SPECTRAL ERROR BOUNDS FOR HERMITIAN INEXACT KRYLOV METHODS†

UTE KANDLER* AND CHRISTIAN SCHRÖDER*

Abstract. We investigate the convergence behavior of inexact Krylov methods for the approximation of a few eigenvectors or invariant subspaces of a large, sparse Hermitian matrix. Bounds on the distance between an exact invariant subspace and a Krylov subspace and between an exact invariant subspace and a Ritz space are presented. Using the first bound we analyze the question: if a few iteration steps have been taken without convergence, how many more iterations have to be performed to achieve a preset tolerance. The second bound provides a measure on the approximation quality of a computed Ritz space. Traditional bounds of these quantities are particularly sensitive to the gap between the wanted eigenvalues and the remaining spectrum. Here this gap is allowed to be small by considering how well the exact invariant subspace is contained in a slightly larger approximated invariant subspace. Moreover, numerical experiments confirm the applicability of the given bounds.

Key words. Hermitian eigenvalue problem, inexact Krylov method, convergence analysis, Krylov relation, Ritz pair

AMS subject classifications. 65F15, 65G99

1. Introduction. The Hermitian eigenvalue problem consists of solving the equation

$$Ax = \lambda x$$

for $\lambda \in \mathbb{R}$, $x \in \mathbb{C}^n \setminus \{0\}$, where $A = A^H \in \mathbb{C}^{n\times n}$ is given. Here we consider the case when A is large and data-sparse and only a small subset of exterior eigenvalues and the corresponding invariant subspace are desired. Under these conditions the most prominent iterative methods are Krylov subspace methods that search in *Krylov subspaces*

$$\mathcal{K}_k := \mathcal{K}_k(A, v_1) := \operatorname{span}(v_1, Av_1, A^2v_1, \ldots, A^{k-1}v_1)$$

for approximations of eigenvectors or invariant subspaces.

The application we have in mind is the identification of ground states of quantum systems. Mathematically this amounts to is an Hermitian eigenvalue problem. A ground state is an eigenvector corresponding to the smallest eigenvalue, whereas eigenvectors corresponding to the second (third, forth,...) smallest eigenvalues are called first (second, third,...) excited states. Note that these eigenvalues can be very close to each other which can lead to complications in the determination of the ground state [25]. The dimension of these quantum systems is often extremely large; it is 2^d where d is the number of particles of the system (and $d > 50$ is not uncommon) [11,25]. Hence, already a single vector might not fit into memory of even a large computing cluster when stored in standard element-by-element format. For that reason vectors have to be stored in a data sparse way, such as, e.g., in the tensor train or the hierarchical tensor format [9, 10, 21]. These formats, however, entail the drawback that vector operations like matrix-vector multiplication, vector addition and vector scaling are

†This work was supported by German research council, DFG under project "Scalable Numerical Methods for Adiabatic Quantum Preparation"

*TU Berlin, Str. des 17. Juni 136, 10623 Berlin, GERMANY, ({kandler,schroed}@math.tu-berlin.de)

inexact, i.e., only approximations of the intended quantities are available. Using inexact vector operations inside Krylov methods result in *inexact Krylov methods*. Other occurrences include inexact solves in a shift-invert setting (when $A = (\mathcal{A} - \sigma I)^{-1}$) or mixed precision arithmetic (when computations are carried out in double precision, but vectors are stored in single precision).

Backward error analysis shows that inexact Krylov subspace methods can be interpreted as exact Krylov methods applied to a perturbed matrix $A + E$ [14, 29, 33]. In general, an inexact Krylov method will generate orthonormal $V_{k+1} = [V_k, v_{k+1}] = [v_1, \ldots, v_{k+1}] \in \mathbb{C}^{n\times k+1}$, $b_{k+1} \in \mathbb{C}^k$, and $B_k = B_k^H \in \mathbb{C}^{k\times k}$ such that a *Krylov relation* of the form

$$(A + E)V_k = V_k B_k + v_{k+1} b_{k+1}^H \tag{1.1}$$

holds. Equation (1.1) takes the role of the familiar Arnoldi relation $AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T$. The backward error matrix $E = E^H \in \mathbb{C}^{n\times n}$ is unknown, only a bound on its norm $\|E\|_2$ is usually available. Note that using the Lanczos process B_k would be tridiagonal, but other methods generate a full Hermitian B_k [14]. Relation (1.1) implies that V_k is a basis of a Krylov subspace

$$\tilde{\mathcal{K}}_k := \mathcal{K}_k(A + E, \tilde{v}_1)$$

of the Hermitian matrix $A + E$ close to A for some $\tilde{v}_1 \in \operatorname{ran}(V_k)$.

In this paper the following questions are addressed:

- *How well is a desired invariant subspace $\mathcal{X}$ of the original matrix A approximated by the Krylov subspace $\tilde{\mathcal{K}}_k$ of a perturbed matrix $A + E$ or by a Ritz space $\tilde{\mathcal{Y}}$ in $\tilde{\mathcal{K}}_k$?*
  In our notation a tilde indicates a quantity that corresponds to $A+E$ instead of to A. E.g.,$\tilde{\mathcal{K}}_k$ is a Krylov subspace of $A+E$ and $\tilde{\mathcal{Y}}$ is a Ritz space of $A+E$ in $\tilde{\mathcal{K}}_k$.
- *How can the sensitivity of the error bounds with respect to a small gap between the wanted and the remaining eigenvalues be avoided?*
  We allow the approximation $\tilde{\mathcal{Y}}$ to be of larger dimension than $\mathcal{X}$. In other words, we alter the question from "How close is $\mathcal{X}$ to $\tilde{\mathcal{Y}}$?" to "How well is $\mathcal{X}$ contained in $\tilde{\mathcal{Y}}$?". This enables us in contrast traditional bounds, e.g., [7, 13, 31], to treat small gaps between the wanted and the remaining eigenvalues.
- *What is a suitable measure for the quality of the approximate invariant subspace?*
  We use the *angle of inclusion of $\mathcal{X}$ in $\tilde{\mathcal{Y}}$* for nonzero subspaces $\mathcal{X}$, $\tilde{\mathcal{Y}} \subset \mathbb{C}^n$ which is defined by

  $$\angle^{\frown}_{\max}(\mathcal{X}, \tilde{\mathcal{Y}}) := \max_{x\in\mathcal{X},\, x\neq 0} \angle(x, \tilde{\mathcal{Y}}). \tag{1.2}$$

  A small angle $\angle^{\frown}_{\max}(\mathcal{X}, \tilde{\mathcal{Y}})$ does not mean that $\mathcal{X}$ is close to $\tilde{\mathcal{Y}}$, but indicates that $\mathcal{X}$ is almost contained in $\tilde{\mathcal{Y}}$. In numerous applications this is all that is needed [1]. When $\dim(\mathcal{X}) \le \dim(\tilde{\mathcal{Y}})$, the angle $\angle^{\frown}_{\max}(\mathcal{X}, \tilde{\mathcal{Y}})$ coincides with the well-known maximal canonical angle [5, 37], otherwise (i.e., when $\dim(\mathcal{X}) > \dim(\tilde{\mathcal{Y}})$) it is $\pi/2$. This unsymmetric formulation is reasonable, because for example a 2-dimensional space can be approximately contained in a 3-dimensional space but obviously not vice versa. As a nice consequence it also means that our bounds hold in the trivial case when $\dim(\mathcal{X}) > \dim(\tilde{\mathcal{Y}})$.

The bounds presented in this paper use (1.1) as a starting point. We distinguish between bounds on the distance to Krylov subspaces and to Ritz spaces. In the first case we achieve an a priori result (requiring $\|E\|_2$ only) and in the second case an a posteriori result (requiring V_k, B_k and b_{k+1} in addition to $\|E\|_2$). In both cases the bounds also require the exact eigenvalues of A. Next to being interesting in their own right the bounds may also be beneficial inside algorithms as stopping criterion when combined with certain spectral estimation techniques, cf. [2, 22].

Our theory heavily relies on spectral perturbation theory. Classical works in this field include [7,12,22,24,34]. The distance of an approximate invariant subspaces from an exact one is considered in [13, 20] (based on the residual) and in [31] (based on the quality of a surrounding search space). For the a priori setting we will generalize a classic result [23, Theorem 6.3] that bounds the angle between an eigenvector and the kth Krylov subspace. Generalizations for block Krylov methods of this classic result can be found in [18, 27]. [3] presents bounds for the angle between an invariant subspace of A and the Krylov subspace $\mathcal{K}_k(A, v_1)$ generated by a Krylov method with polynomial restarts. Other works considering inexact Krylov subspace methods include [19, 28, 29, 36], where only the matrix vector multiplication is assumed to be inexact. Finally we mention [33] where the backward error for an approximate Krylov subspace is analyzed.

The paper is structured as follows: Section 2 introduces more notation and basic or preliminary results. The main results of the paper are presented in sections 3 and 4. In Section 3 we present bounds on the distance to Krylov subspaces of a perturbed matrix $A+E$, while in Section 4 a bound on the approximation quality of Ritz spaces is discussed. Numerical examples illustrating our findings are presented in Section 5. Finally we offer some concluding remarks in Section 6.

2. Notation and preliminary results. In this section we introduce some notation and collect basic results, grouped by topic. Throughout this paper $\mathbb{C}^{m\times n}$ denotes the set of complex $m \times n$ matrices and $\mathbb{C}^n$ the n-dimensional complex vector space. A matrix norm $\|\cdot\|$ is called *unitarily invariant* if $\|UAQ\| = \|A\|$ for any matrix $A \in \mathbb{C}^{n\times m}$ and any unitary matrices $U \in \mathbb{C}^{m\times m}$ and $Q \in \mathbb{C}^{n\times n}$. Prominent unitarily invariant matrix norms are the 2-norm denoted by $\|.\|_2$ and the Frobenius norm denoted by $\|.\|_F$.

For a set $\Lambda \subset \mathbb{R}$ and $A \in \mathbb{C}^{n\times n}$ we define

$$\operatorname{spread}(\Lambda) := \max_{\lambda_1,\lambda_2\in\Lambda} |\lambda_1 - \lambda_2| \quad \text{and} \quad \operatorname{ran}(A) := \{Ax \,|\, x \in \mathbb{C}^n\},$$

where the range of A equals the column space of the matrix A. We indicate by 0 the null matrix, by I_n the $n \times n$ identity matrix and by e_k its kth column. The *cardinality* of a discrete set S, denoted $|S|$, is the number of elements in S. The *envelope* $\operatorname{env}(\mathcal{M})$ of a set $\mathcal{M} \subset \mathbb{R} \cup \{\infty, -\infty\}$ is defined by the smallest interval that contains $\mathcal{M}$.

2.1. The angle of inclusion. For nonzero subspaces $\mathcal{X}$, $\mathcal{Y} \subset \mathbb{C}^n$ the angle of inclusion of $\mathcal{X}$ in $\mathcal{Y}$ is defined by (1.2) (we replace $\tilde{\mathcal{Y}}$ by $\mathcal{Y}$ here for ease of notation) where

$$\angle(x, \mathcal{Y}) := \min_{\substack{y\in\mathcal{Y}\\ y\neq 0}} \angle(x, y), \quad \text{and} \quad \angle(x, y) := \arccos\left(\frac{|x^H y|}{\|x\|_2\|y\|_2}\right) \in \left[0, \tfrac{\pi}{2}\right].$$

For matrices X, Y we define $\angle^{\frown}_{\max}(X, Y) := \angle^{\frown}_{\max}(\operatorname{ran}(X), \operatorname{ran}(Y))$.

Intuitively, the angle of inclusion (1.2) provides a measure of how well $\mathcal{X}$ is contained in $\mathcal{Y}$. If the angle of inclusion is small then for every $x \in \mathcal{X}$ there is a $y \in \mathcal{Y}$ that is close to x, i.e., $\|x-y\|_2/\|x\|_2$ is small. In particular, we have $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}) = 0$ if and only if $\mathcal{X} \subset \mathcal{Y}$ and $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y})$ reaches $\pi/2$ if $\mathcal{X}$ contains a direction orthogonal to $\mathcal{Y}$. In particular the latter is the case whenever $\dim(\mathcal{X}) > \dim(\mathcal{Y})$. Using these rationale the angle of inclusion can be extended to zero-dimensional spaces: $\angle^{\frown}_{\max}(\{0\},\mathcal{Y}) := 0$ for any (zero or non-zero) $\mathcal{Y}$, and $\angle^{\frown}_{\max}(\mathcal{X},\{0\}) := \pi/2$ for any non-zero $\mathcal{X}$.

We also stress that while the angle of inclusion is non-symmetric in general , i.e., $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}) \neq \angle^{\frown}_{\max}(\mathcal{Y},\mathcal{X})$, it is symmetric whenever $\dim(\mathcal{X}) = \dim(\mathcal{Y})$. In the this case $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}) = \angle^{\frown}_{\max}(\mathcal{Y},\mathcal{X})$ coincides with the maximal canonical angle between $\mathcal{X}$ and $\mathcal{Y}$.

As a side note we mention that there is also a minimal angle, which is defined by $\angle_{\min}(\mathcal{X},\mathcal{Y}) := \min_{x\in\mathcal{X},x\neq 0} \angle(x,\mathcal{Y})$, but this concept does not play a role in this paper. For more details on subspace angles see [6,15,37]. We state some useful properties of inclusion angles in the following lemma and theorem.

