

CONVERGENCE OF INTEGRATION-BASED METHODS FOR THE SOLUTION OF STANDARD AND GENERALIZED HERMITIAN EIGENVALUE PROBLEMS*

LUKAS KRÄMER[†] AND BRUNO LANG[†]

Abstract. Recently, methods based on spectral projection and numerical integration have been proposed in the literature as candidates for reliable high-performance eigenvalue solvers. The key ingredients of this type of eigenvalue solver are a Rayleigh-Ritz process and a routine to compute an approximation to the desired eigenspace. The latter computation can be performed by numerical integration of the resolvent. In this article we investigate the progress of the Rayleigh-Ritz process and the achievable quality of the computed eigenpairs for the case that an upper bound for the normwise difference between the currently used subspace and the desired eigenspace is available. Then, such bounds are derived for the Gauß-Legendre rule and the trapezoidal rule.

Key words. eigenvalue solver, FEAST, spectral projection method, Rayleigh-Ritz, Gauß-Legendre rule, trapezoidal rule

AMS subject classifications. 65F15, 65D30, 30B10

1. Introduction. This work is dealing with techniques for the solution of generalized eigenvalue problems

$$(1.1) \quad AX = BX\Lambda,$$

where $A, B \in \mathbb{C}^{n \times n}$, A is Hermitian, and B is Hermitian positive definite. We will refer to such a matrix pair (A, B) as a *definite* matrix pair. The matrix $X \in \mathbb{C}^{n \times m}$ is a matrix whose columns form m eigenvectors of (A, B) , and Λ is a diagonal matrix of order m containing the m corresponding eigenvalues of (A, B) . It is well known that the pair (A, B) has real eigenvalues and that the eigenvectors can be normalized to be B -orthogonal, i.e., $X^*BX = I$ [16]. In this text we suppose that the matrices A and B are large and sparse and that only a part of the eigensystem is wanted.

Widely used methods for the solution of this type of problem include methods based on Krylov subspaces, i.e., of Arnoldi and Lanczos type, e.g., [1, 14, 23]. For an overview, see, e.g., [19]. Another class of algorithms is based on Jacobi-Davidson techniques [22]; see, e.g., [18] for recent HPC developments.

In recent years, a new class of algorithms have been proposed based on approximations of spectral projectors, i.e., projectors onto eigenspaces. They circumvent one drawback of solvers based on Krylov subspaces, namely that the subspace dimension grows with each iteration of the method, at least if no restarts are used. Suppose that we have a B -orthogonal projector P onto the subspace $\mathcal{X} = \text{span}(X)$ at hand, meaning its image and null space are orthogonal to each other with respect to the scalar product induced by B ; see below for details. Let $U := PY$ for some $n \times m$ matrix Y . If Y is chosen appropriately, then the matrix U spans the eigenspace \mathcal{X} , i.e., $\text{span}(U) = \mathcal{X}$. If the projector P is available only approximately, which is the case in practice, then an iterative process can be used to compute a basis for \mathcal{X} . One way to form approximations to P is by using contour integrals around a contour \mathcal{C} . This technique was introduced by the name FEAST algorithm by Polizzi [17]. Another technique using contour integration is the method by Sakurai and Sugiura [20].

*Received October 10, 2014. Accepted April 24, 2018. Published online on June 14, 2018. Recommended by Martin Gutknecht. This work was partially supported by the Deutsche Forschungsgemeinschaft through the priority programme 1648 “Software for Exascale Computing.” It was carried out while L. Krämer was with the Faculty of Mathematics and Natural Sciences at the University of Wuppertal.

[†]University of Wuppertal, Faculty of Mathematics and Natural Sciences, D-42097 Wuppertal, Germany (lang@math.uni-wuppertal.de).

Some advances have been made concerning the analysis of the method [26]. In this reference, emphasis was put on the convergence of the “outer” iteration of the algorithm; see the loop in Algorithm 2.1 below. In contrast to the present work, in [26] the convergence was analyzed in terms of ratios of the values of a certain function at the eigenvalues. The convergence of the “outer” algorithm is *not* part of the present article. In [13], we presented a simple analysis of the method explaining certain aspects with numerical experiments.

The aim of the present work is to analyze the convergence from another point of view. We will show under which conditions the numerical integration process converges and provide explicit error bounds in terms of the distance of the eigenvalues to the contour \mathcal{C} . Further, we analyze the implications of this convergence to the errors in the computed eigenvalues and eigenvectors. The paper summarizes and extends results from the first author’s PhD. thesis [12].

The document is structured as follows. Section 2 gives an overview on the basic FEAST algorithm. In Section 3 we investigate under which conditions the computed eigenvectors and eigenvalues converge towards their exact counterparts. The derived error bounds are given in terms of the distance $U - \tilde{U}$, where U is a basis of the eigenspace and \tilde{U} is an approximation of this basis. In Section 4 we state error bounds for certain norms of this distance $U - \tilde{U}$ for the case that \tilde{U} is obtained by the trapezoidal or the Gauß-Legendre rules. Some numerical results are reported in Section 5, and in Section 6 we conclude the article.

Throughout the article, the symbol $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix, respectively.

2. FEAST algorithm. The FEAST algorithm aims at computing eigenpairs of (A, B) in a given interval $I_\lambda = [\underline{\lambda}, \bar{\lambda}]$. As pointed out in [13], its main ingredients are a Rayleigh-Ritz process and numerical integration of the resolvent $(zB - A)^{-1}B$. The Rayleigh-Ritz process is a framework for extracting eigenvectors and eigenvalues of a matrix (pair) from a subspace \tilde{U} . The following steps are performed:

1. Find a suitable basis \tilde{U} for a space \tilde{U} .
2. Compute the Rayleigh quotients $A_{\tilde{U}} := \tilde{U}^* A \tilde{U}$, $B_{\tilde{U}} := \tilde{U}^* B \tilde{U}$.
3. Compute the primitive Ritz pairs $(\tilde{\Lambda}, \tilde{W})$ of $A_{\tilde{U}} W = B_{\tilde{U}} W \Lambda$.
4. Return the approximate *Ritz pairs* $(\tilde{\Lambda}, \tilde{U}\tilde{W})$ of $AX = BX\Lambda$.
5. Test for convergence; if the convergence criterion is not satisfied, then go back to step 1.

In step 4, a column of the matrix $\tilde{U}\tilde{W}$ is called a *Ritz vector* and the corresponding entry of $\tilde{\Lambda}$ a *Ritz value*. The question under which conditions Ritz values converge towards eigenvalues and Ritz vectors converge towards eigenvectors is discussed in Section 3.

In the FEAST algorithm, the matrix \tilde{U} is computed by an approximation of the spectral projector P onto the eigenspace belonging to the eigenvalues in I_λ . It can be shown that the exact projector is given by [13]

$$(2.1) \quad P = \frac{1}{2\pi i} \int_{\mathcal{C}} (zB - A)^{-1} B dz.$$

In (2.1), the symbol \mathcal{C} denotes a simple closed curve in \mathbb{C} that encircles exactly the m eigenvalues (counted with multiplicity) of (A, B) that reside in I_λ and no other.

The subspace \tilde{U} in the Rayleigh-Ritz procedure can be chosen as the span of $U := PY$, where Y is a full-rank matrix with m columns. If $Y^* B X$ has full rank, where X is a basis of the eigenspace belonging to the eigenvalues of (A, B) residing in I_λ , then we have that $\text{span}(U) = \mathcal{X}$. The integral (2.1) and therefore the matrix $U = PY$ can be approximated by

numerical integration. The resulting basic FEAST algorithm uses an approximation \tilde{U} to U and reads as in Algorithm 2.1. It is basically a Rayleigh-Ritz procedure with a special choice of \tilde{U} .

Algorithm 2.1 Skeleton of the FEAST algorithm [12, 13].

Input: An interval $I_\lambda = [\lambda, \bar{\lambda}]$ and an estimate \tilde{m} of the number of eigenvalues in I_λ .

Output: $\hat{m} \leq \tilde{m}$ eigenpairs with eigenvalues in I_λ .

- 1: Choose $Y \in \mathbb{C}^{n \times \tilde{m}}$ of rank \tilde{m} .
- 2: Compute an approximation \tilde{U} to

$$U := \frac{1}{2\pi i} \int_{\mathcal{C}} (zB - A)^{-1} B Y \, dz.$$

- 3: Form the Rayleigh quotients $A_{\tilde{U}} := \tilde{U}^* A \tilde{U}$, $B_{\tilde{U}} := \tilde{U}^* B \tilde{U}$.
 - 4: Solve the size- \tilde{m} generalized eigenproblem $A_{\tilde{U}} \tilde{W} = B_{\tilde{U}} \tilde{W} \tilde{\Lambda}$.
 - 5: Compute the approximate Ritz pairs $(\tilde{\Lambda}, \tilde{X} := \tilde{U} \tilde{W})$.
 - 6: If convergence is not reached, then go to Step 2 with $Y := \tilde{X}$.
-

In [9] and [26] it was shown that FEAST can be seen as subspace iteration with the matrix P . In an actual implementation it is performed with a perturbed version of P ; an analysis from this point of view can be found in [26].

