A FRAMEWORK FOR ROBUST EIGENVALUE AND EIGENVECTOR ERROR ESTIMATION AND RITZ VALUE CONVERGENCE ENHANCEMENT

RANDOLPH E. BANK, LUKA GRUBIŠIĆ, AND JEFFREY S. OVALL

Abstract. We present a general framework for the *a posteriori* estimation and enhancement of error in eigenvalue/eigenvector computations for symmetric and elliptic eigenvalue problems, and provide detailed analysis of a specific and important example within this framework—finite element methods with continuous, affine elements. A distinguishing feature of the proposed approach is that it provides provably efficient and reliable error estimation under very realistic assumptions, not only for single, simple eigenvalues, but also for clusters which may contain degenerate eigenvalues. We reduce the study of the eigenvalue/eigenvector error estimators to the study of associated boundary value problems, and make use of the wealth of knowledge available for such problems. Our choice of *a posteriori* error estimator, computed using hierarchical bases, very naturally offers a means not only for estimating error in eigenvalue/eigenvector computations, but also cheaply accelerating the convergence of these computations—sometimes with convergence rates which are nearly twice that of the unaccelerated approximations.

1. Introduction

The purposes of this paper are to present a general framework for the *a posteriori* estimation and enhancement of error in eigenvalue/eigenvector computations for symmetric and elliptic eigenvalue problems, and to provide detailed analysis of a specific and important example within this framework—finite element methods with continuous, affine elements. Our approach to error estimation is based on the Frobenius-Schur factorization of the resolvent of the underlying block-matrix operator. The use of finite elements for elliptic eigenvalue problems is not new (cf. the eigenvalue section of [5, Part 1] and references therein), and efforts continue to be made in this area—the following contributions are representative of much of the literature during roughly the past ten years which is most readily compared to our own [22, 19, 27, 24, 26, 12, 18, 13, 33]. A distinguishing feature of the proposed approach is that it provides provably efficient and reliable error estimation under very realistic assumptions, not only for single, simple eigenvalues, but also for clusters which may contain degenerate eigenvalues. We reduce the study of the eigenvalue/eigenvector error estimators to the study of associated boundary value problems, and make use of the wealth of knowledge available for such problems. Our choice of *a posteriori* error estimator, computed using hierarchical bases, very naturally offers a means not only for estimating error in eigenvalue/eigenvector computations, but also cheaply accelerating the convergence of these computations—sometimes with convergence rates which are nearly twice that of the unaccelerated approximations.

The use of hierarchical bases has a long and productive history in the field of numerical partial differential equations [8, 3, 1, 4, 28]. In addition to a posteriori error estimation, hierarchical basis plays an important role in iterative methods for solving the linear systems arising from finite element discretizations. Like their wavelet counterparts [9, 7, 6], hierarchical bases capture the high frequencies that are the major components of the error in finite element solutions. These very oscillatory subspaces have a number of interesting properties. Among the most important for our work here, in terms of computational cost, is the comparability of the element and global mass matrices to their diagonals—a property which is not enjoyed by the linear finite element stiffness matrix. This implies that simple preconditioners (e.g., Jacobi, Symmetric Gauss-Seidel) or *no* preconditioner, can be used to efficiently and optimally solve such linear systems. If one prefers not to assemble and solve a global system, other versions of hierarchical a posteriori error estimators require only the solution of local problems, often on a single element—though we do not pursue those options in the present work.

2000 *Mathematics Subject Classification.* Primary: 65N30, Secondary: 65N25, 65N15.
Key words and phrases. eigenvalue problem, finite element method, a posteriori error estimates .

The paper is organized as follows:

Section 2 contains basic definitions and the problem statement, and sets the scene for our analysis, which appears in Sections 3 and 4. In Section 3 we derive or basic a posteriori error estimates. One unusual feature of our analysis is that we derive a posteriori estimates for *eigenspaces* rather than individual eigenvectors. This is important in cases where one is interested a single eigenvalue with multiplicity greater than one, or in a cluster of eigenvalues in a given interval. A second important aspect of estimators is that they consist of solving boundary value problems, rather than eigenvalue problems, on the hierarchical space.

In Section 4 we show how our errors estimates can be combined with computed eigenvalues and eigenspaces in order to produces higher order approximations of both. This possibility raises again a classical dilemma related to such procedures. On one hand, one could "accept" the original computed eigenvalues and eigenspaces, equipped with very precise a posteriori error estimates. On the other hand one could accept the more accurate enhanced values, but with less complete knowledge of the actual error. The best choice may well depend on the particular situation. Thus, while we have not resolved this dilemma, we think the most important point to emphasize here is that our a posteriori error estimates are sufficiently precise to present such alternatives.

In Section 5 we present some numerical illustrations of our theory, including acceleration of eigenvalues associated with singular eigenfunctions, and degenerate eigenvalues.

Since much of the analysis is quite detailed and technical, we have not included proofs of some preliminary results and technical lemmas in the main manuscript, but rather have included them in two appendices, A and B, for the benefit of interested readers. Appendix A contains the technical results used in the proof of Theorem 3.2, as well as discussion relevant to the cost of our error estimators. In Appendix B, the derivation of our abstract framework for error estimation and enhancement is presented in greater generality than in the main body of the paper—highlighting its broader applicability.

