$, and $\big(\mu_{\sigma(j)},(-1)^j\big)$, $j=3,\dots,k$, with the indices $\sigma(j)\in J$ in increasing order.

Proof. If $\mathcal{K}$ is H-invariant, then one can easily show that $\mu(y')=\mu_1$. Thus, (3.6) holds trivially. If $\mathcal{K}$ is not H-invariant, then Theorem 3.1 provides the auxiliary estimate (3.1). Because of $\mu_1=\alpha_1>\mu_2\geq\alpha_2$ and $\mu(y')=\theta_1$, we have

$$\left(\frac{\mu_1-\mu(y')}{\mu(y')-\mu_2}\right)\left(\frac{\mu_1-\mu(y)}{\mu(y)-\mu_2}\right)^{-1}\leq\left(\frac{\alpha_1-\theta_1}{\theta_1-\alpha_2}\right)\left(\frac{\alpha_1-\mu(y)}{\mu(y)-\alpha_2}\right)^{-1}\leq[\,p(\alpha_1)\,]^{-2}=[\,p(\mu_1)\,]^{-2}.$$

The interpolation conditions $p(\alpha_j)=(-1)^j$, $j=2,\dots,k+1$, together with the relation $\mu_1=\alpha_1>\alpha_2>\cdots>\alpha_{k+1}$ imply that $p(\cdot)$ has all its roots in (α_{k+1},α_2). Then $p(\mu_1)$ has the same sign as $p(\alpha_2)$ so that $p(\mu_1)>0$. Thus, we can maximize $[\,p(\mu_1)\,]^{-2}$ by minimizing $p(\mu_1)$. To prove (3.6) we consider the function $f(\alpha_2,\alpha_{k+1}):=p(\mu_1)$ and compare the values $f(\alpha_2,\alpha_{k+1})$, $f(\mu_2,\alpha_{k+1})$, and $f(\mu_2,\mu_m)$ on condition that $\mu_1>\mu_2\geq\alpha_2>\cdots>\alpha_{k+1}\geq\mu_m$.

In order to show $f(\alpha_2,\alpha_{k+1})\geq f(\mu_2,\alpha_{k+1})$, we use the Newton form of the interpolating polynomial $p(\cdot)$ derived from the table

$$\begin{array}{cccccc}
\alpha_{k+1} & (-1)^{k+1} & & & & \\
\alpha_k & (-1)^k & \delta^{(1)}_{k,k+1} & & & \\
\vdots & \vdots & \vdots & \ddots & & \\
\alpha_3 & -1 & \delta^{(1)}_{3,4} & \ddots & \ddots & \\
\alpha_2 & 1 & \delta^{(1)}_{2,3} & \delta^{(2)}_{2,4} & \cdots & \delta^{(k-1)}_{2,k+1}
\end{array}$$

with the divided differences

$$\delta^{(1)}_{i,i+1}:=\frac{2\,(-1)^i}{\alpha_i-\alpha_{i+1}},\qquad \delta^{(j+1)}_{i,i+j+1}:=\frac{\delta^{(j)}_{i,i+j}-\delta^{(j)}_{i+1,i+j+1}}{\alpha_i-\alpha_{i+j+1}}. \tag{3.7}$$

The first $k-1$ rows define a polynomial $q(\cdot)$ whose coefficients do not depend on α_2. Furthermore, the polynomial $p(\cdot)$ has the form $p(\alpha)=q(\alpha)+\delta^{(k-1)}_{2,k+1}\prod_{i=3}^{k+1}(\alpha-\alpha_i)$ so that

$$f(\alpha_2,\alpha_{k+1})=p(\mu_1)=q(\mu_1)+\delta^{(k-1)}_{2,k+1}\prod_{i=3}^{k+1}(\mu_1-\alpha_i).$$

To represent $f(\mu_2,\alpha_{k+1})$ in a similar way, we substitute α_2 with μ_2 in the Newton table. Then only the last row is changed. The new quotients are

$$\widetilde{\delta}^{(1)}_{2,3}:=\frac{2}{\mu_2-\alpha_3},\qquad \widetilde{\delta}^{(j+1)}_{2,j+3}:=\frac{\widetilde{\delta}^{(j)}_{2,j+2}-\delta^{(j)}_{3,j+3}}{\mu_2-\alpha_{j+3}}$$

so that

$$f(\mu_2,\alpha_{k+1})=q(\mu_1)+\widetilde{\delta}^{(k-1)}_{2,k+1}\prod_{i=3}^{k+1}(\mu_1-\alpha_i).$$

Because of $\mu_1>\alpha_2>\cdots>\alpha_{k+1}$, the product $\prod_{i=3}^{k+1}(\mu_1-\alpha_i)$ is positive. Thus, we can verify $f(\alpha_2,\alpha_{k+1})\geq f(\mu_2,\alpha_{k+1})$ by means of the inequality $\delta^{(k-1)}_{2,k+1}\geq\widetilde{\delta}^{(k-1)}_{2,k+1}$. This is shown below.

First, we get by induction that $\delta^{(j)}_{i,i+j} > 0$ for even i and $\delta^{(j)}_{i,i+j} < 0$ for odd i. In particular, it holds that $\delta^{(j)}_{3,j+3} < 0$. Next, a successive comparison of the two versions of the last row leads to

$$\mu_2 \geq \alpha_2 > \alpha_3 \quad \Rightarrow \quad \delta^{(1)}_{2,3} = \frac{2}{\alpha_2 - \alpha_3} \geq \frac{2}{\mu_2 - \alpha_3} = \widetilde{\delta}^{(1)}_{2,3} > 0,$$

$$\left.\begin{array}{r} \mu_2 \geq \alpha_2 > \alpha_{j+3} \\ \delta^{(j)}_{2,j+2} \geq \widetilde{\delta}^{(j)}_{2,j+2} > 0 \\ \delta^{(j)}_{3,j+3} < 0 \end{array}\right\} \quad \Rightarrow \quad \delta^{(j+1)}_{2,j+3} = \frac{\delta^{(j)}_{2,j+2} - \delta^{(j)}_{3,j+3}}{\alpha_2 - \alpha_{j+3}} \geq \frac{\widetilde{\delta}^{(j)}_{2,j+2} - \delta^{(j)}_{3,j+3}}{\mu_2 - \alpha_{j+3}} = \widetilde{\delta}^{(j+1)}_{2,j+3} > 0.$$

This proves that $\delta^{(k-1)}_{2,k+1} \geq \widetilde{\delta}^{(k-1)}_{2,k+1}$ inductively.

In order to show $f(\mu_2, \alpha_{k+1}) \geq f(\mu_2, \mu_m)$, we modify the Newton table in the form

$$\begin{array}{cccccc} \mu_2 & 1 & & & & \\ \alpha_3 & -1 & \delta^{(1)}_{2,3} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ \alpha_k & (-1)^k & \delta^{(1)}_{k-1,k} & \ddots & \ddots & \\ \alpha_{k+1} & (-1)^{k+1} & \delta^{(1)}_{k,k+1} & \delta^{(2)}_{k-1,k+1} & \cdots & \delta^{(k-1)}_{2,k+1} \end{array} .$$

Here we use again (3.7) to form the divided differences and substitute α_2 by μ_2. As above, we compare the last row with its variant corresponding to μ_m. The variant has the new quotients

$$\widehat{\delta}^{(1)}_{k,k+1} := \frac{2\,(-1)^k}{\alpha_k - \mu_m}, \qquad \widehat{\delta}^{(j+1)}_{k-j,k+1} := \frac{\delta^{(j)}_{k-j,k} - \widehat{\delta}^{(j)}_{k-j+1,k+1}}{\alpha_{k-j} - \mu_m}.$$

Then $f(\mu_2, \alpha_{k+1}) \geq f(\mu_2, \mu_m)$ is verified by showing $\delta^{(k-1)}_{2,k+1} \geq \widehat{\delta}^{(k-1)}_{2,k+1}$. Here it holds again that $\delta^{(j)}_{i,i+j} > 0$ for even i and $\delta^{(j)}_{i,i+j} < 0$ for odd i. In the case of even k, we have

$$\alpha_k > \alpha_{k+1} \geq \mu_m \quad \Rightarrow \quad \delta^{(1)}_{k,k+1} = \frac{2}{\alpha_k - \alpha_{k+1}} \geq \frac{2}{\alpha_k - \mu_m} = \widehat{\delta}^{(1)}_{k,k+1} > 0,$$

$$\delta^{(j+1)}_{k-j,k+1} = \frac{\delta^{(j)}_{k-j,k} - \delta^{(j)}_{k-j+1,k+1}}{\alpha_{k-j} - \alpha_{k+1}} \leq \frac{\delta^{(j)}_{k-j,k} - \widehat{\delta}^{(j)}_{k-j+1,k+1}}{\alpha_{k-j} - \mu_m} = \widehat{\delta}^{(j+1)}_{k-j,k+1} < 0 \text{ for odd } j, \text{ and}$$

$$\delta^{(j+1)}_{k-j,k+1} = \frac{\delta^{(j)}_{k-j,k} - \delta^{(j)}_{k-j+1,k+1}}{\alpha_{k-j} - \alpha_{k+1}} \geq \frac{\delta^{(j)}_{k-j,k} - \widehat{\delta}^{(j)}_{k-j+1,k+1}}{\alpha_{k-j} - \mu_m} = \widehat{\delta}^{(j+1)}_{k-j,k+1} > 0 \text{ for even } j$$

so that $\delta^{(k-1)}_{2,k+1} \geq \widehat{\delta}^{(k-1)}_{2,k+1}$. For odd k a similar induction yields the same result.