In this work we will present a convergence analysis of FEAST that is based on the error in the numerical integration of (2.1), i.e., the difference $U - \tilde{U}$, where \tilde{U} denotes an approximation to U .

3. Perturbation of eigenvalues and eigenvectors. In this section we will give a short overview of the error bounds for eigenvalues and eigenvectors obtained by the Rayleigh-Ritz approach. The errors depend on the normwise difference $\delta := \|U - \tilde{U}\|$, where U denotes a basis for the desired eigenspace and \tilde{U} is an approximation. If the matrix Y in (2.1) is chosen properly, then U obtained by (2.1) is a basis for the eigenspace. However, we do not require \tilde{U} to be obtained by an integration process. For instance, \tilde{U} could be obtained by a polynomial approximation process or some other matrix function approach; see, e.g., [12, 21].

This kind of error analysis is not very common. In most of the well-studied eigenvalue methods, e.g., those based on Krylov subspaces, one is not seeking the exact eigenspace. In those methods bounds of the form $\|U - \tilde{U}\|$ are typically not available, hence the presented analysis does not apply.

For any $n \times m$ -matrix X , let A_X denote the Rayleigh quotient $X^* A X$ corresponding to the space $\text{span}(X)$. Let B_X be defined accordingly.

3.1. Perturbation of eigenvalues. For an $n \times m$ -matrix \tilde{U} we may make the ansatz $\tilde{U}^* A \tilde{U} = A_{\tilde{U}} + E$ for an error matrix E and obtain

$$(3.1) \quad E = A_{\tilde{U}-\tilde{U}} + (\tilde{U} - U)^* A U + U^* A (\tilde{U} - U).$$

Hence, $\|E\| \leq \|A\| (\delta^2 + 2\delta \|U\|)$. Then, as a consequence of Weyl's theorem [29] (see [25, Corollary 4.10] for a modern formulation), the following result can be obtained. It makes use of the fact that Ritz values belonging to an eigenspace of the matrix A are also eigenvalues of this matrix.

THEOREM 3.1 ([12, Theorem 2.7]). *Let A be Hermitian, and consider the standard equation $AX = X\Lambda$. Suppose that an error bound for the subspace $\delta := \|\tilde{U} - U\|$ is at hand. Let $\tilde{\lambda}_1, \dots, \tilde{\lambda}_m$ denote the Ritz values of A with respect to \tilde{U} and $\lambda_1, \dots, \lambda_m$ the eigenvalues of A belonging to the space $\text{span}(U)$, both sorted in ascending order. Then we have*

$$\max_j \left| \tilde{\lambda}_j - \lambda_j \right| \leq \|A\| (\delta^2 + 2\delta \|U\|).$$

In many situations, the above $\mathcal{O}(\delta)$ -error can be sharpened to $\mathcal{O}(\delta^2)$. We start by considering a single vector. Note that, if u is a column of U and \tilde{u} is a column of \tilde{U} , then it holds that $\|u - \tilde{u}\|_2 \leq \|U - \tilde{U}\|_2$. This plays a role when combining the following theorem with the results from Section 4.

THEOREM 3.2. *Let (u, λ) be an eigenpair of the Hermitian matrix A with $\|u\| = 1$, and let $\tilde{u} = u + \Delta u$ with a small perturbation Δu . Then*

$$\left| \tilde{\lambda} - \lambda \right| = \mathcal{O}(\|A\| \delta^2),$$

where $\tilde{\lambda} = \rho(\tilde{u}) := \tilde{u}^* A \tilde{u} / \tilde{u}^* \tilde{u}$ and $\delta = \|u - \tilde{u}\| = \|\Delta u\|$.

Proof. Let (u_i, λ_i) , $i = 1, \dots, n$, be an eigensystem of A with orthonormal eigenvectors u_i such that $u_1 = u$, and let

$$\Delta u = \sum_{j=1}^n \xi_j u_j.$$

Then, $\delta^2 = \|\Delta u\|^2 = \sum_{j=1}^n \bar{\xi}_j \xi_j$. For the Rayleigh quotient $\rho(\tilde{u})$ belonging to the perturbed eigenvector $\tilde{u} = u_1 + \Delta u$, it holds that

$$(3.2) \quad \rho(\tilde{u}) = \frac{\tilde{u}^* A \tilde{u}}{\tilde{u}^* \tilde{u}} = \frac{\lambda (1 + \bar{\xi}_1 + \xi_1)}{1 + \bar{\xi}_1 + \xi_1 + \delta^2} + \frac{\sum_{j=1}^n \bar{\xi}_j \xi_j \lambda_j}{1 + \bar{\xi}_1 + \xi_1 + \delta^2} =: \alpha + \beta.$$

Now, let $\delta \leq \frac{1}{4}$. Then $|\bar{\xi}_1 + \xi_1| \leq 2|\xi_1| \leq 2\delta \leq \frac{1}{2}$. Noting that $\bar{\xi}_1 + \xi_1 \in \mathbb{R}$, we have $\frac{1}{2} < 1 + \bar{\xi}_1 + \xi_1 < \frac{3}{2}$, and therefore the first summand in (3.2) can be estimated as follows:

$$(3.3) \quad \alpha = \frac{\lambda}{1 + \frac{\delta^2}{1 + \bar{\xi}_1 + \xi_1}} \in \left[\frac{\lambda}{1 + 2\delta^2}, \frac{\lambda}{1 + \frac{2}{3}\delta^2} \right].$$

By using a Taylor series expansion of $1/(1 + 2\delta^2)$ and $1/(1 + \frac{2}{3}\delta^2)$ around 0, respectively, it can be seen that both interval bounds in (3.3) can be written as $1 + \mathcal{O}(\delta^2)$ for $\delta \rightarrow 0$. This shows that

$$\alpha = \lambda + \mathcal{O}(\|A\| \delta^2)$$

(recall that $|\lambda| \leq \|A\|$). For the second summand in (3.2) it holds that

$$\beta \leq \frac{\max_j |\lambda_j| \sum_{j=1}^n \bar{\xi}_j \xi_j}{\frac{1}{2}} = \mathcal{O}(\|A\| \delta^2).$$

Together, we obtain $\rho(\tilde{u}) - \lambda = \mathcal{O}(\|A\| \delta^2)$. \square

Now let \mathcal{U} be an invariant subspace with basis U , and let $\tilde{U} = U + \Delta U$ with a small perturbation ΔU . Without loss of generality we may assume the columns of U to be eigenvectors of A , $Au_j = u_j \lambda_j$. (Otherwise consider an eigendecomposition of the Rayleigh quotient,

$U^*AU = W\Lambda W^*$ with $W^*W = I$, and replace U with UW whose columns are eigenvectors of A . Similarly, $\tilde{U} \mapsto \tilde{U}W$ and $\Delta U \mapsto \Delta U W$; these transformations leave the 2-norms invariant.) Applying Theorem 3.2 to the normalized columns of U yields

$$|\rho(\tilde{u}_j) - \lambda_j| = \left| \rho\left(\frac{\tilde{u}_j}{\|\tilde{u}_j\|}\right) - \lambda_j \right| = \mathcal{O}\left(\|A\| \cdot \left(\frac{\|\Delta u_j\|}{\|\tilde{u}_j\|}\right)^2\right).$$

In particular, if U is orthonormal, then we have

$$\max_j |\rho(\tilde{u}_j) - \lambda_j| = \mathcal{O}\left(\|A\| \cdot \|\tilde{U} - U\|^2\right)$$

due to $\|u - \tilde{u}\|_2 \leq \|U - \tilde{U}\|_2$. For the generalized eigenvalue problem involving a definite matrix pair (A, B) , an analogous bound is valid when replacing $\|\Delta u\|$ by $\|\Delta u\|_B = \sqrt{\Delta u^* B \Delta u}$.

In the generalized case we first can interpret the Rayleigh quotient $B_{\tilde{U}}$ as a perturbed Rayleigh quotient of B_U , where U fulfills $AU = BUH$ for some square matrix H . Similarly to (3.1), we obtain

$$F = B_{U-\tilde{U}} + (\tilde{U} - U)^* B U + U^* B (\tilde{U} - U)$$

with $\|F\| \leq \|B\| \cdot (\delta^2 + 2\delta \|U\|)$.