2. Problem statement and basic properties

Let $\Omega \subset \mathbb{R}^2$ be a bounded *polygonal region*, possibly with re-entrant corners, and let $\partial\Omega_D \subset \partial\Omega$ have positive (1D) Lebesgue measure. We define the space $\mathcal{H} = \{v \in H^1(\Omega) : \ v = 0 \text{ on } \partial\Omega_D \text{ in the sense of trace}\}$. We are interested in the eigenvalue problem:

$$\text{Find } (\lambda, \psi) \in \mathbb{R} \times \mathcal{H} \text{ so that } B(\psi, v) = \lambda(\psi, v) \text{ and } \psi \neq 0 \text{ for all } v \in \mathcal{H} \ . \tag{2.1}$$

Here we have assumed

$$B(w, v) = \int_\Omega A\nabla w \cdot \nabla v + cwv \, dx, \tag{2.2}$$

and

$$(w, v) = \int_\Omega wv \, dx \tag{2.3}$$

is the standard L^2 inner-product. We will also assume that $A \in [L^\infty(\Omega)]^{2\times 2}$ is uniformly positive definite a.e., and that $c \in L^\infty(\Omega)$ is non-negative. These assumptions guarantee that there are constants $c_0, c_1 > 0$ such that $B(v, w) \leq c_1\|v\|_1\|w\|_1$ and $|\!|\!|v|\!|\!|^2 =: B(v, v) \geq c_0\|v\|_1^2$ for all $v, w \in \mathcal{H}$. In other words, the "energy"-norm $|\!|\!|\cdot|\!|\!|$ induced by the inner-product $B(\cdot, \cdot)$ is equivalent to $\|\cdot\|_1$. As a practical matter, we will further assume that A and c are piecewise-smooth on some polygonal partition of Ω.

Here and elsewhere, we use the following standard notation for norms and seminorms: for $k \in \mathbb{N}$ and $S \subset \Omega$ we denote the standard norms and semi-norms on the Hilbert spaces $H^k(S)$ by

$$\|v\|_{k,S}^2 = \sum_{|\alpha|\leq k} \|D^\alpha v\|_S^2 \qquad\qquad |v|_{k,S}^2 = \sum_{|\alpha|=k} \|D^\alpha v\|_S^2 \ , \tag{2.4}$$

where $\|\cdot\|_S$ denotes the L^2 norm on S. When $S = \Omega$, we omit it from the subscript. For other real r we also use $H^r(\Omega)$ to denote the standard interpolation spaces, together with their norms $\|\cdot\|_{r,S}$ (cf. [11]). In most cases we use the notation $\|\cdot\|$ (with no subscript) to denote the L^2-norm on Ω, and use the aforementioned subscripts for this norm only when it is useful to clarify the distinction between norms in a specific argument or claim.

2.1. **Tools from the eigenvalue/eigenvector theory.** The variational eigenvalue problem (2.1)–(2.3) is attained by the positive sequence of eigenvalues

$$0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_q \le \cdots \tag{2.5}$$

and the sequence of eigenvectors $(\psi_i)_{i\in\mathbb{N}}$ such that

$$B(\psi_i, v) = \lambda_i(\psi_i, v), \qquad \forall v \in \mathcal{H}, \qquad \text{and } (\psi_i, \psi_j) = \delta_{ij} \ .$$

Here we have counted the eigenvalues according to their multiplicity and we will also use the notation $\psi_i \perp \psi_j$ when $(\psi_i, \psi_j) = 0$ (when $i \neq j$). Furthermore, the sequence $(\lambda_i)_{i\in\mathbb{N}}$ has no finite accumulation point; and due to the Peron-Frobenius theorem we know that, in the case in which Ω is a path-wise connected domain, the inequality $\lambda_1 < \lambda_2$ holds and the eigenvector ψ_1 can be chosen so that ψ_1 is continuous and $\psi_1 > 0$ holds pointwise in Ω.

For various calculations it will be necessary to have the operator based formulation of the eigenvalue problem, too. To this end, let us note that according to [20, Theorem VI-2.23, pp. 331] the positive definite symmetric form B with its domain of definition $\mathcal{H}$ defines in $L^2(\Omega)$ the self-adjoint operator $\mathcal{A}$ such that $\mathcal{H} = \mathrm{Dom}(\mathcal{A}^{1/2})$—the domain of definition of the operator $\mathcal{A}^{1/2}$—and

$$B(\psi, \phi) = (\mathcal{A}^{1/2}\psi, \mathcal{A}^{1/2}\phi), \quad \psi, \phi \in \mathcal{H}. \tag{2.6}$$

The spectrum of the operator $\mathcal{A}$ is the set

$$\mathrm{Spec}(\mathcal{A}) := \{\zeta \in \mathbb{C} \ : \ \mathcal{A} - \zeta\mathbf{I} \text{ is not invertible } \} = \{\lambda_i \ : \ i \in \mathbb{N}\}$$

and the resolvent set is defined as $\rho(\mathcal{A}) = \mathbb{C} \setminus \mathrm{Spec}(\mathcal{A})$.

In the quantitative study of the spectrum the central role is played by the resolvent. The resolvent of $\mathcal{A}$ is the operator valued function $\zeta \mapsto (\mathcal{A} - \zeta\mathbf{I})^{-1}$, $\zeta \in \rho(\mathcal{A})$. For the operator $\mathcal{A}$ which is defined by the form (2.2) we have that, for all $\zeta \in \rho(\mathcal{A})$, the resolvent takes values in the set of compact operators. That means that—using the spectral theorem for the compact operators—the set $\mathrm{Spec}(\mathcal{A})$ is countable and that there exists a sequence of orthogonal projections E_{λ_i}, $i = 1, 2, \cdots$ such that $\sum_{i=1}^{\infty} E_{\lambda_i} = \mathbf{I}$ and the range spaces $\mathsf{R}(E_{\lambda_i})$ and $\mathsf{R}(E_{\lambda_j})$ are mutually orthogonal for for all i and j such that $\lambda_i \neq \lambda_j$. Further we get that

$$B(\psi, \phi) = \sum_{\lambda\in\mathrm{Spec}(\mathcal{A})} \lambda(\psi, E_\lambda\phi), \qquad \psi, \phi \in \mathcal{H}$$

and so we obtain an alternative representation of the energy norm

$$|\!|\!|\psi|\!|\!|^2 = \sum_{\lambda\in\mathrm{Spec}(\mathcal{A})} \lambda(\psi, E_\lambda\psi). \tag{2.7}$$