In summary, we have shown that $f(\alpha_2, \alpha_{k+1}) \geq f(\mu_2, \alpha_{k+1}) \geq f(\mu_2, \mu_m)$. In order to improve the lower bound $f(\mu_2, \mu_m)$, we can consider it as a function of $\alpha_3, \ldots, \alpha_k$ and determine its discrete minimum in the set $\{\mu_3, \ldots, \mu_{m-1}\}^{k-2} \subset \mathbb{R}^{k-2}$ under the constraint $\alpha_3 > \cdots > \alpha_k$. This results in the interpolating polynomial $p_J(\cdot)$ in the assertion. □

An optimal set J in $\big(\min_J \, p_J(\mu_1)\big)^{-2}$ can be determined by solving a discrete minimization problem. Then the analytical approach of Theorem 3.1 is applicable by considering an invariant subspace corresponding to J instead of the subspace $\mathcal{U}$. The resulting bound is sharp. This additionally shows that the slowest convergence can be attained in a $(k{+}1)$-dimensional invariant subspace. For the case $\mu(y) \in (\mu_{i+1}, \mu_i)$, a similar analysis can be formulated with some additional assumptions on y and the Ritz value α_2.

LEMMA 3.3. *Let $\mu_1 > \mu_2 > \cdots > \mu_m$ be the distinct eigenvalues of the symmetric and positive definite matrix $H \in \mathbb{R}^{n\times n}$, and let $\mu(\cdot)$ be the Rayleigh quotient with respect to H.*

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Consider the Krylov subspace $\mathcal{K} := \mathcal{K}_H^k(y) = \operatorname{span}\{y, Hy, \ldots, H^{k-1}y\}$ with $y \in \mathbb{R}^n \backslash \{0\}$ and $k > 2$, where the H-projection w_i of y onto the eigenspace associated with μ_i is nonzero. If $\mathcal{K}$ is not H-invariant and $\mu(y) \in (\mu_{i+1}, \mu_i)$, then the subspace $\mathcal{U} = \operatorname{span}\{w_i, \mathcal{K}\}$ is $(k{+}1)$-dimensional and has $k{+}1$ Ritz values of H which are denoted by $\alpha_1 \geq \cdots \geq \alpha_{k+1}$. Let y' be a Ritz vector associated with the largest Ritz value θ_1 of H in $\mathcal{K}$. If $\mu_{i+1} \geq \alpha_2$, then the estimate (3.1) holds. Furthermore, $\mu(y') \geq \mu_i$ holds trivially for $i > m - k$. In the case $i \leq m - k$ it holds that

$$\frac{\mu_i - \mu(y')}{\mu(y') - \mu_{i+1}} \leq \Big(\min_J \, p_J(\mu_i)\Big)^{-2} \, \frac{\mu_i - \mu(y)}{\mu(y) - \mu_{i+1}},$$

where J runs through all $(k{-}2)$-element subsets of $\{i{+}2, \ldots, m{-}1\}$. Further, $p_J(\cdot)$ is a polynomial of degree $k{-}1$ that interpolates the pairs $\big(\mu_{i+1}, 1\big)$, $\big(\mu_m, (-1)^{k+1}\big)$, and $\big(\mu_{\sigma(j)}, (-1)^j\big)$, $j = 3, \ldots, k$, with the indices $\sigma(j) \in J$ in increasing order.

Proof. With the additional assumptions $w_i \neq 0$ and $\mu_{i+1} \geq \alpha_2$, minor modifications of Lemma 2.1 and Theorem 3.1 result in (3.1). The new estimate can be proved by modifying the proof of Lemma 3.2 and using $\mu_i = \alpha_1 > \mu_{i+1} \geq \alpha_2$. ∎

3.3. Explicit and J-minimization-free estimates. An obvious drawback of the estimates of Lemma 3.2 and Lemma 3.3 is that the bounds are non-explicit, i.e., they depend on a minimization with respect to the index set J. We prefer a description that explicitly depends on the relevant eigenvalues. Such explicit estimates can be proved at least for three-dimensional Krylov subspaces. In order to avoid the additional assumptions on y and the Ritz value α_2 in Lemma 3.3, we start with the case $\mu(y) \in (\mu_2, \mu_1)$. First we adapt Lemma 3.2.

LEMMA 3.4. *In the setting of Lemma 3.2 for $k = 3$, it holds that*

$$\frac{\mu_1 - \mu(y')}{\mu(y') - \mu_2} \leq \big[\, q(\mu_1)\,\big]^{-2} \, \frac{\mu_1 - \mu(y)}{\mu(y) - \mu_2}.$$

Here $q(\cdot)$ is a quadratic polynomial which interpolates the pairs $\big(\mu_2, 1\big)$, $\big(\mu_\xi, -1\big)$, and $\big(\mu_m, 1\big)$, and μ_ξ is an eigenvalue which has the smallest distance to $(\mu_2 + \mu_m)/2$ among the eigenvalues $\mu_3, \ldots, \mu_{m-1}$, where we select the larger one as μ_ξ if there are two such eigenvalues nearest to $(\mu_2 + \mu_m)/2$.

Proof. We start with the estimate (3.6) from Lemma 3.2. In the case $k = 3$, the polynomial $p_J(\cdot)$ in the convergence factor $\big(\min_J \, p_J(\mu_1)\big)^{-2}$ is a quadratic polynomial interpolating the pairs $\big(\mu_2, 1\big)$, $\big(\mu_m, 1\big)$, and $\big(\mu_\xi, -1\big)$ with $\mu_\xi \in \{\mu_3, \ldots, \mu_{m-1}\}$. In order to determine the optimal $p_J(\cdot)$, we define

$$f(\mu_\xi) := p_J(\mu_1) = \frac{2\mu_1^2 - 2(\mu_2 + \mu_m)\mu_1 + (\mu_2\mu_\xi + \mu_\xi\mu_m + \mu_2\mu_m - \mu_\xi^2)}{(\mu_2 - \mu_\xi)(\mu_\xi - \mu_m)}.$$

The derivative

$$f'(\mu_\xi) = \frac{-2(\mu_1 - \mu_2)(\mu_1 - \mu_m)(\mu_2 - 2\mu_\xi + \mu_m)}{(\mu_2 - \mu_\xi)^2(\mu_\xi - \mu_m)^2}$$

is a negative multiple of $(\mu_2 - 2\mu_\xi + \mu_m)$. Hence, $f(\cdot)$ is decreasing in $(\mu_m, \widetilde{\mu})$ and increasing in $(\widetilde{\mu}, \mu_2)$ with $\widetilde{\mu} := (\mu_2 + \mu_m)/2$. Moreover, $f(\cdot)$ is symmetric with respect to $\widetilde{\mu}$ within the interval (μ_m, μ_2), i.e., $f(\widetilde{\mu} - t) = f(\widetilde{\mu} + t) \;\; \forall\, t \in [0,\, \mu_2 - \widetilde{\mu})$. The minimizer of $f(\cdot)$ with respect to $\{\mu_3, \ldots, \mu_{m-1}\} \subset (\mu_m, \mu_2)$ is therefore given by an element nearest to $\widetilde{\mu}$. Consequently, the optimal $p_J(\cdot)$ coincides with the polynomial $q(\cdot)$ from the assertion. ∎

We remark that the value $q(\mu_1)$ is a discrete minimum of the function $f(\cdot)$, whereas the continuous minimum of $f(\cdot)$ in the interval (μ_m, μ_2) is $T_2(1{+}2\gamma_1)$ with $\gamma_1 = (\mu_1{-}\mu_2)/(\mu_2{-}\mu_m)$. Thus, $q(\mu_1) \geq T_2(1 + 2\gamma_1)$ which means that the bound $\big[\, q(\mu_1)\,\big]^{-2}$ is smaller than the Chebyshev bound $\big[\, T_2(1 + 2\gamma_1)\,\big]^{-2}$. This comparison inspires us to extend Lemma 3.4 to general eigenvalue intervals (μ_{i+1}, μ_i) by using the proof technique in [19] since the latter analysis has adapted the Chebyshev-type estimates to general eigenvalue intervals. For this sake we use Lemma 3.2 in [19], which is restated next.

