

# NEW *A PRIORI* FEM ERROR ESTIMATES FOR EIGENVALUES

ANDREW V. KNYAZEV \* AND JOHN E. OSBORN †

**Abstract.** We analyze the Ritz–Galerkin method for symmetric eigenvalue problems and prove *a priori* eigenvalue error estimates. For a simple eigenvalue, we prove an error estimate that depends mainly on the approximability of the corresponding eigenfunction and provide explicit values for all constants. For a multiple eigenvalue we prove, in addition, apparently the first truly *a priori* error estimates that show the levels of the eigenvalue errors depending on approximability of eigenfunctions in the corresponding eigenspace. These estimates reflect a known phenomenon that different eigenfunctions in the corresponding eigenspace may have different approximabilities, thus resulting in different levels of errors for the approximate eigenvalues. For clustered eigenvalues, we derive eigenvalue error bounds that do not depend on the width of the cluster. Our results are readily applicable to the classical Ritz method for compact symmetric integral operators and to finite element method eigenvalue approximation for symmetric positive definite differential operators.

**Key words.** eigenvalue problem, operator, invariant subspace, multiple eigenvalues, clustered eigenvalues, approximation, Ritz method, Ritz value, finite element method, *a priori* error estimates, angles between subspaces

**AMS(MOS) subject classifications.** 65F35

July 28, 2005

**1. Introduction.** We revisit the classical subject of *a priori* eigenvalue error estimates for the Ritz–Galerkin approximation of symmetric eigenvalue problems, with application to Finite Element Methods (FEM) eigenvalue approximation. *A priori* estimates have traditionally been used to prove the convergence of FEM eigenvalue approximation, and to determine the convergence rate when the mesh is refined. These estimates are typically based on the approximability of the eigenfunctions by the FEM subspace, and can be used to explain certain interesting features of eigenvalue approximation. For example, [1–4], why the third vibration mode of an L-shaped membrane is easier to approximate than the first two, and why two Ritz values approximating a double eigenvalue may converge at different rates.

The main result of the present paper — briefly stated — is that the eigenvalue errors depend mainly on just the approximability of the corresponding invariant subspaces, whether the eigenvalues are well-separated, multiple, or clustered. Our results differ from those in [2–4] in particular in that the information required by the new estimates is minimal and is covered by *explicitly given constants* that can be relatively easily obtained *a posteriori* from approximate eigenvalues and eigenfunctions. The question of whether our theorems give completely *computable eigenvalue bounds* thus is reduced to explicitly estimating the main factor, namely the approximability of invariant subspaces.

Computing the approximability is, however, difficult except in fairly trivial situations. One traditional approach is to use approximation theory results based on the smoothness of the eigenfunctions. Many results of this type are known, but usually the constants in the estimates are generic and not easily computable in practice. Assessing the smoothness of the eigenfunctions, meaning obtaining an estimate for

---

\*Department of Mathematics, University of Colorado at Denver. Electronic mail address: knyazev@na-net.ornl.gov. WWW home page URL: <http://www-math.cudenver.edu/~aknyazev>. Partially supported by the National Science Foundation award DMS 0208773.

†Department of Mathematics, University of Maryland. Electronic mail address: jeo@math.umd.edu. WWW home page URL: <http://www.wam.umd.edu/~jeo>. Partially supported by the National Science Foundation award DMS 0341982.

an appropriate higher Sobolev norm of the eigenfunction in question, can be done using an appropriate regularity theory for the underlying partial differential equation. Much is known about the regularity of the eigenfunctions, but, again, the constants are typically generic and cannot be easily estimated, with rather trivial exceptions. Nevertheless, *a priori* eigenvalue error analysis is a classical approach that has proved to be useful.

Early examples of *a priori* eigenvalue error estimates could be found, e.g., in [17]. Later, it became clear that the eigenvalue error was governed by the approximability of the exact eigenfunctions by the approximation space. In [5], Birkhoff, de Boor, Swartz, and Wendroff showed that the error for the  $j$ th eigenvalue was bounded by a constant times the sum of the norms squared of the approximation errors of the all eigenfunctions corresponding to the first  $j$  eigenvalues. In [22], Weinberger improved this result, showing that in the estimate for the relative eigenvalue error the constant simply equals to one; see Remark 2.1 for the exact formulation. Knyazev in [11], see also [8], further improved this result by replacing the norms of the approximation errors of individual eigenfunctions with the angle that measures the approximability of the invariant subspace spanned by these eigenfunctions. We reproduce this latter result by Knyazev in the present paper, in Theorem 2.4, and show that it is sharp.

The estimates of [5, 11, 22] suggest that the  $j$ th eigenvalue error depends on the approximability of all the eigenfunctions in the corresponding eigenspace, as well as of all the eigenfunctions corresponding to the previous eigenvalues. In reality, this is not the case. Numerical experiments for the  $L$ -shaped membrane eigenvalue problem show that the accuracy of approximation for the third eigenvalue is significantly better than for the first two. This can be explained as follows. We first note that the first two eigenfunctions of the  $L$ -shaped membrane eigenvalue problem are singular because of the reentrant corner, but the third eigenfunction is analytic because of symmetry, and hence easily approximated, especially by the p-method (see [1]). Secondly, Vainikko [16] and Chatelin [7] derive estimates of the eigenvalue error mainly in terms of just the approximability of the eigenfunctions in the corresponding eigenspace. Coupling this approximability result with this eigenvalue error estimate, we obtain the accurate eigenvalue approximation for the third eigenvalue.

Moreover, Vainikko [16] and Chatelin [7] show that the multiplicative constant in the estimate of the relative eigenvalue error approaches 1 under the approximability assumption on the family of the approximating spaces; see Section 3.2 for details. In [3], Babuška and Osborn determine that the closeness of the constant to 1 depends on the approximability of the operator of the original problem by the Ritz method; again, see Section 3.2.

Our first main results — Theorems 2.7 and 3.2 — clarify the estimate of [3] and improve the constant. All our constants are explicitly given, and no asymptotic assumptions are being made. In the FEM context, our results are readily applicable for a fixed mesh without making the traditional assumption that the mesh size is small enough.

When the eigenvalue of interest is of multiplicity  $q > 1$ , different eigenfunctions in the corresponding eigenspace may have different approximabilities, thus resulting in different levels of error for the approximate eigenvalues, i.e. the  $q$  Ritz values, corresponding to the multiple eigenvalue, may approach the eigenvalue with different rates. It is important to have eigenvalue error estimates that capture this phenomenon.

The error bounds of Vainikko [16] and Chatelin [7] effectively require approximability of all eigenfunctions in the corresponding eigenspace that provides an estimate

for the largest eigenvalue error only. In [2–4], Babuška and Osborn perform analysis that differentiate levels of eigenvalue error depending on approximability of different eigenfunctions in the eigenspace, but their estimates are not truly *a priori*, except for the estimate for the smallest eigenvalue error, which depends mainly on the approximability of the most easily approximated eigenfunction within the eigenspace.

Our results for multiple eigenvalues — Theorems 2.11 and 3.3 — clarify and improve these results of [2–4]. For example, if the eigenspace is spanned by three eigenfunctions of different approximation qualities, our results estimate the corresponding quality of each of the three Ritz values.

Error estimates for clustered eigenvalues are not well examined in the literature. The results presented in this paper are valid for clustered eigenvalues, as well as for multiple eigenvalues, and give error estimates that do not depend on the width of the cluster. Ovtchinnikov, in [19], independently derives similar estimates, which he calls “cluster robust.” Our estimates, compared to those of [19], are more compact and use less information.

In our proofs, we heavily use approximation error estimates for eigenspaces and invariant subspaces obtained by Knyazev in [13].

The paper intentionally contains some material that may be considered redundant, in order to improve readability. A critic once wrote about Beethoven’s Symphony No.2 in D major op.36 that it “would surely benefit from the abbreviation of some passages and the deletion of others.” If we are allowed to use musical terms in our defense and to compare our paper to a symphony, it consists of four movements:

The first, fast, movement is Subsections 2.1-2.5. Section 2.1 sets the stage of an abstract setting of a compact symmetric operator on a Hilbert space. We briefly introduce the angles instruments in the development Subsection 2.2 and then, in Subsection 2.3, the main theme, *a priori* estimates for eigenvalues. Subsection 2.4 is the most important in the first movement — it brings us the main theme in its most “ideal” form in Theorem 2.7, without a proof. Theorem 2.7 is an error estimate for a  $j$ -th eigenvalue mainly in terms of the approximation error of the corresponding eigenfunctions. In Subsection 2.5, the theme appears with slight variations for multiple and clustered eigenvalues. It becomes apparent that a major development needs to come.

The second, slow, movement is the massive Subsection 2.6. The main theme is significantly extended and generalized, with a complete vigorous proof, to carry a considerable improvement, in Theorem 2.11, for multiple and clustered eigenvalues. A number of possible variations surface at the end of the second movement.

The third, dance-like, movement is Subsection 3.1, which is a brief reminiscence of the first two movements. The same theme is essentially repeated, but in a different key, for the variational Galerkin method in a context applicable for FEM eigenvalue approximation for second order symmetric positive definite differential operators. Our last main results — Theorem 3.2 and Theorem 3.3 — appear in this subsection.

The fast Finale, Subsection 3.2, takes the material of the previous movement and contrasts it from earlier work. It opens in a relaxed manner and cites several well known results. In the closing, it reaches the climax by showing how to obtain differential levels of eigenvalue error depending on approximability of different eigenfunctions in the eigenspace.

It would, of course, be appealing to have practical numerical examples of our *a priori* analysis providing computable eigenvalue bounds, e.g., for the Laplacian in a polygonal domain. It is known that the eigenfunctions of the Laplacian are

smooth (analytic, in fact) inside the domain, but are generally singular in the corners. In certain cases, however, the eigenfunction is smooth, e.g., already discussed third eigenfunction of the Laplacian in the L-shaped domain is smooth. It is very interesting to try to use this kind of information to compute eigenvalue bounds. But due to the difficulties of computing the approximability of invariant subspaces, discussed in the first paragraphs of the Introduction, such a project lies beyond the scope of the paper. For some computational examples we refer to [3, 4].

Preliminary results of this paper were presented at the meeting State Of The Art In Finite Element Method at the City University of Hong Kong in 1998. An extended version of the paper was published as a technical report of the Center for Computational Mathematics at the CU-Denver [15].

