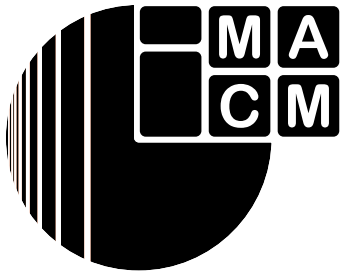

Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 14/30

Lukas Krämer

Convergence of integration based methods for the solution of standard and generalized Hermitian eigenvalue problems

October 10, 2014

http://www.math.uni-wuppertal.de

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

LUKAS KRÄMER†

Abstract. Recently, methods based on spectral projection and numerical integration came up 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 study the convergence of the resulting method. We investigate the convergence of the Rayleigh–Ritz process in dependence of the normwise difference between the used subspace and the desired eigenspace. Then, upper bounds for these norms 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

$$\mathsf{A}\mathsf{X} = \mathsf{B}\mathsf{X}\Lambda \tag{1.1}$$

where $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{n\times n}$, 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 $\mathsf{X} \in \mathbb{C}^{n\times m}$ 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., $\mathsf{X}^\star\mathsf{B}\mathsf{X} = \mathsf{I}$ [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 came up, 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, we have a B-orthogonal projector P onto the subspace $\mathcal{X} = \operatorname{span}(\mathsf{X})$ on 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 $\mathsf{U} := \mathsf{P}\mathsf{Y}$ for some $n \times m$ matrix Y. If Y was chosen with full rank and without any components being B-orthogonal to $\mathcal{X}$, the matrix U spans the eigenspace $\mathcal{X}$, i. e., $\operatorname{span}(\mathsf{U}) = \mathcal{X}$. If the projector P is available only approximately, which is the case in practice, 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. 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].

*This work was partially supported by the Deutsche Forschungsgemeinschaft through the priority programme 1648 "Software for Exascale Computing" (SPPEXA) under project ESSEX.

†Bergische Universität Wuppertal, Fachbereich C – Mathematik und Naturwissenschaften, 42097 Wuppertal, Germany, lkraemer@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 Algorithm 2.1 below). 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. Further we analyze the implications of this convergence to the errors in the computed eigenvalues and eigenvectors.
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 $\mathsf{U} - \widetilde{\mathsf{U}}$, where U is a basis of the eigenspace and $\widetilde{\mathsf{U}}$ is an approximation to it. In Section 4 we state error bounds for certain norms of this distance $\mathsf{U} - \widetilde{\mathsf{U}}$ for the case that $\widetilde{\mathsf{U}}$ os obtained by the trapezoidal and Gauß–Legendre rule, respectively. Some numerical results are reported in Section 5 and in Section 6 we conclude the article.
Throughout the article, the symbol $\|\cdot\|$ will denote the Euclidean norm of a vector or matrix, respectively.

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

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

In step 4, a column of the matrix $\widetilde{\mathsf{U}}\widetilde{\mathsf{W}}$ is called a *Ritz vector*, the corresponding entry of $\widetilde{\Lambda}$ is called *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 $\widetilde{\mathsf{U}}$ is computed by an approximation to 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]

$$\mathsf{P} = \frac{1}{2\pi\mathbf{i}}\int_{\mathcal{C}} (z\mathsf{B} - \mathsf{A})^{-1}\mathsf{B}\mathrm{d}z. \tag{2.1}$$

In (2.1), the symbol $\mathcal{C}$ denotes a simply closed curve in $\mathbb{C}$ that encircles exactly the m eigenvalues (counted with multiplicity) of $(\mathsf{A},\ \mathsf{B})$ that reside in I_λ and no other.
The subspace $\widetilde{\mathcal{U}}$ in the Rayleigh–Ritz procedure can be chosen as the span of $\mathsf{U} := \mathsf{PY}$, where Y is a full rank matrix with m columns. If $\mathsf{Y}^\star\mathsf{BX}$ has full rank, where X is a basis of the eigenspace belonging to the eigenvalues of $(\mathsf{A},\ \mathsf{B})$ residing in I_λ, we have that $\mathrm{span}(\mathsf{U}) = \mathcal{X}$. The integral (2.1) and therefore the matrix $\mathsf{U} = \mathsf{PY}$ can be approximated by numerical integration. The resulting basic FEAST algorithm uses an approximation $\widetilde{\mathsf{U}}$ to U and reads as in Algorithm 2.1. It is basically a Rayleigh–Ritz procedure with a special choice of $\widetilde{\mathsf{U}}$.
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 $\mathsf{U} - \widetilde{\mathsf{U}}$, where $\widetilde{\mathsf{U}}$ denotes an approximation to U.

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

Input: An interval $I_\lambda = \left[\underline{\lambda}, \overline{\lambda}\right]$ and an estimate $\widetilde{m}$ of the number of eigenvalues in I_λ.
Output: $\hat{m} \leq \widetilde{m}$ eigenpairs with eigenvalues in I_λ.