LEMMA 3.5. *In the setting of Lemma 3.2, let* $\widetilde{y} = \sum_{j=1}^{m} \widetilde{w}_j$ *be the expansion of* $\widetilde{y} \in \mathbb{R}^n \backslash \{0\}$ *in terms of its orthogonal projections* $\widetilde{w}_j$ *to the eigenspaces of* H *for the* m *distinct eigenvalues* μ_j*. If* $\mu(\widetilde{y}) \in [\mu_{i+1}, \mu_i]$*, then the reweighted vector* $\widetilde{z} = \sum_{j=1}^{m} \alpha_j \widetilde{w}_j$ *satisfies*

(a) $\mu(\widetilde{z}) \geq \mu(\widetilde{y})$ *if* $|\alpha_j| \geq 1 \; \forall \, j \leq i$ *and* $|\alpha_j| \leq 1 \; \forall \, j > i$*,*

(b) $\mu(\widetilde{z}) \leq \mu(\widetilde{y})$ *if* $|\alpha_j| \leq 1 \; \forall \, j \leq i$ *and* $|\alpha_j| \geq 1 \; \forall \, j > i$*.*

The combination of Lemma 3.4 and Lemma 3.5 yields the following sharp Ritz value estimate for three-dimensional Krylov subspaces.

THEOREM 3.6. *Let* $\mu_1 > \mu_2 > \cdots > \mu_m$ *be the distinct eigenvalues of the symmetric and positive definite matrix* $H \in \mathbb{R}^{n \times n}$*, and let* $\mu(\cdot)$ *be the Rayleigh quotient with respect to* H*. Let* y' *be a Ritz vector associated with the largest Ritz value of* H *in the Krylov subspace* $\mathcal{K} := \mathcal{K}_H^3(y) = \operatorname{span}\{y, Hy, H^2 y\}$ *with* $y \in \mathbb{R}^n \backslash \{0\}$*. If* $\mu(y) \in (\mu_{i+1}, \mu_i)$*, then* $\mu(y') \geq \mu_i$ *holds trivially for* $i > m-3$*. Note that this case is not relevant for the intended high-dimensional* H *whose largest eigenvalue is to be computed. In the case* $i \leq m-3$ *it holds that*

$$\frac{\mu_i - \mu(y')}{\mu(y') - \mu_{i+1}} \leq [\, q(\mu_i) \,]^{-2} \, \frac{\mu_i - \mu(y)}{\mu(y) - \mu_{i+1}}. \tag{3.8}$$

Here $q(\cdot)$ *is a quadratic polynomial which interpolates the pairs* $\big(\mu_{i+1}, 1\big)$*,* $\big(\mu_\xi, -1\big)$*, and* $\big(\mu_m, 1\big)$*, and* μ_ξ *is an eigenvalue which has the smallest distance to* $(\mu_{i+1} + \mu_m)/2$ *among the eigenvalues* $\mu_{i+2}, \ldots, \mu_{m-1}$*, where we select the larger one as* μ_ξ *if there are two such eigenvalues nearest to* $(\mu_{i+1} + \mu_m)/2$*. Equality in (3.8) can be attained in the limit case that* y *belongs to the invariant subspace associated with the eigenvalues* $\mu_i, \mu_{i+1}, \mu_\xi$*, and* μ_m *together with* $\mu(y) \to \mu_i$*.*

Proof. The three interpolation conditions

$$q(\mu_{i+1}) = 1, \quad q(\mu_\xi) = -1, \text{ and } \quad q(\mu_m) = 1$$

with $\mu_1 > \cdots > \mu_m$ imply that $q(\cdot)$ is strictly increasing in $[\mu_{i+1}, \infty)$ so that

$$\min_{j=1,\ldots,i} |q(\mu_j)| = q(\mu_i) > q(\mu_{i+1}) = 1. \tag{3.9}$$

Moreover, $q(\cdot)$ is symmetric with respect to $\widetilde{\mu} := (\mu_{i+1} + \mu_m)/2$ in the interval (μ_m, μ_{i+1}). Since μ_ξ is an eigenvalue closest to $\widetilde{\mu}$, no eigenvalues are contained in the interval $(\widetilde{\mu} - \delta, \widetilde{\mu} + \delta)$ with $\delta := |\widetilde{\mu} - \mu_\xi|$. Combining this with the symmetry and the monotonicity of $q(\cdot)$ shows that

$$\max_{j=i+1,\ldots,m} |q(\mu_j)| = 1. \tag{3.10}$$

The case $\mu(y') \geq \mu_i$ is trivial since then the left-hand side of (3.8) is non-positive. In the other case $\mu(y') < \mu_i$, we begin with the obvious relation $\mu_i > \mu(y') \geq \mu(y) > \mu_{i+1}$ and select an auxiliary vector z satisfying $\mu(y') \geq \mu(z) \geq \mu(y)$. The rest of the proof is similar to the analysis for the Chebyshev-type estimate in [19, Section 3.2]. For convenience, we present here a concise version of the proof. Based on the eigenspace expansion $y = \sum_{j=1}^{m} w_j$ of y suggested in Lemma 3.5, we define $z := q(\mu_i) y_1 + y_2$ with $y_1 = \sum_{j=1}^{i} w_j$ and $y_2 = \sum_{j=i+1}^{m} w_j$. Then

$$\left(\frac{\mu(y_1) - \mu(z)}{\mu(z) - \mu(y_2)} \right) \left(\frac{\mu(y_1) - \mu(y)}{\mu(y) - \mu(y_2)} \right)^{-1} = [\, q(\mu_i) \,]^{-2}. \tag{3.11}$$

Note that $\mu(y_1)$ and $\mu(y_2)$ are the Ritz values of H in the subspace $\operatorname{span}\{y_1, y_2\}$. The properties (3.9) and (3.10) allow us to apply Lemma 3.5 (a) to $\widetilde{y} = y$ and $\widetilde{z} = z$, which yields $\mu(z) \geq \mu(y)$. Further inequalities can be derived using the vector $q(H)y = \sum_{j=1}^{m} q(\mu_j) w_j \in \mathcal{K}$. Applying Lemma 3.5 (a) to $\widetilde{y} = y$ and $\widetilde{z} = q(H)y$ implies that $\mu\big(q(H)y\big) \geq \mu(y)$. Therefore, we have

$\mu\big(q(H)y\big) \in [\mu(y), \mu(y')] \subseteq [\mu_{i+1}, \mu_i]$. Then Lemma 3.5 (b) applied to $\widetilde{y} = q(H)y$ and $\widetilde{z} = z$ shows that $\mu(z) \leq \mu\big(q(H)y\big)$. In summary, we have

$$\mu(y_1) \geq \mu_i > \mu(y') \geq \mu\big(q(H)y\big) \geq \mu(z) \geq \mu(y) > \mu_{i+1} \geq \mu(y_2)$$

so that

$$\left(\frac{\mu_i - \mu(y')}{\mu(y') - \mu_{i+1}}\right)\left(\frac{\mu_i - \mu(y)}{\mu(y) - \mu_{i+1}}\right)^{-1} \leq \left(\frac{\mu(y_1) - \mu(z)}{\mu(z) - \mu(y_2)}\right)\left(\frac{\mu(y_1) - \mu(y)}{\mu(y) - \mu(y_2)}\right)^{-1}.$$

This result together with (3.11) completes the proof of the estimate (3.8). The sharpness of the bound can be proved by applying the analysis in Theorem 3.1 to the invariant subspace spanned by the orthogonal projections w_i, w_{i+1}, w_ξ, and w_m of y. □

The convergence factor in (3.8) can easily be shown to fulfill $[\,q(\mu_i)\,]^{-2} \leq [\,T_2(1+2\gamma_i)\,]^{-2}$ with the gap ratio $\gamma_i := (\mu_i - \mu_{i+1})/(\mu_{i+1} - \mu_m)$. To illustrate this inequality we consider the simple example $H = \operatorname{diag}(10, 9, 8, 7, 6.5, 3, 2, 1)$ and select 98 equidistant values in each of the intervals $(7, 8)$, $(8, 9)$, and $(9, 10)$. By using random vectors in the corresponding level sets of the Rayleigh quotient $\mu(\cdot)$ and additionally by using equiangular vectors (with respect to spherical coordinates) from the invariant subspace associated with $\mu_i, \mu_{i+1}, \mu_\xi$, and μ_m, we compute numerical upper bounds of $\left(\frac{\mu_i - \mu(y')}{\mu(y') - \mu_{i+1}}\right)\left(\frac{\mu_i - \mu(y)}{\mu(y) - \mu_{i+1}}\right)^{-1}$ and compare them with the new estimated bound $[\,q(\mu_i)\,]^{-2}$ and the known Chebyshev bound $[\,T_2(1+2\gamma_i)\,]^{-2}$ based on [8] and [19]; see Figure 3.2.