## 2. Estimates for a compact symmetric operator.

**2.1. An abstract eigenvalue problem.** We consider in this section a compact symmetric positive definite operator  $T$  defined on a real separable Hilbert space  $H$ , with inner product  $(u, v)$  and norm  $\|u\| = \sqrt{(u, u)}$ . The spectral theory of such operators is well known; see e.g. [9]. The spectrum consists of nonzero eigenvalues of finite multiplicity, together with 0, which is in the continuous spectrum. The eigenvectors can be chosen to be orthonormal. We denote the eigenvalues and corresponding eigenvectors of  $T$  by  $\mu_1 \geq \mu_2 \geq \dots > 0$  and  $u_1, u_2, \dots$ , where  $(u_i, u_j) = \delta_{ij}$ . We are interested in approximating the eigenpairs  $(\mu_i, u_i)$  of  $T$  by the Ritz method. Given a finite dimensional subspace  $\tilde{U}$  of  $H$ , referred to as the trial subspace, the Ritz approximation to  $T$  is the operator  $\tilde{T} = (\tilde{Q}T)|_{\tilde{U}}$ , where  $\tilde{Q}$  is the orthogonal projector onto  $\tilde{U}$ . The operator  $\tilde{T}$  is symmetric positive definite. The eigenpairs of  $\tilde{T}$  are called the Ritz pairs of  $T$ ; we regard them as approximations of the eigenpairs of  $T$ . We denote the eigenvalues and corresponding eigenvectors of  $\tilde{T}$  by  $\tilde{\mu}_1 \geq \tilde{\mu}_2 \geq \dots \geq \tilde{\mu}_n > 0$ , where  $n = \dim \tilde{U}$ , and  $\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_n$ , where  $(\tilde{u}_i, \tilde{u}_j) = \delta_{ij}$ . The numbers  $\tilde{\mu}_i$  are called the Ritz values and the vectors  $\tilde{u}_i$  are called the Ritz vectors. In this paper we are specifically concerned with approximating the eigenvalues of  $T$  by Ritz values:  $\mu_i \approx \tilde{\mu}_i$ . It is an immediate consequence of the max-min characterization of eigenvalues that  $\tilde{\mu}_i \leq \mu_i$ ,  $i = 1, \dots, n$ .

We make the assumptions that operator  $T$  is positive definite and compact just to simplify the arguments. The majority of our results in this section can be easily modified to hold without these assumptions. The most important modification is a replacement of the ratios like  $(\mu_j - \tilde{\mu}_j)/\mu_j$  that appear in the left-hand sides of most of our estimates below with  $(\mu_j - \tilde{\mu}_j)/(\mu_j - \mu_{inf})$ , where  $\mu_{inf}$  is the algebraically smallest point of the spectrum of  $T$  (when  $T$  is positive definite, evidently  $\mu_{inf} = 0$ ). Such a modification makes our estimates invariant with respect to a scalar shift  $T - \alpha I$  in  $T$  for any real scalar  $\alpha$ .

**2.2. Principal angles between subspaces.** If  $M$  and  $N$  are nontrivial finite dimensional subspaces of  $H$ , we will quantify the approximability of  $M$  by  $N$  using the sine of the largest principal angle from  $M$  to  $N$ , which is defined by

$$(2.1) \quad \sin \angle \{M; N\} = \sup_{u \in M, \|u\|=1} \text{dist}(u, N) = \sup_{u \in M, \|u\|=1} \inf_{v \in N} \|u - v\|.$$

For nonzero vectors  $u$  and  $v$ , if  $M = \text{span}\{u\}$ , we write  $\sin \angle \{u; N\}$  for  $\sin \angle \{M; N\}$ ; and if  $M = \text{span}\{u\}$  and  $N = \text{span}\{v\}$ , we write  $\sin \angle \{u; v\}$  for  $\sin \angle \{M; N\}$ .

It is immediate that  $0 \leq \sin \angle \{M; N\} \leq 1$  and that  $\sin \angle \{M; N\} = 0$  if and only if  $M \subseteq N$ . If  $\dim M > \dim N$ , then  $\sin \angle \{M; N\} = 1$ . If  $\dim M = \dim N < \infty$ , then

$\sin \angle\{M; N\} = \sin \angle\{N; M\}$ . In the remainder of this paper, we will typically have  $\dim M \leq \dim N$ .

We will need the following simple observations; cf. Lemma 3.4 of [6].

LEMMA 2.1. *Let the subspace  $M$  be split into an orthogonal sum of subspaces  $M = M_1 \oplus M_2$ . Then, see [15],*

$$(2.2) \quad \sin^2 \angle\{M; N\} \leq \sin^2 \angle\{M_1; N\} + \sin^2 \angle\{M_2; N\}.$$

Applying (2.2) recursively, we immediately obtain

COROLLARY 2.2. *Let vectors  $\{u_i, i = 1, \dots, \dim M\}$  form an orthogonal basis for the subspace  $M$ . Then*

$$(2.3) \quad \sin^2 \angle\{M; N\} \leq \sum_i \sin^2 \angle\{u_i; N\}.$$

We call angle  $\angle\{M; N\}$  the largest since it is also well known, e.g., [14], that smaller angles between subspaces can be defined as follows. Using  $P$  and  $Q$ , the orthogonal projectors onto  $M$  and  $N$ , respectively, the sine of the largest angle equals to the largest singular value of the operator  $(I - Q)P$ . Introducing the notation  $s_1((I - Q)P) \geq s_2((I - Q)P) \geq \dots \geq s_{\dim M}((I - Q)P)$  for the  $\dim M$  largest singular values of the operator  $(I - Q)P$ , we define the  $i$ th angle from subspace  $M$  to subspace  $N$  using its sine:  $\sin \angle_i\{M; N\} = s_{\dim M - i + 1}((I - Q)P)$ ,  $i = 1, \dots, \dim M$ , assuming that all angles lie on the closed interval  $[0, \pi/2]$ . The complete set of  $\dim M$  angles from subspace  $M$  to subspace  $N$  gives detailed information of approximability of  $M$  by  $N$ , e.g., if the smallest angle vanishes, the subspaces  $M$  and  $N$  have a nontrivial intersection.

Later in the paper we use the following property of angles (see [14])

$$(2.4) \quad \angle_j\{M; N\} = \inf_{L \subseteq M, \dim L = j} \angle\{L; N\}, \quad j = 1, \dots, \dim M.$$

Finally, we will also need the following generalization of Corollary 2.2.

LEMMA 2.3. *Let vectors  $\{u_i, i = 1, \dots, \dim M\}$  form an orthogonal basis for the subspace  $M$  and be arranged in such a way that*

$$\angle\{u_1; N\} \leq \dots \leq \angle\{u_{\dim M}; N\}.$$

Then

$$(2.5) \quad \sin^2 \angle_j\{M; N\} \leq \sum_{i=1}^j \sin^2 \angle\{u_i; N\}, \quad j = 1, \dots, \dim M.$$

*Proof.* We deduce from (2.4) that

$$\sin^2 \angle_j\{M; N\} \leq \sin^2 \angle\{\text{span}\{u_1, \dots, u_j\}; N\}.$$

Now, the statement of the lemma, (2.5), immediately follows from (2.3) applied to  $M = \text{span}\{u_1, \dots, u_j\}$ .  $\square$

**2.3. Estimates based on the approximability of all previous eigenvectors.** Sharp eigenvalue error estimates are usually derived under the assumption that the eigenvector corresponding to the eigenvalue being estimated is well approximated by the trial subspace.

We derive an estimate for the error in approximating  $\mu_j$ , the  $j$ th eigenvalue of  $T$ , by  $\tilde{\mu}_j$ , the  $j$ th Ritz value of  $T$ , i.e., the  $j$ th eigenvalue of  $\tilde{T}$ . Let  $U_{1,\dots,j}$  denote the span of the eigenvectors  $u_1, \dots, u_j$ , and let  $P_{1,\dots,j}$  be the orthogonal projector onto  $U_{1,\dots,j}$ . For  $u \neq 0$ , let  $\mu(u) = (Tu, u)/(u, u) = (u, u)_T/(u, u)$  be the Rayleigh quotient associated with  $T$ . Here  $(\cdot, \cdot)_T$  is a second inner product on  $H$ . We will refer to orthogonality in  $(\cdot, \cdot)_T$  as  $T$ -orthogonality. Note that  $\mu(u) > 0$  since  $T$  is positive definite.

Our first theorem is known; it was proved in [11] and reproduced in [8]. For the particular case  $j = \dim \tilde{U}$ , a different proof was then suggested in [10, 12].

**THEOREM 2.4.** *For  $j = 1, 2, \dots, n = \dim \tilde{U}$  we have*

$$(2.6) \quad 0 \leq \frac{\mu_j - \tilde{\mu}_j}{\mu_j} \leq \sin^2 \angle \{U_{1,\dots,j}; \tilde{U}\} = \|(I - \tilde{Q})P_{1,\dots,j}\|^2.$$

The estimate (2.6) is sharp, see [15].

**REMARK 2.1.** *By Corollary 2.2 we have*

$$\sin^2 \angle \{U_{1,\dots,j}; \tilde{U}\} \leq \sum_{i=1}^j \sin^2 \angle \{u_i; \tilde{U}\} = \sum_{i=1}^j \|(I - \tilde{Q})u_i\|^2,$$

therefore, the estimate

$$(2.7) \quad \frac{\mu_j - \tilde{\mu}_j}{\mu_j} \leq \sum_{i=1}^j \|(I - \tilde{Q})u_i\|^2$$

follows directly from Theorem 2.4. Estimate (2.7) is well-known (see, e.g., [20, 22]); on the right-hand side we have the sum of the squares of the approximation errors for the eigenvectors  $u_1, \dots, u_j$ . If  $j = 1$ , the estimates (2.6) and (2.7) are identical.

**2.4. Estimates based mainly on the approximability of the target eigenvector.** Theorem 2.4 has a major weakness; namely, the right-hand side of estimate (2.6) for the target eigenvalue  $\mu_j$  involves the approximability of all functions in  $U_{1,\dots,j}$ . The result thus suggests that the eigenvalue error  $(\mu_j - \tilde{\mu}_j)/\mu_j$  depends on the approximation errors for all eigenfunctions  $u_1, \dots, u_j$ . We now mention two results suggesting that this is not the case; that, in fact, the ratio  $(\mu_j - \tilde{\mu}_j)/\mu_j$  depends mainly on just the approximation error for  $u_j$ , the target eigenfunction. First, consider the following

**LEMMA 2.5.** *For  $j = 1, 2, \dots, n = \dim \tilde{U}$ , the estimate*

$$(2.8) \quad \begin{aligned} \frac{\mu_j - \tilde{\mu}_j}{\mu_j} &= \|(I - \tilde{P}_j)u_j\|^2 - \frac{1}{\mu_j}((I - P_j)\tilde{u}_j, T(I - P_j)\tilde{u}_j) \\ &\leq \sin^2 \angle \{u_j, \tilde{u}_j\} \end{aligned}$$

holds, where  $\tilde{P}_j$  is the orthogonal projector onto  $\text{span}\{\tilde{u}_j\}$ .