1: Choose $\mathsf{Y} \in \mathbb{C}^{n \times \widetilde{m}}$ of rank $\widetilde{m}$ and compute an approximation $\widetilde{\mathsf{U}}$ to

$$\mathsf{U} := \frac{1}{2\pi \mathbf{i}} \int_{\mathcal{C}} (z\mathsf{B} - \mathsf{A})^{-1} \mathsf{B}\, \mathsf{Y} \mathrm{d}z.$$

2: Form the Rayleigh quotients $\mathsf{A}_{\widetilde{\mathsf{U}}} := \widetilde{\mathsf{U}}^\star \mathsf{A} \widetilde{\mathsf{U}}$, $\mathsf{B}_{\widetilde{\mathsf{U}}} := \widetilde{\mathsf{U}}^\star \mathsf{B} \widetilde{\mathsf{U}}$.
3: Solve the size-$\widetilde{m}$ generalized eigenproblem $\mathsf{A}_{\widetilde{\mathsf{U}}} \widetilde{\mathsf{W}} = \mathsf{b}_{\widetilde{\mathsf{U}}} \widetilde{\mathsf{W}} \widetilde{\Lambda}$.
4: Compute the approximate Ritz pairs $(\widetilde{\Lambda}, \widetilde{\mathsf{X}} := \widetilde{\mathsf{U}} \cdot \widetilde{\mathsf{W}})$.
5: If convergence is not reached then go to Step 1 with $\mathsf{Y} := \widetilde{\mathsf{X}}$.

3. Perturbation of eigenvalues and eigenvectors. In this section we will give a short overview on error bounds for eigenvalues and eigenvectors obtained by the Rayleigh–Ritz approach. The errors depend on the normwise difference $\delta := \|\mathsf{U} - \widetilde{\mathsf{U}}\|$, where U denotes a basis of the desired eigenspace and $\widetilde{\mathsf{U}}$ is an approximation. If the matrix Y in (2.1) was chosen properly (see above), U obtained by (2.1) is a basis for the eigenspace. However, we do not require $\widetilde{\mathsf{U}}$ to be obtained by an integration process. For instance, $\widetilde{\mathsf{U}}$ 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 $\|\mathsf{U} - \widetilde{\mathsf{U}}\|$ are typically not available, hence the presented analysis does not apply.
For any $n \times m$-matrix X, let A_X define the Rayleigh quotient $\mathsf{X}^\star \mathsf{A} \mathsf{X}$ corresponding to the space $\mathrm{span}(\mathsf{X})$. Let B_X be defined accordingly.

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

$$\mathsf{E} = \mathsf{A}_{\mathsf{U} - \widetilde{\mathsf{U}}} + (\widetilde{\mathsf{U}} - \mathsf{U})^\star \mathsf{A} \mathsf{U} + \mathsf{U}^\star \mathsf{A} (\widetilde{\mathsf{U}} - \mathsf{U}). \tag{3.1}$$

Hence, $\|\mathsf{E}\| \leq \|\mathsf{A}\| \, (\delta^2 + 2\delta \, \|\mathsf{U}\|)$. Then, as a consequence of Weyl's theorem [30] (see [25, Cor. 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, Thm. 2.7]). *Let A be Hermitian and consider the standard equation $\mathsf{A}\mathsf{X} = \mathsf{X}\Lambda$. Suppose an error bound for the subspace, $\delta := \|\widetilde{\mathsf{U}} - \mathsf{U}\|$, is at hand. Let $\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_m$ denote the Ritz values of A with respect to $\widetilde{\mathsf{U}}$ and $\lambda_1, \ldots, \lambda_m$ the eigenvalues of A belonging to the space $\mathrm{span}(\mathsf{U})$, both ordered ascendingly. We then have*

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

In the generalized case, involving a definite matrix pair (A, B), we first can interpret the Rayleigh quotient $\mathsf{B}_{\widetilde{\mathsf{U}}}$ as a perturbed Rayleigh quotient of B_{U}, where U fulfills $\mathsf{AU} = \mathsf{BUH}$ for some square matrix H. Similarly to (3.1), we obtain

$$\mathsf{F} = \mathsf{B}_{\mathsf{U}-\widetilde{\mathsf{U}}} + (\widetilde{\mathsf{U}} - \mathsf{U})^\star \mathsf{BU} + \mathsf{U}^\star \mathsf{B}(\widetilde{\mathsf{U}} - \mathsf{U})$$

with $\|\mathsf{F}\| \leq (\delta^2 + 2\delta \|\mathsf{U}\|)$.
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 $\widetilde{\mathsf{U}}$ that depends on δ. The result can be formulated as follows.