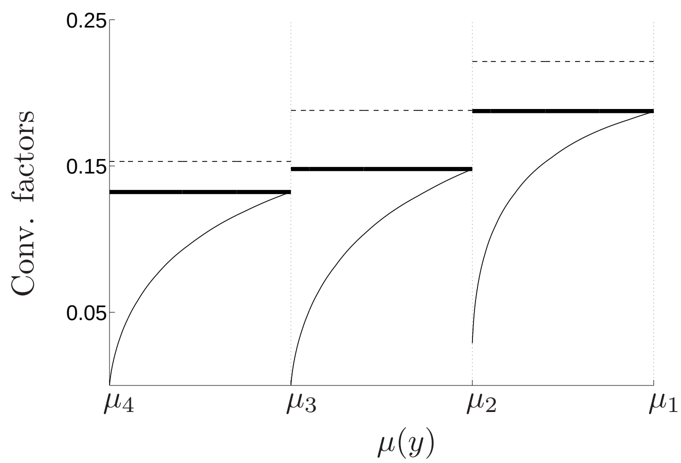


FIG. 3.2. *Comparison of the numerical upper bounds (curve) with the estimates bounds* $[\,q(\mu_i)\,]^{-2}$ *(bold line) and* $[\,T_2(1+2\gamma_i)\,]^{-2}$ *(dashed line) for the ratios* $\left(\frac{\mu_i - \mu(y')}{\mu(y') - \mu_{i+1}}\right)\left(\frac{\mu_i - \mu(y)}{\mu(y) - \mu_{i+1}}\right)^{-1}$.

We remark that the Chebyshev bound $[\,T_2(1+2\gamma_i)\,]^{-2}$ is sharp if $(\mu_{i+1} + \mu_m)/2$ is an eigenvalue of H. The following conclusion is more general.

COROLLARY 3.7. *The Chebyshev bound in the estimate* (1.7) *from Theorem 1.1 is sharp for* $k > 2$ *if all extreme points of the polynomial*

$$p(\alpha) := T_{k-1}\big(1 + 2(\alpha - \mu_{i+1})/(\mu_{i+1} - \mu_m)\big)$$

in the interval (μ_m, μ_{i+1}) *belong to the spectrum of* H.

Proof. Since the polynomial $p(\cdot)$ is given by the transformation of the Chebyshev polynomial $T_{k-1}(\cdot)$ from $[-1, 1]$ to $[\mu_m, \mu_{i+1}]$, there are $k-2$ extreme points in (μ_m, μ_{i+1}). If these are eigenvalues of H, then they can be denoted by $\mu_{\sigma(3)} > \cdots > \mu_{\sigma(k)}$. By applying the analysis

ETNA
Kent State University and
Johann Radon Institute (RICAM)

in Theorem 3.1 to the case that y belongs to an invariant subspace associated with the $k+1$ eigenvalues $\mu_i > \mu_{i+1} > \mu_{\sigma(3)} > \cdots > \mu_{\sigma(k)} > \mu_m$, we have the sharp estimate (3.1) with $\alpha_1 = \mu_i$, $\alpha_2 = \mu_{i+1}$, $\alpha_3 = \mu_{\sigma(3)}, \ldots, \alpha_k = \mu_{\sigma(k)}$, $\alpha_{k+1} = \mu_m$. Then (3.1) can be rewritten in the form of (1.7) by using simple substitutions. This means that the Chebyshev bound in (1.7) is attainable for a proper y. □

The eigenvalue μ_ξ in Theorem 3.6 can be interpreted as an eigenvalue nearest to a certain extreme point of a Chebyshev polynomial. However, for $k > 3$ it is difficult to describe such supporting eigenvalues by the distances to an extreme point, since the level sets of the objective function of the associated optimization problem are not ellipsoidal. As an example, we discuss the supporting eigenvalues for the sharp estimate in the case $k = 4$ according to Lemma 3.2. First, we define the function $f(\mu_\xi, \mu_\eta) := p(\mu_1)$ with the cubic polynomial $p(\cdot)$ which interpolates the pairs $(\mu_2, 1)$, $(\mu_\xi, -1)$, $(\mu_\eta, 1)$, and $(\mu_m, -1)$ on condition that $\mu_1 > \mu_2 > \mu_\xi > \mu_\eta > \mu_m$. The derivatives f_{μ_ξ} and f_{μ_η} are nonzero multiples of $(\mu_\eta - \mu_m)(\mu_2 - 2\mu_\xi + \mu_\eta) + (\mu_2 - \mu_\xi)^2$ and $(\mu_2 - \mu_\xi)(\mu_\xi - 2\mu_\eta + \mu_m) - (\mu_\eta - \mu_m)^2$, respectively. Setting f_{μ_ξ} and f_{μ_η} equal to zero gives a stationary point $(\mu_\xi^*, \mu_\eta^*)^T$ with $\mu_\xi^* = (3\mu_2 + \mu_m)/4$ and $\mu_\eta^* = (\mu_2 + 3\mu_m)/4$ where the minimum of the continuous problem is attained. Moreover, the coordinates μ_ξ^*, μ_η^* are extreme points of the transformed cubic Chebyshev polynomial in the interval (μ_m, μ_2). The discrete minimum of $f(\cdot,\cdot)$ in $\{\mu_3, \ldots, \mu_{m-1}\}^2$, however, is not always given by a pair with the smallest distance to $(\mu_\xi^*, \mu_\eta^*)^T$. In order to illustrate this, we set $m = 6$ and $(\mu_1, \ldots, \mu_6) = (10, 9, 8, 6.5, 5, 1)$ and draw several level sets of the corresponding function $f(\mu_\xi, \mu_\eta)$ surrounding its extreme point $P_1 = (7, 3)^T$; see Figure 3.3. The discrete minimum is attained in $P_2 = (8, 5)^T$, whereas the nearest point to P_1 is another point $P_3 = (6.5, 5)^T$.

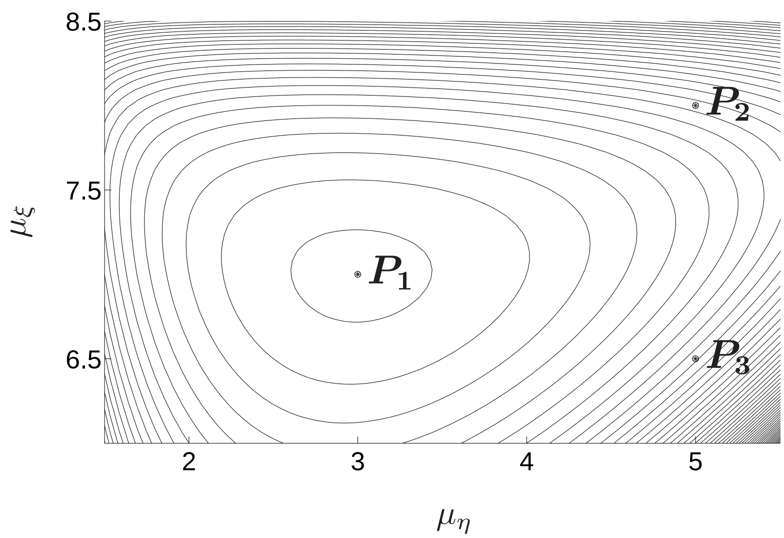


FIG. 3.3. *Level sets of $f(\mu_\xi, \mu_\eta)$ associated with the interior supporting eigenvalues for the sharp estimate in four-dimensional Krylov subspaces. The continuous minimum is attained in $P_1 = (7, 3)^T$. Although P_3 is close to P_1 compared to P_2, the function takes the larger function value in P_3.*

3.4. Estimates for the generalized eigenvalue problem. Because of the importance of the generalized eigenvalue problem $Ax = \lambda Mx$, we explicitly state for this problem the main results obtained in Section 3 on the convergence of the restarted Krylov subspace iteration (1.3). This reformulation is a direct consequence of the previous results together with the substitutions based on $y = A^{1/2}x$ and $H = A^{-1/2}MA^{-1/2}$; see [19] for the details of the transformation.

THEOREM 3.8. *Let $\lambda_1 < \lambda_2 < \cdots < \lambda_m$ be the distinct eigenvalues of the generalized eigenvalue problem $Ax = \lambda Mx$, where $A, M \in \mathbb{R}^{n\times n}$ are symmetric and positive definite matrices.*

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Further, let $\rho(\cdot)$ be the Rayleigh quotient defined in (1.1) and x' be a Ritz vector associated with the smallest Ritz value of (A, M) in the Krylov subspace $\operatorname{span}\{x, A^{-1}Mx, \ldots, (A^{-1}M)^{k-1}x\}$ with $x \in \mathbb{R}^n \backslash \{0\}$ and $k > 2$. Recall the convergence measure $\Delta_{i,i+1}(\cdot)$ given in (1.4).