2.2. **Tools from the finite element approximation theory.** We will approximate collections of eigenpairs using piecewise linear Lagrange finite elements on a family of conforming *shape-regular* triangular meshes, $\mathcal{F} = \{\mathcal{T}\}$. We assume that the edges of the triangles in $\mathcal{T}$ align themselves with any discontinuities of A and c. To define the subspaces V, $W \subset \mathcal{H}$ in which we will approximate eigenfunctions and eigenfunction errors, we first introduce mesh-related notation:

- $\mathcal{V}$ = non-Dirichlet vertices; $\mathcal{V}_D$ = Dirichlet boundary vertices; $\overline{\mathcal{V}} = \mathcal{V} \cup \mathcal{V}_D$
- $\mathcal{E}$ = non-Dirichlet edges; $\mathcal{E}_D$ = Dirichlet boundary edges; $\overline{\mathcal{E}} = \mathcal{E} \cup \mathcal{E}_D$
- For $z \in \overline{\mathcal{V}}$, $\ell_z \in C(\Omega)$ is defined by the relations: $\ell_z|_T \in \mathbb{P}_1$ for each $T \in \mathcal{T}$ and $\ell_z(z') = \delta_{zz'}$ for all $z' \in \overline{\mathcal{V}}$
- For $e \in \overline{\mathcal{E}}$, $b_e \in C(\Omega)$ is defined by $b_e = 4\ell_z\ell_{z'}$, where z and z' are the endpoints of e; we note that $b_e|_T \in \mathbb{P}_2$ for each $T \in \mathcal{T}$

Here and elsewhere, $\mathbb{P}_k$ denotes the space of polynomials of total degree k on T. We note that

$$\{\ell_z\}_{z\in\overline{\mathcal{V}}}, \quad \{b_e\}_{e\in\overline{\mathcal{E}}} \quad \text{and} \quad \{\ell_z\}_{z\in\overline{\mathcal{V}}} \cup \{b_e\}_{e\in\overline{\mathcal{E}}}$$

are each linearly independent sets. Two subspaces of interest, V, $W \subset \mathcal{H}$, related to the partition $\mathcal{T}$ are defined by

$$V = \{v \in \mathcal{H} \cap C(\Omega) : \ v|_T \in \mathbb{P}_1\} = \mathrm{span}\{\ell_z\}_{z\in\mathcal{V}} \ , \tag{2.8}$$

$$W = \{w \in \mathcal{H} \cap C(\Omega) : \ w|_T \in \mathbb{P}_2, \, w(z) = 0 \text{ for } z \in \bar{\mathcal{V}}\} = \mathrm{span}\{b_e\}_{e\in\mathcal{E}} \ . \tag{2.9}$$

In what follows we will consider a discrete version of (2.1):

$$\text{Find } (\hat{\lambda}, \hat{\psi}) \in \mathbb{R} \times V \text{ such that } B(\hat{\psi}, v) = \hat{\lambda}(\hat{\psi}, v) \text{ for all } v \in V \ . \tag{2.10}$$

We also assume, without further comment, that the solutions are ordered and indexed as in (2.5), with $(\hat{\psi}_i, \hat{\psi}_j) = \delta_{ij}$. In what follows, we will only notationally emphasize the dependence of the finite element spaces and approximate solutions on the underlying triangulation $\mathcal{T}$ when it is necessary for clarification.

The actual problem we wish to address in the present work is the following. Suppose that (a, b) is an interval containing m eigenvalues of B, counting multiplicities. We will denote the set of eigenvalues by $s_m = \{\mu_k\}_{k=1}^m$ and the corresponding invariant subspace by $S_m = \text{span}\{\phi_k\}_{k=1}^m$. Our computed approximations of s_m, S_m (via (2.10)) will be denoted by $\hat{s}_m = \{\hat{\mu}_k\}_{k=1}^m$ and $\hat{S}_m = \text{span}\{\hat{\phi}_k\}_{k=1}^m$. We are interested in the following related problems:

(1) Reliably estimate *a posteriori* how well $\hat{s}_m$ and $\hat{S}_m$ approximate s_m and S_m.
(2) Use the estimates of eigenvalue and eigenvector error to cheaply enhance the quality of $\hat{s}_m$ and $\hat{S}_m$.

We addressed the first of these issues for the Dirichlet Laplacian in [18], but the present work significantly generalizes and extends our efforts there. We emphasize that our error estimation and enhancement techniques do not assume that the eigenvalues of interest are simple, and clusters of eigenvalues are treated just as readily as single eigenvalues.

We point out the shift from the (λ_k, ψ_k)-notation, which reflects the global ordering and enumeration of the eigenpairs, and the (μ_k, ϕ_k)-notation, which reflects the same ordering but a local enumeration of a cluster of eigenpairs. Let us also mention that some of the results will only hold under the assumption that we are approximating the lowermost set of eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_m < \lambda_{m+1}$. In this case we will exceptionally use the notation $\hat{\lambda}_1 \le \cdots \le \hat{\lambda}_m$ for the approximations of $\lambda_1 \le \cdots \le \lambda_m$ in V.

Remark 2.1. Although λ_1 is simple, i.e. $\lambda_1 < \lambda_2$ holds, for the class of problems we consider numerically in the present work, much of the theory carries over to problems where Ω is not pathwise connected, or the boundary conditions are periodic (as examples). In these cases the Peron-Frobenius theorem does not apply, and it is quite possible that the smallest eigenvalue is degenerate.