THEOREM 3.2 ([12, Thm. 2.9]). *Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ denote the eigenvalues of $(\mathsf{A}_\mathsf{U}, \mathsf{B}_\mathsf{U})$. Let $\|\mathsf{F}\| < \lambda_{\min}(\mathsf{B}_\mathsf{U})$. Then, the perturbed pair $(\mathsf{A}_\mathsf{U} + \mathsf{E}, \mathsf{B}_\mathsf{U} + \mathsf{F})$ is Hermitian definite and its eigenvalues $\widetilde{\lambda}_1 \leq \widetilde{\lambda}_2 \leq \cdots \leq \widetilde{\lambda}_m$ satisfy*

$$\max_j \left|\lambda_j - \widetilde{\lambda}_j\right| \leq \frac{\|\mathsf{A}\| \cdot \Delta}{\lambda_{\min}(\mathsf{B}_\mathsf{U})} + \frac{\|\mathsf{A}_\mathsf{U}\| + \|\mathsf{A}\| \cdot \Delta}{\lambda_{\min}(\mathsf{B}_\mathsf{U})\,(\lambda_{\min}(\mathsf{B}_\mathsf{U}) - \|\mathsf{B}\| \cdot \Delta)} \cdot \|\mathsf{B}\| \cdot \Delta, \tag{3.2}$$

where $\Delta := (\delta^2 + 2\delta \|\mathsf{U}\|)$.
The right hand side of (3.2) converges to 0 as $\delta \to 0$.
Other eigenvalue bounds depending on the norm $\|\mathsf{U} - \widetilde{\mathsf{U}}\|$ can be found in [11]. They are also consequences of Weyl's theorem and yield upper bounds on the approximation error of similar order as those stated above. In [12, Cor. 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 that 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 $\mathsf{x}, \mathsf{y} \in \mathbb{C}^n$ the angle $\theta := \angle(\mathsf{x}, \mathsf{y})$ is defined implicitly by the relation

$$\cos\theta = \frac{|\mathsf{x}^\star \mathsf{y}|}{\|\mathsf{x}\|\,\|\mathsf{y}\|}.$$

The well-definition follows from the Cauchy-Schwartz inequality.
Angles can also be measured in the scalar product defined by a Hermitian positive definite matrix B, that is, $\mathsf{x}^\star \mathsf{B}\mathsf{y}$. This scalar product first induces a norm, the so called B-norm $\|\mathsf{x}\|_\mathsf{B} := \sqrt{\mathsf{x}^\star \mathsf{B}\mathsf{x}}$. Then the angle between x and y in the B scalar product, $\angle_\mathsf{B}(\mathsf{x}, \mathsf{y})$, is defined implicitly by the equation

$$\cos\angle_\mathsf{B}(\mathsf{x}, \mathsf{y}) = \frac{|\mathsf{x}^\star \mathsf{B}\mathsf{y}|}{\|\mathsf{x}\|_\mathsf{B}\,\|\mathsf{y}\|_\mathsf{B}}.$$

The usual way to measure angles between a vector and a subspace or between two subspaces is by means of the so called *principal angles*, see [8, Sec. 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 $\angle(\mathcal{U}, \mathcal{V})$. It can be shown that $\cos\angle(\mathcal{U}, \mathcal{V}) = \|\mathsf{U}^\star \mathsf{V}\|$, given orthonormal bases U and V of $\mathcal{U}$ and $\mathcal{V}$, respectively. For the B-scalar product we then have $\cos\angle_\mathsf{B}(\mathcal{U}, \mathcal{V}) = \|\mathsf{U}^\star \mathsf{B}\mathsf{V}\|$, if U and V are orthonormal with respect to that scalar product [10].

For any two spaces $\mathcal{U} = \operatorname{span}(\mathsf{U}), \widetilde{\mathcal{U}} = \operatorname{span}(\widetilde{\mathsf{U}})$ of same dimension it can be shown that [10, Lem. 5.5]

$$\sin \angle_{\mathsf{B}}(\mathcal{U}, \widetilde{\mathcal{U}}) \leq \kappa_{\mathsf{B}}(\mathsf{U}) \frac{\left\| \mathsf{B}^{1/2}(\mathsf{U} - \widetilde{\mathsf{U}}) \right\|}{\left\| \mathsf{B}^{1/2}\mathsf{U} \right\|}, \tag{3.3}$$

where $\kappa_{\mathsf{B}} = \sigma_{\max}(\mathsf{B}^{1/2}\mathsf{U})/\sigma_{\min}(\mathsf{B}^{1/2}\mathsf{U})$ is the condition number of U with respect to the B-norm. The symbol $\mathsf{B}^{1/2}$ 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 $\mathsf{K}^\star\mathsf{K} = \mathsf{B}$.
Thinking of $\mathcal{U}$ as an eigenspace of $(\mathsf{A},\ \mathsf{B})$ with B-orthonormal basis U, (3.3) boils down to

$$\sin \angle_{\mathsf{B}}(\mathcal{U}, \widetilde{\mathcal{U}}) \leq \left\| \mathsf{B}^{1/2}(\mathsf{U} - \widetilde{\mathsf{U}}) \right\|, \tag{3.4}$$

since $\mathsf{B}^{1/2}\mathsf{U}$ 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, statements on the largest principal angle between those two spaces can be made. Note that for small numbers θ we have $\theta \approx \sin\theta$. It is clear that for a vector $\mathsf{u} \in \mathcal{U}$ we have

$$\angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathsf{U}}) \leq \angle_{\mathsf{B}}(\mathcal{U}, \widetilde{\mathsf{U}}). \tag{3.5}$$

This follows from the min-max definition of angles, see [3].
Next, we will explain how bounds for the angle $\angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathsf{u}})$ for a an eigenvector u and a vector $\widetilde{\mathsf{u}} \in \widetilde{\mathcal{U}}$ can be obtained. From the convergence of Ritz values and a small angle $\angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathcal{U}})$ it does not follow that there is vector $\widetilde{\mathsf{u}} \in \widetilde{\mathcal{U}}$ such that $\angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathsf{u}})$ is small. See the discussion in [24]. The convergence of Ritz vectors depends on the distance of Ritz values.