LEMMA 2.1. *Let $\mathcal{X},\mathcal{Y},\mathcal{Z}$ be subspaces of $\mathbb{C}^n$. Then*

i. $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Z}) \leq \angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}) + \angle^{\frown}_{\max}(\mathcal{Y},\mathcal{Z})$ *(triangle inequality);*

ii. $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}) \geq \angle^{\frown}_{\max}(\mathcal{X},\mathcal{Z})$
whenever $\mathcal{Y} \subset \mathcal{Z}$;

iii. $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}) = \angle^{\frown}_{\max}(\mathcal{Y}^\perp,\mathcal{X}^\perp)$
where $\mathcal{X}^\perp,\mathcal{Y}^\perp \subset \mathbb{C}^n$ are the orthogonal complements of $\mathcal{X},\mathcal{Y}$ in $\mathbb{C}^n$;

iv. $$\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}) = \begin{cases} 0 & \text{if} \quad \dim(\mathcal{X}) = 0 \\ \arccos\left(\sigma_{\dim(\mathcal{X})}(X^H Y)\right) & \text{if} \quad 1 \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y}), \\ \frac{\pi}{2} & \text{if} \quad \dim(\mathcal{X}) > \dim(\mathcal{Y}) \end{cases}$$
where X,Y are any orthonormal bases of $\mathcal{X},\mathcal{Y}$ respectively, i.e. $\mathcal{X} := \operatorname{ran}(X)$, $\mathcal{Y} := \operatorname{ran}(Y)$, and $\sigma_i(X^H Y)$ denotes the i-th largest singular value of $X^H Y$;

v. $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y}^\perp \cap \mathcal{Z}) \leq \angle^{\frown}_{\max}(\mathcal{X} \oplus \mathcal{Y},\mathcal{Z})$
whenever $\mathcal{X} \perp \mathcal{Y}$.

Proof. i. We distinguish three cases. a) It is easy to check that the inequality holds whenever at least one of $\mathcal{X},\mathcal{Y},\mathcal{Z}$ is zero. b) If $\dim(\mathcal{X}) > \dim(\mathcal{Y})$ or $\dim(\mathcal{Y}) > \dim(\mathcal{Z})$ at least one angle on the right-hand side reaches $\frac{\pi}{2}$ and there is nothing to prove. c) Otherwise, i.e., if $1 \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y}) \leq \dim(\mathcal{Z})$ the definition of $\angle^{\frown}_{\max}(\mathcal{X},\mathcal{Y})$ coincides with the definition of angles between subspaces in [37]. Hence in this case the proof in [37, pp. 275] applies to our definition as well.

ii. This is a special case of part i. with $\angle^{\frown}_{\max}(\mathcal{Y},\mathcal{Z}) = 0$.

iii. We distinguish three cases. a) If $\mathcal{X} \subset \mathcal{Y}$ or, equivalently, $\mathcal{Y}^\perp \subset \mathcal{X}^\perp$ then both angles are zero. (This covers the cases that $\mathcal{X} = 0$ or $\mathcal{Y}^\perp = 0$.) b) If $\dim(\mathcal{X}) > \dim(\mathcal{Y})$ or, equivalently, $\dim(\mathcal{Y}^\perp) > \dim(\mathcal{X}^\perp)$ then both angles are $\pi/2$. (This covers the cases that $\mathcal{Y} = 0$ or $\mathcal{X}^\perp = 0$.) c) If $1 \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y})$ see [16].

iv. The first case ($\dim(\mathcal{X}) = 0$) holds by definition. The last case ($\dim(\mathcal{X}) > \dim(\mathcal{Y})$) was discussed above. For the middle case ($1 \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y})$) see [17, p. 2010].

v. W.l.o.g. $\mathcal{X}$ is non-zero (otherwise the angle on the left hand side is zero and there is nothing to prove). Also, w.l.o.g. $\dim(\mathcal{X}) \leq \dim(\mathcal{Y}^\perp \cap \mathcal{Z})$ (otherwise $\dim(\mathcal{X}) > \dim(\mathcal{Y}^\perp \cap \mathcal{Z})$ implies

$$\begin{aligned} \dim(\mathcal{X} \oplus \mathcal{Y}) = \dim(\mathcal{X}) + \dim(\mathcal{Y}) &> \dim(\mathcal{Y}^\perp \cap \mathcal{Z}) + \dim(\mathcal{Y}) \\ &\geq (\dim(\mathcal{Y}^\perp) + \dim(\mathcal{Z}) - n) + \dim(\mathcal{Y}) = \dim(\mathcal{Z}), \end{aligned}$$

so both angles in the claimed inequality are $\pi/2$ and there is nothing to prove). So $1 \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y}^\perp \cap \mathcal{Z})$ which implies that $\mathcal{Z}$ and $\mathcal{Y}^\perp$ are non-zero. W.l.o.g., also $\mathcal{Y}$ is non-zero (otherwise both angles in the claimed inequality reduce to $\angle^{\frown}_{\max}(\mathcal{X}, \mathcal{Z})$ and there is nothing to prove). Finally we can assume w.l.o.g. that $\dim(\mathcal{X} \oplus \mathcal{Y}) = \dim(\mathcal{X}) + \dim(\mathcal{Y}) \leq \dim(\mathcal{Z})$ (otherwise the angle on the right-hand side in the claimed inequality is $\pi/2$ and there is nothing to prove).

Let $X \in \mathbb{C}^{n \times \dim(\mathcal{X})}$, $Y \in \mathbb{C}^{n \times \dim(\mathcal{Y})}$, $Z =\in \mathbb{C}^{n \times \dim(\mathcal{Z})}$ be orthonormal bases of $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$, respectively, where $Z = [Z_1, Z_2]$ is chosen such that $\operatorname{ran}(Z_1) = \mathcal{Y}^\perp \cap \mathcal{Z}$. Then $[X, Y]$ is an orthonormal basis of $\mathcal{X} \oplus \mathcal{Y}$, since $\mathcal{X} \perp \mathcal{Y}$. Moreover, $W := [X, Y]^H [Z_1, Z_2]$ is of the form $\left[\begin{smallmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{smallmatrix}\right]$, where $W_{11} = X^H Z_1$ does not have less columns than rows (since $1 \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y}^\perp \cap \mathcal{Z})$) and $W_{22} = Y^H Z_2$ has full column rank. Also W does not have less columns than rows (since $1 \leq \dim(\mathcal{X} \oplus \mathcal{Y}) \leq \dim(\mathcal{Z})$). Hence, with part iv. we have

$$\cos \angle^{\frown}_{\max}(\mathcal{X}, \mathcal{Y}^\perp \cap \mathcal{Z}) = \sigma_{\dim(\mathcal{X})}(W_{11}) \quad \text{and} \quad \sigma_{\dim(\mathcal{X} \oplus \mathcal{Y})}(W) = \cos \angle^{\frown}_{\max}(\mathcal{X} \oplus \mathcal{Y}, \mathcal{Z}).$$

Using the monotonically decreasing behavior of the cosine in the interval $[0, \frac{\pi}{2}]$ all that remains to prove is $\sigma_{\dim(\mathcal{X})}(W_{11}) \geq \sigma_{\dim(\mathcal{X} \oplus \mathcal{Y})}(W)$. We distinguish two cases. If W_{22} has more rows than columns, then the rows of W must be linearly dependent, i.e., $\sigma_{\dim(\mathcal{X} \oplus \mathcal{Y})}(W) = \sigma_{\min}(W) = 0$. Otherwise W_{22} is square and we have by the interlacing property of eigenvalues of Hermitian matrices [22] that

$$\begin{aligned} \sigma_{\dim(\mathcal{X})}(W_{11})^2 = \lambda_{\dim(\mathcal{X})}(W_{11}^H W_{11}) &\geq \lambda_{\dim(\mathcal{X})+\dim(\mathcal{Y})} \left(\begin{bmatrix} W_{11}^H W_{11} & * \\ * & * \end{bmatrix} \right) \\ &= \lambda_{\dim(\mathcal{X})+\dim(\mathcal{Y})}(W^H W) = \sigma_{\dim(\mathcal{X} \oplus \mathcal{Y})}(W)^2. \end{aligned}$$

□

THEOREM 2.2. *[30, Theorem 2.7] Let $X \in \mathbb{C}^{n \times m}, X_\perp \in \mathbb{C}^{n \times n-m}, Z \in \mathbb{C}^{n-m \times m}$ such that $[X, X_\perp]$ is unitary. Let $\mathcal{X} := \operatorname{ran}(X)$ and $\tilde{\mathcal{X}} := \operatorname{ran}(X + X_\perp Z)$. Then $\tan \angle^{\frown}_{\max}(\mathcal{X}, \tilde{\mathcal{X}}) = \|Z\|_2$.*

2.2. The spectrum and its perturbation. The set of eigenvalues of a matrix A, i.e., its *spectrum*, is denoted by $\operatorname{eig}(A)$. A subspace $\mathcal{X} \in \mathbb{C}^n$ is called an *invariant subspace* of $A \in \mathbb{C}^{n \times n}$ if $A\mathcal{X} \subset \mathcal{X}$. When A is Hermitian, choosing orthonormal bases $X \in \mathbb{C}^{n \times m}, X_\perp \in \mathbb{C}^{n \times n-m}$ of an invariant subspace $\mathcal{X}$ and its orthogonal complement $\mathcal{X}^\perp$, respectively, gives rise to a *block spectral decomposition* of A of the form

$$[X, X_\perp]^H A [X, X_\perp] = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \tag{2.1}$$

with $A_{11} \in \mathbb{C}^{m \times m}$, $A_{22} \in \mathbb{C}^{n-m \times n-m}$ and $\operatorname{eig}(A) = \operatorname{eig}(A_{11}) \cup \operatorname{eig}(A_{22})$. The invariant subspace $\mathcal{X}$ is called *simple* if $\operatorname{eig}(A_{11}) \cap \operatorname{eig}(A_{22}) = \emptyset$.

The following result is known as Weyl's theorem for the 2-norm, e.g., [34, Corollary 4.10], [12], and as Hoffman-Wielandt theorem for the Frobenius norm, e.g., [12, 38].

THEOREM 2.3. *Let $A, E \in \mathbb{C}^{n \times n}$ be Hermitian. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of A and $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ the eigenvalues of $A + E$. For $* \in \{2, F\}$ we have*

$$\| \operatorname{diag}(\lambda_1, \ldots, \lambda_n) - \operatorname{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n) \|_* \leq \|E\|_*.$$

The following theorem is a generalization of the well-known Davis-Kahan $\tan(\theta)$ theorem which imposes, compared to the original formulation [7, Theorem 6.3] relaxed conditions on the spectrum.

THEOREM 2.4. *[20, Theorem 2] Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and let $X = [X_1, X_2, X_3]$ be unitary so that $X^H AX = \operatorname{diag}(A_{11}, A_{22}, A_{33})$ is block diagonal, where $X_i \in \mathbb{C}^{n\times n_i}$, $A_{ii} \in \mathbb{C}^{n_i\times n_i}$ for $i = 1, 2, 3$ with $n_1+n_2+n_3 = n$. Let $\tilde{X}_3 \in \mathbb{C}^{n\times n_3}$ have orthogonal columns and let $R = A\tilde{X}_3 - \tilde{X}_3\tilde{A}_3$, where $\tilde{A}_3 = \tilde{X}_3^H A\tilde{X}_3$. Suppose that $\operatorname{eig}(A_{11})$ lies in $[a, b]$ and $\operatorname{eig}(\tilde{A}_3)$ lies in the union of $(-\infty, a-\delta]$ and $[b+\delta, \infty)$. Then*

$$\tan \angle^{\frown}_{\max}(\tilde{X}_3, [X_2, X_3]) \le \frac{\|R\|_2}{\delta}. \tag{2.2}$$

REMARK 2.5. In [20, Theorem 2] $X^H AX$ was assumed diagonal (instead of just block diagonal). Our formulation holds, because the change of bases that diagonalizes A_{11}, A_{22}, and A_{33} does not influence the subspace angles.

The standard Davis-Kahan $\tan(\theta)$ theorem is obtained for $n_2 = 0$.

2.3. Gap between eigenvalues and separation of subspaces. We define the *gap* between closed sets Λ_1, $\Lambda_2 \subset \mathbb{R}$ and square matrices A, B respectively, by

$$\operatorname{gap}(\Lambda_1, \Lambda_2) := \min_{\lambda_1\in\Lambda_1, \lambda_2\in\Lambda_2} |\lambda_1 - \lambda_2| \quad \text{and} \quad \operatorname{gap}(A, B) := \operatorname{gap}(\operatorname{eig}(A), \operatorname{eig}(B)).$$

Note that with (2.1) $\operatorname{gap}(A_{11}, A_{22}) = \operatorname{gap}(\operatorname{eig}(A_{11}), \operatorname{eig}(A) \setminus \operatorname{eig}(A_{11}))$ if $\operatorname{ran}(X)$ is a simple invariant subspace. The gap provides a natural way of describing the distance between two spectra. A similar quantity, also measuring in some sense the distance between the spectra of two arbitrary square matrices $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{m\times m}$, is given by the *separation*

$$\operatorname{sep}_*(A, B) := \min_{\substack{Z\in\mathbb{C}^{m\times n}\\ \|Z\|_*=1}} \|AZ - ZB\|_*, \quad \text{where} \quad * \in \{2, F\}. \tag{2.3}$$

Furthermore, the separation between two subspaces $\mathcal{X}, \mathcal{Y} \in \mathbb{C}^n$ with respect to a matrix $A \in \mathbb{C}^{n\times n}$ is defined by

$$\operatorname{sep}_{A,*}(\mathcal{X}, \mathcal{Y}) := \operatorname{sep}_*(X^H AX, Y^H AY),$$

where X and Y are any orthonormal bases for $\mathcal{X}$ and $\mathcal{Y}$. This quantity is well defined because the norms used in (2.3) are unitarily invariant, cf. [31]. For Hermitian matrices the sep and the gap operators are related as follows.