(i) If $\rho(x) \in (\lambda_1, \lambda_2)$, then it holds that

$$\Delta_{1,2}\big(\rho(x')\big) \leq \Big(\min_J p_J(\lambda_1^{-1})\Big)^{-2} \Delta_{1,2}\big(\rho(x)\big)$$

where J runs through all $(k-2)$-element subsets of $\{3, \ldots, m-1\}$. Further, $p_J(\cdot)$ is a polynomial of degree $k-1$ interpolating the pairs $\big(\lambda_2^{-1}, 1\big)$, $\big(\lambda_m^{-1}, (-1)^{k+1}\big)$, and $\big(\lambda_{\sigma(j)}^{-1}, (-1)^j\big)$, $j = 3, \ldots, k$, with the indices $\sigma(j) \in J$ in increasing order; cf. Lemma 3.2. A similar estimate for the case $\rho(x) \in (\lambda_i, \lambda_{i+1})$ can be derived by adding assumptions on x and a certain Ritz value as in Lemma 3.3.

(ii) If $k = 3$ and $\rho(x) \in (\lambda_i, \lambda_{i+1})$, then $\rho(x') \leq \lambda_i$ holds trivially for $i > m-3$. In the case $i \leq m-3$ it holds that

$$\Delta_{i,i+1}\big(\rho(x')\big) \leq [\,q(\lambda_i^{-1})\,]^{-2}\, \Delta_{i,i+1}\big(\rho(x)\big).$$

Here $q(\cdot)$ is a quadratic polynomial interpolating the pairs $\big(\lambda_{i+1}^{-1}, 1\big)$, $\big(\lambda_\xi^{-1}, -1\big)$, and $\big(\lambda_m^{-1}, 1\big)$, and λ_ξ^{-1} is an element nearest to $(\lambda_{i+1}^{-1} + \lambda_m^{-1})/2$ in the set $\{\lambda_{i+2}^{-1}, \ldots, \lambda_{m-1}^{-1}\}$ where we select the larger one as λ_ξ^{-1} if there are two such elements. Equality can be attained in the limit case that x belongs to the invariant subspace associated with the relevant eigenvalues and $\rho(x) \to \lambda_i$; cf. Theorem 3.6.

(iii) The Chebyshev bound in the estimate (1.6) is sharp if all extreme points of the polynomial

$$p(\alpha) := T_{k-1}\big(1 + 2(\alpha - \lambda_{i+1}^{-1})/(\lambda_{i+1}^{-1} - \lambda_m^{-1})\big)$$

in the interval $(\lambda_m^{-1}, \lambda_{i+1}^{-1})$ are reciprocals of the eigenvalues of the matrix pair (A, M); cf. Corollary 3.7.

The representation $\min_J\, p_J(\lambda_1^{-1})$ in (i) cannot be compared with the bound [22, eq. (6.45)] by Saad, which is related to the distance between an eigenvector and a Krylov subspace. The advantage of the estimate in (i) is that it is applicable to a restarted method. Moreover, the two decisive indices 2 and m of p_J have explicitly been determined, and the interpolation conditions describe the monotonicity of p_J. This allows us to interpret it in a similar way as a Chebyshev bound.

4. Numerical experiments. We consider a high-dimensional matrix eigenvalue problem and illustrate the sharpness of the new estimates for restarted Krylov subspace iterations by comparing the analytical bounds with the numerically observed bounds. The matrix eigenvalue problem is derived from a finite element discretization of the Laplacian eigenproblem

$$-\Delta u = \lambda u \tag{4.1}$$

on a 2D domain with the boundary $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ where

$$\Gamma_1 = \Big\{\big(r(t)\cos(t),\, r(t)\sin(t)\big)^T \text{ with } r(t) := |\cos(1.5\,t)| + |\sin(1.5\,t)|\,;\ t = [0, 2\pi)\Big\},$$

$$\Gamma_2 = \big\{(1-t,\, 0)^T\,;\ t = (0,1)\big\}, \text{ and } \Gamma_3 = \big\{(t,\, 0)^T\,;\ t = [0,1)\big\};$$

see Figure 4.1. Homogeneous Dirichlet boundary conditions are imposed on $\Gamma_1 \cup \Gamma_3$ and homogeneous Neumann boundary conditions are considered on Γ_2. For the computation of the smallest eigenvalues and the associated eigenfunctions we use our Adaptive-Multigrid-Preconditioned (AMP) eigensolver software; available at http://www.math.uni-rostock.de/ampe. The code combines adaptive finite element discretizations of self-adjoint and elliptic partial differential operators

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 4.1
Ritz approximations θ_1 of the smallest eigenvalue $\lambda_1 \approx 5.598520$ computed by using the AMP eigensolver software with a block version of (1.3). More than $23{\cdot}10^6$ degrees of freedom are used on the finest grid with the level index 72.

level	1	17	38	51	63	72
nodes	26	2768	69118	676518	5018199	23295537
d.o.f.	12	2582	67870	672500	5006817	23271083
θ_1	10.44194	5.637456	5.599702	5.598635	5.598534	5.598522

with gradient-type eigensolvers for the solution of the discretized matrix eigenvalue problems. Here, we use linear finite elements on triangle meshes. The residual-based error estimator from [13] controls the grid refinement by using additionally quadratic elements. The eigenfunction of (4.1) which is associated with the smallest eigenvalue has an unbounded derivative at $(0, 0)^T$. Hence, the adaptive grid generation results in a high depth of the triangulation around the origin; see the initial grid in Figure 4.1 and three finer triangle meshes in Figure 4.2. The eigensolver is a block version (with three-dimensional subspace iterates) of the restarted Krylov subspace iteration.

For all numerical computations we have used a personal computer with a single core Intel Xeon 3.2GHz CPU with 31.4GiB RAM. In Figure 4.3 the left subfigure illustrates the computational costs versus the degrees of freedom. The cumulative computation times until a certain grid level is reached are drawn by the solid curve, and the computation times within the current level are displayed by the curve with markers. In the central subfigure we show the convergence history of $\theta_i - \lambda_i$ for $i \in \{1, 2, 3\}$. The curve with markers represents $i = 1$, the dashed curve stands for $i = 2$, and the solid curve for $i = 3$. The limit values λ_i are taken approximately from a grid with more than 39 million (39,247,401) nodes. The error indicators of the residual-based error estimator are plotted in the right subfigure. The solid curve shows the norm of the estimated residual vector corresponding to quadratic elements. The absolute values of the components of this residual vector are used to steer the grid refinement process. The dashed curve contains a modified residual norm suggested in [13, Section 4] which is used to define the stopping criterion of the eigensolver. The curve with markers shows the A^{-1}-norm of the residual of the eigenfunction approximation computed with linear elements. Additionally, the Ritz approximations θ_1 of the smallest eigenvalue on six exemplary levels are listed in Table 4.1.

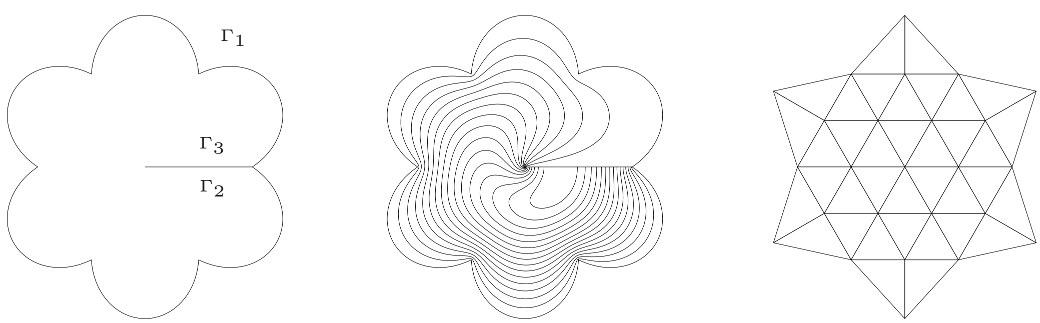


FIG. 4.1. *Left: the 2D domain for the Laplacian eigenvalue problem (4.1). Center: the contour lines of an eigenfunction corresponding to the smallest eigenvalue. Right: the initial grid.*

Experiment I. We use an FE grid with depth 63. The associated discretization matrices of the matrix eigenvalue problem $Ax = \lambda Mx$ have the dimension 5,006,817. The four smallest eigenvalues are $\lambda_1 \approx 5.598534$, $\lambda_2 \approx 8.727609$, $\lambda_3 \approx 12.24183$, and $\lambda_4 \approx 16.01582$. We verify the statements of Theorem 3.8 by numerical tests with the restarted Krylov subspace iteration (1.3) in the form $x' \leftarrow \mathrm{RR}_{\min}\big(\mathcal{K}^k(x)\big)$ for $k \in \{3, 4\}$; see Figure 4.4. To this end we fix 98 equidistant values in each of the three intervals $(\lambda_i, \lambda_{i+1})$, $i \in \{1, 2, 3\}$. For each value we select

ETNA
Kent State University and
Johann Radon Institute (RICAM)