Recently, Nakatsukasa [15] published a theorem in the spirit of Weyl's classic theorem for the eigenvalues of a definite matrix pair. As a consequence, we can obtain an error bound for the Ritz values of (A, B) corresponding to \tilde{U} that depends on δ . The result can be formulated as follows:

THEOREM 3.3 ([12, Theorem 2.9]). *Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ denote the eigenvalues of (A_U, B_U) . Let $\|F\| < \lambda_{\min}(B_U)$. Then the perturbed pair $(A_U + E, B_U + F)$ is Hermitian definite, and its eigenvalues $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \dots \leq \tilde{\lambda}_m$ satisfy*

$$(3.4) \quad \max_j \left| \lambda_j - \tilde{\lambda}_j \right| \leq \frac{\|A\| \cdot \Delta}{\lambda_{\min}(B_U)} + \frac{\|A_U\| + \|A\| \cdot \Delta}{\lambda_{\min}(B_U) (\lambda_{\min}(B_U) - \|B\| \cdot \Delta)} \cdot \|B\| \cdot \Delta,$$

where $\Delta := (\delta^2 + 2\delta \|U\|)$.

The right-hand side of (3.4) converges to 0 as $\delta \rightarrow 0$.

Other eigenvalue bounds depending on the norm $\|U - \tilde{U}\|$ can be found in [11]. They are also consequences of Weyl's theorem and yield upper bounds for the approximation error of similar order as those stated above. In [12, Corollary 2.6] also a version for the generalized eigenvalue problem can be found.

3.2. Perturbation of eigenvectors. The convergence theory for eigenvectors is more complicated than the one for eigenvalues. First, an appropriate measure for the "distance" of vectors has to be defined. This distance is typically defined as the angle between vectors; see, e.g., [10]. For two nonzero vectors $x, y \in \mathbb{C}^n$, the angle $\theta := \angle(x, y)$ is defined implicitly by the relation

$$\cos \theta = \frac{|x^* y|}{\|x\| \|y\|}.$$

The fact that θ is well-defined follows from the Cauchy-Schwarz inequality.

Angles can also be measured in the scalar product defined by a Hermitian positive definite matrix B , that is, $x^* B y$. This scalar product first induces a norm, the so-called B -norm

$\|x\|_B := \sqrt{x^*Bx}$. Then the angle between x and y in the B scalar product, $\angle_B(x, y)$, is defined implicitly by the equation

$$\cos \angle_B(x, y) = \frac{|x^*By|}{\|x\|_B \|y\|_B}.$$

The usual way to measure angles between a vector and a subspace or between two subspaces is by means of so-called *principal angles*; see [8, Section 6.4.3]. In [10], a comprehensive overview on principal angles defined by the scalar product induced by B can be found.

Given two m -dimensional subspaces $\mathcal{U}, \mathcal{V} \subset \mathbb{C}^n$, there are m principal angles between those subspaces. The largest canonical angle will be denoted by $\angle(\mathcal{U}, \mathcal{V})$. It can be shown that $\cos \angle(\mathcal{U}, \mathcal{V}) = \|U^*V\|$, given orthonormal bases U and V of \mathcal{U} and \mathcal{V} , respectively. For the B -scalar product we then have $\cos \angle_B(\mathcal{U}, \mathcal{V}) = \|U^*BV\|$ if U and V are orthonormal with respect to that scalar product [10].

For any two spaces $\mathcal{U} = \text{span}(U)$, $\tilde{\mathcal{U}} = \text{span}(\tilde{U})$ of the same dimension, it can be shown that [10, Lemma 5.5]

$$(3.5) \quad \sin \angle_B(\mathcal{U}, \tilde{\mathcal{U}}) \leq \kappa_B(U) \frac{\|B^{1/2}(U - \tilde{U})\|}{\|B^{1/2}U\|},$$

where $\kappa_B = \sigma_{\max}(B^{1/2}U)/\sigma_{\min}(B^{1/2}U)$ is the condition number of U with respect to the B -norm. The symbol $B^{1/2}$ denotes the square root of B ; see [8]. In this and any other statement where the square root of B appears, it can also be replaced by its Cholesky factor [8] or any matrix K fulfilling $K^*K = B$.

Thinking of \mathcal{U} as an eigenspace of (A, B) with a B -orthonormal basis U , (3.5) boils down to

$$(3.6) \quad \sin \angle_B(\mathcal{U}, \tilde{\mathcal{U}}) \leq \|B^{1/2}(U - \tilde{U})\|$$

since $B^{1/2}U$ is a matrix with orthonormal columns in this case. Hence, a way to limit the sine of the largest canonical angle between two subspaces in terms of the bases of those subspaces has been found. In particular, if an upper bound for the normwise distance between bases of the eigenspace and an approximation is known, then statements on the largest principal angle between those two spaces can be made. Note that for θ close to 0, we have $\theta \approx \sin \theta$. It is clear that for a vector $u \in \mathcal{U}$ we have

$$(3.7) \quad \angle_B(u, \tilde{U}) \leq \angle_B(\mathcal{U}, \tilde{\mathcal{U}}).$$

This follows from the min-max definition of angles; see [3].

Next, we will explain how bounds for the angle $\angle_B(u, \tilde{u})$ for an eigenvector u and a vector $\tilde{u} \in \tilde{\mathcal{U}}$ can be obtained. From the convergence of Ritz values and a small angle $\angle_B(u, \tilde{\mathcal{U}})$ it does not follow that there is a vector $\tilde{u} \in \tilde{\mathcal{U}}$ such that $\angle_B(u, \tilde{u})$ is small; see the discussion in [24]. The convergence of Ritz vectors depends on the distance of the Ritz values. The following theorem is an extension of [19, Theorem 4.6] to the case of the generalized eigenvalue problem.

THEOREM 3.4 ([12]). *Let (u, λ) be any eigenpair of (A, B) . Let $\tilde{\lambda}$ be an approximate eigenvalue extracted from $\tilde{\mathcal{U}}$, and let η be the distance between λ and the approximate eigenvalues other than $\tilde{\lambda}$. Then there is an approximate eigenvector \tilde{u} associated with $\tilde{\lambda}$ such that*

$$\sin \angle_B(u, \tilde{u}) \leq \sin \angle_B(u, \tilde{\mathcal{U}}) \sqrt{1 + \frac{\gamma^2}{\eta^2}},$$

where $\gamma \leq \|A\|$.

Proof. We write the generalized eigenvalue equation in the form of a standard eigenvalue equation as

$$B^{-1/2}AB^{-1/2}y = y\lambda,$$

obtaining $B^{1/2}u = y$ as eigenvector of the pair (A, B) corresponding to the eigenvalue λ . Similarly, we obtain the Ritz vector $B^{1/2}\tilde{u}$ belonging to the space $B^{1/2}\tilde{U}$. Applying a corresponding theorem for the standard eigenvalue problem [19, Theorem 4.6] yields

$$\sin \angle(B^{1/2}u, B^{1/2}\tilde{u}) \leq \sin \angle(B^{1/2}u, B^{1/2}\tilde{U}) \sqrt{1 + \frac{\gamma^2}{\eta^2}}$$

for some number γ . Next, let \tilde{U} be a B -orthonormal basis of \tilde{U} . For γ we obtain, similarly to the theorem in [19],

$$\gamma = \|PA(I - P)\|,$$

where $P = (B^{1/2}\tilde{U})^*B^{1/2}\tilde{U}$ is the orthogonal projector onto the space $B^{1/2}\tilde{U}$. Hence, $\gamma \leq \|A\|$. Using $\angle(B^{1/2}u, B^{1/2}\tilde{u}) = \angle_B(u, \tilde{u})$ and $\angle(B^{1/2}u, B^{1/2}\tilde{U}) = \angle_B(u, \tilde{U})$ (see [10]) finishes the proof. \square

Combining Theorem 3.4 with (3.6) and (3.7) yields convergence of the Ritz vectors towards the eigenvectors if $\tilde{U} \rightarrow U$. More precisely, if U is a B -orthonormal matrix, then, as $\tilde{U} \rightarrow U$, we obtain

$$\sin \angle_B(u, \tilde{u}) \leq \sin \angle_B(u, \tilde{U}) \cdot C \leq \sin \angle_B(U, \tilde{U}) \cdot C \leq \|B^{1/2}(U - \tilde{U})\| \cdot C \rightarrow 0,$$

where $C := \sqrt{1 + \gamma^2/\eta^2}$. Note that γ is always bounded by $\|A\|$.

4. Convergence of integration schemes. In this section we investigate the numerical integration process of

$$(4.1) \quad U = PY = \frac{1}{2\pi\mathbf{i}} \int_{\mathcal{C}} (zB - A)^{-1}BY \, dz.$$

Suppose that a parametrization $\varphi : [0, 2\pi] \rightarrow \mathbb{C}$ of the contour \mathcal{C} is given. By φ we can express the integral (4.1) as

$$(4.2) \quad U = \frac{1}{2\pi\mathbf{i}} \int_0^{2\pi} \varphi'(t)(\varphi(t)B - A)^{-1}BY \, dt.$$

The integration interval can also be chosen differently from $[0, 2\pi]$, but this choice sometimes simplifies the presentation.