LEMMA 2.6. *Let $A_{11} \in \mathbb{C}^{n\times n}$ and $A_{22} \in \mathbb{C}^{m\times m}$ be Hermitian. Then*

$$\frac{2}{\pi}\operatorname{gap}(A_{11}, A_{22}) \le \operatorname{sep}_2(A_{11}, A_{22}) \le \operatorname{gap}(A_{11}, A_{22}) = \operatorname{sep}_F(A_{11}, A_{22}).$$

Proof. The first inequality is discussed in [4, p.15] leveraging a function-theoretical result in [35]. The second inequality is obtained by restricting Z in (2.3) to the form $Z = u_i v_j^H$, where $u_i \in \mathbb{C}^n$ and $v_j \in \mathbb{C}^m$ are a unit (right) eigenvector of A_{11} and a unit (left) eigenvector of A_{22}, respectively. The last relation is stated in [34, Theorem 3.1]. ∎

REMARK 2.7. The second inequality of Lemma 2.6 is also valid in the case of non-Hermitian matrices A_{11} and A_{22}. In that case the words "right" and "left" put

in parentheses in the above proof become important. The first and the last relation in Lemma 2.6 also hold for normal A_{11} and A_{22} (if $\frac{2}{\pi}$ is replaced by $\frac{4}{\pi}$), but do not carry over to the non-normal case.

The following theorem is a corollary of Theorem 2.4 and provides a measure how well the invariant subspace $\mathrm{ran}(X_1)$ of A is contained in the invariant subspace of larger dimension $\mathrm{ran}([\tilde{X}_1, \tilde{X}_2])$ of a perturbed matrix $A+E$. Note that it is even applicable if the eigenvalues corresponding to $\mathrm{ran}(X_1)$ are not well separated from the remaining ones.

THEOREM 2.8. *Let $A, E \in \mathbb{C}^{n\times n}$ be Hermitian. Let $X_i, \tilde{X}_i \in \mathbb{C}^{n\times n_i}$ for $i = 1,2,3$ with $n_1+n_2+n_3 = n$ be such that $X := [X_1, X_2, X_3], \tilde{X} := [\tilde{X}_1, \tilde{X}_2, \tilde{X}_3]$ are unitary and*

$$X^H A X = \mathrm{diag}(A_{11}, A_{22}, A_{33}), \quad \tilde{X}^H (A+E)\tilde{X} = \mathrm{diag}(\tilde{A}_{11}, \tilde{A}_{22}, \tilde{A}_{33}) \tag{2.4}$$

are block diagonal.

i) If $\mathrm{gap}(\mathrm{env}(\mathrm{eig}(A_{11})), \mathrm{eig}(\tilde{A}_{33})) > \|E\|_2$ *then*

$$\tan \angle^{\frown}_{\max}\left(X_1, [\tilde{X}_1, \tilde{X}_2]\right) \le \frac{\|E\|_2}{\mathrm{gap}(A_{11}, \tilde{A}_{33}) - \|E\|_2}.$$

ii) If $\mathrm{gap}(\mathrm{env}(\mathrm{eig}(A_{11})), \mathrm{eig}(A_{33})) > 2\|E\|_2$ *and* $\max\limits_{\tilde{\lambda}\in\mathrm{eig}(\tilde{A}_{33})} \min\limits_{\lambda\in\mathrm{eig}(A_{33})} |\tilde{\lambda}-\lambda| \le \|E\|_2$ *then*

$$\tan \angle^{\frown}_{\max}\left(X_1, [\tilde{X}_1, \tilde{X}_2]\right) \le \frac{\|E\|_2}{\mathrm{gap}(A_{11}, A_{33}) - 2\|E\|_2}.$$

iii) If $\mathrm{gap}(\mathrm{env}(\mathrm{eig}(\tilde{A}_{11})), \mathrm{eig}(\tilde{A}_{33})) > 2\|E\|_2$ *and* $\max\limits_{\lambda\in\mathrm{eig}(A_{11})} \min\limits_{\tilde{\lambda}\in\mathrm{eig}(\tilde{A}_{11})} |\tilde{\lambda}-\lambda| \le \|E\|_2$ *then*

$$\tan \angle^{\frown}_{\max}\left(X_1, [\tilde{X}_1, \tilde{X}_2]\right) \le \frac{\|E\|_2}{\mathrm{gap}(\tilde{A}_{11}, \tilde{A}_{33}) - 2\|E\|_2}.$$

Proof. Defining $\tilde{A}_3 := \tilde{X}_3^H A \tilde{X}_3 = \tilde{X}_3^H (A+E)\tilde{X}_3 - \tilde{X}_3^H E \tilde{X}_3 = \tilde{A}_{33} - \tilde{X}_3^H E \tilde{X}_3$ we get

$$\begin{aligned} R := A\tilde{X}_3 - \tilde{X}_3\tilde{A}_3 &= (A+E)\tilde{X}_3 - E\tilde{X}_3 - \tilde{X}_3(\tilde{A}_{33} - \tilde{X}_3^H E\tilde{X}_3) \\ &= -E\tilde{X}_3 + \tilde{X}_3\tilde{X}_3^H E\tilde{X}_3 = -(I - \tilde{X}_3\tilde{X}_3^H)E\tilde{X}_3, \end{aligned}$$

which implies $\|R\|_2 \le \|E\|_2$. Since $\tilde{A}_3$ and $\tilde{A}_{33}$ differ by $\tilde{X}_3^H E\tilde{X}_3$ their eigenvalues differ by at most $\|E\|_2$ by Theorem 2.3, implying $\mathrm{gap}(\mathrm{eig}(A_{11}), \mathrm{eig}(\tilde{A}_3)) \ge \mathrm{gap}(\mathrm{eig}(A_{11}), \mathrm{eig}(\tilde{A}_{33})) - \|E\|_2$. Together with the assumption

$$\mathrm{gap}\left(\mathrm{env}(\mathrm{eig}(A_{11})), \mathrm{eig}(\tilde{A}_{33})\right) > \|E\|_2$$

we deduct that there is no eigenvalue of $\tilde{A}_3$ in the interval $[\lambda_{\min}(A_{11}) - \mathrm{gap}(A_{11}, \tilde{A}_3), \lambda_{\max}(A_{11}) + \mathrm{gap}(A_{11}, \tilde{A}_3)]$, where $\lambda_{\min}(A_{11}) := \min(\mathrm{eig}\,(A_{11}))$ and $\lambda_{\max}(A_{11}) := \max(\mathrm{eig}(A_{11}))$. Hence Theorem 2.4 is applicable and leads to

$$\angle^{\frown}_{\max}(\tilde{X}_3, [X_2, X_3]) \le \frac{\|R\|_2}{\mathrm{gap}(A_{11}, \tilde{A}_3)} \le \frac{\|E\|_2}{\mathrm{gap}(A_{11}, \tilde{A}_3)} \le \frac{\|E\|_2}{\mathrm{gap}(A_{11}, \tilde{A}_{33}) - \|E\|_2}.$$

From Lemma 2.1.iii it follows that

$$\angle^{\frown}_{\max}(X_1, [\tilde{X}_1, \tilde{X}_2]) = \angle^{\frown}_{\max}([\tilde{X}_1, \tilde{X}_2]^{\perp}, X_1^{\perp}) = \angle^{\frown}_{\max}(\tilde{X}_3, [X_2, X_3]),$$

which concludes the proof of part i).

For part ii) we use that by assumption $\max_{\tilde{\lambda}\in\mathrm{eig}(\tilde{A}_{33})} \min_{\lambda\in\mathrm{eig}(A_{33})} |\tilde{\lambda}-\lambda| \le \|E\|_2$ for every eigenvalue of $\tilde{A}_{33}$ there exists an eigenvalue of A_{33} that is no further away than $\|E\|_2$. This implies $\mathrm{gap}(A_{11}, A_{33}) \ge \mathrm{gap}(A_{11}, \tilde{A}_{33}) - \|E\|_2$. Together with part i) we obtain the assertion. Analogously, also part iii) follows from part i). □

REMARK 2.9. The assumption in part ii) that $\max_{\tilde{\lambda}\in\mathrm{eig}(\tilde{A}_{33})} \min_{\lambda\in\mathrm{eig}(A_{33})} |\tilde{\lambda} - \lambda| \le \|E\|_2$ and the similar condition in part iii) are not very restrictive. By Theorem 2.3, the eigenvalues of A differ from the corresponding ones of $A+E$ by at most $\|E\|_2$. So the assumption is mainly a restriction on the distribution of $\mathrm{eig}(A+E)$ into $\mathrm{eig}(\tilde{A}_{11})$, $\mathrm{eig}(\tilde{A}_{22})$, and $\mathrm{eig}(\tilde{A}_{33})$.

2.4. Ritz and Krylov subspaces. Let $A \in \mathbb{C}^{n\times n}$, $\mathcal{V} \subset \mathbb{C}^n$ be a subspace, $Y \in \mathbb{C}^{n\times k}$, and $M \in \mathbb{C}^{k\times k}$. The pair (M, Y) is called a *Ritz pair* of A with respect to $\mathcal{V}$ if it satisfies the *Galerkin* condition

$$AY - YM \perp \mathcal{V} \quad \text{and} \quad \mathrm{ran}(Y) \subset \mathcal{V}.$$

A subspace $\mathcal{Y}$ is called a *Ritz space* of A in $\mathcal{V}$ if for some basis $Y \in \mathbb{C}^{n\times k}$ of $\mathcal{Y}$ there is an $M \in \mathbb{C}^{\dim(\mathcal{Y})\times\dim(\mathcal{Y})}$ such that (M, Y) is a Ritz pair of A with respect to $\mathcal{V}$. The eigenvalues of such an M are well-defined and are called *Ritz values* of A corresponding to $\mathcal{Y}$.

A *Krylov relation* of A of order k is a relation of the form

$$AV_k = V_k B_k + v_{k+1} b_{k+1}^H, \tag{2.5}$$

where $B_k \in \mathbb{C}^{k\times k}$, $b_{k+1} \in \mathbb{C}^k$, and the columns of $[V_k, v_{k+1}] \in \mathbb{C}^{n\times k+1}$ are orthonormal.

Relation (2.5) implies that $\mathrm{ran}(V_k)$ is a Krylov subspace of A. But it does not imply that $\mathrm{ran}(V_i) = \mathrm{ran}([v_1, \ldots, v_i])$ is a Krylov subspace of A for $i < k$. It does if B_k is Hessenberg and $b_{k+1} = b_{k+1,k} e_k$. The following two theorems are the main basis of our results.

THEOREM 2.10. *[24, Theorem 6.3] [23] Let the eigenvalues λ_i of the Hermitian matrix A be ordered decreasingly. Then the angle between the exact eigenvector z_i associated with λ_i and the k-th Krylov subspace $\mathcal{K}_k$ satisfies the inequality*

$$\tan\angle(z_i, \mathcal{K}_k) \le \frac{\theta_i}{\psi_{k-i}(1+2\eta_i)} \tan\angle(v_1, z_i), \tag{2.6}$$

where

$$\theta_1 = 1, \quad \theta_i = \prod_{j=1}^{i-1} \frac{\lambda_j - \lambda_n}{\lambda_j - \lambda_i} \quad \textit{for} \quad i > 1, \quad \eta_i = \frac{\lambda_i - \lambda_{i+1}}{\lambda_{i+1} - \lambda_n} \tag{2.7}$$

and ψ_{k-i} is the Chebychev polynomial of degree $k-i$.

THEOREM 2.11. *[31, Theorem 2] Let $\mathcal{X}$ be an eigenspace of $A \in \mathbb{C}^{n\times n}$ and let $\mathcal{K}$ be a subspace in $\mathbb{C}^n$. Let $\mathcal{Y}$ be a Ritz space in $\mathcal{K}$ and let $\mathcal{Y}^{\perp}$ be the orthogonal complement of $\mathcal{Y}$ in $\mathcal{K}$. Then*

$$\sin\angle^{\frown}_{\max}(\mathcal{X}, \mathcal{Y}) \le \sin\angle^{\frown}_{\max}(\mathcal{X}, \mathcal{K}) \sqrt{1 + \frac{\|PA(I-P)\|_2^2}{\mathrm{sep}_{A,2}(\mathcal{Y}_{\perp}, \mathcal{X})^2}}, \tag{2.8}$$

where P is the orthogonal projector onto $\mathcal{K}$.