2.3. **Approximation defects.** Let the subspaces V, $W \subset \mathcal{H}$ be given and let $\hat{s}_m$ and $\hat{S}_m$ be the approximations which are computed from V. We define the *approximation defects* in $\hat{s}_m, \hat{S}_m$ as:

$$\eta_i^2(\hat{S}_m) = \max_{\substack{\mathcal{S} \subset \hat{S}_m \\ \dim \mathcal{S} = m-i+1}} \ \min_{\substack{f \in \mathcal{S} \\ f \neq 0}} \frac{|\!|\!| u(f) - \hat{u}(f) |\!|\!|^2}{|\!|\!| u(f) |\!|\!|^2} \ , \tag{2.11}$$

where $u(f)$ and $\hat{u}(f)$ satisfy:

$$B(u(f), v) = (f, v) \text{ for every } v \in \mathcal{H} \tag{2.12}$$

$$B(\hat{u}(f), v) = (f, v) \text{ for every } v \in V \ . \tag{2.13}$$

We will argue below that such approximation defects are very useful for estimating the error in $\hat{s}_m$ as an approximation of s_m (and $\hat{S}_m$ as an approximation of S_m), and can be used naturally to accelerate the convergence of the computed approximations—our two stated objectives.

Of course, $u(f)$, and hence η_i, cannot be computed, so we must efficiently and reliably estimate these quantities. However, this formulation has reduced the problem to estimating error (and functions) for associated boundary value problems, and we take advantage of that well-developed theory. Our computable estimates of the approximation defects are

$$\tilde{\eta}_i^2(\hat{S}_m) = \max_{\substack{\mathcal{S} \subset \hat{S}_m \\ \dim \mathcal{S} = m-i+1}} \ \min_{\substack{f \in \mathcal{S} \\ f \neq 0}} \frac{|\!|\!| \varepsilon(f) |\!|\!|^2}{|\!|\!| \varepsilon(f) |\!|\!|^2 + |\!|\!| \hat{u}(f) |\!|\!|^2} \ , \tag{2.14}$$

where the approximate error function $\varepsilon(f) \in W$ is the solution of

$$B(\varepsilon(f), v) = (f, v) - B(\hat{u}(f), v) \text{ for every } v \in V \ . \tag{2.15}$$

Recognizing that the right-hand side in (2.15) is $B(u(f) - \hat{u}(f), v)$, it is clear that $\varepsilon(f)$ is the projection of $u(f) - \hat{u}(f)$ into W in the energy inner-product $B(\cdot, \cdot)$. In Section 3 we will discuss in detail matters relevant to these estimates, such as: a practical means for solving the max-min problems (2.14), the effectivity of $\tilde{\eta}_i$

as an estimator of η_i, and computational cost. The more technical aspects of the effectivity proofs are given in Appendix A.

Let us now justify the term *approximation defects*. We do so by stating some of the key results of our prior work [18], in order to indicate the sorts of results we will prove here in a more general setting. The approximation defects are related to the eigenvalue error in the following way. Assume that $\hat{S}_m$ is the span of first $m \in \mathbb{N}$ eigenvectors of (2.10) then we have the following efficiency and reliability result.

Theorem 2.2. *Let $B(\cdot,\cdot)$ be the standard Dirichlet form which generates the Dirichlet Laplace operator and let $\lambda_m < \lambda_{m+1}$. If $\hat{S}_m = \operatorname{span}\{\hat{\psi}_1, \cdots, \hat{\psi}_m\}$ is such that $\frac{\eta_m(\hat{S}_m)}{1-\eta_m(\hat{S}_m)} < \frac{\lambda_{m+1}-\hat{\lambda}_m}{\lambda_{m+1}+\hat{\lambda}_m}$ then*

$$\frac{\hat{\lambda}_1}{2\hat{\lambda}_m}\sum_{i=1}^{m}\eta_i^2(\hat{S}_m) \leq \sum_{i=1}^{m}\frac{\hat{\lambda}_i-\lambda_i}{\hat{\lambda}_i} \leq C_m\sum_{i=1}^{m}\eta_i^2(\hat{S}_m). \tag{2.16}$$

The constant C_m depends solely on the shape regularity of $\mathcal{T}$ and the relative distance to the unwanted component of the spectrum (e.g. $\frac{\lambda_m-\lambda_{m+1}}{\lambda_m+\lambda_{m+1}}$).

The constant C_m is given by an explicit formula which is a reasonable practical overestimate, see [18] for details. A similar results holds for the eigenvectors. We point the interested reader to [18, Theorem 4.1 and equation (3.10)].

Remark 2.3. This result also holds for more general domains and boundary conditions, which allow λ_1 to be degenerate. In this case, if $\lambda_1 = \lambda_m$, then the constant $\hat{\lambda}_1/2\hat{\lambda}_m$ in (2.16) can be replaced by 1.

An important feature of these estimates is that they are asymptotically exact, both as eigenvector as well as eigenvalue estimators.