Let $c := (\underline{\lambda} + \bar{\lambda})/2$ and $r := (\bar{\lambda} - \underline{\lambda})/2$ denote center and radius of the interval I_λ , respectively, and assume that $\underline{\lambda}, \bar{\lambda} \notin \text{spec}(A, B)$. If we set

$$(4.3) \quad \varphi : [0, 2\pi] \rightarrow \mathbb{C}, \quad \varphi(t) = c + r \exp(\mathbf{i}t),$$

then we obtain for (4.2)

$$U = \frac{1}{2\pi\mathbf{i}} \int_0^{2\pi} \mathbf{i} \exp(\mathbf{i}t) ((c + r \exp(\mathbf{i}t))B - A)^{-1}BY \, dt.$$

By an *integration scheme of order p* we denote a sequence of pairs $(\omega_j, t_j)_{j=0, \dots, p}$ defined by integration points $t_j \in [0, 2\pi]$ and integration weights $\omega_j \in \mathbb{R}$. The integral (4.2) then is approximated by

$$(4.4) \quad \tilde{U}_p := \frac{1}{2\pi i} \sum_{j=0}^p \omega_j \varphi'(t_j) (\varphi(t_j) B - A)^{-1} B Y.$$

The goal of this section is to derive error bounds for certain norms of $U - \tilde{U}_p$. We will do so for the trapezoidal and the Gauß-Legendre integration rules. Most of the results in this section are taken from [12].

4.1. Error bounds for the trapezoidal rule. The integrand of the integral (4.2) is periodic if the parametrization function φ is periodic, e.g., a circle as in (4.3). It is well known that the trapezoidal rule delivers more accurate results if the integrand is periodic [7, 28]. It was Beyn [2] who designed an integration-based algorithm for nonlinear eigenvalue problems. In [2], an error analysis for the trapezoidal rule and a periodic parametrization can be found. In [12] the analysis was adapted to the special case of the generalized Hermitian eigenvalue problem (1.1). Recently, an article by Trefethen and Weideman [27] was published that gives a thorough overview of the convergence of the trapezoidal rule applied to periodic functions.

For a 2π -periodic function f defined on $[0, 2\pi]$, the trapezoidal rule of order p reads

$$\frac{1}{2\pi} \int_0^{2\pi} f(t) dt \approx T_p(f) := \frac{1}{p} \sum_{j=0}^{p-1} f\left(\frac{2\pi j}{p}\right)$$

since we have $f(0) = f(2\pi)$.

The following theorem based on [2, Theorem 4.1] shows that the error of the trapezoidal rule applied to a scalar function decays exponentially with p if the function is periodic and analytic in an open strip containing the real line. A detailed proof can be found in the first author's thesis [12].

THEOREM 4.1 ([2]). *Let $s_- < 0 < s_+$, and let*

$$S = S(s_-, s_+) := \{z \in \mathbb{C} : s_- < \operatorname{Im}(z) < s_+\}$$

denote a strip containing the real line. Let $f : S(s_-, s_+) \rightarrow \mathbb{C}$ be 2π -periodic and analytic on S . Then, for all $0 > \sigma_- > s_-$ and $0 < \sigma_+ < s_+$, the error of the trapezoidal sum

$$E_{T_p}(f) := \frac{1}{2\pi} \int_0^{2\pi} f(x) dx - \underbrace{\frac{1}{p} \sum_{j=0}^{p-1} f\left(\frac{2\pi j}{p}\right)}_{T_p(f)}$$

satisfies

$$(4.5) \quad |E_{T_p}(f)| \leq \max_{\operatorname{Im}(z)=\sigma_+} |f(z)| F(\exp(-p\sigma_+)) + \max_{\operatorname{Im}(z)=\sigma_-} |f(z)| F(\exp(p\sigma_-)),$$

where $F(t) = \frac{t}{1-t}$ for $t \neq 1$.

When investigating the integration of the resolvent $G(z) := (zB - A)^{-1}B$, the scalar-valued functions $r_\lambda(z) := (z - \lambda)^{-1}$ plays a prominent role. If the matrix pair (A, B) has

eigenvalues $\lambda_1, \dots, \lambda_n$ with corresponding B-orthonormal eigenvectors x_1, \dots, x_n , then we can write

$$G(z) = \sum_{j=1}^n r_{\lambda_j}(z) x_j x_j^* B.$$

This leads to the following representation of (4.1),

$$\begin{aligned} U &= \frac{1}{2\pi i} \int_{\mathcal{C}} (zB - A)^{-1} B Y \, dz = \frac{1}{2\pi i} \int_{\mathcal{C}} \sum_{j=1}^m r_{\lambda_j}(z) x_j x_j^* B Y \, dz \\ (4.6) \quad &= \frac{1}{2\pi i} \sum_{j=1}^m \left(\int_{\mathcal{C}} r_{\lambda_j}(z) \, dz \right) x_j x_j^* B Y, \end{aligned}$$

where we suppose an eigenvalue ordering such that $\lambda_1, \dots, \lambda_m \in I_\lambda$, $\lambda_{m+1}, \dots, \lambda_n \notin I_\lambda$. Hence, computing (4.1) numerically amounts to computing the integrals in (4.6) numerically.

A parametrized form of the integrals in (4.6) is

$$\int_{\mathcal{C}} r_\lambda(z) \, dz = \int_0^{2\pi} r_\lambda(\varphi(t)) \varphi'(t) \, dt.$$

Putting the integrand of the last integral into (4.5), we obtain

$$\begin{aligned} (4.7) \quad |E_{T_p}(r_\lambda)| &\leq \max_{\text{Im}(z)=\sigma_+} |\varphi'(z)| |r_\lambda(\varphi(z))| F(\exp(-p\sigma_+)) \\ &\quad + \max_{\text{Im}(z)=\sigma_-} |\varphi'(z)| |r_\lambda(\varphi(z))| F(\exp(-p\sigma_-)), \end{aligned}$$

for certain $0 > \sigma_- > s_-$, $0 < \sigma_+ < s_+$. The following lemma is a special case of [2, Lemma 4.6] and substantiates (4.7). The version stated here is taken from the first author's thesis [12].

LEMMA 4.2. *Let φ be defined on S and 2π -periodic. Further let $\varphi(z) \in \text{Int}(\mathcal{C})$ for $\text{Im}(z) > 0$ and $\varphi(z) \in \text{Ext}(\mathcal{C})$ for $\text{Im}(z) < 0$. Let $\text{dist}(\lambda, \mathcal{C}) = \min_{z \in \mathcal{C}} |\lambda - z|$. Then there are constants $C_1, C_2, C_3 > 0$ such that*

$$|E_{T_p}(r_\lambda)| \leq C_1 \text{dist}(\lambda, \mathcal{C})^{-1} \exp(-C_2 p \text{dist}(\lambda, \mathcal{C}))$$

for $\text{dist}(\lambda, \mathcal{C}) \leq C_3$. The constants are independent of λ and p .

We obtain

$$\begin{aligned} (4.8) \quad \|U - \tilde{U}_p\| &\leq m \cdot C_1 d(\mathcal{C})^{-1} \exp(-C_2 p d(\mathcal{C})) \max_{j=1, \dots, m} \|x_j x_j^* B\| \|Y\|, \\ d(\mathcal{C}) &:= \min_{\lambda \in \text{spec}(A, B)} \text{dist}(\lambda, \mathcal{C}) \end{aligned}$$

by applying Lemma 4.2 to every term of (4.6) and using the triangular inequality. This shows that the normwise error in the computed basis decays exponentially with the integration order. The following theorem summarizes this result and substantiates it for the special case of φ being a circle. Again, it is a special version of a theorem of Beyn [2] that can be found in [12].

THEOREM 4.3 (Beyn ([2, Theorem 4.7]), [12]). *Let the curve φ be the parametrization of \mathcal{C} and fulfill the conditions of Lemma 4.2. Then there are constants C_1, C_2 as defined above such that (4.8) holds with $d(\mathcal{C}) = \min_{\lambda \in \text{spec}(A, B)} \text{dist}(\lambda, \mathcal{C})$. If $\varphi(t) = c + r \exp(it)$, then we have*

$$\|U - \tilde{U}_p\| \leq m \cdot C_1 (\alpha_-^p + \alpha_+^p) \max_{j=1, \dots, m} \|x_j x_j^* B\| \|Y\|$$

with

$$\alpha_- = \max_{\lambda \in \text{spec}(A, B), |\lambda - c| < r} \frac{|\lambda - c|}{r}, \quad \alpha_+ = \max_{\lambda \in \text{spec}(A, B), |\lambda - c| > r} \frac{r}{|\lambda - c|}.$$

Suppose that the eigenvectors $x_j, j = 1, \dots, n$, form a B -orthonormal system. Then we have $\|x_j x_j^* B\| = 1$ for all j since the matrices $x_j^* x_j B$ are projectors.

4.2. Error bounds for the Gauß-Legendre rule. Another widely used integration rule is the Gauß-Legendre rule; see, e.g., [7]. It has proven to be a reliable integration method in the context of the FEAST algorithm [13, 17].