REMARK 2.12. In the original formulation of Theorem 2.11 in [31] it is required that $\mathcal{X}$ and $\mathcal{Y}$ are of the same dimension. However, the proof given there is actually valid for $\dim(\mathcal{X}) \neq \dim(\mathcal{Y})$.

3. Distance to inexact Krylov subspaces. Here we present our main result concerning how well an invariant subspace $\mathcal{X}_1$ of A is contained in a search space $\tilde{\mathcal{K}}_k$ which is a Krylov subspace of a perturbed matrix $A + E$. We consider a situation where l iterations of an inexact Krylov subspace method have been carried out but the desired eigenpair approximations cannot yet be considered converged. This leads to the question of how many more iterations have to be carried out until convergence can be expected. The theorem uses nested subsets of eigenvalues. An illustration is provided in Figure 1.

THEOREM 3.1. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Let $J_1 \subseteq J_2 \subseteq J_3 \subseteq J_4$ be nested nonempty subsets of $\{1, 2, \ldots, n\}$ so that $\max(J_3) < n$ and J_2, J_3 consist of consecutive integers.*
Denote by $J_L := \{1, 2, \ldots, \min(J_3)-1\}$ the leading and by $J_T := \{\max(J_3)+1, \ldots, n\}$ the trailing indices of J_3.
Let $\Lambda_i := \{\lambda_j : j \in J_i\}$ and $\Lambda_{-i} = \{\lambda_j : j \in \{1, \ldots n\} \setminus J_i\}$ for $i \in \{1, \ldots, 4, L, T\}$.
Let $\mathcal{X}_1$ and $\mathcal{X}_4$ be invariant subspaces of A corresponding to Λ_1 and Λ_4, respectively.
Let $k > \max(J_3)$.
For $j = 1, \ldots, k$ let $\tilde{\mathcal{K}}_j := \mathcal{K}_j(A + E, \tilde{v}_1)$ for some Hermitian $E \in \mathbb{C}^{n\times n}$ and some $\tilde{v}_1 \in \mathbb{C}^n$.
For $i \in \{2, 3\}$ let $\tilde{\mathcal{X}}_i$ be an invariant subspace of $A + E$ corresponding to $\{\lambda_j(A + E) : j \in J_i\}$ such that $\tilde{X}_2 \subset \tilde{X}_3$.
If $2\|E\|_2 < \min\{\operatorname{gap}(\Lambda_1, \Lambda_{-2}), \operatorname{gap}(\Lambda_2, \Lambda_{-3}), \operatorname{gap}(\Lambda_3, \Lambda_{-4})\}$ then for every $l = 1, \ldots, k - |J_L|$, we have

$$\begin{aligned} \angle^{\frown}_{\max}\left(\mathcal{X}_1, \tilde{\mathcal{K}}_k\right) &\leq \angle^{\frown}_{\max}\left(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k\right) + \delta_{1,2} \\ &\leq \arctan\left(\varrho_{k,l} \cdot \tan \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)\right) + \delta_{1,2} \qquad (3.1) \\ &\leq \arctan\left(\varrho_{k,l} \cdot \tan_{\leq \frac{\pi}{2}}(\angle^{\frown}_{\max}(\mathcal{X}_4, \tilde{\mathcal{K}}_l) + \delta_{3,4})\right) + \delta_{1,2} \qquad (3.2) \end{aligned}$$

where $\tan_{\leq \frac{\pi}{2}}(\alpha) := \tan(\min\{\alpha, \frac{\pi}{2}\})$ and

(3.3)
$$\varrho_{k,l} := \left(\sum_{i\in J_2} \frac{\tilde{\theta}_i^2}{\psi_{k-l-|\Lambda_L|}(1 + 2\tilde{\eta}_i)^2}\right)^{\frac{1}{2}}, \quad \delta_{i,j} := \arctan\left(\frac{\|E\|_2}{\operatorname{gap}(\Lambda_i, \Lambda_{-j}) - 2\|E\|_2}\right),$$

(3.4)
$$\tilde{\theta}_i := \prod_{j\in J_L} \frac{|\lambda_j - \lambda_n| + 2\|E\|_2}{|\lambda_j - \lambda_i| - 2\|E\|_2} > 0, \qquad \tilde{\eta}_i := \frac{\operatorname{gap}(\lambda_i, \Lambda_T) - 2\|E\|_2}{\operatorname{spread}(\Lambda_T) + 2\|E\|_2} > 0,$$

where ψ_j denotes the Chebychev polynomial of degree j.

Proof. The proof consists of three parts, proving that
i) $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k) \leq \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k) + \delta_{1,2}$,
ii) $\tan \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k) \leq \varrho_{k,l} \cdot \tan \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)$, and
iii) $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l) \leq \min(\frac{\pi}{2}, \angle^{\frown}_{\max}(\mathcal{X}_4, \tilde{\mathcal{K}}_l) + \delta_{3,4})$, respectively.
Then the claim follows because of the monotonicity of the tangent in $[0, \frac{\pi}{2}]$.

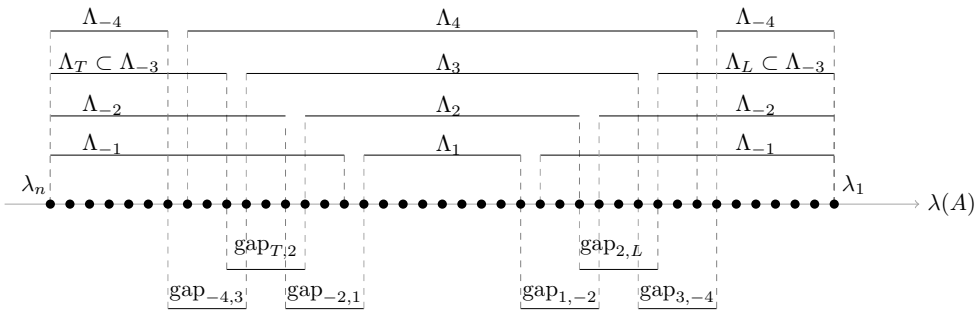


FIG. 1. *Illustration of the nested spectral subsets, here* $\mathrm{gap}(\Lambda_i, \Lambda_j) = \min\{\mathrm{gap}_{ij}, \mathrm{gap}_{ji}\}$

Part i) Using Theorem 2.8 ii) and the definition of $\delta_{1,2}$, respectively, we have

$$\tan \angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{X}}_2) \leq \frac{\|E\|_2}{\mathrm{gap}(\Lambda_1, \Lambda_{-2}) - 2\|E\|_2} = \tan \delta_{1,2}. \tag{3.5}$$

Hence, by the triangle inequality, Lemma 2.1.i, we get

$$\angle^{\frown}_{\max}\left(\mathcal{X}_1, \tilde{\mathcal{K}}_k\right) \leq \angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{X}}_2) + \angle^{\frown}_{\max}\left(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k\right) \leq \delta_{1,2} + \angle^{\frown}_{\max}\left(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k\right). \tag{3.6}$$

Part iii) The proof of part iii) is similar to that of part i) and also uses a triangle inequality, this time in the form of

$$\angle^{\frown}_{\max}\left(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l\right) \leq \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \mathcal{X}_4) + \angle^{\frown}_{\max}\left(\mathcal{X}_4, \tilde{\mathcal{K}}_l\right) \leq \delta_{3,4} + \angle^{\frown}_{\max}\left(\mathcal{X}_4, \tilde{\mathcal{K}}_l\right)$$

where we used Theorem 2.8 iii) for the second inequality. Also, any subspace angle can at most be $\frac{\pi}{2}$ by definition.

Part ii) If $\dim(\tilde{\mathcal{K}}_l) < \dim(\tilde{\mathcal{X}}_3) = |J_3|$ then $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l) = \frac{\pi}{2}$ and the right-hand side is infinite. Thus there is nothing to prove in this case. Hence we may assume that $\dim(\tilde{\mathcal{K}}_l) \geq |J_3|$ in the following.

We start by considering the angles between $\tilde{\mathcal{K}}_k$ and eigenvectors of $A + E$. So, let $\tilde{x}_1, \ldots, \tilde{x}_n$ be unit eigenvectors of $A + E$ corresponding to the eigenvalues $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n$ such that $\tilde{\mathcal{X}}_3 = \mathrm{span}\{\tilde{x}_j : j \in J_3\}$ and $\tilde{\mathcal{X}}_2 = \mathrm{span}\{\tilde{x}_j : j \in J_2\}$. Let $\tilde{\Lambda}_T := \{\tilde{\lambda}_j : j \in J_T\}$.

For $i \in J_2$ let $\mathcal{V}_i := \mathrm{span}\{\tilde{x}_j : j \in J_3, j \neq i\} = \tilde{\mathcal{X}}_3 \cap \mathrm{span}(\tilde{x}_i)^{\perp}$. We note that $\tilde{\mathcal{K}}_l \cap \mathcal{V}_i^{\perp}$, being an intersection of a $(\geq |J_3|)$-dimensional and an $(n - |J_3| + 1)$-dimensional subspace, is at least one-dimensional. Thus there exists a nonzero vector $v_i \in \tilde{\mathcal{K}}_l \cap \mathcal{V}_i^{\perp}$ such that $\angle(\tilde{x}_i, v_i) = \angle(\tilde{x}_i, \tilde{\mathcal{K}}_l \cap \mathcal{V}_i^{\perp})$. Let

$$A_i := A + E - \sum_{\substack{j \in J_3 \\ j \neq i}} (\tilde{\lambda}_j - \tilde{\lambda}_n) \tilde{x}_j \tilde{x}_j^H.$$

Note that the matrices A_i and $A+E$ have the same eigenvectors. Also the eigenvalues $\tilde{\lambda}_j$ for $j \in (\{1, \ldots, n\} \setminus J_3) \cup \{i\}$ coincide. The remaining eigenvalues of $A + E$ are moved to $\tilde{\lambda}_n \in \tilde{\Lambda}_T$. Then Theorem 2.10, applied to A_i and v_i, yields

$$\tan \angle(\tilde{x}_i, \mathcal{K}_{k-l+1}(A_i, v_i)) \leq \rho_i := \frac{\theta_i}{\psi_{k-l-|J_T|}(1 + 2\eta_i)} \tan \angle(\tilde{x}_i, v_i), \tag{3.7}$$

with

$$\theta_i = \prod_{j\in J_L} \frac{|\tilde{\lambda}_j - \tilde{\lambda}_n|}{|\tilde{\lambda}_j - \tilde{\lambda}_i|}, \quad \eta_i = \frac{|\tilde{\lambda}_i - \tilde{\lambda}_{\min(J_T)}|}{|\tilde{\lambda}_{\min(J_T)} - \tilde{\lambda}_n|} = \frac{\operatorname{gap}(\tilde{\lambda}_i, \tilde{\Lambda}_T)}{\operatorname{spread}(\tilde{\Lambda}_T)}. \tag{3.8}$$

Let $g_i \in \mathcal{K}_{k-l+1}(A_i, v_i)$ be such that $\angle(\tilde{x}_i, g_i) = \angle(\tilde{x}_i, \mathcal{K}_{k-l+1}(A_i, v_i))$. W.l.o.g. g_i is scaled such that $\tilde{x}_i^H g_i = 1$ (otherwise, if $\tilde{x}_i^H g_i \neq 0$ we can rescale g_i, if $\tilde{x}_i^H g_i = 0$ it follows with Lemma 2.1.v that $\pi/2 = \angle(\tilde{x}_i, g_i) = \angle(\tilde{x}_i, \tilde{\mathcal{K}}_l \cap \mathcal{V}_i^\perp) \leq \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)$, i.e., $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l) = \pi/2$ such that the right-hand side of part ii) is infinite, i.e., in this case is nothing to prove). Note that since $v_i \perp \mathcal{V}_i$, we have $A_i v_i = (A+E)v_i$ and $A_i v_i \perp \mathcal{V}_i$. Thus by induction $\mathcal{K}_j(A_i, v_i) = \mathcal{K}_j(A+E, v_i)$ and $\mathcal{K}_j(A_i, v_i) \perp \mathcal{V}_i$ for every $j = 1, 2, \ldots$. Hence, $g_i \perp \mathcal{V}_i$ thus there exist $g_{L,i} \in \mathbb{C}^{|J_L|}$ and $g_{T,i} \in \mathbb{C}^{|J_T|}$ such that

$$g_i = [\tilde{x}_1, \ldots, \tilde{x}_{|J_L|}] g_{L,i} + \tilde{x}_i + [\tilde{x}_{\min(J_T)}, \ldots, \tilde{x}_n] g_{T,i}.$$