Theorem 2.4. *Let $B(\cdot,\cdot)$ be the standard Dirichlet form which generates the Dirichlet Laplace operator and let $\lambda_{q-1} < \lambda_q = \lambda_{q+m-1} < \lambda_{q+m}$. Let $\hat{S}_m = \hat{S}_m(\mathcal{T}) = \operatorname{span}(\hat{\phi}_k) \subset V = V(\mathcal{T})$ be the computed approximation of the invariant subspace corresponding to λ_q. Then, taking the pairing of eigenvectors ϕ_i and Ritz vectors $\hat{\phi}_i$ as in [18], we have*

$$\lim_{h_{\mathcal{T}}\to 0}\frac{\sum_{i=1}^{m}\frac{|\hat{\mu}_i-\lambda_q|}{\hat{\mu}_i}}{\sum_{i=1}^{m}\eta_i^2(\hat{S}_m)} = 1 \quad , \quad \lim_{h_{\mathcal{T}}\to 0}\frac{\sum_{i=1}^{m}\frac{\|\nabla\hat{\phi}_i-\nabla\phi_i\|^2}{\|\nabla\phi_i\|^2}}{\sum_{i=1}^{m}\eta_i^2(\hat{S}_m)} = 1 \ , \tag{2.17}$$

where $h_{\mathcal{T}}$ is the diameter of the largest triangle in $\mathcal{T}$. Furthermore, if $\hat{S}_m$ is such that $\frac{\eta_m(\hat{S}_m)}{1-\eta_m(\hat{S}_m)} < \gamma_q := \min\left\{\frac{\lambda_{q+m}-\hat{\mu}_m}{\lambda_{q+m}+\hat{\mu}_m}, \frac{\hat{\mu}_1-\lambda_{q-1}}{\hat{\mu}_1+\lambda_{q-1}}\right\}$ holds, then

$$1 \leq \frac{\sum_{i=1}^{m}\frac{|\hat{\mu}_i-\lambda_q|}{\hat{\mu}_i}}{\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m)} \leq C_1 \quad , \quad 1 \leq \frac{\sum_{i=1}^{m}\frac{\|\nabla\hat{\phi}_i-\nabla\phi_i\|^2}{\|\nabla\phi_i\|^2}}{\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m)} \leq C_2 \ , \tag{2.18}$$

for constants $C_1, C_2 < \infty$ depending only on the shape regularity of $\mathcal{T}$.

Furthermore, the experiments which were performed in [18] indicate that results like (2.17) may also hold for practically relevant estimator $\tilde{\eta}_i(\hat{S}_m)$. In particular, it is not unreasonable to expect to see

$$\lim_{h_{\mathcal{T}}\to 0}\frac{\sum_{i=1}^{m}\frac{|\hat{\mu}_i-\lambda_q|}{\hat{\mu}_i}}{\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m)} = 1 \quad , \quad \lim_{h_{\mathcal{T}}\to 0}\frac{\sum_{i=1}^{m}\frac{\|\nabla\hat{\phi}_i-\nabla\phi_i\|^2}{\|\nabla\phi_i\|^2}}{\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m)} = 1 \tag{2.19}$$

in practice. This optimal behavior can be partially explained using results from [28] for boundary value problems, in which it is shown, under certain assumptions on the regularity of the solution and the meshes, that $\|\varepsilon\|/\|u-\hat{u}\| \to 1$ as $h_{\mathcal{T}} \to 0$. It is a nice feature, though not well-understood, that this behavior is also often seen in situations where the solution does not enjoy this regularity. Let us note that even the obtained inequalities (2.18) are particularly good news for adaptive algorithms. The aim of the following section is to prove such a result for general divergence type self-adjoint and elliptic form $B(\cdot,\cdot)$. We will present a general framework for such analysis in Appendix B. This framework will be applied to analyze a practically relevant result in Sections 3 and 4.

3. Estimating the approximation defects

We recall the expressions for the approximation defects in $\hat{s}_m, \hat{S}_m$ and our estimates of them:

$$\eta_k^2(\hat{S}_m) = \max_{\substack{\mathcal{S}\subset S_m \\ \dim \mathcal{S}=m-k+1}} \min_{\substack{f\in\mathcal{S} \\ f\neq 0}} \frac{\|u(f)-\hat{u}(f)\|^2}{\|u(f)\|^2} \ , \tag{3.1}$$

$$\tilde{\eta}_k^2(\hat{S}_m) = \max_{\substack{\mathcal{S}\subset S_m \\ \dim \mathcal{S}=m-k+1}} \min_{\substack{f\in\mathcal{S} \\ f\neq 0}} \frac{\|\varepsilon(f)\|^2}{\|\varepsilon(f)\|^2+\|\hat{u}(f)\|^2} \ , \tag{3.2}$$

where $u(f)$, $\hat{u}(f)$ and $\varepsilon(f)$ satisfy

$$B(u(f),v) = (f,v) \text{ for every } v\in\mathcal{H} \tag{3.3}$$

$$B(\hat{u}(f),v) = (f,v) \text{ for every } v\in V \tag{3.4}$$

$$B(\varepsilon(f),v) = (f,v) - B(\hat{u}(f),v) \text{ for every } v\in W \ . \tag{3.5}$$

3.1. **The Computation of $\tilde{\eta}_k$.** The spaces $\hat{S}_m$ are given in terms of a Ritz basis $\{\hat{\phi}_j\}_{j=1}^m$: $B(\hat{\phi}_j,v) = \hat{\mu}_j(\hat{\phi}_j,v)$ for all $v\in V$, and $(\hat{\phi}_i,\hat{\phi}_j)=\delta_{ij}$. We define $\hat{u}_j = \hat{u}(\hat{\phi}_j) = (\hat{\mu}_j)^{-1}\,\hat{\phi}_j$ and $\varepsilon_j = \varepsilon(\hat{\phi}_j)$. By the Courant-Fischer Theorem, we see that solving the max-min problem (3.2) is equivalent to solving the small $(m\times m)$ generalized eigenvalue problem

$$E\mathbf{x} = \tilde{\eta}^2 G\mathbf{x} \ \text{ for } E_{ij} = B(\varepsilon_j,\varepsilon_i) \text{ and } G = E + \mathrm{diag}((\hat{\mu}_j)^{-1}) \ . \tag{3.6}$$

The cost of solving (3.6) is small relative to that of solving

$$B(\varepsilon_k,w) = (\hat{\phi}_k,w) - (\hat{\mu}_k)^{-1}B(\hat{\phi}_k,w) \text{ for all } w\in W \tag{3.7}$$

for each k, and we argue next that the cost of computing ε_k is small in comparison to that of computing $\hat{\phi}_k$.