Error bounds for the Gauß-Legendre rule typically include high-order derivatives of the function at unknown points of the integration interval [7, p. 98]. Therefore, they are not very useful in practice.

In this section we will derive a derivative-free bound for $\|U - \tilde{U}_p\|$ that shows exponential decay with p . We use an older and (in the authors' impression) not well-known result of Davis [6]. The key ingredient is that Gauß-Legendre rules of order p integrate polynomials of order $\leq 2p + 1$ exactly.

LEMMA 4.4 ([6]). *Let the (scalar-valued) function f be analytic on $[0, 2\pi]$ and analytically continuable throughout the interior of an ellipse whose foci are at 0 and 2π and whose sum of the semi-axes is γ . Then, for every $\epsilon > 0$, there is an integer p_ϵ such that for all integers $p > p_\epsilon$ we have*

$$|E_{G_p}(f)| \leq 4\pi \left(\frac{\pi}{\gamma} + \epsilon \right)^{2p+1}.$$

The functions $t \mapsto \varphi'(t) r_\lambda(\varphi(t))$ from the previous section fulfill the hypothesis of the lemma under certain conditions. The parametrization must have a continuation to an ellipse with foci 0 and 2π , and φ must not achieve the value λ , i.e., $\varphi(z) \neq \lambda$ for all values z from the ellipse. We will make statements on the structure of the ellipse below. First, let us state the main result for the error in the Gauß-Legendre rule applied to $(zB - A)^{-1}B$.

THEOREM 4.5 (Error of Gauß-Legendre applied to $(zB - A)^{-1}B$, [12]). *Let (A, B) be a definite matrix pair, and let $X = [x_1, \dots, x_n]$ be its full eigenvector matrix consisting of B -orthonormal eigenvectors ($X^* B X = I$). Let a, b denote the semi-axes of an ellipse with foci 0 and 2π that is chosen such that for all eigenvalues $\lambda \in \text{spec}(A, B)$ the functions $t \mapsto \varphi'(t) r_\lambda(\varphi(t))$ are analytic in the interior of the ellipse. Suppose $\gamma := a + b > \pi$. Suppose further that only the eigenvalues $\lambda_1, \dots, \lambda_m$ reside in $\text{Int}(\mathcal{C})$. Then, for every $\epsilon > 0$, there is a number $p_\epsilon \in \mathbb{Z}_{\geq 0}$ such that for all $p > p_\epsilon$ we have*

$$(4.9) \quad \|U - \tilde{U}_p\|_2 \leq 2\kappa(X) \cdot \left(\frac{\pi}{\gamma} + \epsilon \right)^{2p+1} \cdot \|Y\|_2$$

and

$$(4.10) \quad \|U - \tilde{U}_p\|_{B^2} \leq 2m \cdot \left(\frac{\pi}{\gamma} + \epsilon \right)^{2p+1} \cdot \|Y\|_{B^2},$$

where \tilde{U}_p denotes the approximation (4.4) of U via the Gauß-Legendre method of order p . The norm in (4.10) is defined for an $n \times m$ -matrix M as $\|M\|_{B_2} := \|B^{1/2}M\|_2$.

Proof. To prove (4.9) we first write

$$\begin{aligned} h(t) &= \varphi'(t)(\varphi(t)B - A)^{-1}BY = \varphi'(t)(\varphi(t)I - B^{-1}A)^{-1}Y \\ &= \varphi'(t)X \cdot \text{diag}(r_{\lambda_1}(\varphi(t)), \dots, r_{\lambda_n}(\varphi(t))) \cdot X^{-1}Y \\ &= X \cdot \text{diag}(\varphi'(t)r_{\lambda_1}(\varphi(t)), \dots, \varphi'(t)r_{\lambda_n}(\varphi(t))) \cdot X^{-1}Y, \end{aligned}$$

where $r_\lambda(z) = (z - \lambda)^{-1}$. Define $g_j(t) := \varphi'(t)r_{\lambda_j}(\varphi(t))$, $j = 1, \dots, n$. Then, for every j , the function g_j can analytically be continued to the interior of the ellipse. Hence, the conditions of Lemma 4.4 are fulfilled. For every $\epsilon > 0$ and every j we can find a number $p_j(\epsilon)$ such that

$$\left| E_{G_{p_j}}(g_j) \right| \leq 4\pi \left(\frac{\pi}{\gamma} + \epsilon \right)^{2p_j+1}, \quad p_j \geq p_j(\epsilon), \quad j = 1, \dots, n.$$

Set $p := \max_j(p_j(\epsilon))$. Then, for every j we have

$$(4.11) \quad \left| E_{G_p}(g_j) \right| \leq 4\pi \cdot \left(\frac{\pi}{\gamma} + \epsilon \right)^{2p+1}.$$

It follows that (note the factor $1/(2\pi i)$ in the integral (4.1))

$$\begin{aligned} \|U - \tilde{U}_p\|_2 &= \left\| \frac{1}{2\pi i} E_{G_p}(h) \right\|_2 = \frac{1}{2\pi} \|E_{G_p}(\varphi'(t)(\varphi(t)B - A)^{-1}B)Y\|_2 \\ &= \frac{1}{2\pi} \|X \cdot \text{diag}(E_{G_p}(\varphi'(t)r_{\lambda_1}(\varphi(t))), \dots, E_{G_p}(\varphi'(t)r_{\lambda_n}(\varphi(t)))) \cdot X^{-1}Y\|_2 \\ &\leq \frac{1}{2\pi} \kappa(X) \|Y\|_2 \cdot \|\text{diag}(E_{G_p}(g_1), \dots, E_{G_p}(g_n))\|_2 \\ &= \frac{1}{2\pi} \kappa(X) \|Y\|_2 \cdot \max_j |E_{G_p}(g_j)| \leq \frac{1}{2\pi} \cdot 4\pi \cdot \kappa(X) \cdot \left(\frac{\pi}{\gamma} + \epsilon \right)^{2p+1} \cdot \|Y\|_2 \\ &= 2\kappa(X) \cdot \left(\frac{\pi}{\gamma} + \epsilon \right)^{2p+1} \cdot \|Y\|_2, \end{aligned}$$

where the last inequality is due to (4.11).

In order to prove the other inequality (4.10) we use the eigenvector expansion (4.6), which induces

$$U = \frac{1}{2\pi i} \sum_{j=1}^m \int_0^{2\pi} \varphi'(t)r_{\lambda_j}(\varphi(t)) dt x_j x_j^* BY$$

since $\lambda_{m+1}, \dots, \lambda_n \notin \text{Int}(\mathcal{C})$. Consequently, because the error $E_{G_p}(\cdot)$ is a linear operator,

we have with $p = \max_j(p_j(\epsilon))$

$$\begin{aligned}
 \|U - \tilde{U}_p\|_{\mathbb{B}2} &= \|\mathbb{B}^{1/2}(U - \tilde{U}_p)\|_2 = \|\mathbb{B}^{1/2}E_{G_p}(h)\|_2 \\
 &= \frac{1}{2\pi} \left\| \sum_{j=1}^k [E_{G_p}(\varphi'(t)r_{\lambda_j}(\varphi(t)))] \mathbb{B}^{1/2}x_j x_j^* \mathbb{B}Y \right\|_2 \\
 &\leq \frac{1}{2\pi} \sum_{j=1}^k |E_{G_p}(g_j)| \cdot \|\mathbb{B}^{1/2}x_j x_j^* \mathbb{B}^{1/2} \mathbb{B}^{1/2}Y\|_2 \\
 &\leq \frac{1}{2\pi} \cdot k \cdot 4\pi \left(\frac{\pi}{\gamma} + \epsilon\right)^{2p+1} \cdot \max_j \|\mathbb{B}^{1/2}x_j\|_2^2 \cdot \|\mathbb{B}^{1/2}Y\|_2 \\
 &= 2 \cdot k \cdot \left(\frac{\pi}{\gamma} + \epsilon\right)^{2p+1} \cdot \|Y\|_{\mathbb{B}2}.
 \end{aligned}$$

The last equality follows from $\|\mathbb{B}^{1/2}x_j\|_2 = \|x_j\|_{\mathbb{B}} = 1$. \square

The theorem shows that the norm of $U - \tilde{U}_p$ decays asymptotically with the factor $(\pi/\gamma + \epsilon)^{2p+1}$, hence exponentially with the factor $(\pi/\gamma + \epsilon)$. This factor can be chosen to be smaller than 1 with, e.g., $\epsilon = (1 - \pi/\gamma)/2$.