Using Theorem 2.2, the definition of g_i, and (3.7) leads to $\|[g_{L,i}^T, g_{T,i}^T]^T\|_2 = \tan\angle(g_i, \tilde{x}_i) \leq \rho_i$. Since the subspaces $\tilde{\mathcal{K}}_j$ are nested and $v_i \in \tilde{\mathcal{K}}_l$, we have $\mathcal{K}_{k-l+1}(A+E, v_i) \subset \mathcal{K}_k(A+E, \tilde{v}_1) = \tilde{\mathcal{K}}_k$. Thus $g_i \in \tilde{\mathcal{K}}_k$. Since $\tilde{\mathcal{K}}_k$ is independent of i, it contains g_i for all $i \in J_2$. Thus it contains the subspace $\operatorname{ran}(G)$ where

$$G := [g_{\min(J_2)}, \ldots, g_{\max(J_2)}] = [\tilde{x}_1, \ldots, \tilde{x}_n] \begin{bmatrix} G_L \\ 0 \\ I_{|J_2|} \\ 0 \\ G_T \end{bmatrix}$$

with $G_L := [g_{L,\min(J_2)}, \ldots, g_{L,\max(J_2)}] \in \mathbb{C}^{|J_L|\times|J_2|}$, $G_T := [g_{T,\min(J_2)}, \ldots, g_{T,\max(J_2)}] \in \mathbb{C}^{|J_T|\times|J_2|}$ and the upper and lower zero blocks are of format $(\min(J_2) - \min(J_3)) \times |J_2|$ and $(\max(J_3) - \max(J_2)) \times |J_2|$, respectively. Using Lemma 2.1.ii and Theorem 2.2 we have

$$\tan\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k) \leq \tan\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \operatorname{ran}(G)) = \left\| \begin{bmatrix} G_L \\ G_T \end{bmatrix} \right\|_2. \tag{3.9}$$

Further transformations yield

$$\left\| \begin{bmatrix} G_L \\ G_T \end{bmatrix} \right\|_2 \leq \left\| \begin{bmatrix} G_L \\ G_T \end{bmatrix} \right\|_F \leq \left(\sum_{i\in J_2} \rho_i^2 \right)^{\frac{1}{2}}. \tag{3.10}$$

Applying Theorem 2.3, we have for $i \neq j$: $|\lambda_i - \lambda_j| - 2\|E\|_2 \leq |\tilde{\lambda}_i - \tilde{\lambda}_j| \leq |\lambda_i - \lambda_j| + 2\|E\|_2$. Thus $\theta_i \leq \tilde{\theta}_i$ and $\eta_i \geq \tilde{\eta}_i$, for θ_i, η_i as in (3.8) and $\tilde{\theta}_i, \tilde{\eta}_i$ as in (3.4). Also, $\tilde{\theta}_i$ and $\tilde{\eta}_i$ are positive by the assumed bound on $\|E\|_2$. Moreover, $\psi_{k-l-|J_L|}(1+2\eta_i) > \psi_{k-l-|J_L|}(1+2\tilde{\eta}_i)$, because Chebychev polynomials are positive and monotonically increasing in $[1;\infty)$. Furthermore, since $\tilde{\mathcal{X}}_3 = \operatorname{span}(\tilde{x}_i) \oplus \mathcal{V}_i$ we have by Lemma 2.1.v that $\angle(\tilde{x}_i, \tilde{\mathcal{K}}_l \cap \mathcal{V}_i^\perp) \leq \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)$, hence $\angle(\tilde{x}_i, v_i) \leq \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)$. Thus

$$\rho_i \leq \frac{\tilde{\theta}_i}{\psi_{k-l-|J_L|}(1+2\tilde{\eta}_i)} \tan\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l). \tag{3.11}$$

Combining (3.9), (3.10), and (3.11) we have

$$\tan \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k) \leq \left(\sum_{i \in J_2} \frac{\tilde{\theta}_i^2}{\psi_{k-l-|J_L|}(1+2\tilde{\eta}_i)^2} \right)^{\frac{1}{2}} \tan \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l) \\ = \varrho_{k,l} \tan \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l). \tag{3.12}$$

This concludes the proof of part ii) and of the whole theorem. □

The following remarks are in order.

Remark 3.2. Theorem 3.1 somewhat resembles Theorem 2.10, but holds several improvements. i) The search space is chosen as Krylov subspace of a perturbed matrix $A + E$ (instead of A itself). Thus the setting of inexact Krylov methods is covered. ii) Instead of just eigenvectors we consider invariant subspaces. This allows to treat clusters of eigenvalues as a whole. iii) The dimension l of the Krylov subspace on the right-hand side is allowed to be larger than one. (Note that $\angle(v_1, z_i)$ in (2.6) could be written as $\angle(v_i, \mathcal{K}_1)$.) This is necessary for the theorem to be meaningful as for $l < \dim(\mathcal{X}_4)$ the right-hand side is infinite. Moreover, this might be useful if information about the angle $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)$ or $\angle^{\frown}_{\max}(\mathcal{X}_4, \tilde{\mathcal{K}}_l)$ is available for some l. iv) Theorem 3.1 is still useful even if the wanted eigenvalues Λ_1 are not well separated from the rest of the spectrum, i.e., if $\operatorname{gap}(\Lambda_1, \Lambda_{-1}) - 2\|E\|_2$ is tiny or even negative. This is achieved by considering four possibly different (but nested) sets of eigenvalues, $\Lambda_1, \ldots, \Lambda_4$. All that matters is that the gaps between Λ_i and the complement of Λ_{i+1}, i.e., $\operatorname{gap}(\Lambda_1, \Lambda_{-2})$, $\operatorname{gap}(\Lambda_2, \Lambda_{-3})$, $\operatorname{gap}(\Lambda_3, \Lambda_{-4})$, are well larger than $2\|E\|_2$. We stress in particular that $\operatorname{gap}(\Lambda_i, \Lambda_{-i})$ may even be zero, i.e., $\mathcal{X}_i$ is not required to be simple. Finally, Theorem 3.1 is a generalization of Theorem 2.10 and reduces to the latter in case $\|E\| = 0$, $l = 1$, $J_1 = J_2 = J_3 = J_4 = \{i\}$.

Remark 3.3. The constant $\varrho_{k,l}$ decreases exponentially fast as k grows. Indeed, from $\psi_k(x) = \cosh(k \cdot \operatorname{arcosh}(x))$ for $|x| \geq 1$, $\cosh(x) = \frac{1}{2}(e^x + e^{-x}) > \frac{1}{2}e^x$, and $\operatorname{arcosh}(x) = \ln(x + \sqrt{x^2 - 1})$ for $x \geq 1$ we deduct that $\psi_k(1 + 2\eta) > \frac{1}{2}(1 + 2\eta + 2\sqrt{\eta + \eta^2})^k$. This indicates at least linear convergence of $\varrho_{k,l}$ towards zero.

Remark 3.4. In order to get an idea of the qualitative behavior of bound (3.2) we consider the following. For small arguments $0 < a \ll 1$ we have $\arctan(a) \approx a$ and $\arctan(a) \leq a$. Since in the later stages of the iteration $\varrho_{k,l}$ is small we may replace $\arctan(\cdot)$ by the identity in (3.2). Doing so, taking logarithms, and substituting $\gamma_i := 1 + 2\tilde{\eta}_i + 2\sqrt{\tilde{\eta}_i + \tilde{\eta}_i^2}$, $\tau_l := \tan_{\leq \frac{\pi}{2}}\left(\angle^{\frown}_{\max}(\mathcal{X}_4, \tilde{\mathcal{K}}_l) + \delta_{3,4}\right)$ results in

$$\log\left(\angle^{\frown}_{\max}\left(\mathcal{X}_1, \tilde{\mathcal{K}}_k\right)\right) \leq \log\left(\tau_l \left(\sum_{i \in J_2} \frac{\tilde{\theta}_i^2}{\frac{1}{4}\gamma_i^{2(k-l-|\Lambda_L|)}} \right)^{\frac{1}{2}} + \delta_{1,2} \right).$$

Moreover, using $\log(a + b) \approx \max(\log(a), \log(b))$ and some algebraic transformations the right hand side can be approximated by

$$\max\left(\log(\delta_{1,2}), \log(\tau_l) + \max_{i \in J_2}\left(\log(\tilde{\theta}_i) - \log\left(\tfrac{1}{2}\right) + (l + |\Lambda_L|)\log(\gamma_i) - k\log(\gamma_i)\right)\right).$$

On the other hand, in the beginning of the iteration the argument of the arctan in (3.2) is usually large and the only save bound for $\angle^{\frown}_{\max}\left(\mathcal{X}_1, \tilde{\mathcal{K}}_k\right)$ is $\frac{\pi}{2}$. Together

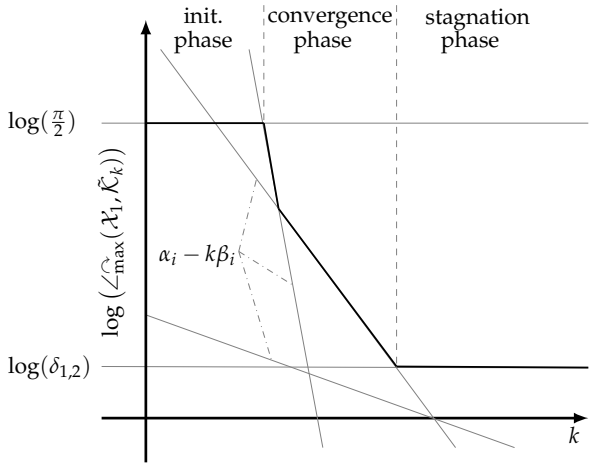


FIG. 2. *Qualitative shape of the bound* (3.2)*, see Remark 3.4*

with the previous consideration we get

$$\log\left(\angle_{\max}\left(\mathcal{X}_1,\tilde{\mathcal{K}}_k\right)\right) \lessapprox \min\left(\log\left(\tfrac{\pi}{2}\right), \max\left(\log(\delta_{1,2}), \max_{i\in J_2}(\alpha_i - k\beta_i)\right)\right) \tag{3.13}$$

with $\alpha_i = \log(\tau_l) + \log(\tilde{\theta}_i) - \log(\frac{1}{2}) + (l + |\Lambda_L|)\log(\gamma_i)$, $\beta_i = \log(\gamma_i) > 0$, and $a \lessapprox b$ meaning that "a is usually smaller than b, it can be a bit larger, but not much".

So, the logarithm of $\angle_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ can be approximately bounded by a set of straight lines, two horizontal and a few decreasing ones, see Figure 2 for an illustration. We can identify three phases. i) During the first iteration steps all the decreasing lines lie above the $\frac{\pi}{2}$-level. Hence the minimum is assumed by $\log\left(\frac{\pi}{2}\right)$. We call this phase the initialization phase. ii) In the following iteration steps the decreasing lines fall below the $\frac{\pi}{2}$-level and become the important terms. Here the largest of the decreasing lines defines the bound. We obtain a strict monotonically decreasing, piecewise linear, convex shape. This phase is called the convergence phase. It is well possible that only one line dominates for the whole convergence phase, i.e., there is no bend, see the numerical examples in Section 5. iii) In the last phase the maximum is assumed by $\log(\delta_{1,2})$. We denote the last section by stagnation phase because the angle of inclusion does not decrease anymore.

REMARK 3.5. Theorem 3.1 offers some freedom in choosing the index sets J_i. J_1 is fixed to the indices of the wanted eigenvalues. J_2, J_3, J_4 are free and could thus be chosen to minimize the bounds (3.1), (3.2). Here the various relations between the index sets J_1, J_2, J_3, J_4 influence the constants δ_{12},$\tilde{\theta}_i$, $\tilde{\eta}_i$ and δ_{34} in different ways. For example extending J_2 leads to a larger $\mathrm{gap}(\Lambda_1, \Lambda_{-2})$ and thus to a decreased δ_{12}. of (3.2). Similar extending J_3 with respect to J_2 will improve $\tilde{\theta}_i$ and $\tilde{\eta}_i$. Finally extending J_4 with respect to J_3 will improve δ_{34}.

REMARK 3.6. Theorem 3.1 still holds if the eigenvalues are sorted in increasing order, $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. This can be seen by applying the theorem to $-A$ and using that all the constants $\tilde{\theta}_i, \tilde{\eta}_i, \delta_{i,j}$ and $\varrho_{k,l}$ are invariant under negating the eigenvalues.

REMARK 3.7. Although this case is irrelevant in practice, we mention that Theorem 3.1 still holds if the denominator of $\tilde{\eta}_i$ is zero, i.e., if $\|E\|_2 = 0 = \mathrm{spread}(\Lambda_T)$. In this case $\tilde{\eta}_i = \infty$ for all $i \in J_2$ and $\varrho_{k,l} = 0$. Then $\angle_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k) \leq \delta_{1,2}$ for all

$k > l + |J_L|$.

REMARK 3.8. Theorem 3.1 relates four angles to one another: $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$, $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k)$, $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)$, and $\angle^{\frown}_{\max}(\mathcal{X}_4, \tilde{\mathcal{K}}_l)$. The relation "looks nicest" when restricted to the first and the last of these. (Especially since the middle two angles involve invariant subspaces of $A + E$ which seem hard to obtain in practice.) The reason to include all four angles in Theorem 3.1 is that i) $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k)$ appears in the following Theorem 4.1 and ii) $\angle^{\frown}_{\max}(\tilde{\mathcal{X}}_3, \tilde{\mathcal{K}}_l)$ may actually be bounded in practice.