The first line of (2.8) follows from the chain of identities in the proof of Lemma 3.5 of [6].

Next consider

LEMMA 2.6. *If  $(\tilde{u}_j, u_j) \neq 0$ , the estimate*

$$(2.9) \quad \begin{aligned} \frac{\mu_j - \tilde{\mu}_j}{\mu_j} &= \|(I - \tilde{Q})u_j\|^2 + \frac{1}{\mu_j} \left( T(I - \tilde{Q})u_j, \frac{(I - P_j)\tilde{u}_j}{\|P_j\tilde{u}_j\|} \right) \\ &\leq \left( 1 + \frac{\|(I - \tilde{Q})T\| \tan \angle\{u_j, \tilde{u}_j\}}{\mu_j \sin \angle\{u_j, \tilde{U}\}} \right) \sin^2 \angle\{u_j, \tilde{U}\} \end{aligned}$$

holds, where  $P_j$  is the orthogonal projector onto  $\text{span}\{u_j\}$ .

The identity in the first line of (2.9) is based on an argument from the proof of Theorem 4.1 in [3] (see also [18]). For a complete proof, see [15].

It is informative to compare (2.8) with (2.9). The first term on the right-hand side of the first line of (2.8) is larger than that of (2.9). However, the second term in the first line of (2.8) is negative, and thus is dropped in the second line of (2.8). The second term on the right-hand side in the first line of (2.9), while generally not negative, in typical applications (when  $\|(I - \tilde{Q})T\|$  is small) is significantly smaller compared to the first term, in other words, the term added to 1 in the second line of (2.9) in such applications is small because of the multiplier  $\|(I - \tilde{Q})T\|$ . We conclude that both (2.8) and (2.9) suggest that  $(\mu_j - \tilde{\mu}_j)/\mu_j$  depends mainly on the approximation error for  $u_j$ .

Both estimates (2.8) and (2.9), in addition to being dependent on the eigenfunction  $u_j$ , depend explicitly on the approximate eigenfunction  $\tilde{u}_j$ : (2.8) in the main term and (2.9) in the constant. Our next theorem is based on a novel alternative technique, where the approximate eigenfunction  $\tilde{u}_j$  is not used in the proof and does not appear in the theorem statement.

THEOREM 2.7. *For a fixed index  $j$  such that  $1 \leq j \leq n = \dim \tilde{U}$ , suppose that*

$$(2.10) \quad \min_{i=1, \dots, j-1} |\tilde{\mu}_i - \mu_j| \neq 0.$$

Then

$$(2.11) \quad \begin{aligned} 0 \leq \frac{\mu_j - \tilde{\mu}_j}{\mu_j} &\leq \|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-1})u_j\|^2 \\ &\leq \left( 1 + \frac{\|(I - \tilde{Q})T\tilde{P}_{1, \dots, j-1}\|^2}{\min_{i=1, \dots, j-1} |\tilde{\mu}_i - \mu_j|^2} \right) \sin^2 \angle\{u_j, \tilde{U}\}, \end{aligned}$$

where  $\tilde{P}_{1, \dots, j-1}$  is the orthogonal projector onto  $\tilde{U}_{1, \dots, j-1} = \text{span}\{\tilde{u}_1, \dots, \tilde{u}_{j-1}\}$  (if  $j = 1$ , we define  $\tilde{P}_{1, \dots, j-1} = 0$  and do not use (2.10)).

To satisfy the page limit, we do not prove the theorem here, instead referring to [15] and to the proof of the Theorem 2.11 later in the paper, which is a generalization of Theorem 2.7.

Since  $\|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-1})u_j\| \leq \|(I - \tilde{P}_j)u_j\|$ , our new estimate (2.11) clearly improves (2.8). A direct comparison of the constants in (2.9) and (2.11) in a general case does not appear to be simple because of the unresolved dependence of (2.9) on  $\tilde{u}_j$ . However, we have

$$\begin{aligned} \frac{\|(I - \tilde{Q})T\tilde{P}_{1, \dots, j-1}\|^2}{\min_{i=1, \dots, j-1} |\tilde{\mu}_i - \mu_j|^2} &\leq \frac{\|(I - \tilde{Q})T\|^2}{\min_{i=1, \dots, j-1} |\tilde{\mu}_i - \mu_j|^2} \\ &\leq \frac{\|(I - \tilde{Q})T\|}{\mu_j}, \end{aligned}$$

assuming

$$(2.12) \quad \|(I - \tilde{Q})T\| \leq \frac{\min_{i=1, \dots, j-1} |\tilde{\mu}_i - \mu_j|^2}{\mu_j}.$$

Since  $\tan \angle \{u_j, \tilde{u}_j\} \geq \sin \angle \{u_j, \tilde{U}\}$ , we can conclude that our estimate (2.11) is sharper than (2.9) under the assumption (2.12). We note that in the FEM context assumption (2.12) is realistic as for typical problems  $\|(I - \tilde{Q})T\|$  vanishes when the mesh parameter tends to zero.

Let us finally comment that the ratio

$$\frac{\|(I - \tilde{Q})T\tilde{P}_{1, \dots, j-1}\|^2}{\min_{i=1, \dots, j-1} |\tilde{\mu}_i - \mu_j|^2} = \frac{\|(I - \tilde{Q})(T/\mu_j)\tilde{P}_{1, \dots, j-1}\|^2}{\min_{i=1, \dots, j-1} |\tilde{\mu}_i/\mu_j - 1|^2}$$

in (2.11) is “dimensionless,” i.e. invariant with respect to scaling of  $T$ . Here, the quantity in the denominator,  $\min_{i=1, \dots, j-1} (\tilde{\mu}_i/\mu_j - 1)$ , in the limit, where all  $\tilde{\mu}_i \rightarrow \mu_i$ , turns into  $\mu_{j-1}/\mu_j - 1$ , the one-sided relative gap in the spectrum at  $\mu_j$ .

**2.5. Corollaries of Theorems 2.4 and 2.7 for multiple eigenvalues.** Here we address in details the case when the eigenvalue  $\mu_j$  is multiple of multiplicity  $q > 1$ . Our Theorems 2.4 and 2.7 hold for multiple eigenvalues since we never assumed the eigenvalues were simple. However, the case of multiple eigenvalues has special features, which we want to highlight. Let us start with the simplest case, where we are interested only in estimates for the largest eigenvalue  $\mu_1$ . From Theorem 2.4 we easily derive

COROLLARY 2.8. *Let*

$$\mu_1 = \mu_2 = \dots = \mu_q > \mu_{q+1}$$

and  $q \leq n = \dim \tilde{U}$ . For  $j = 1, 2, \dots, q$  we have

$$(2.13) \quad \begin{aligned} 0 \leq \frac{\mu_1 - \tilde{\mu}_j}{\mu_1} &\leq \inf_{\substack{U_{1, \dots, j} \subset U_{1, \dots, q} \\ \dim U_{1, \dots, j} = j}} \sin^2 \angle \{U_{1, \dots, j}; \tilde{U}\} \\ &= \sin^2 \angle_j \{U_{1, \dots, q}; \tilde{U}\}. \end{aligned}$$

Estimate (2.13) has two important properties. First, it controls the error for *every* Ritz values corresponding to the first eigenvalue  $\mu_1$ . Second, it shows that different Ritz values may have different approximation qualities, depending on approximability of the eigenspace  $U_{1, \dots, q}$  by the trial subspace  $\tilde{U}$  of the Ritz method, where the approximability is measured by the angles from  $U_{1, \dots, q}$  to  $\tilde{U}$  and, thus, can be estimated *a priori*.

In general, the multiple eigenvalue of interest may not be the largest:

$$(2.14) \quad \mu_{p-1} > \mu_p = \mu_{p+1} = \dots = \mu_j = \dots = \mu_{p+q-1} > \mu_{p+q}.$$

Applying Theorem 2.4, we obtain

COROLLARY 2.9. *Suppose (2.14) is satisfied and  $p + q - 1 \leq n$ . For any index  $j = p, p + 1, \dots, p + q - 1$  we have*

$$0 \leq \frac{\mu_p - \tilde{\mu}_j}{\mu_p} \leq \inf_{\substack{U_{1, \dots, p-1} \subset U_{1, \dots, j} \subset U_{1, \dots, p+q-1} \\ \dim U_{1, \dots, j} = j}} \sin^2 \angle \{U_{1, \dots, j}; \tilde{U}\}.$$

*Proof.* The subspace  $U_{1,\dots,j}$  has a fixed part  $U_{1,\dots,p-1} \subset U_{1,\dots,j}$ , but the rest of it we can choose within  $U_{p,\dots,\min\{p+q-1,n\}}$  as we like.  $\square$

Corollary 2.9 preserves the desired properties of Corollary 2.8, i.e. it provides a different estimate for each Ritz value of interest, but it requires approximability of all previous eigenvectors.

Let us now turn our attention to Theorem 2.7. The only relevant assumption in Theorem 2.7 is that (2.10) is satisfied so that the denominator in the constant in Theorem 2.7 is not zero. Let us analyze the likely behavior of this constant for the particular case  $q = 2$  so that

$$(2.15) \quad \mu_{p-1} > \mu_p = \mu_{p+1} > \mu_{p+2}.$$

There are two relevant possibilities for  $j$  in Theorem 2.7:  $j = p$  or  $j = p+1$ . Assuming that all Ritz values  $\tilde{\mu}_i$  approximate the corresponding eigenvalues  $\mu_i$ , which is typical for FEM applications (see Section 3.2 for details), we observe that in (2.10)

$$\min_{i=1,\dots,j-1} |\tilde{\mu}_i - \mu_j| \approx \mu_{j-1} - \mu_j.$$

Thus, if  $j = p$ , the denominator is asymptotically positive; specifically, it is asymptotically equal to  $\mu_{p-1} - \mu_p$ , and the estimate of Theorem 2.7 is asymptotically valid; while if  $j = p+1$ , the denominator in the constant in Theorem 2.7 asymptotically vanishes. This discussion demonstrates that Theorem 2.7 provides an asymptotically valid estimate only for one out of the  $q = 2$  Ritz values. On the positive side, however, we can freely choose the eigenvector  $u_j$  within the eigenspace corresponding to  $\mu_p$  to minimize the right hand side of (2.11). Let us reformulate Theorem 2.7 to reflect these observations.