THEOREM 3.3. *Let (u, λ) be any eigenpair of $(\mathsf{A},\ \mathsf{B})$. Let $\widetilde{\lambda}$ be an approximate eigenvalue extracted from $\widetilde{\mathcal{U}}$ and let η be the distance between λ and the approximate eigenvalues other than $\widetilde{\lambda}$. Then there is an approximate eigenvector $\widetilde{\mathsf{u}}$ associated with $\widetilde{\lambda}$ such that*

$$\sin \angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathsf{u}}) \leq \sin \angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathcal{U}}) \sqrt{1 + \frac{\gamma^2}{\eta^2}},$$

where $\gamma \leq \|\mathsf{A}\|$.

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

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

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

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

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

$$\gamma = \|\mathsf{P}\mathsf{A}(\mathsf{I} - \mathsf{P})\|,$$

where $\mathsf{P} = (\mathsf{B}^{1/2}\widetilde{\mathsf{U}})^\star \mathsf{B}^{1/2}\widetilde{\mathsf{U}}$ is the orthogonal projector onto the space $\mathsf{B}^{1/2}\widetilde{\mathcal{U}}$. Hence, $\gamma \leq \|\mathsf{A}\|$. Using $\angle(\mathsf{B}^{1/2}\mathsf{u}, \mathsf{B}^{1/2}\widetilde{\mathsf{u}}) = \angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathsf{u}})$ and $\angle(\mathsf{B}^{1/2}\mathsf{u}, \mathsf{B}^{1/2}\widetilde{\mathcal{U}}) = \angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathcal{U}})$ (see [10]) finishes the proof. □

Combining Theorem 3.3 with (3.4) and (3.5) yields convergence of Ritz vectors towards eigenvectors, if $\widetilde{\mathsf{U}} \longrightarrow \mathsf{U}$. More precisely, if U is a B-orthonormal matrix, then, as $\widetilde{\mathsf{U}} \longrightarrow \mathsf{U}$ we obtain

$$\begin{aligned} \sin \angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathsf{u}}) &\leq \sin \angle_{\mathsf{B}}(\mathsf{u}, \widetilde{\mathcal{U}}) \cdot C \\ &\leq \sin \angle_{\mathsf{B}}(\mathsf{U}, \widetilde{\mathcal{U}}) \cdot C \\ &\leq \left\| \mathsf{B}^{1/2}(\mathsf{U} - \widetilde{\mathsf{U}}) \right\| \cdot C \\ &\longrightarrow 0, \end{aligned}$$

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

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

$$\mathsf{U} = \mathsf{P}\mathsf{Y} = \frac{1}{2\pi \mathbf{i}} \int_{\mathcal{C}} (z\mathsf{B} - \mathsf{A})^{-1} \mathsf{B}\mathsf{Y} \mathrm{d}z. \tag{4.1}$$

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

$$\mathsf{U} = \frac{1}{2\pi \mathbf{i}} \int_0^{2\pi} \varphi'(t)(\varphi(t)\mathsf{B} - \mathsf{A})^{-1} \mathsf{B}\mathsf{Y} \mathrm{d}t. \tag{4.2}$$

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

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

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

we obtain for (4.2)

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

By an *integration scheme of order* p we denote a sequence of pairs $(\omega_j, t_j)_{j=0,\ldots,p}$, 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

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

The goal of this section is to derive error bounds for certain norms of $\mathsf{U} - \widetilde{\mathsf{U}}_p$. 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, 29]. 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 [28] that gives a thorough overview on the convergence of the trapezoidal rule applied to periodic functions was puplished.

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

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

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

The following theorem shows that the error of the trapezoidal rule, applied to a scalar function, decays exponentially with p if the function is periodic in an open strip containing the real line. A detailed proof can also be found in [12].

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

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

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

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

satisfies

$$\left|E_{T_p}(f)\right| \le \max_{\mathrm{Im}(z)=\sigma_+} |f(z)|F(\exp(-p\sigma_+)) + \max_{\mathrm{Im}(z)=\sigma_-} |f(z)|F(\exp(p\sigma_-)), \tag{4.5}$$

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

When investigating the integration of the resolvent $G(z) := (z\mathsf{B} - \mathsf{A})^{-1}\mathsf{B}$, the scalar valued functions $r_\lambda(z) := (z - \lambda)^{-1}$ will play a prominent role. If the matrix pair (A, B) has eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding B-orthonormal eigenvectors $\mathsf{x}_1 \ldots, \mathsf{x}_n$, we can write

$$G(z) = \sum_{j=1}^{n} r_{\lambda_j}(z)\mathsf{x}_j\mathsf{x}_j^\star\mathsf{B}.$$