REMARK 3.9. The formulation of Theorem 3.1 and Figure 1 may seem to suggest that the desired eigenvalues can be in the center of the spectrum (and the theorem does hold in this case). However, the constants $\tilde{\theta}_i$, $i \in J_2$ rapidly grow for an increasing number of leading eigenvalues $|J_L|$. Hence the theorem is useful only if fairly exterior eigenvalues are wanted (say, when $|J_L|$ is not larger than two or three). For example considering the matrix constructed by the MATLAB command $\mathtt{A} = \mathrm{diag}(\mathtt{100} : -\mathtt{1} : \mathtt{1})$, $\|E\|_2 = 10^{-10}$ and choose $\Lambda_2 = \Lambda_3 = \Lambda_4 = \Lambda_1$. Then for

- $\Lambda_1 = \{90, 91, \ldots, 100\}$ we get $\tilde{\theta}_{\max} = 1$,
- $\Lambda_1 = \{90, 91\}$ we get $\tilde{\theta}_{\max} = 1.731 \cdot 10^{12}$,
- $\Lambda_1 = \{50, 51\}$ we get $\tilde{\theta}_{\max} = 5.045 \cdot 10^{28}$

with $\tilde{\theta}_{\max} := \max_{i \in J_2} \tilde{\theta}_i$. Hence, already if we are interested in the 9th and 10th largest eigenvalue of A Theorem 3.1 becomes little meaningful. Note, if largest Λ_3 contains all the largest eigenvalues then $\tilde{\theta}_{\max} = 1$.

4. Distance to inexact Ritz spaces. So far we have considered how well the exact eigenspace $\mathcal{X}_1$ is contained in the search space $\tilde{\mathcal{K}}_k$. If $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ is small then there is a subspace of $\tilde{\mathcal{K}}_k$ that is close to $\mathcal{X}_1$. However, in practice this subspace in $\tilde{\mathcal{K}}_k$ is not known and $\mathcal{X}_1$ is approximated by a Ritz space $\tilde{\mathcal{Y}}$ in $\tilde{\mathcal{K}}_k$. Hence, in this section we address the question: How much worse is the distance to a Ritz space compared to the distance to the whole search space?

THEOREM 4.1. *Let $A, E, \Lambda_1, \Lambda_2, \Lambda_{-2}, \mathcal{X}_1, \tilde{\mathcal{X}}_2$ and $\delta_{1,2}$ be defined as in Theorem 3.1.*
Let $B_k \in \mathbb{C}^{k\times k}$, $b_{k+1} \in \mathbb{C}^k$, and $V_k \in \mathbb{C}^{n\times k}$ be such that the Krylov relation (2.5) for $A + E$ is satisfied and $\tilde{\mathcal{K}}_k = \mathrm{ran}(V_k)$.
Let $\tilde{\mathcal{Y}}$ be a Ritz space of $A + E$ in $\tilde{\mathcal{K}}_k$.
Let $\tilde{\mathcal{M}}$ be the Ritz values corresponding to $\tilde{\mathcal{Y}}$ and let $\tilde{\mathcal{M}}_-$ be the set of $k - \dim(\tilde{\mathcal{Y}})$ remaining Ritz values, with $\|E\|_2 < \mathrm{gap}(\tilde{\mathcal{M}}_-, \Lambda_2)$.
Then

(4.1)
$$\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}}) \le \delta_{1,2} + \arcsin_{\le 1}\left(\sqrt{1 + \frac{\pi^2 \|b_{k+1}\|_2^2}{4(\mathrm{gap}(\tilde{\mathcal{M}}_-, \Lambda_2) - \|E\|_2)^2}}\, \sin \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k)\right)$$

where $\arcsin_{\le 1}(x) := \arcsin(\min\{1, x\})$.

Proof. By Lemma 2.1.i we have

$$\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}}) \le \angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{X}}_2) + \angle^{\frown}_{\max}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{Y}}). \tag{4.2}$$

Now we treat each term of the right-hand side separately. Using Theorem 2.8 ii) and (3.3), respectively, as in Theorem 3.1, the first one is bounded by

$$\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{X}}_2) \le \arctan \frac{\|E\|_2}{\mathrm{gap}(\Lambda_1, \Lambda_{-2}) - 2\|E\|_2} = \delta_{1,2}. \tag{4.3}$$

Applying Theorem 2.11 to the second right-hand side angle in (4.2) results in

$$\angle_{\max}^{\frown}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{Y}}) \leq \arcsin_{\leq 1}\left(\sqrt{1 + \frac{\|P(A+E)(I-P)\|_2^2}{\mathrm{sep}_{A+E,2}(\tilde{\mathcal{Y}}^{\perp}, \tilde{\mathcal{X}}_2)^2}}\sin\angle_{\max}^{\frown}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k)\right), \tag{4.4}$$

where P is the orthogonal projector onto $\tilde{\mathcal{K}}_k$ and $\tilde{\mathcal{Y}}^{\perp}$ is the orthogonal complement of $\tilde{\mathcal{Y}}$ in $\tilde{\mathcal{K}}_k$. Choosing $V_{\perp} \in \mathbb{C}^{n\times n-k}$ such that $[V_k, V_{\perp}]$ is unitary, we have for the numerator $\|P(A+E)(I-P)\|_2^2$ that

$$\begin{aligned}\|P(A+E)(I-P)\|_2^2 = \|V_k^H(A+E)V_{\perp}\|_2^2 = \|V_{\perp}^H(A+E)V_k\|_2^2 \\ = \|V_{\perp}^H V_k B_k + V_{\perp}^H v_{k+1} b_{k+1}^H\|_2^2 = \|b_{k+1}\|_2^2,\end{aligned} \tag{4.5}$$

where relation (1.1) and the fact that $A+E$ is Hermitian was used. Moreover, applying Lemma 2.6 to the denominator $\mathrm{sep}_{A+E,2}(\tilde{\mathcal{Y}}^{\perp}, \tilde{\mathcal{X}}_2)^2$ leads to

$$\mathrm{sep}_{A+E,2}(\tilde{\mathcal{Y}}_{\perp}, \tilde{\mathcal{X}}_2)^2 \geq \frac{4}{\pi^2}\,\mathrm{gap}(\tilde{\mathcal{M}}_-, \tilde{\Lambda}_2)^2 \geq \frac{4}{\pi^2}\left(\mathrm{gap}(\tilde{\mathcal{M}}_-, \Lambda_2) - \|E\|_2\right)^2,$$

with $\tilde{\Lambda}_2 := \{\lambda_j(A+E) : j \in J_2\}$ where J_2 is defined as in Theorem 3.1. Together with (4.5) we obtain

$$\angle_{\max}^{\frown}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{Y}}) \leq \arcsin_{\leq 1}\left(\sqrt{1 + \frac{\pi^2\|b_{k+1}\|_2^2}{4\left(\mathrm{gap}(\tilde{\mathcal{M}}_-, \Lambda_2) - \|E\|_2\right)^2}}\sin\angle_{\max}^{\frown}(\tilde{\mathcal{X}}_2, \tilde{\mathcal{K}}_k)\right). \tag{4.6}$$

Finally, inserting (4.3) and (4.6) into (4.2) completes the proof. □

REMARK 4.2. Again, in order to deal with small gaps between the wanted and unwanted eigenvalues, we have to allow the Ritz space $\tilde{\mathcal{Y}}$ to be of larger dimension than the wanted invariant subspace $\mathcal{X}_1$.

REMARK 4.3. One distinction between the Theorems 3.1 and 4.1 is the amount of information necessary to evaluate the bounds corresponding to the kth step of the Krylov method. Theorem 3.1 is an a priori result as it only needs the information of the lth step (with $l < k$). In contrast, Theorem 4.1 is an a posteriori result, because the vector b_{k+1} and the Ritz values at the kth step are required. It is unclear how an a priori result bounding $\angle_{\max}^{\frown}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ could look like.

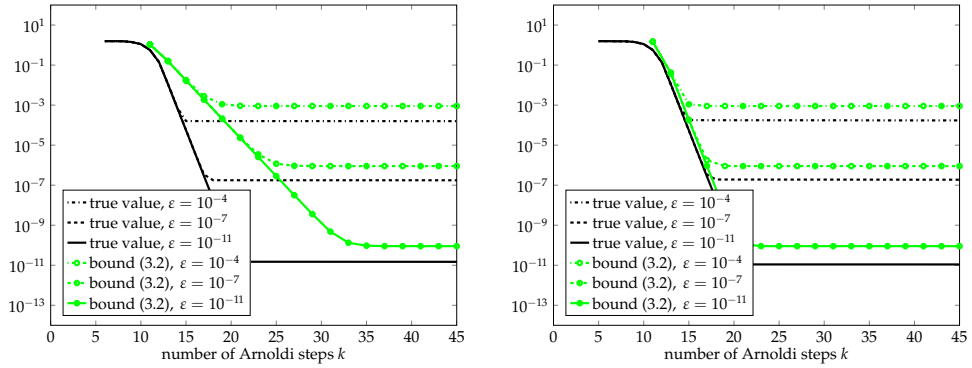
5. Numerical results. In this section we verify the spectral error bounds presented in sections 3 and 4 using two different test matrices. The first one is constructed to evaluate how changes in the eigenvalue distribution influence the behavior of the bounds. The second test matrix comes from the application mentioned in the introduction.

To obtain the Krylov relation (1.1) we apply an exact Arnoldi method, cf. [8,32], to a matrix $A + E$, where E is chosen as a Hermitian, random matrix of prescribed norm.

5.1. Academic problem. We use a diagonal matrix A built by the MATLAB command

$$\mathtt{A} = \mathtt{diag}([9, 9-\mathtt{g}, 7, 6, 5, 4, 4-\mathtt{g}, 2, \mathtt{rand}(1, 992)]), \tag{5.1}$$

where $g \in \mathbb{R}$ is a parameter. This is a 1000×1000 matrix with the eigenvalues $9, 9-g, 7, 6, 5, 4, 4-g, 2$ and additionally 992 eigenvalues in the interval $(0, 1)$. We



(a) case (A), $\Lambda_2 = \Lambda_3 = \Lambda_4 = \Lambda_1$ (b) case (B), $\Lambda_1 = \Lambda_2 \subset \Lambda_3 = \Lambda_4$

FIG. 3. *Experiment 1: True angle of inclusion* $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ *and corresponding bound* (3.2) *with* $\varepsilon := \|E\|_2/\|A\|_2 \in \{10^{-3}, 10^{-7}, 10^{-11}\}$

are interested in the eigenspace corresponding to the eigenvalues $\Lambda_1 = \{9-g, 7, 6, 5, 4\}$. Thus for $g \in [0, 2]$, g is the gap between the wanted and the unwanted eigenvalues and if g is small Λ_1 is not well separated from the remaining spectrum of A.

Experiment 1. The first experiment explores the influence of the norm $\|E\|_2$ of the perturbation on the development of $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ over the course of the Arnoldi iteration. We choose $g = 1$ in (5.1), so the wanted and the unwanted eigenvalues are well separated. We distinguish two cases (A) $\Lambda_2 = \Lambda_3 = \Lambda_4 = \Lambda_1$ and (B) $\Lambda_1 = \Lambda_2 \subset \Lambda_3 = \Lambda_4$.

For case (A) $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ is plotted as black lines for $\|E\|_2 \in \{10^{-3}, 10^{-7}, 10^{-11}\}\cdot\|A\|_2$ in Figure 3(a) . It can be observed that the angle decreases during the iteration, although we search for eigenspaces of A in a Krylov subspace of $A+E$ (and not of A). Convergence sets in after an initial phase and is linear with convergence rate independent of $\|E\|_2$. But $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ decreases not to zero but only to a limiting accuracy that is on the order of $\|E\|_2/\|A\|_2$. We see that all the curves agree in the sense that the three phases mentioned in Remark 3.4 can be identified. The lighter lines show the bounds of Theorem 3.1, where we choose $l = 10$, and $\Lambda_2 = \Lambda_3 = \Lambda_4 = \Lambda_1$ because the $\mathrm{gap}(\Lambda_1, \Lambda_{-1})$ is large. The bound (3.2) starts to exist after the initialization phase and correctly predicts the convergence and stagnation phases. In the stagnation phase the bound overestimates the limiting accuracy only by a value of about 6. The convergence rate (the α in $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k) \le \alpha \cdot \angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_{k+1})$) is overestimated to be 0.41 (true value 0.12, both roughly computed from the observed values). This amounts to a slow down factor of $2.5 \approx \log(0.12)/\log(0.41)$, i.e., reducing the angle by a certain amount takes two and a half times less iterations than predicted by the bound.

The convergence rate of bound (3.2) is determined by $\tilde{\eta}_i$. According to Remark 3.5 the constant $\tilde{\eta}_i$ can be improved by extending subset J_3 with respect to J_2. This is confirmed by Figure 3(b) where $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ is plotted for $\Lambda_1 = \Lambda_2$ and $\Lambda_3 = \Lambda_4 = \Lambda_1 \cup \{9, 4-a, 2\}$ and for the same values of $\|E\|_2$ as before. We see that during the convergence phase the convergence rate of the true value and of the bound almost coincide (true: 0.11, bound: 0.12).