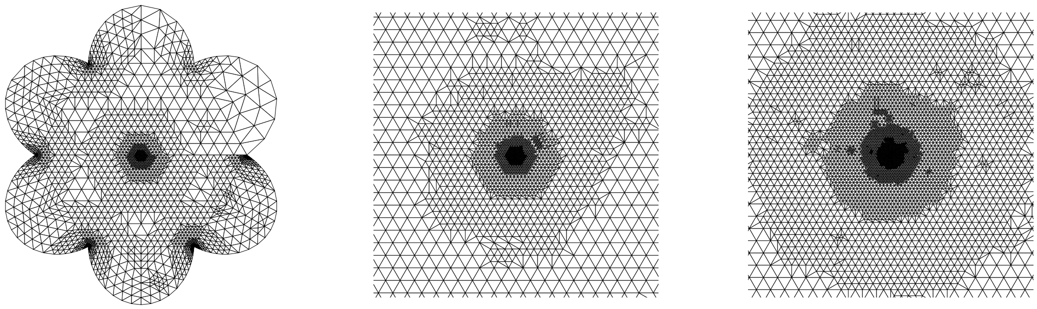


FIG. 4.2. *Triangle meshes with* 2,768, 69,118, *and* 676,518 *nodes and associated numbers of* 2,582, 67,870, *and* 672,500 *inner nodes. The associated depths of the triangulations are* 17, 38, *and* 51. *Square-shaped sectional enlargements around the critical point* $(0,\ 0)^T$ *are drawn for the two finer meshes. The side lengths of these enlargements are* $3 \cdot 10^{-4}$ *and* $6 \cdot 10^{-7}$, *respectively.*

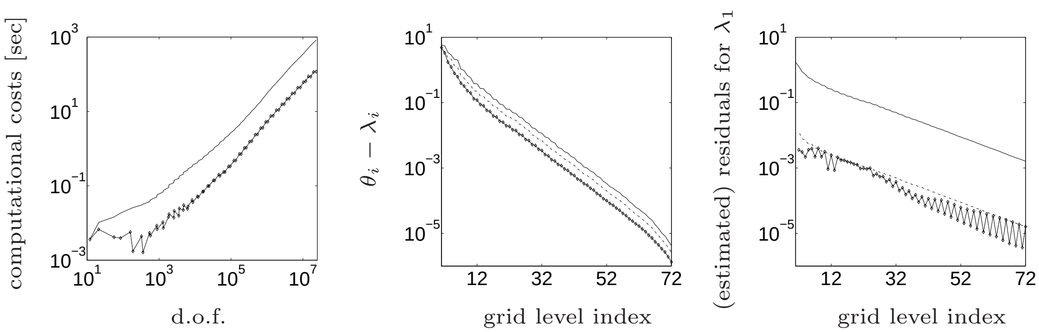


FIG. 4.3. *Computational information on the AMP eigensolver for the approximation of the smallest eigenvalues by a block version of the restarted Krylov subspace iteration* (1.3). *Left: cumulative computation times and computation times on the current level. Center: convergence history of the eigenvalue approximations for the three smallest eigenvalues. Right: error indicators of the residual-based error estimator associated with the smallest eigenvalue* λ_1.

1000 random vectors from the corresponding level set of the Rayleigh quotient $\rho(\cdot)$ together with equiangular vectors from the invariant subspaces associated with the supporting eigenvalues. For these test vectors we compute the numerical values of the ratio

$$\Delta_{i,i+1}(\rho(x'))/\Delta_{i,i+1}(\rho(x)) \tag{4.2}$$

with $\Delta_{i,i+1}(\cdot)$ defined in (1.4). As shown in Figure 4.4, the maxima with respect to the test vectors form in each interval a curve which takes its maximum at the left end-point of the interval. Such a maximum is compared with the analytical bounds given in (i) and (ii) of Theorem 3.8, respectively.

For $k = 3$ the statement (ii) can directly be used in order to determine the supporting eigenvalue λ_ξ for a quadratic polynomial $q(\cdot)$, which is here denoted by q_i according to the cases $\rho(x) \in (\lambda_i, \lambda_{i+1})$, $i \in \{1,\ 2,\ 3\}$. Table 4.2 lists λ_ξ for each q_i together with the distance ratios

$$\left|\lambda_\xi^{-1} - (\lambda_{i+1}^{-1} + \lambda_m^{-1})/2\right| \,/\, \left|\lambda_{i+1}^{-1} - \lambda_m^{-1}\right| \tag{4.3}$$

and the relative accuracy of the Chebyshev bound compared to the new sharp bound. The monotonically increasing relative overestimation by the T_2-bound in dependence of the distance ratio (see columns 3 and 4 in Table 4.2) support the statement (iii) on the sharp case of the Chebyshev bound.

For $k = 4$ we formulate a discrete optimization problem on the eigenvalue set $\{\lambda_3, \ldots, \lambda_{m-1}\}$ according to (i) in Theorem 3.8. An optimization with respect to the subset $\{\lambda_3, \ldots, \lambda_{25}\}$ gives two supporting eigenvalues λ_ξ and λ_η which are sufficient in order to solve the optimization problem. The sharp bounds have the form $[\, q(\lambda_i^{-1})\,]^{-2}$ with a cubic polynomial $q(\cdot)$ interpolating the pairs

ETNA
Kent State University and
Johann Radon Institute (RICAM)

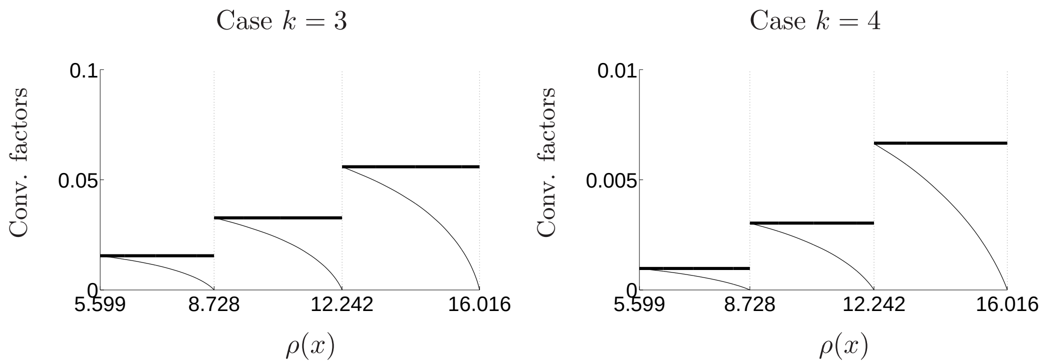


FIG. 4.4. *Slowest convergence of single steps of the restarted Krylov subspace iteration for $k = 3$ (left subplot) and $k = 4$ (right subplot). The abscissa is marked at the eigenvalues $\lambda_1, \ldots, \lambda_4$. The curves show the numerical maxima of the ratio (4.2). The bold lines represent the analytical bounds based on the statements (i), (ii) in Theorem 3.8.*

TABLE 4.2
Supporting eigenvalue λ_ξ in the statement (ii) of Theorem 3.8. The relative overestimation (in percent) of the Chebyshev bound in the estimate (1.6) is given in the last column.

Polynomial	λ_ξ	Distance ratio (4.3)	Relative overestimation by T_2-bound
q_1	λ_4	$4.494 \cdot 10^{-2}$	1.43%
q_2	λ_7	$2.403 \cdot 10^{-2}$	0.38%
q_3	λ_9	$1.591 \cdot 10^{-2}$	0.15%

TABLE 4.3
Supporting eigenvalues λ_ξ, λ_η for the sharp bound $[\,q_i(\lambda_i^{-1})\,]^{-2}$ based on the statement (i) in Theorem 3.8. The overestimation of the Chebyshev bound in the estimate (1.6) is measured in terms of the relative error with respect to the sharp bound.

Polynomial	λ_ξ	λ_η	Distance ratio (4.4)	Relative overestimation by T_3-bound
q_1	λ_3	λ_{11}	$3.731 \cdot 10^{-2}$	2.41%
q_2	λ_4	λ_{14}	$1.526 \cdot 10^{-2}$	0.42%
q_3	λ_6	λ_{21}	$4.480 \cdot 10^{-2}$	3.42%

$\big(\lambda_{i+1}^{-1}, 1\big)$, $\big(\lambda_\xi^{-1}, -1\big)$, $\big(\lambda_\eta^{-1}, 1\big)$, and $\big(\lambda_m^{-1}, -1\big)$. As shown in Figure 4.5, such cubic polynomials, denoted by q_i, $i \in \{1, 2, 3\}$, fulfill

$$\min_{j=1,\ldots,i} |q_i(\lambda_j^{-1})| = q_i(\lambda_i^{-1}) > q_i(\lambda_{i+1}^{-1}) = 1 \quad \text{and} \quad \max_{j=i+1,\ldots,m} |q_i(\lambda_j^{-1})| = 1,$$

where the second property is verified by using the monotonicity of q_i on the interval $(\lambda_m^{-1}, \lambda_{25}^{-1})$. These properties of q_i correspond to the main properties (3.9) and (3.10) used in the proof of Theorem 3.6. A similar further analysis shows that $[\,q_i(\lambda_i^{-1})\,]^{-2}$ is the sharp bound for the ratio (4.2). Table 4.3 lists the supporting eigenvalues λ_ξ and λ_η for the polynomials q_i. Additionally, the distance ratio

$$\|(\lambda_\xi^{-1}, \lambda_\eta^{-1})^T - \big((3\lambda_{i+1}^{-1} + \lambda_m^{-1})/4,\, (\lambda_{i+1}^{-1} + 3\lambda_m^{-1})/4\big)^T\|_2 \,/\, |\lambda_{i+1}^{-1} - \lambda_m^{-1}| \tag{4.4}$$

containing the extreme points of the polynomial

$$p(\alpha) := T_3\big(1 + 2(\alpha - \lambda_{i+1}^{-1})/(\lambda_{i+1}^{-1} - \lambda_m^{-1})\big)$$

is compared with the overestimation of the Chebyshev bound using T_3 with respect to the sharp bound $[\,q_i(\lambda_i^{-1})\,]^{-2}$. Finally, the sharp bounds for $k = 3$ and $k = 4$ are drawn by bold lines in Figure 4.4. The numerical maxima coincide with the sharp bounds.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