This leads to the following representation of (4.1),

$$\begin{aligned}
\mathsf{U} &= \frac{1}{2\pi\mathbf{i}}\int_{\mathcal{C}} (z\mathsf{B} - \mathsf{A})^{-1}\mathsf{B}\mathsf{Y}\mathrm{d}z \\
&= \frac{1}{2\pi\mathbf{i}}\int_{\mathcal{C}} \sum_{j=1}^{m} r_{\lambda_j}(z)\mathsf{x}_j\mathsf{x}_j^\star\mathsf{B}\mathsf{Y}\mathrm{d}z \\
&= \frac{1}{2\pi\mathbf{i}}\sum_{j=1}^{m}\left(\int_{\mathcal{C}} r_{\lambda_j}(z)\mathrm{d}z\right)\mathsf{x}_j\mathsf{x}_j^\star\mathsf{B}\mathsf{Y}, \qquad (4.6)
\end{aligned}$$

where we suppose that the eigenvalues are ordered 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

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

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

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

for certain $0 > \sigma_- > s_-$, $0 < \sigma_+ < s_+$. The following lemma is a special case of [2, Lemma 4.6] and substantiates (4.7)
LEMMA 4.2 (Beyn, [2, Lemma 4.6], [12]). *Let φ be defined on S and 2π-periodic, further let $\varphi(z) \in \mathrm{Int}(\mathcal{C})$ for $\mathrm{Im}(z) > 0$ and $\varphi(z) \in \mathrm{Ext}(\mathcal{C})$ for $\mathrm{Im}(z) < 0$. Let $\mathrm{dist}(\lambda, \mathcal{C}) = \min_{z\in\mathcal{C}} |\lambda - z|$. Then there are C_1, C_2, $C_3 > 0$ such that*

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

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

$$\begin{aligned}(4.8)\qquad \left\|\mathsf{U} - \widetilde{\mathsf{U}}_p\right\| &\leq m \cdot C_1 d(\mathcal{C})^{-1} \exp(-C_2\, p\, d(\mathcal{C})) \max_{j=1,\dots,m} \left\|\mathsf{x}_j\mathsf{x}_j^\star\mathsf{B}\right\| \|\mathsf{Y}\|, \\ d(\mathcal{C}) &:= \min_{\lambda\in\mathrm{spec}(\mathsf{A},\,\mathsf{B})} \mathrm{dist}(\lambda, \mathcal{C}),\end{aligned}$$

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

$$\left\|\mathsf{U} - \widetilde{\mathsf{U}}_p\right\| \leq m \cdot C_1(\alpha_-^p + \alpha_+^p) \max_{j=1,\dots,m} \left\|\mathsf{x}_j\mathsf{x}_j^\star\mathsf{B}\right\| \|\mathsf{Y}\|$$

with

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

If we suppose that the eigenvectors x_j, $j = 1, \dots, n$ form a B-orthonormal system, we have $\left\|\mathsf{x}_j\mathsf{x}_j^\star\mathsf{B}\right\| = 1$ for all j since the matrices $\mathsf{x}_j^\star\mathsf{x}_j\mathsf{B}$ are projectors.

4.2. Error bounds for 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 $\|\mathsf{U} - \widetilde{\mathsf{U}}_p\|$ that shows exponential decay with p. We use an older and (in the author's 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 continuable analytically throughout the interior of an ellipse whose foci are at 0 and 2π and whose sum of 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*

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

The functions $t \mapsto \varphi'(t) r_\lambda(\varphi(t))$ from the previous section fulfill the prerequisites of the lemma under certain conditions. The parametrization must have a continuation to an ellipse with foci 0 and 2π, and φ must not hit 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 on the error in the Gauß–Legendre rule applied to $(z\mathsf{B} - \mathsf{A})^{-1}\mathsf{B}$.

THEOREM 4.5 (Error of Gauß–Legendre applied to $(z\mathsf{B} - \mathsf{A})^{-1}\mathsf{B}$, [12]). *Let (A, B) be a definite matrix pair and let $\mathsf{X} = [\mathsf{x}_1, \ldots, \mathsf{x}_n]$ be its full eigenvector matrix, consisting of B-orthonormal eigenvectors ($\mathsf{X}^\star\mathsf{B}\mathsf{X} = \mathsf{I}$). Let a, b denote the semi-axes of an ellipse with foci 0 and 2π that is chosen such that for all eigenvalues $\lambda \in \operatorname{spec}(\mathsf{A}, \mathsf{B})$ 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 $\lambda_1, \ldots, \lambda_m$ reside in $\operatorname{Int}(\mathcal{C})$. 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*

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

and

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

where $\widetilde{\mathsf{U}}_p$ 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 $\|\mathsf{M}\|_{\mathsf{B}2} := \left\|\mathsf{B}^{1/2}\mathsf{M}\right\|_2$.