In the following, we will discuss the structure of the ellipse of Theorem 4.5. For simplicity, suppose that the parametrization φ is given by $\varphi(t) := c + r \exp(it)$, a circle centered at the midpoint of the interval I_λ . For any of the real eigenvalues λ , the equation

$$\lambda = \varphi(z) = c + r \exp(iz)$$

can easily be solved for the complex number z . We have

$$z = \varphi^{-1}(\lambda) = \begin{cases} -i \log\left(\frac{\lambda-c}{r}\right) & , \lambda > c, \\ i \log\left(\frac{r}{c-\lambda}\right) + \pi & , \lambda < c. \end{cases}$$

The number γ from Theorem 4.5 can be computed in dependence of the eigenvalues λ of (A, B) . It is defined as the sum of the semi-axes a, b of the ellipse from the theorem, which can be determined by

$$\begin{aligned}
 \eta_1 &= \min \{ |\operatorname{Im}(\varphi^{-1}(\lambda))| : \lambda \in \operatorname{spec}(A, B), \lambda > c \}, \\
 \eta_2 &= \min \{ |\operatorname{Im}(\varphi^{-1}(\lambda))| : \lambda \in \operatorname{spec}(A, B), \lambda < c \};
 \end{aligned}$$

see Figure 4.1. By η_1 , an ellipse with foci 0 and 2π and semi-axes a_1, b_1 is defined such that $0 + i\eta_1$ is outside of the ellipse. Similarly, by η_2 an ellipse with foci 0 and 2π and semi-axes a_2, b_2 is defined such that $\pi + i\eta_2$ is outside of the ellipse. An ellipse according to Theorem 4.5 can consequently be chosen with semi-axes $a < \min(a_1, a_2)$, $b < \min(b_1, b_2)$.

The numbers a_1, a_2, b_1, b_2 can be computed by means of elementary geometry; see, e.g., [4, pp. 221–222]. For the ellipse defined by η_2 (the height of the ellipse over π), we have $\pi = \sqrt{a_2^2 - b_2^2}$, with $b_2 = \eta_2$, hence $a_2 = \sqrt{\pi^2 + \eta_2^2}$.

The numbers a_1, b_1 are a little harder to track; we can compute them from the equations

$$(4.12) \quad \eta_1 = \frac{b_1^2}{a_1},$$

$$(4.13) \quad 1 = \left(\frac{x - \pi}{a_1}\right)^2 + \left(\frac{y}{b_1}\right)^2,$$

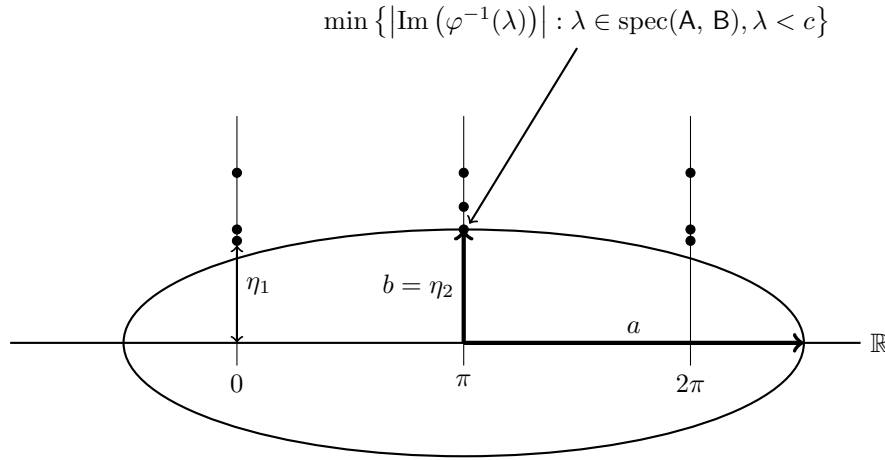


FIG. 4.1. Location of the ellipse from Theorem 4.5. The semi-axes are denoted a and b and marked by the arrows. The dots on the vertical lines denote the absolute values of $\varphi^{-1}(\lambda)$ for eigenvalues λ .

where now (x, y) denotes any point on the ellipse. Solving (4.12) for a_1 and inserting it into (4.13) at $x = 0$ yields the equation of fourth order in b_1 ,

$$\frac{\pi^2}{(b_1^2/\eta_1)^2} + \frac{\eta_1^2}{b_1^2} = 1.$$

The positive solution of this equation is

$$b_1 = +\sqrt{\frac{\eta_1^2}{2} + \sqrt{\left(\frac{\eta_1^2}{2}\right)^2 + \pi^2\eta_1^2}}.$$

The main semi-axis a_1 then can be computed from (4.12).

5. Numerical examples. To illustrate the theory of the previous sections, some numerical examples will examine the connections between the normwise error in the subspace, the approximation error in the eigenvalues, and the canonical angles between the computed subspaces.

EXPERIMENT 5.1 ([12]). We choose a symmetric matrix A of size $n = 100$ at random by setting $\tilde{A} = \text{randn}(n)$, $A := \tilde{A} + \tilde{A}^*$ in MATLAB (version R2013a) and perform essentially the steps of one FEAST iteration with a random orthonormal starting basis $Y \in \mathbb{R}^{100 \times 50}$. First, we measure the normwise errors $\|U - \tilde{U}_p\|$.

If Y has as many columns as the subspace spanned by U , then the theory in Sections 4.1 and 4.2 basically ensures that the errors $\|U - \tilde{U}_p\|$ converge to zero. In practice, however, the dimension of this subspace is not known in advance. Indeed, the subspace spanned by Y should be chosen larger than the dimension of $\text{span}(U)$ [13, 17]. Nevertheless, in this case a subset of the columns of \tilde{U}_p would converge. To avoid the problem of a subspace of wrong dimension, we enforce in this example $\dim(\text{span}(U)) = \dim(\text{span}(Y)) = 50$ by choosing the curve \mathcal{C} such that it encircles the first 50 eigenvalues of A . We have $\lambda_{51} - \lambda_{50} \approx 0.82$, hence we may choose the curve \mathcal{C} such that $d := \text{dist}(\mathcal{C}, \text{spec}(A)) = 0.41$, which would be a fairly large number in practice. The errors $\|U - \tilde{U}_p\|$ are shown in Figure 5.1. Further, the error

bound $2(\pi/\gamma)^{2p+1}$ is displayed for each p that was used. In this test we found $\pi/\gamma \approx 0.99075$, hence a value being close to 1 even though the curve \mathcal{C} has a comfortable distance of about 0.41 to the closest eigenvalue.

Note that the integration orders used are extremely large. They range up to $p = 2000$, while those used in practice in the context of the FEAST algorithm are, e.g. $p = 8$ or $p = 16$ [13, 17]. For these comparatively small values taken from practice, the errors in our experiment were still of order 1. However, the computed subspaces were already able to deliver reasonable eigenvalue approximations. The first 50 exact eigenvalues of A (computed by Matlab's `eig`) as well as the Ritz values belonging to the subspaces computed by the Gauß-Legendre and the trapezoidal rule, each of order $p = 16$, are displayed in the top plot of Figure 5.2. Of course, this figure does not report a small absolute error; the approximation error of each Ritz value is still about 0.9 on average. However, it can be seen that the Ritz values are of similar order as the eigenvalues. The process described here corresponds to one single FEAST iteration with a dimension of the search space \tilde{U}_p being exactly the dimension of the desired eigenspace, which was shown to be problematic [13]. Hence, with a larger subspace, the eigenvalues typically converge faster.

Some of the components of \tilde{U}_p also move into the correct direction for modest values of p . The 50 canonical angles between U and \tilde{U}_p are presented in the bottom plot of Figure 5.2 for $p = 16$. For this value of p , the largest canonical angle is still close to $\pi/2$. For $p = 2000$, the largest canonical angles for both integration schemes were of order comparable to $\|U - \tilde{U}_p\|$ as is stated by the theory; cf. equation (3.6).

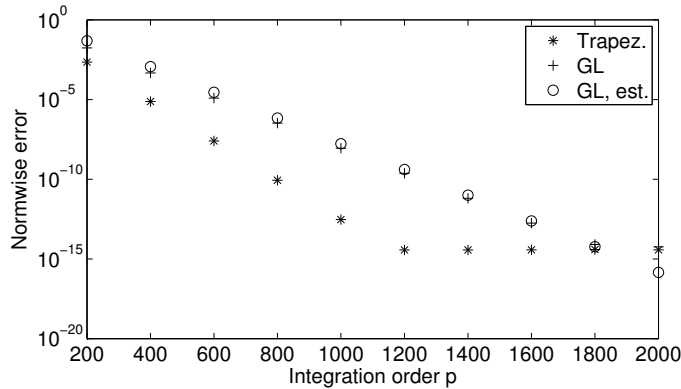


FIG. 5.1. Results for Experiment 5.1. Normwise integration error $\|U - \tilde{U}_p\|$ for the trapezoidal and the Gauß-Legendre rule and the estimated error for the Gauß-Legendre rule.