Experiment 2. Here we investigate the sensitivity of bound (3.2) with respect to the gap between the wanted and the remaining eigenvalues. We use the same Λ_1,

case	g in (5.1)	$\delta_{1,2}$ (lower is better)	$\delta_{3,4}$ (lower is better)	$\tilde{\theta}_{\max}$ (lower is better)	$\tilde{\eta}_{\min}$ (higher is better)	overest. of lim. accuracy	slow down factor
(A)	$10^4 \cdot \|E\|_2$	10^{-4}	10^{-4}	$9 \cdot 10^6$	$2 \cdot 10^{-7}$	0.33	1.96
(A)	$10^2 \cdot \|E\|_2$	10^{-2}	10^{-2}	$9 \cdot 10^8$	$2 \cdot 10^{-9}$	0.40	2.41
(A)	$2.1 \cdot \|E\|_2$	1.47	1.47	$9 \cdot 10^{11}$	$2 \cdot 10^{-12}$	1.00	$1.5 \cdot 10^5$
(B)	$2.1 \cdot \|E\|_2$	$5 \cdot 10^{-11}$	$5 \cdot 10^{-11}$	1.00	1.00	0.23	1.47
(B)	$10^{-2} \cdot \|E\|_2$	$5 \cdot 10^{-11}$	$5 \cdot 10^{-11}$	1.00	1.00	0.23	1.49
(C)	$10^{-2} \cdot \|E\|_2$	$5 \cdot 10^{-11}$	$9 \cdot 10^{-11}$	1.00	3.00	0.11	1.00

TABLE 1

Experiment 2: Constants of the bound (3.2) *for case (A) ($\Lambda_1 = \Lambda_2 = \Lambda_3 = \Lambda_4$), case (B) ($\Lambda_1 \subset \Lambda_2 = \Lambda_3 = \Lambda_4$) and case (C) $\Lambda_1 \subset \Lambda_2 \subset \Lambda_3 = \Lambda_4$ (the numbers are correct to leading digit)*

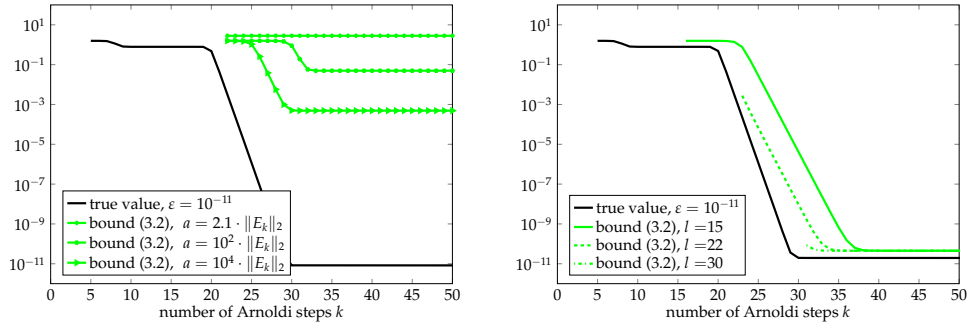
$g \in \{10^{-2}, 2.1, 10^2, 10^4\} \cdot \|E\|_2$ and distinguish between the three cases (A) $\Lambda_1 = \Lambda_2 = \Lambda_3 = \Lambda_4$, (B) $\Lambda_1 \subset \Lambda_2 = \Lambda_3 = \Lambda_4$ and (C) $\Lambda_1 \subset \Lambda_2 \subset \Lambda_3 = \Lambda_4$.

For the first case because of the small gap between the wanted and the unwanted eigenvalues the bound (3.2) can be expected to be very loose. This is confirmed by Figure 4(a), where the true angle $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ and the bound (3.2) are depicted for $\|E\|_2 = 10^{-11}\|A\|_2$, $l = 20$ and $g \in \{2.1, 10^2, 10^4\} \cdot \|E\|_2$ for $g = 10^{-2}\|E\|_2$ the assumptions of Theorem 3.1 are not fulfilled and the bound does not exist. To be precise, as we consider three different matrices we would have to plot three different true angles. However, (apart from a minimal deviation in the starting point of the convergence phase by $2-3$ iteration steps) the three true angles behave the same. So for the sake of readability, only one true angle (for $g = 2.1 \cdot \|E\|_2$) is depicted. The figure illustrates that in case of a small gap choosing $\Lambda_2, \Lambda_3, \Lambda_4$ to coincide with Λ_1 results in a rather loose bound. In particular, we see that the smaller the gap between the wanted and the remaining eigenvalues the less meaningful bound (3.2) becomes. The reason for the poor quality of the bound are the large constants $\delta_{1,2}$, $\delta_{3,4}$, $\tilde{\theta}_{\max}$ and the small $\tilde{\eta}_{\min}$ (see Table 5.1 where $\tilde{\theta}_{\max} := \max_{i \in J_2}\{\tilde{\theta}_i\}$ and $\tilde{\eta}_{\min} := \min_{i \in J_2}\{\tilde{\eta}_i\}$. As mentioned in Remark 3.5 $\Lambda_2, \ldots, \Lambda_4$ may be chosen to optimize $\delta_{1,2}$, $\delta_{3,4}$, $\tilde{\theta}_{\max}$ and $\tilde{\eta}_{\min}$ in (3.2).

In case (B) we consider again $g = 2.1 \cdot \|E\|_2$, but choose now $\Lambda_2 = \Lambda_3 = \Lambda_4 = \Lambda_1 \cup \{9, 4-g\} \supset \Lambda_1$. The constants stated in the second part of Table 5.1 improve substantially and lead to a much sharper bound (3.2) compared to case (A), see Figure 4(b). The bound predicts the actual behavior of $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$, i.e., all three phases are reflected by the bound. We use this setting to analyze the influence of the parameter l on our bound. Choosing $l \in \{15, 22, 30\}$ we see that the larger the number of iteration steps l that already have been carried out, the sharper the corresponding bound in the convergence phase. In this phase the convergence rate of the bound underestimates the true value (≈ 0.11) to be 0.23, which amounts to a slow down factor of 1.47. Afterwards in the stagnation phase the bounds overestimate the actual value only by a constant factor of about 3 (independent of l).

The constants of bound (3.2), stated in the second row of case (B) in Table 5.1, indicate that even for a matrix where the gap between the wanted and the remaining eigenvalues is smaller than the norm of the perturbation, i.e., $g = 10^{-2}\|E\|_2$ the bound (3.2) is meaningful.

We note that in case (C) the bound is further improved choosing Λ_1, Λ_2 as in



(a) $g \in \{2.1, 10^2, 10^4\} \cdot \|E\|_2$, $l = 20$ and $\Lambda_2 = \Lambda_3 = \Lambda_4 = \Lambda_1$

(b) $g = 2.1 \cdot \|E\|_2$, $l \in \{15, 22, 30\}$ and $\Lambda_2 = \Lambda_3 = \Lambda_4 \supset \Lambda_1$.

FIG. 4. *Experiment 2: True angle of inclusion* $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ *and its bounds* (3.2)

case (B) and $\Lambda_3 = \Lambda_4 = \Lambda_2 \cup \{2\}$. As stated in Remark 3.5 this improves $\tilde{\eta}_i$ which leads to more accurate convergence rate of the bound. Here the convergence rate of the bound and of the true value almost coincide, i.e., the slow down factor is about 1.00.

Experiment 3. In the third experiment we investigate how well the exact invariant subspace $\mathcal{X}_1$ is included in a Ritz space $\tilde{\mathcal{Y}}$ using Theorem 4.1, i.e., we are interested in the quality of the computed subspace $\tilde{\mathcal{Y}}$. The subspace of interest $\mathcal{X}_1$ still corresponds to the eigenvalues $\Lambda_1 := \{9 - g, 7, 6, 5, 4\}$. We choose $\Lambda_2 := \Lambda_1 \cup \{9, 4 - g\}$, $\|E\|_2 = 10^{-13}\|A\|_2$ and $g = 10^{-11}$, i.e., the gap between the wanted and the remaining eigenvalues is very small. The approximate eigenpair $(\tilde{M}, \tilde{\mathcal{Y}})$ is obtained via calculating a Ritz pair of $A + E$.

In this experiment we again distinguish between two cases: In case (A) we consider the Ritz space corresponding to the 2nd to 6th largest Ritz values, i.e., the subspace is of the same dimension as $\mathcal{X}_1$ and the wanted eigenvalues Λ_1 are not well separated from the unselected Ritz values $\tilde{\mathcal{M}}_-$. This leads to a very slow convergence of the angle $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$. Referring to Figure 5 $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ does not start to converge until the 27th iteration step and decreases only to 10^{-3} afterwards. Moreover, the angle $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ wiggles around after the convergence phase, which is also caused by the insufficient separation of the wanted and remaining Ritz values. The bound (4.1) captures all three phases correctly, although it lags behind 5 iteration steps. After the convergence phase the bound overestimates the true value by a factor of about 1.5.

In case (B) the subspace $\tilde{\mathcal{Y}}$ corresponds to the 8 largest Ritz values. Now, there is a sufficiently large gap between the wanted eigenvalues Λ_1 und the unselected Ritz values $\tilde{\mathcal{M}}_-$. We see in Figure 5 that now the angle $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ starts to converge earlier after 22 iteration steps and decreases afterwards to limiting accuracy on the order of $\|E\|_2/\|A\|_2$. The bound (4.1) captures all three phases initialization, convergence and stagnation very well. During the convergence phase the convergence rates of the bound and the true value differ only by 0.005 and afterwards the bound overestimates the true value by a constant factor of about 35.

5.2. Ising model. Returning to the motivation of this paper we consider as a second example the one-dimensional quantum Ising model in the presence of a transverse field as defined in [25, 26]. The Hamiltonian of a system of d two-level

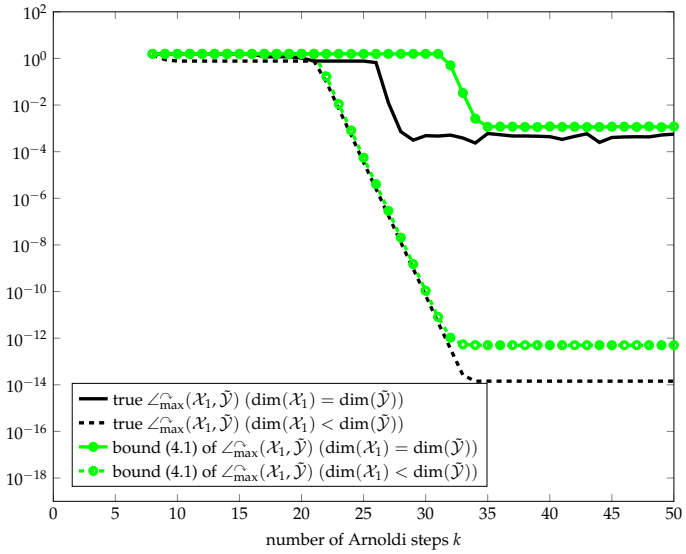


FIG. 5. *Experiment 3: True angles of inclusion* $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ *and corresponding bounds* (4.1)

subsystems (qubits) is defined by

$$A := (1-s)\sum_{i=1}^{d} A_i + s\sum_{i=1}^{d} B_i B_{i+1} \tag{5.2}$$

where $s \in [0,1]$ and for $i = 1, \ldots, d$

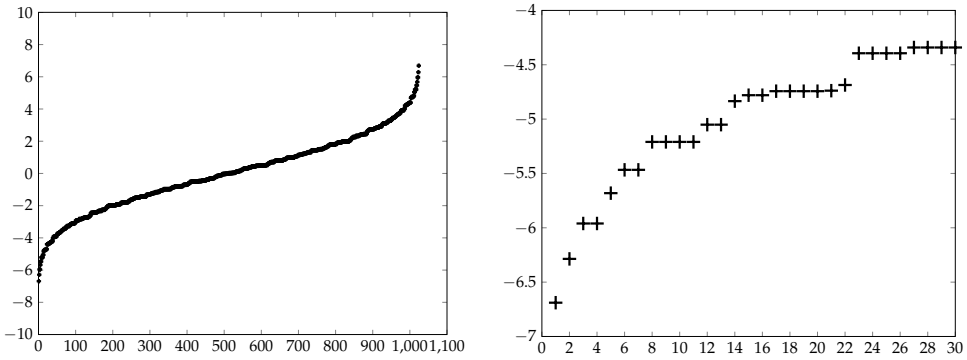
$$A_i := -I_{2^{(i-1)}} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes I_{2^{(d-i)}}, \qquad B_i := -I_{2^{(i-1)}} \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes I_{2^{(d-i)}}$$

and $B_{d+1} := B_1$. The dimension of the Hamiltonian A is $n = 2^d$ growing exponentially in the number of qubits. In the following experiments we choose $d = 10$, $n = 1024$ and $s = 0.4$. The spectrum for these values is depicted in Figure 6. We are interested in the ground state as well as in the first and second excited state, i.e., we want to approximate the subspace $\mathcal{X}_1$ corresponding to Λ_1 consisting of the three smallest eigenvalues of A. So $J_1 = \{1, 2, 3\}$.

In Figure 6(b), we see that there is no gap between the third smallest and the fourth smallest eigenvalue of A, i.e. Λ_1 is not well separated from the remaining eigenvalues.