In [28, 1], for example, it is argued that the matrix associated with computing ε is spectrally equivalent to its diagonal *independent of the mesh scaling* for shape-regular families of meshes. Because of this the computation of ε will require few iterations of a Krylov solver (CG, GMRES) to sufficiently converge—either with no preconditioning at all, or with (symmetric) diagonal preconditioning. In other words, although we do solve a (global) system which is larger than that for computing $\hat{u}$, it is actually cheaper to compute ε than $\hat{u}$. In order to make this paper more self-contained, we have included a brief explanation of this fact in Appendix A.

3.2. **Effectivity of Approximation Defect Estimates.** The first step in analyzing the effectivity of $\tilde{\eta}_k$ as an estimate of η_k is to determine the effectivity of $\|\varepsilon(f)\|$ as an estimate of $\|u(f)-\hat{u}(f)\|$ for the types of data f we will encounter. Because f is allowed to vary, an ideal estimate would be one in which the relevant constants do not depend on f. Such estimates are provided by the following analysis, which we discuss here briefly, and elaborate upon the more technical details in Appendix A. In the exposition below, we suppress the notation indicating the dependence on f.

For any $v\in\mathcal{H}$, $\hat{v}\in V$ and $w\in W$, we have

$$\begin{aligned} B(u-\hat{u},v) &= B(\varepsilon,w) + B(u-\hat{u},v-\hat{v}-w) \\ &= B(\varepsilon,w) + \sum_{z\in\bar{\mathcal{V}}} B(u-\hat{u},v\ell_z - v_z - w_z) \\ &= B(\varepsilon,w) + \sum_{z\in\bar{\mathcal{V}}} \int_{\omega_z} R(v\ell_z - v_z - w_z)\,dV \\ &\qquad + \sum_{e\in\mathcal{E}} \int_e r(v-\hat{v}-w)\,dS \ , \end{aligned}$$

where $v_z\in V$, $w_z\in W$, $\mathrm{supp}(v_z),\mathrm{supp}(w_z)\subset\mathrm{supp}(\ell_z)\doteq\omega_z$, $\hat{v}=\sum_{z\in\bar{\mathcal{V}}} v_z$ and $w=\sum_{z\in\bar{\mathcal{V}}} w_z$. The standard element and edge residuals, R and r are piecewise-defined on triangles T and edges e, respectively by

$$R|_T = f - c\hat{u} + \nabla\cdot A\nabla\hat{u} \ , \tag{3.8}$$

$$r|_e = -(A\nabla\hat{u}\cdot\mathbf{n}_T)|_T - (A\nabla\hat{u}\cdot\mathbf{n}_{T'})|_{T'} \ , \tag{3.9}$$

where r is taken to be 0 for $e \notin \mathcal{E}$, T and T' are the two triangles adjacent to $e \in \mathcal{E}$, and $\mathbf{n}_T$ and $\mathbf{n}_{T'}$ are their outward unit normals (so $\mathbf{n}_T = -\mathbf{n}_{T'}$ on e).

In Appendix A, we will select $\hat{v}$ and w in such a way as to prove

Lemma 3.1. *There are scale-invariant constants $K_1 = K_1(\mathcal{T}, B)$ and $K_2 = K_2(\mathcal{T})$ such that*

$$B(u - \hat{u}, v) \le K_1 |\!|\!| \varepsilon |\!|\!| \, |\!|\!| v |\!|\!| + K_2 \operatorname{osc}(R, r) |v|_1 \ ,$$

$$[\operatorname{osc}(R, r)]^2 = \sum_{z \in \bar{\mathcal{V}}} d_z^2 \inf_{R_z \in \mathbb{R}} \|R - R_z\|_{0,\omega_z}^2 + \sum_{e \in \mathcal{E}} |e| \inf_{r_e \in \mathbb{R}} \|r - r_e\|_{0,e}^2 \ ,$$

where d_z is the diameter of ω_z.

From this lemma it is readily deduced that

Theorem 3.2. *There are scale-invariant constants $K_1 = K_1(\mathcal{T}, B)$ and $C_2 = C_2(\mathcal{T}, B)$ such that*

$$|\!|\!| \varepsilon |\!|\!| \le |\!|\!| u - \hat{u} |\!|\!| \le K_1 |\!|\!| \varepsilon |\!|\!| + C_2 \operatorname{osc}(R, r) \ .$$

Remark 3.3. We will see in Appendix A that this result is unaffected by the presence of a convection term $\mathbf{b} \cdot \nabla u$ in B, or non-trivial Neumann conditions $A \nabla u \cdot \mathbf{n} = g$ on $\partial\Omega_N = \partial\Omega \setminus \partial\Omega_D$. All that need be changed is to replace $|\!|\!| \cdot |\!|\!|$ with $\| \cdot \|_1$ in both Lemma 3.1 and Theorem 3.2. Although this level of generality is not needed for the eigenvalue problems under consideration, the proofs for either case are the same, so we opt to analyze the more general boundary value problems in Appendix A.