EXPERIMENT 5.2. Next, we repeat the experiment in a similar fashion for the generalized eigenvalue problem. The matrices A and B are constructed in an analogous way as in Experiment 5.1, and B is forced to be positive definite. Further, A now has complex entries. The results of the integration errors are displayed in the bottom plot of Figure 5.3 together with the theoretical upper bound $m(\pi/\gamma)^{2p+1}$ from (4.10). We used (4.10) for the measured and estimated errors since it does not require the computation of $\kappa(X)$ for the full eigenvector matrix X . For the eigenvalue approximation, the results were comparable to the results from Experiment 5.1. Here, we had $\pi/\gamma \approx 0.9954692$.

The experiment has shown that a small normwise error in the subspace is not necessary for the convergence of the subspaces measured by the canonical angles or for the eigenvalue convergence.

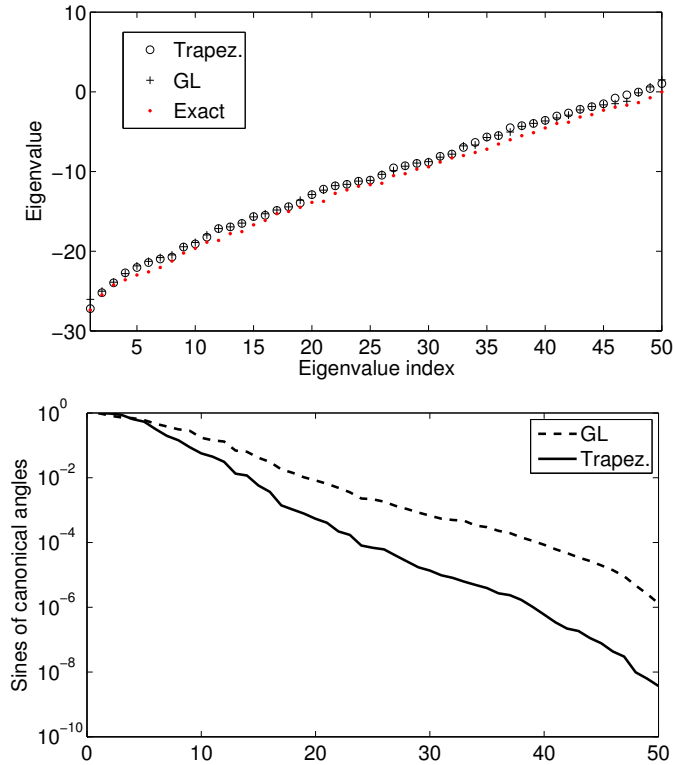


FIG. 5.2. Results for Experiment 5.1. Approximation of the eigenvalues computed by the Gauß-Legendre and the trapezoidal rule, respectively (top). Sines of canonical angles between U and \tilde{U}_p (bottom). We used $p = 16$ integration points in all cases.

The very slow convergence of the subspaces \tilde{U}_p towards U can be justified theoretically. In case of the Gauß-Legendre rule, we have, according to Theorem 4.5, an error bound depending on the ratio π/γ . The number γ depends on the size of the region of analyticity of the resolvent and is basically determined by the distance of the curve to the closest eigenvalue. In Experiment 5.1, both $\kappa(X)$ and $\|Y\|$ have value 1 since X is the eigenvector matrix of the symmetric matrix A , hence orthonormal, and Y was chosen orthonormal. In the second experiment we can choose Y such that $\|Y\|_{B_2} = 1$. Note that when computing the theoretical error bounds we neglected $\epsilon > 0$ from (4.9), (4.10).

For the trapezoidal rule, things are slightly different than for the Gauß-Legendre rule. We have, according to Theorem 4.3,

$$(5.1) \quad \|U - \tilde{U}_p\| \leq m \cdot C_1 d^{-1} \exp(-C_2 p d)$$

if the eigenvectors and Y are orthonormal. In this equation, C_1 and C_2 denote some positive constants, and m is the number of eigenvalues inside \mathcal{C} . Again we denote by d the distance from the curve \mathcal{C} to the closest eigenvalue. With $d = 0.41$ as in Experiment 5.1, the right-hand side of (5.1) converges extremely fast towards zero when C_1 and C_2 are of order 1. Hence, the constant C_2 from the experiment must be very small (but still positive).

For less artificial examples than in Experiments 5.1 and 5.2, the curve \mathcal{C} typically is much closer to the spectrum. For instance, if we have $d = 10^{-6}$ (which could be a value

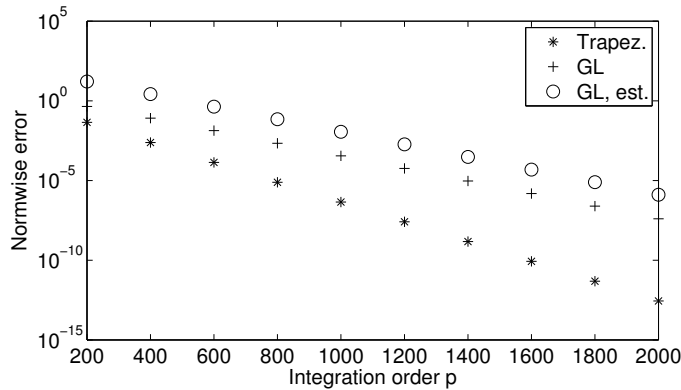


FIG. 5.3. Results for Experiment 5.2. Normwise integration error $\|U - \tilde{U}_p\|$ for the trapezoidal and the Gauß-Legendre rule and the estimated error for the Gauß-Legendre rule.

from practice), we would obtain $\pi/\gamma \approx 0.9999997$, a value whose positive powers converge extremely slowly towards 0. The same holds for the error bound (5.1). However, the statements about the eigenvalue approximation and the canonical angles stay true, at least in a qualitative sense (the convergence of eigenvalues and canonical angles takes place much earlier than the normwise convergence of the subspaces).

In all our experiments with the FEAST algorithm [12, 13], we only use very modest integration orders, $p = 8, 16, 32$. In the literature, however, in actual computations, up to hundreds of thousands ($2^{18} \approx 262,000$) integration points were used in the context of matrix functions [5]. For nonlinear eigenvalue problems, Beyn used at least orders up to $p = 150$; see [2].

As pointed out by one of the referees, one might fear that the quantity $\delta = \|\tilde{U}_p - U\|$ gives a gross overestimation of the error in the computed Ritz values $\rho(\tilde{u})$. Indeed, even assuming that \tilde{U} was obtained by just changing the basis of the subspace, $\tilde{U} = UM$ with a nonsingular matrix M , the quantity δ can be arbitrarily large while the Ritz values of \tilde{U} coincide with those of U . A more precise estimate should be obtainable by considering the action of \tilde{U}_p on the B-orthogonal complement of U . This issue is addressed in the following experiment.

EXPERIMENT 5.3. We consider a total of 140 matrix pairs as follows: For each $n \in \{50, 100, 200, 400\}$ and each $\kappa_B \in \{1, 10^1, 10^2, \dots, 10^6\}$, we generate five size- n matrix pairs (A, B) with random eigenvalues (uniformly distributed in $[-1, +1]$) and $\text{cond}_2(B) = \kappa_B$. For each of these pairs, we search for 10% of the eigenvalues, namely the smallest positive ones (with indices $i_1 : i_2$, say, in ascending order), using a circle through the midpoints of the “boundary gaps,” $(\lambda_{i_1-1} + \lambda_{i_1})/2$ and $(\lambda_{i_2} + \lambda_{i_2+1})/2$, a random B-orthonormal matrix $Y \in \mathbb{R}^{n \times n/10}$, and 17 trapezoidal and Gauß-Legendre integration rules with p points, where $p \in \{8, 12, 16, 24, 32, 48, \dots, 1536, 2048\}$.

First we observe that in all the $140 \times 17 \times 2 = 4760$ runs, the value of the ratio $\|(I - UU^*B)(\tilde{U}_p - U)\|_{B_2} / \|\tilde{U}_p - U\|_{B_2}$ never fell below 0.39, and only 10 of these values were below 0.5; cf. Theorem 4.5 for the definition of the B_2 -norm. Since $I - UU^*B$ is the projector onto the B-orthogonal complement of U , this means that the overall error $\tilde{U}_p - U$ typically contains a significant component in that direction.