Proof. To prove (4.9) we first write

$$\begin{aligned}
h(t) &= \varphi'(t)(\varphi(t)\mathsf{B} - \mathsf{A})^{-1}\mathsf{B}\mathsf{Y} \\
&= \varphi'(t)(\varphi(t)\mathsf{I} - \mathsf{B}^{-1}\mathsf{A})^{-1}\mathsf{Y} \\
&= \varphi'(t)\mathsf{X} \cdot \operatorname{diag}\left(r_{\lambda_1}(\varphi(t)), \ldots, r_{\lambda_n}(\varphi(t))\right) \cdot \mathsf{X}^{-1}\mathsf{Y} \\
&= \mathsf{X} \cdot \operatorname{diag}\left(\varphi'(t) r_{\lambda_1}(\varphi(t)), \ldots, \varphi'(t) r_{\lambda_n}(\varphi(t))\right) \cdot \mathsf{X}^{-1}\mathsf{Y},
\end{aligned}$$

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

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

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

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

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

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

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

$$\mathsf{U} = \frac{1}{2\pi\mathbf{i}} \sum_{j=1}^{m} \int_0^{2\pi} \varphi'(t) r_{\lambda_j}(\varphi(t)) \mathrm{d}t \, \mathsf{x}_j \mathsf{x}_j^\star \mathsf{B}\mathsf{Y},$$

since $\lambda_{m+1}, \ldots, \lambda_n \notin \operatorname{Int}(\mathcal{C})$. Consequently, because the error $E_{G_p}(\cdot)$ is a linear operator, we have with $p = \max_j(p_j(\epsilon))$

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

The last equality follows by $\left\|\mathsf{B}^{1/2}\mathsf{x}_j\right\|_2 = \left\|\mathsf{x}_j\right\|_{\mathsf{B}} = 1$. □

The theorem shows that the norm of $\mathsf{U}-\widetilde{\mathsf{U}}_p$ 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., $\epsilon = (1-\gamma/\pi)/2$.
In the following, we will discuss the structure of the ellipse from Theorem 4.5. For simplicity, suppose that the parametrization φ is given by $\varphi(t) := c + r\exp(\mathbf{i}t)$, a circle centered at the midpoint of the interval I_λ. For any of the real eigenvalues λ, the equation

$$\lambda = \varphi(z) = c + r\exp(\mathbf{i}z)$$

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

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

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

$$\eta_1 = \min\left\{\left|\mathrm{Im}\left(\varphi^{-1}(\lambda)\right)\right| : \lambda \in \mathrm{spec}(\mathsf{A},\ \mathsf{B}),\ \lambda > c\right\},$$
$$\eta_2 = \min\left\{\left|\mathrm{Im}\left(\varphi^{-1}(\lambda)\right)\right| : \lambda \in \mathrm{spec}(\mathsf{A},\ \mathsf{B}),\ \lambda < c\right\},$$

see Fig. 4.1. By η_1, an ellipse with foci 0 and 2π and semi axes a_1, b_1 is defined such that $0 + \mathbf{i}\eta_1$ 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 $\pi + \mathbf{i}\eta_1$ 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

$$\eta_1 = \frac{b_1^2}{a_1}, \tag{4.12}$$

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

where now $(x,\ y)$ denotes any point on the ellipse. Solving (4.12) for a_1 and inserting in (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).

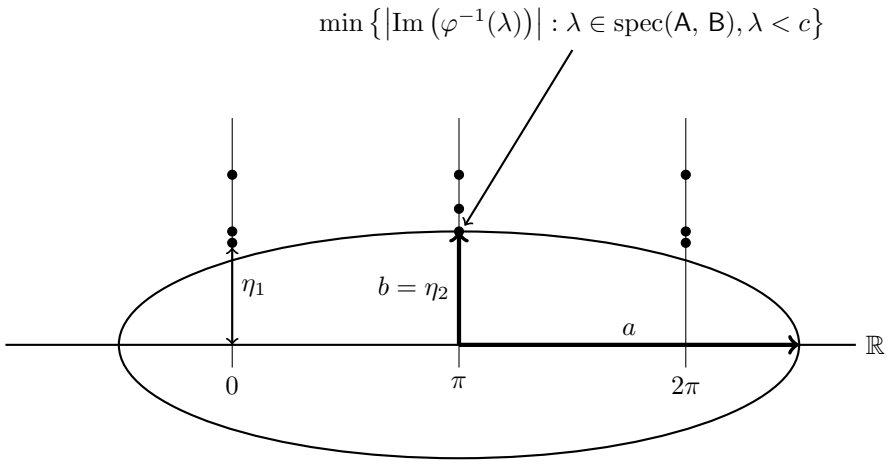

FIG. 4.1. *Location of the ellipse from Theorem 4.5. The semi axes are denoted a and b and marked by the arrows. The dots on the vertical lines denote the absolute values of $\varphi^{-1}(\lambda)$ for eigenvalues λ.*

5. Numerical example. To illustrate the theory from the previous sections a numerical example will examine the connection 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 $\check{\mathsf{A}} = \texttt{randn(n)}$, $\mathsf{A} := \check{\mathsf{A}} + \check{\mathsf{A}}^\star$ in MATLAB [27] and perform essentially the steps of one FEAST iteration with a random orthonormal starting basis $\mathsf{Y} \in \mathbb{R}^{100\times 50}$. First, we measure the normwise errors $\|\mathsf{U} - \widetilde{\mathsf{U}}_p\|$.