Remark 3.4. The oscillation term is computable, or at least conveniently estimable, if desired. In particular, if A is piecewise-constant on the triangulation, then the $\nabla \cdot A \nabla \hat{u}$-term vanishes from R, and there is no edge contribution to osc. In this case we obtain $\inf_{R_z \in \mathbb{R}} \|R - R_z\|_{0,\omega_z} \le C_z d_z \|f\|_{1,\omega_z}$, provided c is not discontinuous on ω_z, because our data f are H^1-functions. More generally, we expect osc to be much smaller than $|\!|\!| \varepsilon |\!|\!|$, at least as the mesh is adaptively refined, so we will use $|\!|\!| \varepsilon |\!|\!| \approx |\!|\!| u - \hat{u} |\!|\!|$ in practice. All constants can also be estimated, or at least bounded, purely in terms of local Poincaré constants and and the equivalence bounds on the norms $m \|v\|_1 \le |\!|\!| v |\!|\!| \le M \|v\|_1$.

Remark 3.5. The result above is a generalization of that in [18], which was argued only for the Laplacian. These results avoid the saturation assumption commonly associated with hierarchical error estimates, instead replacing it, in a sense, with something which can be directly assessed. That argument was itself an effort to apply observations of Dörfler and Nochetto [10] in the eigenvalue context. Similar results, though different in both proof or assumptions, were previously obtained in [4].

Using Theorem 3.2, we deduce that

Theorem 3.6. *There are scale-invariant constants $K_1 = K_1(\mathcal{T}, B)$ and $C_2 = C_2(\mathcal{T}, B)$, such that*

$$1 \le \frac{\eta_k}{\tilde{\eta}_k} \le K_1 + C_2 \max_{\substack{f \in S_m \\ \|f\|_0 = 1}} \frac{\operatorname{osc}(R, r)}{|\!|\!| \varepsilon |\!|\!|} \ . \tag{3.10}$$

Proof. From the definitions of $\hat{u}, \varepsilon$, we have $|\!|\!| u |\!|\!|^2 = |\!|\!| u - \hat{u} |\!|\!|^2 + |\!|\!| \hat{u} |\!|\!|^2$ and $|\!|\!| u |\!|\!|^2 = |\!|\!| u - \hat{u} - \varepsilon |\!|\!|^2 + |\!|\!| \hat{u} |\!|\!|^2 + |\!|\!| \varepsilon |\!|\!|^2$, so

$$\frac{|\!|\!| u - \hat{u} |\!|\!|^2}{|\!|\!| u |\!|\!|^2} = 1 - \frac{|\!|\!| \hat{u} |\!|\!|^2}{|\!|\!| u |\!|\!|^2} \ge 1 - \frac{|\!|\!| \hat{u} |\!|\!|^2}{|\!|\!| \hat{u} |\!|\!|^2 + |\!|\!| \varepsilon |\!|\!|^2} = \frac{|\!|\!| \varepsilon |\!|\!|^2}{|\!|\!| \hat{u} |\!|\!|^2 + |\!|\!| \varepsilon |\!|\!|^2} \ .$$

The left-hand inequality in (3.10) follows directly from this estimate. By Theorem 3.2,

$$\frac{|\!|\!| u - \hat{u} |\!|\!|}{|\!|\!| u |\!|\!|} \le K_1 \frac{|\!|\!| \varepsilon |\!|\!|}{|\!|\!| u |\!|\!|} + C_2 \frac{\operatorname{osc}(R, r)}{|\!|\!| u |\!|\!|} \ ,$$

and the upper bound is only increased by replacing $|\!|\!| u |\!|\!|$ with $\sqrt{|\!|\!| \hat{u} |\!|\!|^2 + |\!|\!| \varepsilon |\!|\!|^2}$ in the denominators. From this estimate, the right-hand inequality in (3.10) is clear. Q.E.D.

Remark 3.7. Suppose that A is a piecewise constant matrix and $c = 0$. The computed eigenvalues $\hat{s}_m$ are contained in an interval $[\tilde{\sigma}_0, \tilde{\sigma}_1]$. If we have a family of meshes indexed by the (longest edge) mesh parameter h, then $|\!|\!| \varepsilon |\!|\!| \sim h$, and $\operatorname{osc}(R, r) \le C \sqrt{\tilde{\sigma}_1}\, h^2$ for $f \in \hat{S}_m$ with $\|f\|_0 = 1$, where C is scale-invariant, and depends only on the bilinear form B and quasi-uniformity parameters. In this case we could, *a priori*, choose

a quasi-uniform mesh based on the portion of the spectrum we wish to compute, for which we can guarantee that the term in (3.10) involving oscillation can be safely ignored for all refinements of this mesh.

3.3. **Eigenvalue/vector estimates for general divergence type operators.** We now present reliability and efficiency results for general self-adjoint divergence-type operators. The following two theorems extend the results of [18], which were summarized here as Theorems 2.2 and 2.4.

Theorem 3.8. *Let $B(\cdot,\cdot)$ be the divergence type sesquilinear form with a self-adjoint boundary condition. Further, let $\lambda_m < \lambda_{m+1}$, and take $\hat{S}_m$ as defined in Theorem 2.2. If $\hat{S}_m$ is such that $\frac{\eta_m(\hat{S}_m)}{1-\eta_m(\hat{S}_m)} < \frac{\lambda_{m+1}-\hat{\lambda}_m}{\lambda_{m+1}+\hat{\lambda}_m}$ then*

$$\frac{\hat{\lambda}_1}{2\hat{\lambda}_m}\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m) \leq \sum_{i=1}^{m}\frac{\hat{\lambda}_i-\lambda_i}{\hat{\lambda}_i} \leq C_{m,B,\mathcal{T}}\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m).$$

The constant $C_{m,B,\mathcal{T}}$ depends solely on the shape regularity of $\mathcal{T}$, the form B and the relative distance to the unwanted component of the spectrum (e.g. λ_m). If $\lambda_1 = \lambda_m$ then we can drop the constant $\hat{\lambda}_1/(2\hat{\lambda}_m)$ from the lower estimate.