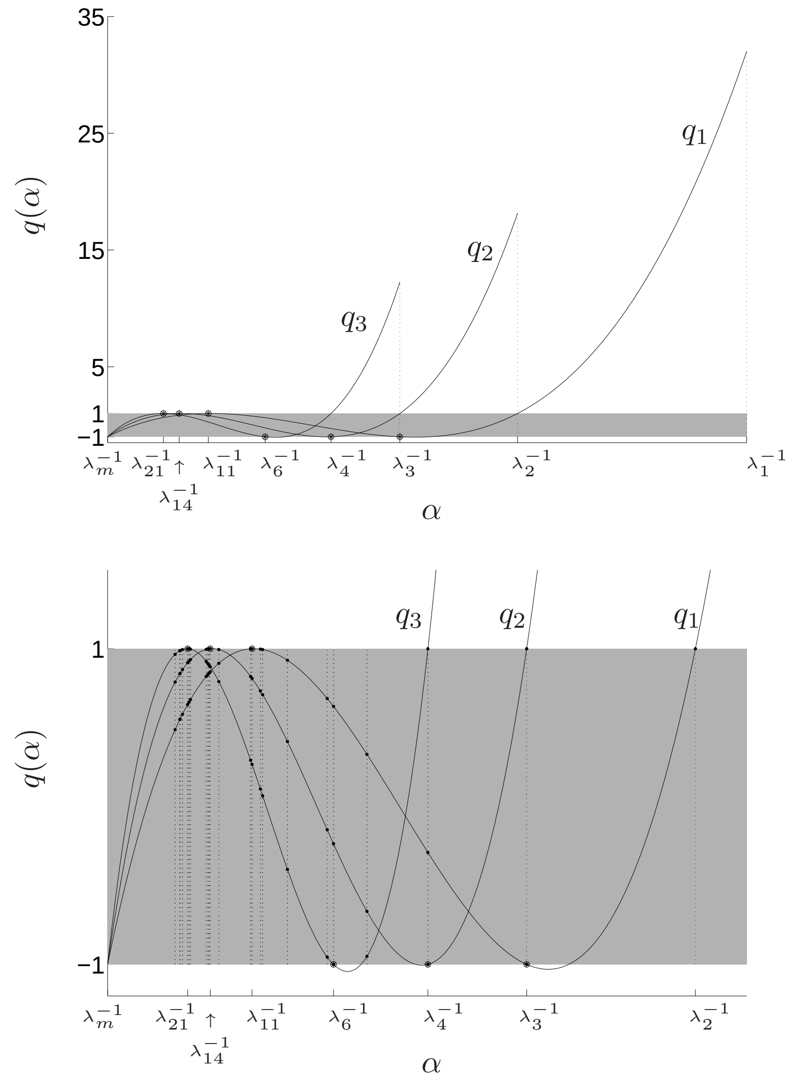


FIG. 4.5. *Cubic polynomials $q(\cdot)$ which are parts of the sharp bounds for single steps of the restarted Krylov subspace iteration with $k = 4$. These polynomials are denoted by q_i for $\rho(x) \in (\lambda_i, \lambda_{i+1})$, $i \in \{1, 2, 3\}$ respectively. The reciprocals of the supporting eigenvalues λ_ξ, λ_η are marked on the curves by filled black circles. The vertical dotted lines in the lower subfigure correspond to the values $\lambda_{25}^{-1}, \ldots, \lambda_2^{-1}$.*

Experiment II. We consider again the eigenvalue problem $Ax = \lambda Mx$ from Experiment I and illustrate the multi-step convergence behavior of the restarted Krylov subspace iteration (1.3). For each $k \in \{3, 4, 5\}$ we select 1000 random initial vectors $x^{(0)}$ from the level set of the Rayleigh quotient $\rho(\cdot)$ with the value $8.294 \in (\lambda_1, \lambda_2)$. The convergence of the eigenvalue approximations $\rho(x^{(\ell)})$ is measured in terms of the ratio $\Delta_{1,2}\big(\rho(x^{(\ell)})\big)$; see (1.4). For the eigenvector approximations $x^{(\ell)}$ we measure the convergence in terms of the tangent value $\tan \angle_M(x^{(\ell)}, \mathcal{E}_1)$ with respect to the eigenspace $\mathcal{E}_1$ associated with λ_1 and the inner product induced by M. The slowest convergence with respect to these two measures is plotted in Figure 4.6 by solid lines.

The first row of Figure 4.6 displays the numerical evaluations of the ratio $\Delta_{1,2}\big(\rho(x^{(\ell)})\big)$, those of the one-step bound $[\,q(\lambda_1^{-1})\,]^{-2}\Delta_{1,2}\big(\rho(x^{(\ell-1)})\big)$, and those of the multi-step bound $[\,q(\lambda_1^{-1})\,]^{-2\ell}\Delta_{1,2}\big(\rho(x^{(0)})\big)$ for the three cases $k \in \{3, 4, 5\}$. For $k = 3$ and $k = 4$ the polynomial $q(\cdot)$ has been determined as q_1 in Experiment I based on the statements (i) and (ii) from Theorem 3.8;

ETNA
Kent State University and
Johann Radon Institute (RICAM)