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 $\|\mathsf{U} - \widetilde{\mathsf{U}}_p\|$ 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 $\mathrm{span}(\mathsf{U})$ [13, 17]. Nevertheless, in this case a subset of the columns of $\widetilde{\mathsf{U}}_p$ would converge. To avoid the problem of a subspace of wrong dimension, we enforce in this example $\dim(\mathrm{span}(\mathsf{U})) = \dim(\mathrm{span}(\mathsf{Y})) = 50$ 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 $d := \mathrm{dist}(\mathcal{C}, \mathrm{spec}(\mathsf{A})) = 0.41$, which would be a fairly large number in practice. The errors $\|\mathsf{U} - \widetilde{\mathsf{U}}_p\|$ are shown in the top plot of 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 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 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 trapezoidal rule, each of order $p = 16$, are shown in 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 value as the eigenvalues. The process described here corresponds to one single FEAST iteration with dimension of the search

space $\widetilde{\mathsf{U}}_p$ 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 $\widetilde{\mathsf{U}}_p$ also move in the correct direction for modest values of p, the 50 canonical angles between U and $\widetilde{\mathsf{U}}_p$ are shown in the bottom plot of Figure 5.1 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 $\|\mathsf{U} - \widetilde{\mathsf{U}}_p\|$ as is stated by the theory, cf. eq. (3.4).

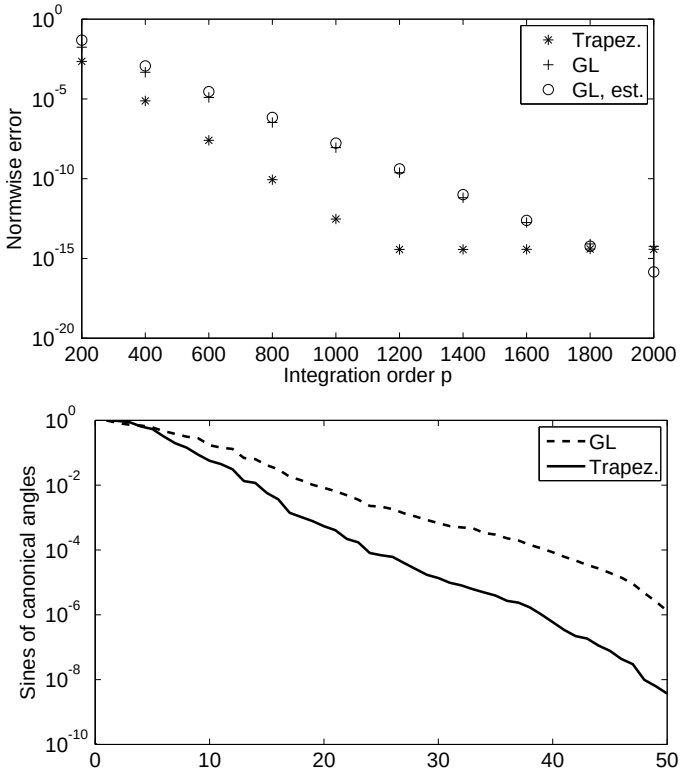

FIG. 5.1. *Results for Experiment 5.1. Normwise integration error* $\|\mathsf{U} - \widetilde{\mathsf{U}}_p\|$ *for trapezoidal and Gauß–Legendre rule and estimated error for Gauß–Legendre rule (top). Sines of canonical angles between* U *and* $\widetilde{\mathsf{U}}_p$ *for* $p = 16$ *(bottom).*

EXPERIMENT 5.2. Next, we repeat the experiment in similar fashion for the generalized eigenvalue problem. The matrices A and B are constructed in a similar way as in 1., where B is forced to be positive definite. Further, A now has complex entries. The results of the integration errors are shown 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 $\kappa(\mathsf{X})$ for the full eigenvector matrix X. For the eigenvalue approximation, the results where 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 convergence of subspaces measured by canonical angles or for eigenvalue convergence.

The very slow convergence of the subspaces $\widetilde{\mathsf{U}}_p$ towards U can be justified theoretically. In case of the Gauß–Legendre rule we have, according to Theorem 4.5, error bound depending

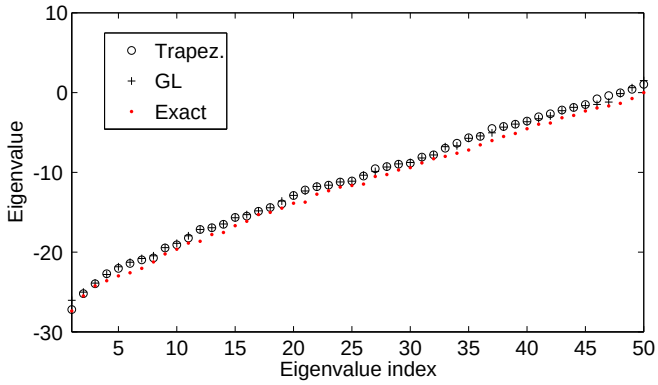

FIG. 5.2. *Approximation of eigenvalues computed by Gauß–Legendre and trapezoidal rule, respectively. We used $p = 16$ integration points in both cases.*