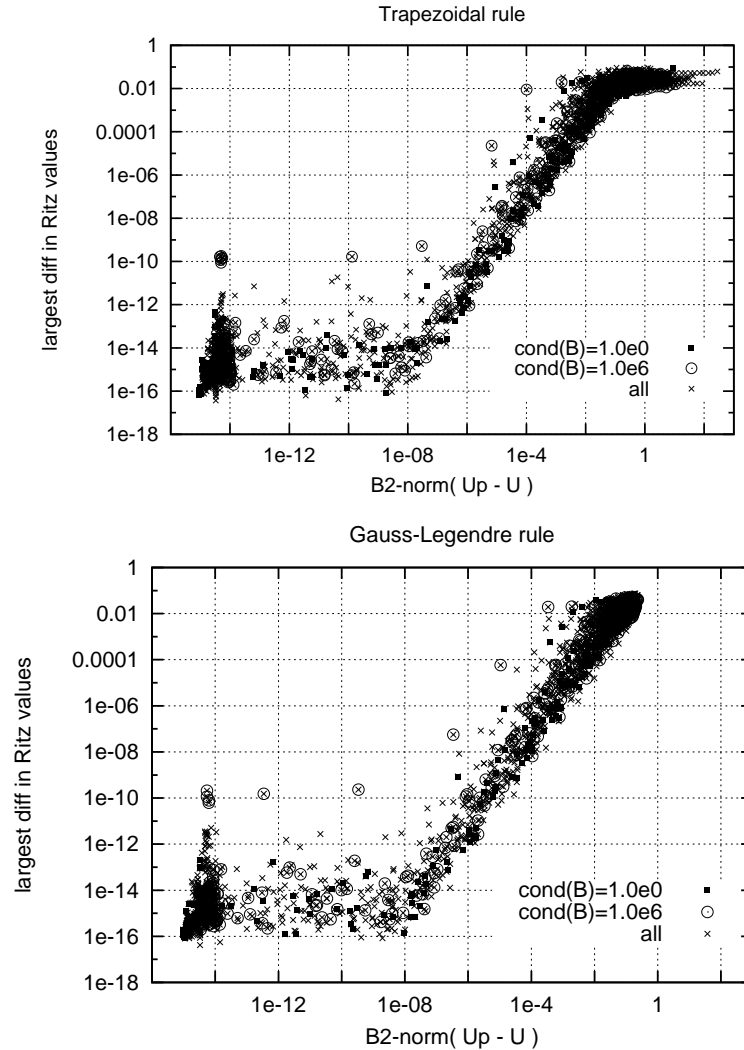


FIG. 5.4. Largest error in the Ritz values, $\max_i |\rho(\tilde{u}_i) - \lambda_i|$, versus the error in the subspace, $\delta = \|\tilde{U}_p - U\|_{B_2}$, for the trapezoidal rule (top) and the Gauß-Legendre rule (bottom).

In Figure 5.4 the maximum error in the computed Ritz values, $\max_{i=1}^{i_2} |\rho(\tilde{u}_i) - \lambda_i|$, is plotted against the error in the subspace $\delta = \|\tilde{U}_p - U\|_{B_2}$. Each plot summarizes $140 \times 17 = 2380$ runs. For $10^{-8} \leq \delta \leq 10^{-2}$ the quadratic approximation of the Ritz values is clearly visible; for $\delta < 10^{-8}$ it can no longer be achieved due to round-off. Since the “extreme” cases $\kappa_B = 1$ (essentially a standard eigenproblem) and $\kappa_B = 10^6$ are shown with different markers we can also see that a large κ_B can have an impact on the quality of the results, but the outliers are rather rare.

Overall, even if the quantity δ cannot be guaranteed to be a good indicator for the error in the Ritz values this seems to be the case in practice.

6. Conclusions. We presented an error analysis of eigensolvers based on integration. First, we investigated the convergence of Ritz vectors and Ritz values towards eigenvectors

and eigenvalues, respectively, in Section 3. The derived error bounds depend on the norm of the error $U - \tilde{U}$ in a certain basis of the eigenspace. The results can also be used independently of how \tilde{U} was computed. This can be done by an integration scheme or by other approximation methods, e.g., based on polynomials [12, 21] or rational approximation [9]. Afterwards, we presented a convergence analysis of two different integration schemes for the computation of \tilde{U} in Section 4. In practice, the search space $\tilde{\mathcal{U}}$ typically has a larger dimension than the actual eigenspace \mathcal{X} . This is even necessary in order to obtain fast convergence of the outer loop of Algorithm 2.1; see [26]. In this case, the given analysis applies to a subspace of $\tilde{\mathcal{U}}$ that has the same dimension as \mathcal{X} .

Taken together, the two kinds of convergences discussed in Sections 3 and 4 imply convergence of the integration-based method with respect to the integration order p , meaning that in every iteration of Algorithm 2.1 we obtain approximate eigenpairs of predictable quality that improve over the iterations.

Acknowledgments. The authors gratefully acknowledge the comments of the anonymous referees.

REFERENCES

- [1] W. E. ARNOLDI, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
- [2] W.-J. BEYN, *An integral method for solving nonlinear eigenvalue problems*, Linear Algebra Appl., 436 (2012), pp. 3839–3863.
- [3] Å. BJÖRCK AND G. H. GOLUB, *Numerical methods for computing angles between linear subspaces*, Math. Comp., 27 (1973), pp. 579–594.
- [4] I. N. BRONSTEIN AND K. A. SEMENDJAJEW, *Taschenbuch der Mathematik*, 21st ed., Harri Deutsch, Frankfurt am Main, 1984.
- [5] P. I. DAVIES AND N. J. HIGHAM, *Computing $f(A)b$ for matrix functions f* , in QCD and Numerical Analysis III, A. Boriçi, A. Frommer, B. Joó, A. Kennedy, and B. Pendleton, eds., vol. 47 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, 2005, pp. 15–24.
- [6] P. J. DAVIS, *On the numerical integration of periodic analytic functions*, in On Numerical Approximation. Proceedings of a Symposium, Madison, April 21-23, 1958, R. E. Langer, eds., Publication No. 1 of the Mathematics Research Center, U.S. Army, the University of Wisconsin, University of Wisconsin Press, Madison, 1959, pp. 45–59.
- [7] P. J. DAVIS AND P. RABINOWITZ, *Methods of Numerical Integration*, 2nd ed., Academic Press, Orlando, 1984.
- [8] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, 2013.
- [9] S. GÜTTEL, E. POLIZZI, P. T. P. TANG, AND G. VIAUD, *Zolotarev quadrature rules and load balancing for the FEAST eigensolver*, SIAM J. Sci. Comput., 37 (2015), pp. A2100–A2122.
- [10] A. V. KNYAZEV AND M. E. ARGENTATI, *Principal angles between subspaces in an A -based scalar product: algorithms and perturbation estimates*, SIAM J. Sci. Comput., 23 (2002), pp. 2008–2040.
- [11] ———, *On proximity of Rayleigh quotients for different vectors and Ritz values generated by different trial subspaces*, Linear Algebra Appl., 415 (2006), pp. 82–95.
- [12] L. KRÄMER, *Integration Based Solvers for Standard and Generalized Hermitian Eigenvalue Problems*, PhD. Thesis, Fachbereich Mathematik und Naturwissenschaften, Bergische Universität Wuppertal, Wuppertal 2014.
- [13] L. KRÄMER, E. DI NAPOLI, M. GALGON, B. LANG, AND P. BIENTINESI, *Dissecting the FEAST algorithm for generalized eigenproblems*, J. Comput. Appl. Math., 244 (2013), pp. 1–9.
- [14] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282.
- [15] Y. NAKATSUKASA, *Absolute and relative Weyl theorems for generalized eigenvalue problems*, Linear Algebra Appl., 432 (2010), pp. 242–248.
- [16] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, SIAM, Philadelphia, 1998.
- [17] E. POLIZZI, *Density-matrix-based algorithm for solving eigenvalue problems*, Phys. Rev. B, 79 (2009), Art. 115112, (6 pages).
- [18] M. RÖHRIG-ZÖLLNER, J. THIES, M. KREUTZER, A. ALVERMANN, A. PIEPER, A. BASERMANN, G. HAGER, G. WELLEIN, AND H. FEHSKE, *Increasing the performance of the Jacobi-Davidson*

- method by blocking*, SIAM J. Sci. Comput., 37 (2015), pp. C697–C722.
- [19] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, SIAM, Philadelphia, 2011.
 - [20] T. SAKURAI AND H. SUGIURA, *A projection method for generalized eigenvalue problems using numerical integration*, J. Comput. Appl. Math., 159 (2003), pp. 119–128.
 - [21] G. SCHOFIELD, J. R. CHELIKOWSKY, AND Y. SAAD, *A spectrum slicing method for the Kohn-Sham problem*, Comput. Phys. Commun., 183 (2012), pp. 497–505.
 - [22] G. L. G. SLEIJPEN AND H. A. VAN DER VORST, *A Jacobi-Davidson iteration method for linear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 401–425.
 - [23] D. C. SORENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.
 - [24] G. W. STEWART, *Matrix Algorithms. Volume II: Eigensystems*, SIAM, Philadelphia, 2001.
 - [25] G. W. STEWART AND J. G. SUN, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
 - [26] P. T. P. TANG AND E. POLIZZI, *FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 354–390.
 - [27] L. N. TREFETHEN AND J. A. C. WEIDEMAN, *The exponentially convergent trapezoidal rule*, SIAM Rev., 56 (2014), pp. 385–458.
 - [28] J. A. C. WEIDEMAN, *Numerical integration of periodic functions: a few examples*, Amer. Math. Monthly, 109 (2002), pp. 21–36.
 - [29] H. WEYL, *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)*, Math. Ann., 71 (1912), pp. 441–479.