**COROLLARY 2.10.** *Suppose that the eigenvalue  $\mu_p$ , where  $p > 1$ , has multiplicity  $q > 1$  so that (2.14) holds, and that  $p + q - 1 \leq n$ , and denote the corresponding eigenspace by  $U_{p,\dots,p+q-1}$ . As in Theorem 2.7, suppose that*

$$\min_{i=1,\dots,p-1} |\tilde{\mu}_i - \mu_p| \neq 0.$$

Then

$$(2.16) \quad \begin{aligned} 0 \leq \frac{\mu_p - \tilde{\mu}_p}{\mu_p} &\leq \min_{u \in U_{p,\dots,p+q-1}, \|u\|=1} \|(I - \tilde{Q} + \tilde{P}_{1,\dots,p-1})u\|^2 \\ &\leq \left( 1 + \frac{\|(I - \tilde{Q})T\tilde{P}_{1,\dots,p-1}\|^2}{\min_{i=1,\dots,p-1} |\tilde{\mu}_i - \mu_p|^2} \right) \min_{u \in U_{p,\dots,p+q-1}, \|u\|=1} \sin^2 \angle \{u; \tilde{U}\} \\ &= \left( 1 + \frac{\|(I - \tilde{Q})T\tilde{P}_{1,\dots,p-1}\|^2}{\min_{i=1,\dots,p-1} |\tilde{\mu}_i - \mu_p|^2} \right) \sin^2 \angle_1 \{U_{p,\dots,p+q-1}; \tilde{U}\}. \end{aligned}$$

*Proof.* We take  $j = p$  in Theorem 2.7 and notice that we can choose  $u_j$  to be any normalized vector in the eigenspace  $U_{p,\dots,p+q-1}$  and finally use (2.4).  $\square$

It is useful to compare Corollary 2.9 with Corollary 2.10. Corollary 2.9 gives different estimates for every Ritz value out of the  $q$  Ritz values corresponding to the multiple eigenvalue  $\mu_p$ , but requires approximability of all previous eigenvectors. In Corollary 2.10, the approximability of previous eigenvectors appears only in the constant, but it gives an estimate only for the largest Ritz value out of the  $q$ .

We want to obtain a result that combines the advantages of Corollary 2.9 and Corollary 2.10 and removes their weaknesses. E.g., if  $q = 3$  and the eigenspace corresponding to the triple eigenvalue  $\mu_p$  is spanned by eigenfunctions of different approximation quality, we want to have three error estimates for  $\mu_p$  reflecting it and not depending strongly on approximability of previous eigenfunctions.

### 2.6. A new estimate that covers multiple and clustered eigenvalues.

Our new result is a generalization of Theorem 2.7 that gives us the desired estimates for a multiple eigenvalue corresponding to an eigenspace spanned by eigenfunctions of different approximation quality. In addition, the new estimate also covers the case of clustered eigenvalues, i.e., the constant in the new estimate does not depend on the width of the eigenvalue cluster.

**THEOREM 2.11.** <sup>1</sup> *For fixed indexes  $j$  and  $m$  satisfying  $1 \leq j \leq n$  and  $1 \leq m \leq j$ , let  $U_{j-m+1, \dots, j}$  be the  $m$ -dimensional invariant subspace corresponding to eigenvalues  $\mu_{j-m+1} \geq \dots \geq \mu_j$  and  $P_{j-m+1, \dots, j}$  be the orthogonal projector on  $U_{j-m+1, \dots, j}$ . If*

$$(2.17) \quad \min_{i=1, \dots, j-m} |\tilde{\mu}_i - \mu_j| \neq 0,$$

then

$$(2.18) \quad \begin{aligned} 0 \leq \frac{\mu_j - \tilde{\mu}_j}{\mu_j} &\leq \|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-m})P_{j-m+1, \dots, j}\|^2 \\ &\leq \left(1 + \frac{\|(I - \tilde{Q})T\tilde{P}_{1, \dots, j-m}\|^2}{\min_{i=1, \dots, j-m} |\tilde{\mu}_i - \mu_j|^2}\right) \|(I - \tilde{Q})P_{j-m+1, \dots, j}\|^2, \end{aligned}$$

where  $\tilde{P}_{1, \dots, j-m}$  is the orthogonal projector onto  $\tilde{U}_{1, \dots, j-m} = \text{span}\{\tilde{u}_1, \dots, \tilde{u}_{j-m}\}$  (if  $j = m$  we set  $\tilde{P}_{1, \dots, j-m} = 0$  and do not use (2.17)). If  $m = j$ , the present theorem turns into Theorem 2.4; if  $m = 1$ , it turns into Theorem 2.7.

*Proof.* The operators  $I - \tilde{Q} + \tilde{P}_{1, \dots, j-m}$  and  $P_{j-m+1, \dots, j}$  are orthogonal projectors; thus,  $\|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-m})P_{j-m+1, \dots, j}\| \leq 1$ . If  $\|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-m})P_{j-m+1, \dots, j}\| = 1$ , the first estimate in (2.18) is trivially true. Now we suppose

$$(2.19) \quad \|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-m})P_{j-m+1, \dots, j}\| < 1.$$

Then, since  $\dim U_{j-m+1, \dots, j} = m$ , the subspace  $(\tilde{Q} - \tilde{P}_{1, \dots, j-m})U_{j-m+1, \dots, j}$  is also  $m$ -dimensional by Theorem 6.34 in Chapter I in [9].

We choose a normalized vector  $\bar{u}$  such that

$$\bar{u} \in (\tilde{Q} - \tilde{P}_{1, \dots, j-m})U_{j-m+1, \dots, j}, \quad \mu(\bar{u}) = \min_{w \in (\tilde{Q} - \tilde{P}_{1, \dots, j-m})U_{j-m+1, \dots, j} \setminus \{0\}} \mu(w),$$

and introduce the orthogonal and  $T$ -orthogonal decomposition

$$\bar{u} = u + v, \quad u \in U_{1, \dots, j}, \quad v \in U_{1, \dots, j}^\perp.$$

Since  $\bar{u} \in (\tilde{Q} - \tilde{P}_{1, \dots, j-m})U_{j-m+1, \dots, j}$ ,  $\|\bar{u}\| = 1$ , and  $u = \bar{u} - v$  is the orthogonal projection of  $\bar{u}$  onto  $U_{1, \dots, j}$ , we see using again Theorem 6.34 in Chapter I in [9] that

$$(2.20) \quad \begin{aligned} \|v\| &= \sin \angle \{\bar{u}; U_{1, \dots, j}\} \\ &\leq \sin \angle \{\bar{u}; U_{j-m+1, \dots, j}\} \\ &\leq \sin \angle \{(\tilde{Q} - \tilde{P}_{1, \dots, j-m})U_{j-m+1, \dots, j}; U_{j-m+1, \dots, j}\} \\ &= \|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-m})P_{j-m+1, \dots, j}\|. \end{aligned}$$

<sup>1</sup>It came to our attention that a similar result is independently obtained in the revised version of [19] accepted to Linear Algebra and Applications, 2005.

It now follows from (2.19) and (2.20) that  $\|v\| < 1$ ; thus,  $u \neq 0$  and  $\mu(u)$  is defined.

We next prove the following chain of inequalities:

$$(2.21) \quad \mu(\bar{u}) \leq \tilde{\mu}_j \leq \mu_j \leq \mu(u).$$

Indeed, the first inequality,

$$\begin{aligned} \mu(\bar{u}) &= \min_{w \in (\tilde{Q} - \tilde{P}_{1, \dots, j-m})U_{j-m+1, \dots, j} \setminus \{0\}} \mu(w) \leq \\ & \max_{\substack{W \subseteq \text{Im}(\tilde{Q} - \tilde{P}_{1, \dots, j-m}) \\ \dim W = m}} \min_{w \in W \setminus \{0\}} \mu(w) = \tilde{\mu}_j, \end{aligned}$$

follows from the min-max principle for Ritz values, since the dimension of the subspace  $(\tilde{Q} - \tilde{P}_{1, \dots, j-m})U_{j-m+1, \dots, j}$  is  $m$ . The second inequality,  $\tilde{\mu}_j \leq \mu_j$ , is an immediate consequence of the max-min principle. The third inequality,  $\mu_j \leq \mu(u)$ , follows from the fact that  $u \in U_{1, \dots, j}$ .

The identity

$$\mu(\bar{u}) = \frac{(Tu, u) + (Tv, v)}{(u, u) + (v, v)}$$

can be rewritten as

$$(2.22) \quad \mu(u) - \mu(\bar{u}) = \begin{cases} [\mu(\bar{u}) - \mu(v)] \frac{(v, v)}{(u, u)}, & v \neq 0 \\ 0, & v = 0 \end{cases}.$$

For  $v \neq 0$ , it follows directly from (2.21) and (2.22) that

$$\begin{aligned} 0 \leq \mu_j - \tilde{\mu}_j &\leq \mu(u) - \mu(\bar{u}) \\ &= [\mu(\bar{u}) - \mu(v)] \frac{(v, v)}{(u, u)} \\ &\leq \tilde{\mu}_j \frac{\|v\|^2}{\|u\|^2} \quad (\text{since } \mu(v) > 0); \end{aligned}$$

hence, since  $\|v\|^2 + \|u\|^2 = 1$  and  $(\mu_j - \tilde{\mu}_j)(1 - \|v\|^2) \leq \tilde{\mu}_j \|v\|^2$ , we get

$$(2.23) \quad 0 \leq \frac{\mu_j - \tilde{\mu}_j}{\mu_j} \leq \|v\|^2.$$

If  $v = 0$ , then from (2.22) we see that  $\mu(u) = \mu(\bar{u})$ , which, together with (2.21), shows that  $\tilde{\mu}_j = \mu_j$ . Thus, estimate (2.23) is also valid for  $v = 0$ .

Combining estimates (2.20) and (2.23), we obtain the first estimate in (2.18).