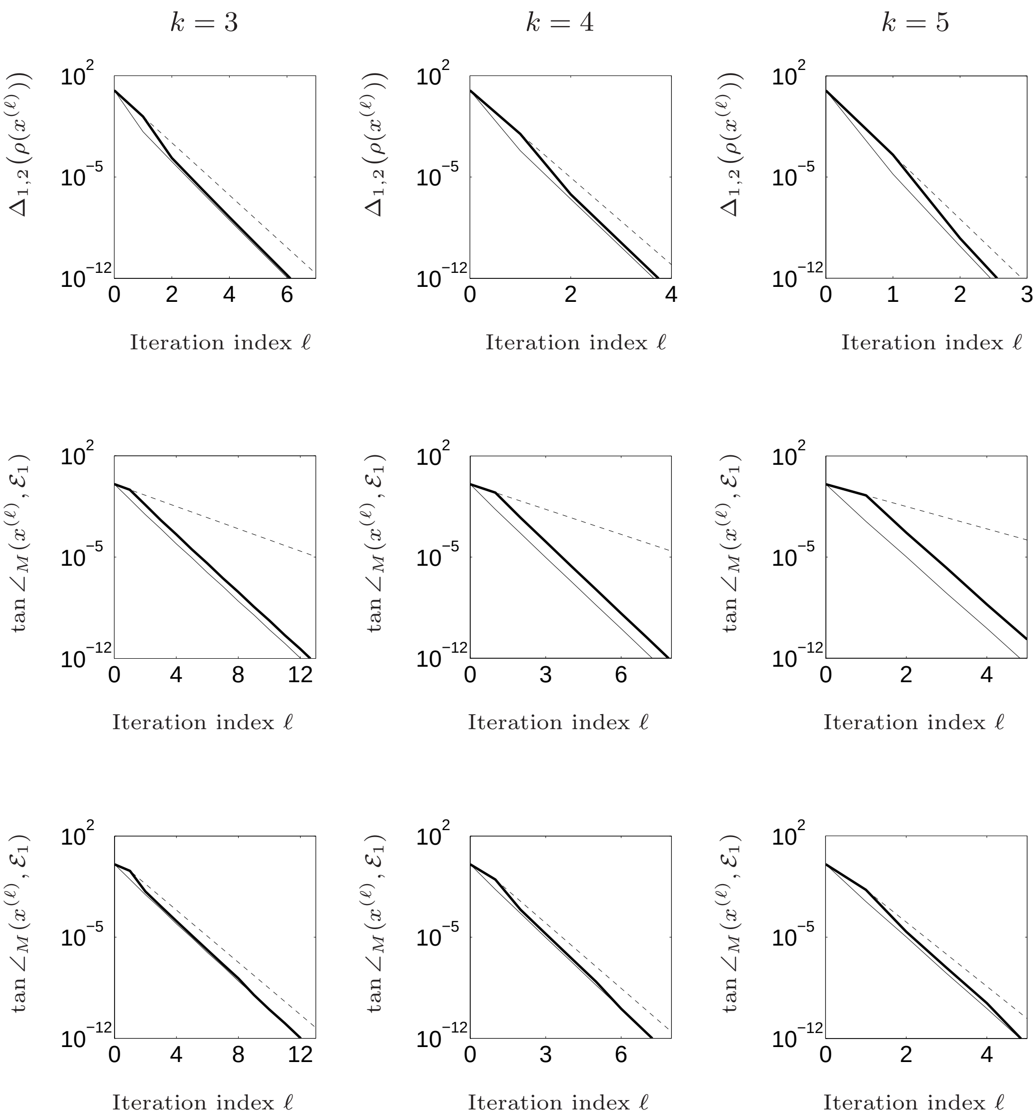


FIG. 4.6. *Multi-step convergence behavior of the restarted Krylov subspace iteration* (1.3)*. The subfigures are arranged in three rows. First row: Ritz value estimates based on Theorem 3.8. Second row: Ritz vector estimates based on Theorem 1.1 (i.e., on the convergence analysis from [19]). Third row: improved Ritz vector estimates with convergence factors of Ritz values from Theorem 3.8. In each row the results are plotted for* $k=3,4,5$*. In each subfigure the numerical data for the slowest convergence (solid lines) are compared with one-step estimates (bold lines) and multi-step estimates (dashed lines).*

cf. Table 4.2 and Table 4.3. For $k=5$ we solve a further discrete optimization problem according to the statement (i). The supporting eigenvalues of $q(\cdot)$ are given by λ_3, λ_5, and λ_{21}. The comparison reflects the fact that the polynomials $q(\cdot)$ which satisfy (3.9) and (3.10) have similar properties as the transformed Chebyshev polynomials and that the value $q(\lambda_1^{-1})$ increases rapidly with the degree of $q(\cdot)$. Therefore, the increasing dimension k causes a smaller convergence factor $[\,q(\lambda_1^{-1})\,]^{-2}$ and guarantees convergence acceleration.

The second row in Figure 4.6 shows a comparative numerical evaluation of the measure $\tan\angle_M(x^{(\ell)},\mathcal{E}_1)$, of the one-step bound $\kappa\tan\angle_M(x^{(\ell-1)},\mathcal{E}_1)$, and of the multi-step bound $\kappa^\ell\tan\angle_M(x^{(0)},\mathcal{E}_1)$ with $\kappa=\prod_{j=1}^{k-1}(\lambda_2^{-1}-\lambda_{m+1-j}^{-1})/(\lambda_1^{-1}-\lambda_{m+1-j}^{-1})$ based on the estimate (1.8) in Theorem 1.1. These κ-bounds, especially the multi-step bound, are shown to cause an overestimation.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

The third row in Figure 4.6 illustrates an improvement of the κ-bounds based on an inequality for any $x \in \mathbb{R}^n$ with $\rho(x) \in (\lambda_1, \lambda_2)$ and the eigenspace expansion $x = \sum_{i=1}^m v_i$. Namely,

$$\tan^2 \angle_M(x, \mathcal{E}_1) = \frac{\sum_{i=2}^m (\lambda_2 - \lambda_1)\|v_i\|_M^2}{(\lambda_2 - \lambda_1)\|v_1\|_M^2} \leq \frac{\sum_{i=1}^m (\lambda_i - \lambda_1)\|v_i\|_M^2}{\sum_{i=1}^m (\lambda_2 - \lambda_i)\|v_i\|_M^2} = \Delta_{1,2}\big(\rho(x)\big).$$

In these plots the numerical data for $\tan \angle_M(x^{(\ell)}, \mathcal{E}_1)$ are compared with the one-step bound $[\, q(\lambda_1^{-1})\,]^{-1}\big(\Delta_{1,2}\big(\rho(x^{(\ell-1)})\big)\big)^{\frac{1}{2}}$ and the multi-step bound $[\, q(\lambda_1^{-1})\,]^{-\ell}\big(\Delta_{1,2}\big(\rho(x^{(0)})\big)\big)^{\frac{1}{2}}$, i.e., the square roots of the bounds for $\Delta_{1,2}\big(\rho(x^{(\ell)})\big)$. This obviously avoids an overestimation.

5. Conclusion. The numerical approximation of the smallest eigenvalues and the associated eigenspaces of self-adjoint and elliptic partial differential operators is a challenging problem. Its successful solution requires efficient discretization techniques, an adaptive grid refinement strategy, a proper (multigrid) preconditioning, and, last but not least, efficient iterative matrix eigensolvers. Sharp estimates for the underlying eigensolvers are of major interest in order to understand these iterations and to develop even more efficient schemes. The present paper focuses on the convergence analysis of restarted Krylov subspace eigensolvers for generalized matrix eigenvalue problems with symmetric and positive definite matrices (A, M). An important ingredient of these Krylov subspace iterations is their underlying A-gradient subspace extension which amounts to an exact-inverse preconditioning. The main results of the new geometry-flavored and ellipsoid-based convergence analysis are sharp Ritz value estimates that improve the Chebyshev-type estimates from the previous work [19]. The new bounds depend on certain interpolating polynomials which have similar properties compared to the Chebyshev polynomials. An important feature of this A-gradient Krylov subspace convergence analysis is that it provides convergence estimates for general preconditioned eigensolvers in the limit that the preconditioner tends to the inverse A^{-1} of the discretization matrix A. This has provided new sharp convergence estimates for the restarted generalized Davidson method.

REFERENCES

[1] M. E. Argentati, A. V. Knyazev, K. Neymeyr, E. E. Ovtchinnikov, and M. Zhou, *Convergence theory for preconditioned eigenvalue solvers in a nutshell*, Found. Comput. Math., 17 (2017), pp. 713–727.
[2] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.
[3] J. Cullum and W. Donath, *A block generalization of the symmetric s-step Lanczos algorithm*, Tech. Report RC 4845, IBM Research, Yorktown Heights, 1974.
[4] E. G. D′yakonov, *Optimization in Solving Elliptic Problems*, CRC Press, Boca Raton, 1996.
[5] G. H. Golub and R. Underwood, *The block Lanczos method for computing eigenvalues*, in Mathematical Software, III, Publ. Math. Res. Center, No. 39, Academic Press, New York, 1977, pp. 361–377.
[6] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, 1996
[7] S. Kaniel, *Estimates for some computational techniques in linear algebra*, Math. Comp., 20 (1966), pp. 369–378.
[8] A. V. Knyazev, *Convergence rate estimates for iterative methods for a mesh symmetric eigenvalue problem*, Soviet J. Numer. Anal. Math. Modelling, 2 (1987), pp. 371–396.
[9] ———, *Preconditioned eigensolvers—an oxymoron?*, Electron. Trans. Numer. Anal., 7 (1998), pp. 104–123. http://etna.ricam.oeaw.ac.at/vol.7.1998/pp104-123.dir/pp104-123.pdf
[10] A. Knyazev and K. Neymeyr, *A geometric theory for preconditioned inverse iteration. III: A short and sharp convergence estimate for generalized eigenvalue problems*, Linear Algebra Appl., 358 (2003), pp. 95–114.
[11] ———, *Efficient solution of symmetric eigenvalue problems using multigrid preconditioners in the locally optimal block conjugate gradient method*, Electron. Trans. Numer. Anal., 15 (2003), pp. 38–55. http://etna.ricam.oeaw.ac.at/vol.15.2003/pp38-55.dir/pp38-55.pdf
[12] 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.
[13] K. Neymeyr, *A posteriori error estimation for elliptic eigenproblems*, Numer. Linear Algebra Appl., 9 (2002), pp. 263–279.
[14] ———, *On preconditioned eigensolvers and invert-Lanczos processes*, Linear Algebra Appl., 430 (2009), pp. 1039–1056.
[15] ———, *A geometric convergence theory for the preconditioned steepest descent iteration*, SIAM J. Numer. Anal., 50 (2012), pp. 3188–3207.

[16] K. NEYMEYR, E. OVTCHINNIKOV, AND M. ZHOU, *Convergence analysis of gradient iterations for the symmetric eigenvalue problem*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 443–456.

[17] K. NEYMEYR AND M. ZHOU, *The block preconditioned steepest descent iteration for elliptic operator eigenvalue problems*, Electron. Trans. Numer. Anal., 41 (2014), pp. 93–108. http://etna.ricam.oeaw.ac.at/vol.41.2014/pp93-108.dir/pp93-108.pdf

[18] ———, *Iterative minimization of the Rayleigh quotient by block steepest descent iterations*, Numer. Linear Algebra Appl., 21 (2014), pp. 604–617.

[19] ———, *Convergence analysis of restarted Krylov subspace eigensolvers*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 955–975.

[20] B. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs, 1980.

[21] Y. SAAD, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706.

[22] ———, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, 1992.

[23] H. A. VAN DER VORST, *A generalized Lanczos scheme*, Math. Comp., 39 (1982), pp. 559–561.