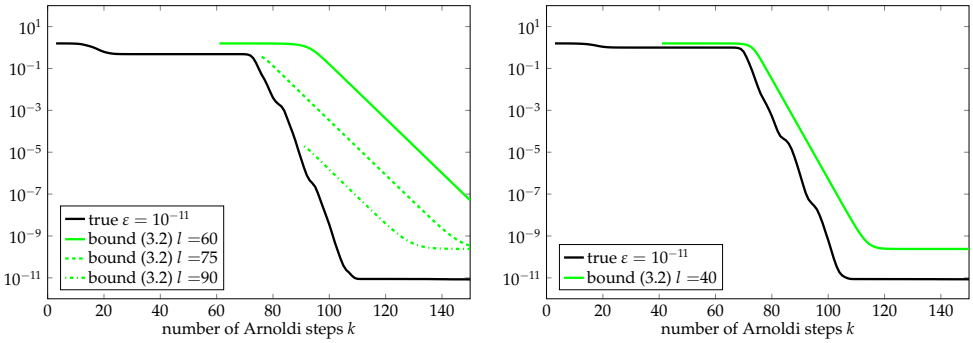
Experiment 4. In this experiment we illustrate the bound of Theorem 3.1. In the first case we choose $\Lambda_2 = \Lambda_3 = \Lambda_4$ to consist of the four smallest eigenvalues of A, i.e., $J_2 = J_3 = J_4 = \{1, \ldots, 4\}$ and $\|E\|_2 = 10^{-11}\|A\|_2$. In Figure 7(a) the true angle of inclusion $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ and the bound (3.2) for $l \in \{60, 75, 90\}$ are plotted.

Qualitatively we obtain the same result as in Experiment 2 (B). The bound (3.2) predicts all three phases initialization, convergence and stagnation of $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$. As for the first test matrix in Experiment 2 we see that the larger the number of previous iterations l, the sharper the corresponding bound in the first iteration steps. The convergence rate is overestimated to be 0.77 (true value: 0.55). This amounts a slow down factor of 2.31. After the convergence phase all three bounds overestimates the true value by a constant factor of approximately 30.



(a) all eigenvalues (b) the 30 smalles eigenvalues

FIG. 6. *Eigenvalue distribution of the Ising model* (5.2) *with* $d = 10$ *and* $s = 0.4$



(a) $\Lambda_1 \subset \Lambda_2 = \Lambda_3 = \Lambda_4$ and $l \in \{60, 75, 90\}$ (b) $\Lambda_1 \subset \Lambda_2 \subset \Lambda_3 \subset \Lambda_4$ and $l = 40$

FIG. 7. *Experiment 4: The angle of inclusion* $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{K}}_k)$ *and its bounds* (3.2)

In the second case we improve the bound by choosing $l = 40$ and $\Lambda_1 \subset \Lambda_2 \subset \Lambda_3 \subset \Lambda_4$, more precisely Λ_1, Λ_2, Λ_3, Λ_4 consists of the 3, 4, 7, 11 smallest eigenvalues of A, respectively. With this choice of the subsets the bound (3.2) is much sharper compared to the first case, see Figure 7(b). Also the convergence rate of the bound improved and the slow down factor is reduced to 1.31.

Experiment 5. In the last experiment we examine how well the exact subspace $\mathcal{X}_1$ of A is included in the computed Ritz space $\tilde{\mathcal{Y}}$, using bound (4.1). Again we distinguish between $\dim(\mathcal{X}_1) = \dim(\tilde{\mathcal{Y}})$ and $\dim(\mathcal{X}_1) < \dim(\tilde{\mathcal{Y}})$, where in the first case the Ritz space corresponds to the three smallest and in the second case to the four smallest Ritz values.

In the first case there is an insufficient separation between the wanted eigenvalues und the unwanted Ritz values. Figure 8 shows that for $\dim(\mathcal{X}_1) = \dim(\tilde{\mathcal{Y}})$ the angle $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ does not converge and starts to wiggle after 100 steps of the algorithm. This behavior is caused by the dense distribution of the eigenvalues of the matrix A. More precisely, since the third and the fourth eigenvalue of A is multiple the exact subspace corresponding to the three smallest eigenvalues can not be approximated by a Ritz space of the perturbed matrix $A + E$.

Since the third and the fourth eigenvalue of A is multiple in general the condition

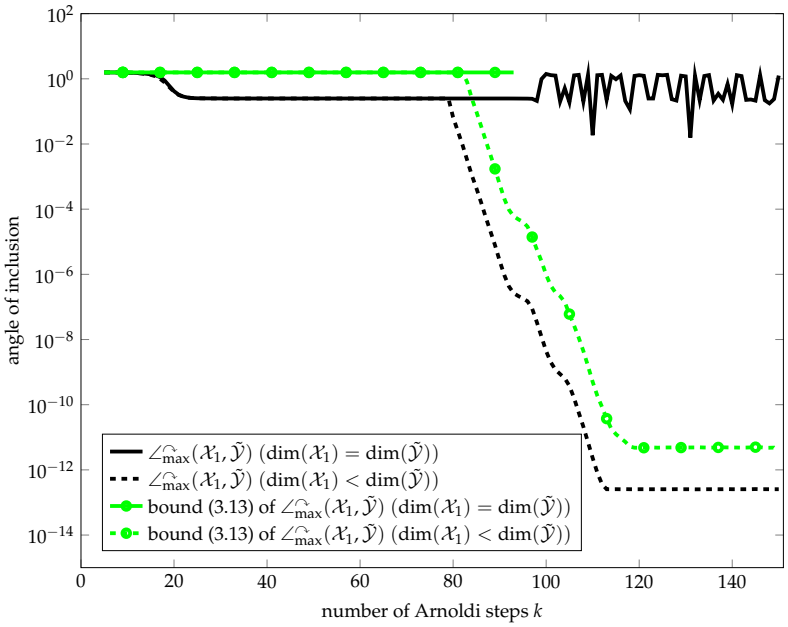


FIG. 8. *Experiment 5: True angles of inclusion* $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ *and corresponding bounds* (4.1)

$\|E\|_2 < \operatorname{gap}(\tilde{\mathcal{M}}_-, \Lambda_2)$ of bound (4.1) is violated. Only during the initial phase the bound (4.1) can be computed because there the approximation of Ritz values is still very inaccurate such that for a few iteration steps the condition on $\operatorname{gap}(\tilde{\mathcal{M}}_-, \Lambda_2)$ holds.

In the second case, the subspace $(\tilde{\mathcal{Y}})$ is slightly enlarged, which leads to an improvement of the convergence behavior of $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$. Referring to Figure 8 the angle $\angle^{\frown}_{\max}(\mathcal{X}_1, \tilde{\mathcal{Y}})$ starts to converge after 83 iteration. The bound (4.1) captures this behavior correctly and overestimates the true value after the convergence phase by a constant factor of 15.

In summary, also for the bound (4.1) we obtain qualitatively the same result as for the first test matrix.

6. Conclusions. We have investigated the convergence of inexact Krylov subspace methods, where we bound the distance of an exact invariant subspace to an inexact Krylov subspace and to an inexact Ritz space therein. The first bound addresses the question: If l iterations have been carried out without convergence, how many more iteration steps of an inexact Krylov subspace method are necessary to ensure convergence to a certain tolerance.

The second bound addresses the question: How much worse is the distance of an invariant subspace to a Ritz space compared to its distance to the whole search space? In particular, we have bounded the angle of inclusion of $\mathcal{X}$ in an inexact Ritz space $\tilde{\mathcal{Y}}$ in an a posteriori setting.

As special features our bounds can handle the presence of perturbations and a small gap between the wanted and the unwanted eigenvalues. This bound needs the information of the l-th step of the iteration, the 2-norm of the perturbation and the exact eigenvalues of A. Here the key idea was to enlarge the Ritz space in order to guarantee a sufficiently large gap between the remaining Ritz values and the wanted

eigenvalues. Finally, in the numerical tests the bounds have been confirmed to correctly predict the trends of the convergence curves.

Acknowledgments. We thank Volker Mehrmann, Michael Karow, and Agnieszka Miedlar (all TU Berlin) for insightful discussions on the topic and constructive comments on early versions of the manuscript.

REFERENCES

[1] A. C. Antoulas, *Approximation of large-scale dynamical systems*, vol. 6 of Advances in Design and Control, SIAM, Philadelphia, PA, 2005.
[2] C. A. Beattie, *Harmonic Ritz and Lehmann bounds*, Electron. Trans. Numer. Anal., 7 (1998), pp. 18–39. Large scale eigenvalue problems (Argonne, IL, 1997).
[3] C. A. Beattie, M. Embree, and D. C. Sorensen, *Convergence of polynomial restart Krylov methods for eigenvalue computations*, SIAM Rev., 47 (2005), pp. 492–515.
[4] R. Bhatia and P. Rosenthal, *How and why to solve the operator equation $AX - XB = Y$*, Bull. London Math. Soc., 29 (1997), pp. 1–21.
[5] Å. Björck, *Numerical methods for least squares problems*, SIAM, Philadelphia, 1996.
[6] S. Börm and C. Mehl, *Numerical Methods for Eigenvalue Problems*, De Gruyter Graduate Lectures, Berlin/Boston, 2012.
[7] C. Davis and W. M. Kahan, *The rotation of eigenvectors by a perturbation. III*, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.
[8] G. H. Golub and C. F. Van Loan, *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, third ed., 1996.
[9] L. Grasedyck, D. Kressner, and C. Tobler, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitt., 36 (2013), pp. 53–78.
[10] S. Holtz, T. Rohwedder, and R. Schneider, *On manifolds of tensors of fixed TT-rank*, Numer. Math., 120 (2012), pp. 701–731.
[11] T. Huckle, K. Waldherr, and T. Schulte-Herbrüggen, *Computations in quantum tensor networks*, Linear Algebra Appl., 438 (2013), pp. 750–781.
[12] I. C. F. Ipsen, *Relative perturbation results for matrix eigenvalues and singular values*, in Acta numerica, 1998, vol. 7 of Acta Numer., Cambridge Univ. Press, Cambridge, 1998, pp. 151–201.
[13] Z. Jia and G. W. Stewart, *An analysis of the Rayleigh-Ritz method for approximating eigenspaces*, Math. Comp., 70 (2001), pp. 637–647.
[14] U. Kandler and C. Schröder, *Backward error analysis of an inexact Arnoldi method using a certain Gram Schmidt variant*, Preprint 10-2013, TU Berlin, 2013.
[15] A. Kiełbasiński and H. Schwetlick, *Numerische lineare Algebra*, vol. 18 of Mathematik für Naturwissenschaft und Technik [Mathematics for Science and Technology], VEB Deutscher Verlag der Wissenschaften, Berlin, 1988. Eine computerorientierte Einführung. [A computer-oriented introduction].
[16] A. Knyazev, A. Jujunashvili, and M. Argentati, *Angles between infinite dimensional subspaces with applications to the Rayleigh-Ritz and alternating projectors methods*, J. Funct. Anal., 259 (2010), pp. 1323–1345.
[17] 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 (electronic).
[18] R. B. Lehoucq, *Implicitly restarted Arnoldi methods and subspace iteration*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 551–562 (electronic).
[19] R. B. Lehoucq and K. Meerbergen, *Using generalized Cayley transformations within an inexact rational Krylov sequence method*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 131–148 (electronic).
[20] Y. Nakatsukasa, *The tan θ theorem with relaxed conditions*, Linear Algebra Appl., 436 (2012), pp. 1528–1534.
[21] I. V. Oseledets, *Tensor-train decomposition*, SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317.
[22] B. N. Parlett, *The symmetric eigenvalue problem*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1980. Prentice-Hall Series in Computational Mathematics.
[23] Y. Saad, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706.
[24] ———, *Numerical methods for large eigenvalue problems*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, rev. ed. ed., 2011.

[25] G. Schaller, *Adiabatic preparation without quantum phase transitions*, Phys. Rev. A, 78 (2008), p. 032328.
[26] R. Schützhold and G Schaller, *Adiabatic quantum algorithms as quantum phase transitions: First versus second order*, Phys. Rev. A, 74 (2006), p. 060304.
[27] V. Simoncini, *Ritz and pseudo-Ritz values using matrix polynomials*, in Proceedings of the Fourth Conference of the International Linear Algebra Society (Rotterdam, 1994), vol. 241/243, 1996, pp. 787–801.
[28] ———, *Variable accuracy of matrix-vector products in projection methods for eigencomputation*, SIAM J. Numer. Anal., 43 (2005), pp. 1155–1174.
[29] V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477 (electronic).
[30] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–764.
[31] ———, *A generalization of Saad's theorem on Rayleigh-Ritz approximations*, Linear Algebra Appl., 327 (2001), pp. 115–119.
[32] ———, *Matrix algorithms. Vol. II*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Eigensystems.
[33] ———, *Backward error bounds for approximate Krylov subspaces*, Linear Algebra Appl., 340 (2002), pp. 81–86.
[34] G. W. Stewart and J.-G. Sun, *Matrix perturbation theory*, Computer Science and Scientific Computing, Academic Press Inc., Boston, MA, 1990.
[35] B. Sz.-Nagy, *Über die Ungleichung von H. Bohr*, Math. Nachr., 9 (1953), pp. 255–259.
[36] J. van den Eshof and G. L. G. Sleijpen, *Inexact Krylov subspace methods for linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 125–153 (electronic).
[37] P. Å. Wedin, *On angles between subspaces of a finite dimensional inner product space*, in Matrix Pencils, vol. 973 of Lecture Notes in Mathematics, Springer Berlin Heidelberg, 1983, pp. 263–285.
[38] H. Wielandt, *An extremum property of sums of eigenvalues.*, Proc. Am. Math. Soc., 6 (1955), pp. 106–110.