Finally, by Lemma 2.1,

$$\|(I - (\tilde{Q} - \tilde{P}_{1, \dots, j-m}))P_{j-m+1, \dots, j}\|^2 \leq \|(I - \tilde{Q})P_{j-m+1, \dots, j}\|^2 + \|\tilde{P}_{1, \dots, j-m}P_{j-m+1, \dots, j}\|^2.$$

The second term can be estimated using Theorem 3.2 of [13]:

$$\|\tilde{P}_{1, \dots, j-m}P_{j-m+1, \dots, j}\| \leq \frac{\|(I - \tilde{Q})T\tilde{P}_{1, \dots, j-m}\|}{\min_{i=1, \dots, j-m} |\tilde{\mu}_i - \mu_j|} \|(I - \tilde{Q})P_{j-m+1, \dots, j}\|.$$

Combining the first estimate in (2.18) with the last two inequalities completes the proof.  $\square$

Alternatively, Lemma 2.1 can be used to estimate  $\|\tilde{P}_{1,\dots,j-m}P_{j-m+1,\dots,j}\|$ , which results in

$$\begin{aligned}\|\tilde{P}_{1,\dots,j-1}P_{j-m+1,\dots,j}\|^2 &= \left\| \sum_{i=1}^{j-1} \tilde{P}_i P_{j-m+1,\dots,j} \right\|^2 \\ &\leq \sum_{i=1}^{j-1} \|\tilde{P}_i P_{j-m+1,\dots,j}\|^2.\end{aligned}$$

Every term  $\|\tilde{P}_i P_{j-m+1,\dots,j}\|^2$  in the sum above can be estimated using results of [13]. For simplicity, let  $m = 1$ , then by Theorem 2.1 in [13]

$$\|\tilde{P}_i P_j\| \leq \frac{\|T\tilde{u}_i - \tilde{\mu}_i \tilde{u}_i\|}{|\tilde{\mu}_i - \mu_j|} \sin \angle\{u_j; \tilde{U}\},$$

where  $\|\tilde{u}_i\| = \|u_j\| = 1$ , so we get

$$0 \leq \frac{\mu_j - \tilde{\mu}_j}{\mu_j} \leq \left( 1 + \sum_{i=1}^{j-1} \frac{\|T\tilde{u}_i - \tilde{\mu}_i \tilde{u}_i\|^2}{|\tilde{\mu}_i - \mu_j|^2} \right) \sin^2 \angle\{u_j; \tilde{U}\},$$

which in some cases may provide a smaller constant compared to that of (2.18) with  $m = 1$ .

REMARK 2.2. A careful examination of the proof of the first estimate in (2.18) of Theorem 2.11 shows that we can replace the orthoprojector  $P_{j-m+1,\dots,j}$  with an orthoprojector  $P_L$  to any  $m$ -dimensional subspace  $L$  of  $U_{1,\dots,j}$ : the argument still holds and the first estimate in (2.18) can be improved:

$$(2.24) \quad 0 \leq \frac{\mu_j - \tilde{\mu}_j}{\mu_j} \leq \inf_{L \subseteq U_{1,\dots,j}, \dim L = m} \|(I - \tilde{Q} + \tilde{P}_{1,\dots,j-m})P_L\|^2.$$

The right-hand side of (2.24) allows a nice geometric interpretation, using the definition (2.4) of the angles between subspaces:

$$\inf_{L \subseteq U_{1,\dots,j}, \dim L = m} \|(I - \tilde{Q} + \tilde{P}_{1,\dots,j-m})P_L\|^2 = \sin^2 \angle_m\{U_{1,\dots,j}; \tilde{U} \cap (\tilde{U}_{1,\dots,j-m})^\perp\}.$$

This may lead to a potential improvement of the second estimate (2.18) — provided one can estimate the right-hand side of (2.24) using terms similar to those of the second estimate in (2.18).

We can derive a simple estimate of the right-hand side of (2.24), using the fact, which follows from dimensionality arguments, that

$$\dim(\tilde{U}_{1,\dots,j-m})^\perp \cap U_{1,\dots,j} \geq m.$$

Since

$$\|(I - \tilde{Q} + \tilde{P}_{1,\dots,j-m})u\| = \|(I - \tilde{Q})u\|, \quad u \in (\tilde{U}_{1,\dots,j-m})^\perp \cap U_{1,\dots,j},$$

restricting the choice of  $L$  to the intersection above in (2.24) we derive that

$$(2.25) \quad 0 \leq \frac{\mu_j - \tilde{\mu}_j}{\mu_j} \leq \inf_{L \subseteq (\tilde{U}_{1,\dots,j-m})^\perp \cap U_{1,\dots,j}, \dim L = m} \|(I - \tilde{Q})P_L\|^2.$$

In FEM applications typically (because of the approximability assumption) we have  $\dim \left( (\tilde{U}_{1,\dots,j-m})^\perp \cap U_{1,\dots,j} \right) = m$  so the inf in (2.25) is then redundant.

Estimate (2.25) improves (2.6). We note that  $m$  is a free parameter in (2.25) and can be chosen arbitrarily,  $1 \leq m \leq j$ . We finally note that (2.25) is not truly an a priori estimate since the right-hand side of it depends on the Ritz vectors  $\tilde{u}_1, \dots, \tilde{u}_{j-m}$  that are not known a priori.

Let us now reformulate Theorem 2.11 in the context of the multiple eigenvalue in order to obtain a generalization of Corollary 2.10. Theorem 2.11 gives us enough flexibility to establish a different error estimate for every of  $q$  Ritz values corresponding to the multiple eigenvalue of multiplicity  $q$ :

COROLLARY 2.12. *Suppose that the eigenvalue  $\mu_p$ , where  $p > 1$ , has multiplicity  $q > 1$ , so that (2.14) holds, and that  $p + q - 1 \leq n$ . Suppose that*

$$\min_{i=1,\dots,p-1} |\tilde{\mu}_i - \mu_p| \neq 0.$$

Then, for  $j = p, \dots, p + q - 1$ , we have

$$(2.26) \quad \begin{aligned} 0 \leq \frac{\mu_p - \tilde{\mu}_j}{\mu_p} &\leq \|(I - \tilde{Q} + \tilde{P}_{1,\dots,p-1})P_{p,\dots,j}\|^2 \\ &\leq \left( 1 + \frac{\|(I - \tilde{Q})T\tilde{P}_{1,\dots,p-1}\|^2}{\min_{i=1,\dots,p-1} |\tilde{\mu}_i - \mu_p|^2} \right) \|(I - \tilde{Q})P_{p,\dots,j}\|^2, \end{aligned}$$

where  $\tilde{P}_{1,\dots,p-1}$  is the orthogonal projector onto  $\tilde{U}_{1,\dots,p-1} = \text{span}\{\tilde{u}_1, \dots, \tilde{u}_{p-1}\}$  and  $P_{p,\dots,j}$  is the orthogonal projector onto any  $j - p + 1$ -dimensional subspace of the eigenspace  $U_{p,\dots,p+q-1}$  corresponding to the eigenvalue  $\mu_p$ . The optimal choice of the projector  $P_{p,\dots,j}$  allows us to replace the term  $\|(I - \tilde{Q})P_{p,\dots,j}\|^2$  in estimate (2.26) with  $\sin^2 \angle_{j-p+1}\{U_{p,\dots,p+q-1}, \tilde{U}\}$ .

*Proof.* We simply take  $m = j - p + 1$  in Theorem 2.11.  $\square$

To see the improvement of Corollary 2.12 over Theorem 2.7, consider the following situation. Suppose  $\mu_2$  has multiplicity 2, so  $p = q = 2$ . Then

$$\min_{i=1,\dots,p-1} |\tilde{\mu}_i - \mu_p| \approx \mu_1 - \mu_2 > 0,$$

provided  $\tilde{\mu}_1$  is close enough to  $\mu_1$ . Taking  $j = 2$  in Corollary 2.12 yields

$$(2.27) \quad \frac{\mu_2 - \tilde{\mu}_2}{\mu_2} \lesssim \left( 1 + \frac{\|(I - \tilde{Q})T\tilde{P}_1\|^2}{(\mu_1 - \mu_2)^2} \right) \|(I - \tilde{Q})P_2\|^2;$$

while taking  $j = 3$  yields

$$(2.28) \quad \frac{\mu_3 - \tilde{\mu}_3}{\mu_3} = \frac{\mu_2 - \tilde{\mu}_3}{\mu_2} \lesssim \left( 1 + \frac{\|(I - \tilde{Q})T\tilde{P}_1\|^2}{(\mu_1 - \mu_2)^2} \right) \|(I - \tilde{Q})P_{2,3}\|^2.$$

In (2.27), the eigenvalues error is bounded by a constant that is slightly larger than 1 times the square of the best approximation error for  $u_2$ ; while in (2.28), we have the square of the best approximation error for  $\text{span}\{u_2, u_3\} =$  the eigenspace for  $\mu_2 = \mu_3$ . Note that estimating  $(\mu_3 - \tilde{\mu}_3)/\mu_3$  with Theorem 2.7 yields no asymptotically valid estimate (cf. the discussion preceding Corollary 2.10).

Results giving different estimates for  $(\mu_p - \tilde{\mu}_j)/\mu_p$ ,  $j = p, \dots, p+q-1$  (cf. Corollaries 2.9 and 2.12) were first proved in [2], see also [3, 4]. Our presentation simplifies and clarifies the analysis in [2, 3], and provides explicit constants. In Section 3.2 we compare these results in details. For an example of a multiple eigenvalue with eigenfunctions of differing approximabilities, see [2, 4].

Let us finally highlight the opportunities that Theorem 2.11 provides for error estimates of clustered eigenvalues in the following situation. Let

$$\mu_1 > \mu_2 \approx \mu_3 > \mu_4,$$

and suppose we are interested in error estimates for  $\mu_2$  and  $\mu_3$ , assuming that  $\tilde{\mu}_1 \approx \mu_1$  and  $\tilde{\mu}_2 \approx \mu_2$ . We do not even need Theorem 2.11 to estimate the error for  $\mu_2$ : Theorem 2.7 with  $j = 2$  already gives us an asymptotically valid estimate (2.27), and the fact that  $\mu_2$  is clustered (or multiple as above) is irrelevant. Theorem 2.7 with  $j = 3$  does not provide an asymptotically valid estimate for the error in  $\mu_3$  since the term  $|\mu_3 - \tilde{\mu}_2| \approx 0$  appears in the denominator.

Applying Theorem 2.11 with  $j = 3$  we have an option to choose the free parameter  $m = 1, 2$ , or 3. Taking  $m = 1$  reduces Theorem 2.11 to Theorem 2.7, which does not work well in this situation as we just discussed. Taking  $m = 2$  yields a good estimate

$$(2.29) \quad \frac{\mu_3 - \tilde{\mu}_3}{\mu_3} \lesssim \left( 1 + \frac{\|(I - \tilde{Q})T\tilde{P}_1\|^2}{(\mu_1 - \mu_3)^2} \right) \|(I - \tilde{Q})P_{2,3}\|^2.$$

Taking  $m = 3$  reduces Theorem 2.11 to Theorem 2.4,

$$(2.30) \quad \frac{\mu_3 - \tilde{\mu}_3}{\mu_3} \leq \|(I - \tilde{Q})P_{1,2,3}\|^2.$$

Comparing the right-hand sides of (2.29) and (2.30), we see that (2.29) provides a sharper estimate than (2.30) if  $\mu_1 - \mu_3$  is large enough and  $u_1$  cannot be well approximated by the trial subspace. To summarize, choosing different  $m$  in Theorem 2.11 allows us to reduce the constants in estimating errors for clustered eigenvalues at the cost of enlarging the invariant subspace that needs to be well approximated by the trial subspace. Note that in neither (2.29) nor (2.30) does the constant depend on the width of the eigenvalue cluster  $\mu_2 \approx \mu_3$ . Ovtchinnikov in [19] calls such estimates “cluster robust.”