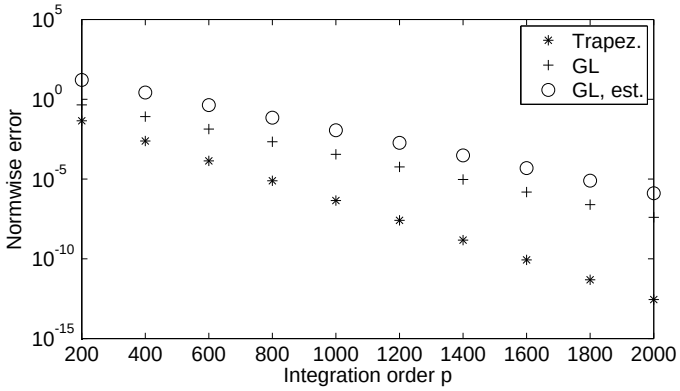

FIG. 5.3. *Results for Experiment 5.2. Normwise integration error $\|\mathsf{U} - \widetilde{\mathsf{U}}_p\|$ for trapezoidal and Gauß–Legendre rule and estimated error for Gauß–Legendre rule.*

on the ratio π/γ. The number γ is a number that depends on the size of the region of analyticity of the resolvent and which is basically determined by the distance of the curve to the next eigenvalue. In Experiment 5.1, both $\kappa(\mathsf{X})$ and $\|\mathsf{Y}\|$ 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 chose Y such that $\|\mathsf{Y}\|_{\mathsf{B2}} = 1$. 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

$$\left\| \mathsf{U} - \widetilde{\mathsf{U}}_p \right\| \leq m \cdot C_1 d^{-1} \exp(-C_2\, p\, d), \tag{5.1}$$

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 thinking of C_1 and C_2 to be 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, 5.2 the curve $\mathcal{C}$ typically passes the

spectrum much more closely. For instance, if we have $d = 10^{-6}$ (which could be a value from practice), we would obtain $\pi/\gamma \approx 0.9999997$, a value whose positive powers converge extremely slow 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 used 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].

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 depended on the norm of the error $\mathsf{U} - \widetilde{\mathsf{U}}$ in a certain basis of the eigenspace. The results can also be used independently of how $\widetilde{\mathsf{U}}$ 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 $\widetilde{\mathsf{U}}$ in Section 4.
Taken together, the two kinds of convergence discussed on Sections 3 and 4 imply convergence of the integration based method with respect to the integration order p.

Acknowledgements. The author likes to thank Bruno Lang for fruitful discussions and comments on earlier drafts of the manuscript.

REFERENCES

[1] W. E. ARNOLDI, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Qart. Appl. Math., 9 (1951), pp. 17–29.
[2] W. 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*, Harri Deutsch, Frankfurt am Main, 21st ed., 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 Lecture Notes in Computational Science and Engineering, Springer Berlin Heidelberg, 2005, pp. 15–24.
[6] P. J. DAVIS, *On the numerical integration of periodic analytic functions*, in On numerical approximation, R. E. Langer, ed., Madison, WI, 1959, The University of Wisconsin Press.
[7] P. J. DAVIS AND P. RABINOWITZ, *Methods of numerical integration*, Academic Press, Orlando, FL, second ed., 1984.
[8] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 4th ed., 2013.
[9] S. GÜTTEL, E. POLIZZI, P. TANG, AND G. VIAUD, *Zolotarev quadrature rules and load balancing for the FEAST eigensolver*. The University of Manchester, MIMS EPrint 2014.39, http://www.manchester.ac.uk/mims/eprints, 2014.
[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. Comp., 23 (2002), pp. 2009–2041.
[11] A. V. KNYAZEV AND M. E. ARGENTATI, *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, Bergische Universität Wuppertal, 2014. published electronically, http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:hbz:468-20140701-112141-6.
[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. Res. Nat. Bur. Stand., 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*, vol. 20 of Classics in Applied Mathematics, SIAM, Philadelphia, PA, Classics ed., 1998. Unabridged, corrected republication of the 1980 original.
[17] E. POLIZZI, *Density-matrix-based algorithm for solving eigenvalue problems*, Phys. Rev. B, 79 (2009), p. 115112.
[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*. to appear, 2014.
[19] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, SIAM, Philadelphia, PA, second ed., 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. Comm., 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. STEWART, *Matrix Algorithms*, vol. II, Eigensystems, SIAM, Philadelphia, PA, 2001.
[25] G. W. STEWART AND J. SUN, *Matrix Perturbation Theory*, Academic Press, San Diego, CA, 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] THE MATHWORKS, INC., *MATLAB, Version R2013a*, 2013.
[28] L. N. TREFETHEN AND J. A. C. WEIDEMAN, *The exponentially convergent trapezoidal rule*, SIAM Rev., 56 (2014), pp. 385–458.
[29] J. A. C. WEIDEMAN, *Numerical integration of periodic functions: A few examples*, Amer. Math. Monthly, 109 (2002), pp. 21–36.
[30] 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.