**3. Application of our abstract results to the variational Galerkin method and comparisons.** We now consider the previous abstract results in two important contexts. First, suppose we have an eigenvalue problem for a symmetric positive compact integral operator  $T$  defined on  $H = L_2$  and apply the classical Ritz method for integral operators. All our results apply immediately and provide relative eigenvalue error estimates for the largest eigenvalues in terms of  $L_2$  approximability of the corresponding eigenfunctions.

Our second application is to the variational Galerkin method for symmetric positive definite differential operators. Here, we essentially need to reformulate our results for the inverse of  $T$ , but the operator  $T$  cannot be just simply replaced with its inverse since this would change the Ritz values. A proper inversion involves a simultaneous change of the scalar product as it is implicitly done in the next subsection. Our estimates of the previous section have to be somewhat rewritten in this context, since they are not invariant with respect to such an inversion.

**3.1. The variational Galerkin method.** Suppose, as above, that  $H$  is a real separable Hilbert space with inner product  $(u, v)$  and norm  $\|u\| = \sqrt{(u, u)}$ , and suppose we are given two symmetric bilinear forms  $B(u, v)$  and  $D(u, v)$  on  $H \times H$ . The bilinear form  $B(u, v)$  is assumed to satisfy

$$(3.1) \quad |B(u, v)| \leq C_1 \|u\| \|v\|, \text{ for all } u, v \in H$$

and

$$(3.2) \quad C_0 \|u\|^2 \leq B(u, u), \text{ for all } u \in H, \text{ with } C_0 > 0.$$

It follows from (3.1) and (3.2) that  $\|u\|_B = \sqrt{B(u, u)}$  and  $\|u\|$  are equivalent norms on  $H$ . For the remainder of this section we use  $B(u, v)$  and  $\|u\|_B$  as the inner product and norm, respectively, on  $H$ , and denote the resulting space by  $H_B$ . We also measure all angles in  $H_B$ , i.e. with respect to  $B(u, v)$ . Regarding  $D(u, v)$  we assume that  $0 < D(u, u)$ , for all nonzero vectors  $u \in H$  and that the unit ball of the norm  $\|\cdot\|_D$  is compact in  $H$ .

We consider the variationally formulated symmetric eigenvalue problem

$$(3.3) \quad \begin{cases} \text{Seek } \lambda \in R \text{ and } 0 \neq u \in H_B \text{ satisfying} \\ B(u, v) = \lambda D(u, v), \text{ for all } v \in H_B. \end{cases}$$

Under our assumptions, problem (3.3) has eigenvalues  $0 < \lambda_1 \leq \lambda_2 \leq \dots \nearrow +\infty$  and corresponding eigenvectors  $u_1, u_2, \dots$ , which satisfy  $B(u_i, u_j) = \lambda_i D(u_i, u_j) = \delta_{ij}$ .

We are interested in approximating the eigenpairs of (3.3) by the variational Ritz method. Toward this end, we suppose we are given a finite dimensional subspace  $\tilde{U}$  of  $H_B$ , and consider the finite dimensional, variationally formulated eigenvalue problem

$$(3.4) \quad \begin{cases} \text{Seek } \tilde{\lambda} \in R \text{ and } 0 \neq \tilde{u} \in \tilde{U} \text{ satisfying} \\ B(\tilde{u}, v) = \tilde{\lambda} D(\tilde{u}, v), \text{ for all } v \in \tilde{U}. \end{cases}$$

Problem (3.4), being a finite dimensional eigenvalue problem, has  $n = \dim \tilde{U}$  positive eigenvalues  $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \dots \leq \tilde{\lambda}_n$ , and corresponding eigenvectors  $\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_n$ , which satisfy  $B(\tilde{u}_i, \tilde{u}_j) = \tilde{\lambda}_i D(\tilde{u}_i, \tilde{u}_j) = \delta_{ij}$ ,  $i, j = 1, \dots, n$ . The Poincaré inequalities  $\lambda_i \leq \tilde{\lambda}_i$ ,  $i = 1, \dots, n$ , and the min-max characterization of eigenvalues of problems (3.3) and (3.4) hold under our assumptions. We then view  $\tilde{\lambda}_i$  as an approximation to  $\lambda_i$ , i.e.  $\lambda_i \approx \tilde{\lambda}_i$ ,  $i = 1, \dots, n$ .

Next we introduce the operator  $T : H_B \rightarrow H_B$  defined by

$$(3.5) \quad \begin{cases} Tf \in H_B \\ B(Tf, v) = D(f, v), \text{ for all } v \in H_B \end{cases}$$

and the operator  $\tilde{T} : \tilde{U} \rightarrow \tilde{U}$  defined by

$$(3.6) \quad \begin{cases} \tilde{T}f \in \tilde{U}, f \in \tilde{U}, \\ B(\tilde{T}f, v) = D(f, v), \text{ for all } v \in \tilde{U}. \end{cases}$$

The operator  $T$  is the solution operator for the ‘‘boundary value problem’’ corresponding to the eigenvalue problem (3.3). By our assumption, the unit ball of  $\|\cdot\|_D$  is compact in  $H$  and, therefore, in  $H_B$ , thus, the operator  $T$  is compact in  $H_B$ . Of course,  $\tilde{T}$ , being an operator on a finite dimensional space, is also compact. It follows directly from the definition (3.5) that  $T$  is symmetric and positive definite on  $H_B$  and

from the definition (3.6) that  $\tilde{T}$  is symmetric and positive definite on  $\tilde{U}$  (with respect to  $B(u, v)$ ). It is easily seen that, if, as above,  $\tilde{Q}$  is the orthogonal projector of  $H_B$  onto  $\tilde{U}$ , then  $\tilde{T} = (\tilde{Q}T)|_{\tilde{U}}$ .

The eigenvalues of problem (3.3) and of the operator  $T$  are reciprocals:  $\lambda_i = 1/\mu_i$ ,  $i = 1, 2, \dots$ ; problem (3.3) and the operator  $T$  have the same eigenvectors  $u_i$ . Likewise, the eigenvalues of problem (3.4) and of the operator  $\tilde{T}$  are reciprocals:  $\tilde{\lambda}_i = 1/\tilde{\mu}_i$ ,  $i = 1, 2, \dots, n$ ; problem (3.4) and the operator  $\tilde{T}$  have the same eigenvectors  $\tilde{u}_i$ . As in the previous section, we choose  $\{u_i\}$  and  $\{\tilde{u}_i\}$  to be orthonormal systems, in the context of the present section, that is in  $H_B$ .

The FEM approximation of eigenvalue problems for symmetric differential operators can be viewed as a variational Ritz method; and the FEM eigenvalue errors can be estimated using the theorems of the previous section.

We can utilize Theorems 2.4 and 2.7, applied to  $T$  and  $\tilde{T}$  on  $H_B$ , to estimate the eigenvalue error  $(\tilde{\lambda}_i - \lambda_i)/\tilde{\lambda}_i$ . Here  $U_{1,\dots,j}$  denotes the span of the eigenvectors  $u_1, \dots, u_j$  and  $P_{1,\dots,j}$  is the  $H_B$  orthogonal projector onto  $U_{1,\dots,j}$ .

**THEOREM 3.1.** *For  $j = 1, \dots, n = \dim \tilde{U}$  we have*

$$(3.7) \quad 0 \leq \frac{\tilde{\lambda}_j - \lambda_j}{\tilde{\lambda}_j} \leq \sin^2 \angle_B \{U_{1,\dots,j}; \tilde{U}\} = \|(I - \tilde{Q})P_{1,\dots,j}\|_B^2.$$

**REMARK 3.1.** *By analogy with Remark 2.1, from Theorem 3.1 we get the following estimate, mathematically equivalent to estimate (2.7):*

$$0 \leq \frac{\tilde{\lambda}_j - \lambda_j}{\tilde{\lambda}_j} \leq \sum_{i=1}^j \|(I - \tilde{Q})u_i\|_B^2,$$

which can be rewritten as

$$(3.8) \quad 0 \leq \frac{\tilde{\lambda}_j - \lambda_j}{\lambda_j} \leq \frac{\sum_{i=1}^j \|(I - \tilde{Q})u_i\|_B^2}{1 - \sum_{i=1}^j \|(I - \tilde{Q})u_i\|_B^2},$$

assuming that the denominator in the latter expression is positive. Estimate (3.8) is well-known (see, e.g., Theorem 2.1, Chapter 4 of [22]); a similar estimate is proved in [5].

To formulate the next theorem — an analog of Theorem 2.7 — we recall that  $\tilde{P}_{1,\dots,j-1}$  is the orthogonal projector of  $H_B$  onto  $\tilde{U}_{1,\dots,j-1} = \text{span}\{\tilde{u}_1, \dots, \tilde{u}_{j-1}\}$ , where  $\tilde{u}_i$  are eigenvectors of (3.4).

**THEOREM 3.2.** *For a fixed index  $j$  such that  $1 \leq j \leq n = \dim \tilde{U}$ , suppose*

$$(3.9) \quad \min_{1,\dots,j-1} |\tilde{\lambda}_i - \lambda_j| \neq 0.$$

Then

$$(3.10) \quad \begin{aligned} 0 \leq \frac{\tilde{\lambda}_j - \lambda_j}{\tilde{\lambda}_j} &\leq \|(I - \tilde{Q} + \tilde{P}_{1,\dots,j-1})u_j\|_B^2 \\ &\leq \left(1 + \max_{i=1,\dots,j-1} \frac{\tilde{\lambda}_i^2 \lambda_j^2}{|\tilde{\lambda}_i - \lambda_j|^2} \|(I - \tilde{Q})T\tilde{P}_{1,\dots,j-1}\|_B^2\right) \sin^2 \angle_B \{u_j; \tilde{U}\}. \end{aligned}$$

Similarly, we can apply Theorem 2.11 to obtain

**THEOREM 3.3.** *For fixed indexes  $j$  and  $m$  satisfying  $1 \leq j \leq n$  and  $1 \leq m \leq j$ , let  $U_{j-m+1, \dots, j}$  be the  $m$ -dimensional invariant subspace corresponding to eigenvalues  $\lambda_j \geq \dots \geq \lambda_{j-m+1}$  and  $P_{j-m+1, \dots, j}$  be the  $H_B$  orthogonal projector on  $U_{j-m+1, \dots, j}$ . If*

$$(3.11) \quad \min_{i=1, \dots, j-m} |\tilde{\lambda}_i - \lambda_j| \neq 0,$$

then

$$(3.12) \quad \begin{aligned} 0 \leq \frac{\tilde{\lambda}_j - \lambda_j}{\tilde{\lambda}_j} &\leq \|(I - \tilde{Q} + \tilde{P}_{1, \dots, j-m})P_{j-m+1, \dots, j}\|_B^2 \\ &\leq \left(1 + \max_{i=1, \dots, j-m} \frac{\tilde{\lambda}_i^2 \lambda_j^2}{|\tilde{\lambda}_i - \lambda_j|^2} \|(I - \tilde{Q})T\tilde{P}_{1, \dots, j-m}\|_B^2\right) \|(I - \tilde{Q})P_{j-m+1, \dots, j}\|_B^2, \end{aligned}$$

where  $\tilde{P}_{1, \dots, j-m}$  is the  $H_B$  orthogonal projector onto  $\tilde{U}_{1, \dots, j-m} = \text{span}\{\tilde{u}_1, \dots, \tilde{u}_{j-m}\}$  (if  $j = m$  we set  $\tilde{P}_{1, \dots, j-m} = 0$  and do not use (3.11)). If  $m = j$ , the present theorem turns into Theorem 3.1; if  $m = 1$ , it turns into Theorem 3.2.

Let us finally reformulate Theorem 3.3 in the context of the multiple eigenvalue by analogy with Corollary 2.12.

**COROLLARY 3.4.** *Suppose that the eigenvalue  $\lambda_p$ , where  $p > 1$ , has multiplicity  $q > 1$ , so that*

$$(3.13) \quad \lambda_{p-1} < \lambda_p = \lambda_{p+1} = \dots = \lambda_{p+q-1} < \lambda_{p+q}$$

holds, and that  $p + q - 1 \leq n$ . Suppose that

$$\min_{i=1, \dots, p-1} |\tilde{\lambda}_i - \lambda_p| \neq 0.$$

Then, for  $j = p, \dots, p + q - 1$ , we have

$$(3.14) \quad \begin{aligned} 0 \leq \frac{\tilde{\lambda}_j - \lambda_p}{\tilde{\lambda}_j} &\leq \|(I - \tilde{Q} + \tilde{P}_{1, \dots, p-1})P_{p, \dots, j}\|_B^2 \\ &\leq \left(1 + \max_{i=1, \dots, p-1} \frac{\tilde{\lambda}_i^2 \lambda_p^2}{|\tilde{\lambda}_i - \lambda_p|^2} \|(I - \tilde{Q})T\tilde{P}_{1, \dots, p-1}\|_B^2\right) \|(I - \tilde{Q})P_{p, \dots, j}\|_B^2, \end{aligned}$$

where  $\tilde{P}_{1, \dots, p-1}$  is the  $H_B$  orthogonal projector onto  $\tilde{U}_{1, \dots, p-1} = \text{span}\{\tilde{u}_1, \dots, \tilde{u}_{p-1}\}$  and  $P_{p, \dots, j}$  is the  $H_B$  orthogonal projector onto any  $j - p + 1$ -dimensional subspace of the eigenspace  $U_{p, \dots, p+q-1}$  corresponding to the eigenvalue  $\lambda_p$ . The main term, the multiplier  $\|(I - \tilde{Q})P_{p, \dots, j}\|_B^2$ , in (3.14) can be replaced with  $\sin^2 \angle_{j-p+1}\{U_{p, \dots, p+q-1}, \tilde{U}\}$  by choosing the projector  $P_{p, \dots, j}$  in the optimal way, where the angle is measured in  $H_B$ .

**3.2. Comparison with known asymptotic estimates for eigenvalues.** Estimate (3.10) should be compared with estimates of Vainikko [16], Chatelin [7], and Babuška and Osborn [3], which address a slightly different context that we now describe.

In addition to all assumptions of the previous subsection, let  $\{U^h\}$  be a family of finite dimensional subspaces of  $H_B$ , depending on a parameter  $h > 0$  called the mesh parameter. For a fixed  $h$ , we use  $U^h = \tilde{U}$  as the trial subspace for the variational Ritz

method. Let  $Q^h = \tilde{Q}$  be the  $H_B$  orthogonal projector on  $U^h$ . We make the following approximability assumption on the family  $\{U^h\}$ :

$$(3.15) \quad \|(I - Q^h)u\|_{H_B} = \inf_{v^h \in U^h} \|u - v^h\|_{H_B} \rightarrow 0 \text{ as } h \rightarrow 0, \text{ for each } u \in H_B.$$

To be consistent with our new  $h$ -based notation, we denote the approximate eigenvalues by  $\lambda_j^h = \tilde{\lambda}_j$  and the corresponding eigenvectors by  $u_j^h = \tilde{u}_j$ . It is well known that under assumption (3.15) we have  $\lambda_j^h \rightarrow \lambda_j$  as  $h \rightarrow 0$  for each fixed  $j$ .

We compare our results to estimates of [3, 7, 16] that are asymptotic,  $h \rightarrow 0$ , upper (and lower) bounds for the ratio  $(\lambda_j^h - \lambda_j)/\lambda_j^h$  in [7, 16] and for the ratio  $(\lambda_j^h - \lambda_j)/\lambda_j$  (notice a slightly different denominator) in [3]. Since

$$\frac{\lambda_j^h - \lambda_j}{\lambda_j} = \frac{\lambda_j^h - \lambda_j}{\lambda_j^h} + \frac{(\lambda_j^h - \lambda_j)^2}{\lambda_j \lambda_j^h},$$

where the second term in the sum on the right can be asymptotically ignored, the results of [7, 16] asymptotically estimate the same eigenvalue error as those of [3]. Results of [3] provide upper bounds for  $(\lambda_j^h - \lambda_j)/\lambda_j$  that trivially also serve as upper bounds for  $(\lambda_j^h - \lambda_j)/\lambda_j^h$ . Moreover, it is possible to show that the lower bounds for  $(\lambda_j^h - \lambda_j)/\lambda_j$  in [3] also hold for  $(\lambda_j^h - \lambda_j)/\lambda_j^h$  without any changes. Here, we will formulate all the results (except for (3.23)) in terms of  $(\lambda_j^h - \lambda_j)/\lambda_j^h$  to be consistent with our estimates.

We start our discussion with the case of a simple eigenvalue  $\lambda_j$  and later turn our attention to the case of multiple eigenvalues. The convergence rate for a simple eigenvalue is determined by the following well known result: let real  $r_j^h$  be defined by

$$(3.16) \quad 0 \leq \frac{\lambda_j^h - \lambda_j}{\lambda_j^h} = (1 + r_j^h) \|(I - Q^h)u_j\|_B^2;$$

then  $r_j^h \rightarrow 0$  as  $h \rightarrow 0$ ; see Subsection 18.6 (pp. 285–286) of [16] and Subsection 6.2 (pp. 315–317) of [7]. Babuška and Osborn [3] showed that

$$(3.17) \quad |r_j^h| \leq d_j \sup_{\|g\|_D=1} \|(I - Q^h)Tg\|_B^2 \rightarrow 0$$

and that, cf. (2.9),

$$(3.18) \quad |r_j^h| \leq d_j \sup_{\|g\|_B=1} \|(I - Q^h)Tg\|_B \rightarrow 0,$$

where  $d_j > 0$  are unknown generic constants.

Our present estimate (3.10) using the  $h$  notation takes the form (3.16) with

$$(3.19) \quad r_j^h \leq \max_{i=1, \dots, j-1} \frac{(\lambda_i^h)^2 \lambda_j^2}{|\lambda_i^h - \lambda_j|^2} \|(I - Q^h)TP_{1, \dots, j-1}^h\|_B^2.$$

The first multiplier in the right hand side of (3.19) is asymptotically (as  $h \rightarrow 0$ ) a constant,

$$\frac{(\lambda_i^h)^2 \lambda_j^2}{|\lambda_i^h - \lambda_j|^2} \rightarrow \frac{\lambda_{j-1}^2 \lambda_j^2}{|\lambda_{j-1} - \lambda_j|^2},$$

provided that the eigenvalue  $\lambda_j$  is simple. The second multiplier is bounded by

$$\begin{aligned} \|(I - Q^h)TP_{1,\dots,j-1}^h\|_B^2 &\leq \|(I - Q^h)T\|_B^2 \\ &= \sup_{\|g\|_B=1} \|(I - Q^h)Tg\|_B^2 \\ &\leq \frac{1}{\lambda_1} \sup_{\|g\|_D=1} \|(I - Q^h)Tg\|_B^2; \end{aligned}$$

thus, our estimate (3.19) is an improvement of both estimates (3.17) (our constant is explicitly written) and (3.18) (we have the small multiplier squared) of [3]. However, our estimate (3.19) provides only an upper bound for  $r_j^h$ , while (3.17) and (3.18) also give the lower bounds because they estimate the absolute value  $|r_j^h|$ . Let us note that the denominator  $|\lambda_{j-1} - \lambda_j|^2$  may be small, but the term in the numerator is bounded from above by a constant times  $\sup_{\|g\|_D=1} \|(I - Q^h)Tg\|_B^2 \rightarrow 0$  as  $h \rightarrow 0$ .

Now suppose eigenvalue  $\lambda_p$  has multiplicity  $q$ , so that (3.13) holds, and let  $P_{p,\dots,p+q-1}$  be the  $H_B$  orthogonal projector on the  $q$ -dimensional eigenspace, corresponding to  $\lambda_p = \lambda_{p+1} = \dots = \lambda_{p+q-1}$  as in Corollary 3.4. Vainikko in Subsection 18.6 (pp. 285–286) of [16] and Chatelin in Subsection 6.2 (pp. 315–317) [7] prove that

$$(3.20) \quad 0 \leq \frac{\lambda_j^h - \lambda_p}{\lambda_j^h} = (1 + r_j^h) \frac{\|(I - Q^h)P_{p,\dots,p+q-1}u_j^h\|_B^2}{\|P_{p,\dots,p+q-1}u_j^h\|_B^2}, \quad j = p, \dots, p + q - 1,$$

where  $r_j^h \rightarrow 0$  as  $h \rightarrow 0$ . An evident difficulty in using estimate (3.20) for *a priori* error analysis is that the approximate eigenfunctions  $u_{j+i-1}^h$  are not known *a priori*. If we consider the worst case, it leads to the estimate, which bounds the error for all  $q$  Ritz values, using

$$(3.21) \quad \|(I - Q^h)P_{p,\dots,p+q-1}\|_B^2 = \sin^2 \angle \{U_{p,\dots,p+q-1}; U_h\}.$$

Let us remind the reader that an angle without an index denotes the largest angle, according to our agreement in Subsection 2.2, and that in this and the previous subsections all angles are measured in  $H_B$ .

In some cases, see [2, 4] for an example, the eigenspace may be spanned by eigenfunctions of different approximation qualities, and it is interesting to analyze how this affects the error for different Ritz values. As mentioned in the Introduction, such results were first proved by Babuška and Osborn in [2]. In [3], they completed such analysis for the smallest of the  $q$  Ritz values, using

$$(3.22) \quad \inf_{u \in U_{p,\dots,p+q-1}, \|u\|_B=1} \|(I - Q^h)u\|_B^2 = \sin^2 \angle_1 \{U_{p,\dots,p+q-1}; U_h\},$$

which depends on the approximability of the most easily approximated eigenfunction in the eigenspace. Thus, estimates based on (3.21) and (3.22) represent two extremes: (3.21) uses the largest angle and serves to estimate the largest error (thus effectively all  $q$  errors at once), while (3.22) uses the smallest angle and can estimate only one, the smallest, eigenvalue error.

For the intermediate multiple eigenvalue error, Babuška and Osborn in [3] established the following result: let for  $j = p, \dots, p + q - 1$  the quantities  $r_j^h$  be redefined

by

$$(3.23) \quad 0 \leq \frac{\lambda_j^h - \lambda_p}{\lambda_p} = (1 + r_p^h) \inf_{\substack{u \in U_{p, \dots, p+q-1}, \\ u \in (U_{p, \dots, j-1}^h)^{\perp_B}, \\ \|u\|_B = 1}} \|(I - Q^h)u\|_B^2;$$

then

$$(3.24) \quad |r_j^h| \leq d_j \sup_{\|g\|_B=1} \|(I - Q^h)Tg\|_B,$$

with generic constants  $d_j > 0$ , and where the orthogonal complement  $(U_{p, \dots, j-1}^h)^{\perp_B}$  is taken in  $H_B$ . In [4], it is shown that  $r_j^h$  in (3.23) are bounded, but more detailed estimates (3.24) appear only in [3].

We note that the constraints on  $u$  in (3.23) are similar to those in (2.25) except that (2.25) involves orthogonalization to all previous Ritz vectors, while (3.23) only needs orthogonalization to previous Ritz vectors corresponding to the multiple eigenvalue under the consideration. Both (2.25) and (3.23) are not truly *a priori* estimates since their right-hand sides depends on Ritz vectors that are not known *a priori*.

In contrast, our estimate (3.14) can be formulated as follows: let for  $j = p, \dots, p+q-1$  the quantities  $r_j^h$  be yet again redefined by

$$(3.25) \quad 0 \leq \frac{\lambda_j^h - \lambda_p}{\lambda_j^h} = (1 + r_j^h) \sin^2 \angle_{j-p+1} \{U_{p, \dots, p+q-1}, U^h\};$$

then

$$r_j^h \leq \max_{i=1, \dots, p-1} \frac{(\lambda_i^h)^2 \lambda_p^2}{|\lambda_i^h - \lambda_p|^2} \|(I - Q^h)TP_{1, \dots, p-1}^h\|_B^2.$$

We have already shown that our upper bound for  $r_p^h$  is better than that given by estimate (3.24): the constant is explicitly written and the  $h$ -dependent part is smaller. Let us turn our attention to the main term of the right-hand side of (3.25), namely, the  $\sin^2 \angle_{j-p+1} \{U_{p, \dots, p+q-1}, U^h\}$  multiplier, and demonstrate that it is smaller than the main term of the right-hand side of (3.23) and that it can be easily estimated from above using (2.5).

We first highlight again that this multiplier can be estimated *a priori* since it does not depend on Ritz vectors, contrary to main term of the estimate (3.23). Second, we can directly compare the main terms in (3.23) and (3.25). Indeed, by (2.4), and since  $\dim\{(U_{p, \dots, j-1}^h)^{\perp} \cap U_{p, \dots, p+q-1}\} \geq j - p + 1$ , we have for  $j = p, \dots, p+q-1$ :

$$\begin{aligned} \sin^2 \angle_{j-p+1} \{U_{p, \dots, p+q-1}, U^h\} &= \inf_{L \subseteq U_{p, \dots, p+q-1}, \dim L = j-p+1} \sin^2 \angle \{L; U^h\} \\ &\leq \sin^2 \angle \{(U_{p, \dots, j-1}^h)^{\perp} \cap U_{p, \dots, p+q-1}; U^h\} \\ &= \inf_{\substack{u \in U_{p, \dots, p+q-1}, \\ u \in (U_{p, \dots, j-1}^h)^{\perp}, \\ \|u\|_B = 1}} \|(I - Q^h)u\|_B^2, \end{aligned}$$

so our estimate (3.25) is sharper than (3.23).

Using the term  $\sin^2 \angle_{j-p+1}\{U_{p,\dots,p+q-1}, U^h\}$  has yet another advantage: namely, it permits the application of (2.5). Suppose the vectors  $\{u_i, i = p, \dots, p+q-1\}$  form an orthogonal basis for the subspace  $U_{p,\dots,p+q-1}$  and are arranged in such a way that

$$\angle\{u_p; U^h\} \leq \dots \leq \angle\{u_{p+q-1}; U^h\}.$$

Then, by (2.5),

$$\sin^2 \angle_{j-p+1}\{U_{p,\dots,p+q-1}; U^h\} \leq \sum_{i=p}^j \sin^2 \angle\{u_i; U^h\}, \quad j = p, \dots, p+q-1.$$

In other words, if the eigenspace  $U_{p,\dots,p+q-1}$  is spanned by eigenfunctions of different approximation qualities, our result assesses the quality of each of the Ritz values corresponding to the multiple eigenvalue.

**Conclusions.** We derive eigenvalue error bounds for the Ritz method that have several novel features:

- For a simple eigenvalue, our estimates improve those previously known and provide explicit values for all constants.
- For a multiple eigenvalue we prove, in addition, apparently the first truly *a priori* error estimates that show the levels of the eigenvalue errors depending on approximability of eigenfunctions in the corresponding eigenspace.
- For clustered eigenvalues, our results provide elegant eigenvalue error bounds that do not depend on the width of the cluster.

In the FEM eigenvalue approximation context, our results improve earlier known results and are readily applicable for a fixed mesh without making the traditional assumption about the mesh size being small enough.

#### References.

- [1] I. Babuška, B. Q. Guo, and J. E. Osborn. Regularity and numerical solution of eigenvalue problems with piecewise analytic data. *SIAM J. Numer. Anal.*, 26(6):1534–1560, 1989.
- [2] I. Babuška and J. E. Osborn. Estimates for the errors in eigenvalue and eigenvector approximation by Galerkin methods, with particular attention to the case of multiple eigenvalues. *SIAM J. Numer. Anal.*, 24(6):1249–1276, 1987.
- [3] I. Babuška and J. E. Osborn. Finite element-Galerkin approximation of the eigenvalues and eigenvectors of selfadjoint problems. *Math. Comp.*, 52(186):275–297, 1989.
- [4] I. Babuška and J. E. Osborn. Eigenvalue problems. In *Handbook of Numerical Analysis, Vol. II*, pages 641–787. North-Holland, Amsterdam, 1991.
- [5] G. Birkhoff, C. de Boor, B. Swartz, and B. Wendroff. Rayleigh-Ritz approximation by piecewise cubic polynomials. *SIAM J. Numer. Anal.*, 3:188–203, 1966.
- [6] J. H. Bramble, J. E. Pasciak, and A. V. Knyazev. A subspace preconditioning algorithm for eigenvector/eigenvalue computation. *Advances in Comp. Math.*, 6(2):159–189, 1996.
- [7] F. Chatelin. *Spectral approximations of linear operators*. Academic Press, New York, 1983.
- [8] E. G. D'yakonov. *Optimization in solving elliptic problems*. CRC Press, Boca Raton, FL, 1996.
- [9] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, New-York, 1976.

- [10] A. V. Knyazev. *Computation of eigenvalues and eigenvectors for mesh problems: algorithms and error estimates*. Dept. Numerical Math. USSR Academy of Sciences, Moscow, 1986. (In Russian).
- [11] A. V. Knyazev. Sharp a priori error estimates of the Rayleigh-Ritz method without assumptions of fixed sign or compactness. *Mathematical Notes*, 38(5–6):998–1002, 1986.
- [12] A. V. Knyazev. Convergence rate estimates for iterative methods for mesh symmetric eigenvalue problem. *Soviet J. Numerical Analysis and Math. Modelling*, 2(5):371–396, 1987.
- [13] A. V. Knyazev. New estimates for Ritz vectors. *Math. Comp.*, 66(219):985–995, 1997.
- [14] A. V. Knyazev and Merico E. Argentati. Principal angles between subspaces in an  $A$ -based scalar product: Algorithms and perturbation estimates. *SIAM J. Sci. Comput.*, 23(6):2009–2041, 2002.
- [15] A. V. Knyazev and J. Osborn. New *a priori* FEM error estimates for eigenvalues. CU-Denver CCM report, UCD-CCM 215, [http : //math.cudenver.edu/ccm/reports/](http://math.cudenver.edu/ccm/reports/), 2004.
- [16] M. A. Krasnosel'skii, G. M. Vainikko, P. P. Zabreiko, Ya. B. Rutitskii, and Y. Ya. Stetsenko. *Approximate Solutions of Operator Equations*. Wolters-Noordhoff, Groningen, 1972. Translated from Russian.
- [17] M. N. Krylov. Les méthodes de solution approchée des problèmes de la physique mathématique. *Mémoires Sci Math. Gauthier-Villars, Paris*, XLIX:68, 1931.
- [18] J. E. Osborn. Spectral approximation for compact operators. *Math. Comput.*, 29:712–725, 1975.
- [19] E. Ovtchinnikov. Cluster robust error estimates for the Rayleigh–Ritz approximation II: Estimates for eigenvalues. Published as CU-Denver CCM report 210, [http : //math.cudenver.edu/ccm/reports/rep210.pdf.gz](http://math.cudenver.edu/ccm/reports/rep210.pdf.gz), 2004.
- [20] B. N. Parlett. *The symmetric eigenvalue problem*. SIAM, Philadelphia, PA, 1998.
- [21] G. Strang and G. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, 1973.
- [22] H. F. Weinberger. *Variational methods for eigenvalue approximation*. SIAM, Philadelphia, Pa., 1974.