Proof. The proof of the statement is a direct combination of the new results from Section 3 and the eigenvalue estimates from [18]. Q.E.D.

Remark 3.9. A practical overestimate of $C_{m,\mathcal{T},B}$ could be obtained, if desired, through a careful reading of Appendix A.

This theorem is a reliability and efficiency results which combines the results of Theorem 3.6 and the main result from [17]. An important further feature of these estimates is that they are asymptotically exact, both as eigenvector as well as as eigenvalue estimators.

Theorem 3.10. *Let $B(\cdot,\cdot)$ be the the divergence type sesquilinear form with a self-adjoint boundary condition and let $\lambda_{q-1} < \lambda_q = \lambda_{q+m-1} < \lambda_{q+m}$. Let $\hat{S}_m = \hat{S}_m(\mathcal{T}) = \mathrm{span}(\hat{\phi}_k) \subset V = V(\mathcal{T})$ be the computed approximation of the invariant subspace corresponding to λ_q. Then, taking the pairing of eigenvectors ϕ_i and Ritz vectors $\hat{\phi}_i$ as in* [18]*, we have*

$$\lim_{h_{\mathcal{T}}\to 0}\frac{\sum_{i=1}^{m}\frac{|\hat{\mu}_i-\lambda_q|}{\hat{\mu}_i}}{\sum_{i=1}^{m}\eta_i^2(\hat{S}_m)} = 1 \quad , \quad \lim_{h_{\mathcal{T}}\to 0}\frac{\sum_{i=1}^{m}\frac{\|\hat{\phi}_i-\phi_i\|^2}{\|\phi_i\|^2}}{\sum_{i=1}^{m}\eta_i^2(\hat{S}_m)} = 1 \ , \tag{3.11}$$

where $h_{\mathcal{T}}$ is the diameter of the largest triangle in $\mathcal{T}$. Furthermore, if $\hat{S}_m$ is such that $\frac{\eta_m(\hat{S}_m)}{1-\eta_m(\hat{S}_m)} < \gamma_q := \min\left\{\frac{\lambda_{q+m}-\hat{\mu}_m}{\lambda_{q+m}+\hat{\mu}_m}, \frac{\hat{\mu}_1-\lambda_{q-1}}{\hat{\mu}_1+\lambda_{q-1}}\right\}$ holds, then

$$1 \leq \frac{\sum_{i=1}^{m}\frac{|\hat{\mu}_i-\lambda_q|}{\hat{\mu}_i}}{\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m)} \leq K_1 \quad , \quad 1 \leq \frac{\sum_{i=1}^{m}\frac{\|\hat{\phi}_i-\phi_i\|^2}{\|\phi_i\|^2}}{\sum_{i=1}^{m}\tilde{\eta}_i^2(\hat{S}_m)} \leq K_2 \ , \tag{3.12}$$

for constants $K_1, K_2 < \infty$ depending only on the shape regularity of $\mathcal{T}$ and the coefficients of B.

4. Enhancing the Ritz value/vector convergence using the approximation defects

In this section, we consider some practical procedures for accelerating the convergence of computed Ritz values $\hat{s}_m = \{\hat{\mu}_k\}$ to the eigenvalues $s_m = \{\mu_k\}$ they approximate. A heuristic motivation for the approaches we consider is given by the approximations

$$\sum_{k=1}^{m}\frac{\hat{\mu}_k-\mu_k}{\hat{\mu}_k} \approx \sum_{k=1}^{m}\eta^2(\hat{S}_m) \approx \sum_{k=1}^{m}\tilde{\eta}^2(\hat{S}_m) \ . \tag{4.1}$$

Since we generally expect these three quantities to be asymptotically equivalent, this suggests that choosing "enhanced Ritz values", $\{\hat{\mu}_k^*\}$, which solve the equation

$$\sum_{k=1}^{m}\frac{\hat{\mu}_k-\hat{\mu}_k^*}{\hat{\mu}_k} = \sum_{k=1}^{m}\tilde{\eta}^2(\hat{S}_m) \ . \tag{4.2}$$

may lead to faster convergence. Obviously there are many ways to satisfy (4.2), and an optimal choice of enhanced Ritz values in a general setting (i.e. a cluster of eigenvalues which may contain degenerate members) is not trivial to determine. Ideally, the enhanced Ritz values would converge to their appropriate eigenvalues more rapidly than their non-enhanced counterparts both individually and collectively—in other words, not just in terms of the trace error. We give an answer of how ideal enhanced Ritz values would look in Appendix B using the Formula (B.10), and we denote them there by $\{\hat{\mu}_k^{\#}\}$ to notationally distinguish them from the practical enhanced Ritz values $\{\hat{\mu}_k^*\}$ we employ in this work.

Returning to (4.2), we use its most obvious solution to define our enhanced Ritz values, namely

$$\hat{\mu}_k^* = (1 - \tilde{\eta}_k^2(\hat{S}_m))\, \hat{\mu}_k \ . \tag{4.3}$$

When $S_m = \{\lambda_q\}$, it makes sense to define a single enhanced Ritz value from $\hat{s}_m$ and $\{\tilde{\eta}(\hat{S}_m)^2\}$, and we do so via

$$\hat{\mu}^* = \frac{\sum_{k=1}^m (1 - \tilde{\eta}_k^2(\hat{S}_m))}{\sum_{k=1}^m 1/\hat{\mu}_k} \ . \tag{4.4}$$

In the case of a single, simple eigenvalue ($m = 1$), the definitions (4.3) and (4.4) clearly coincide. Both of these versions are used in the experiments.

We now give the main theorem of this section. It will be a superconvergence estimate which combines an *a posteriori* part, in which approximation defects feature, and an *a priori* part which we use to assert that the part of the Ritz value residual which we cannot compute is of higher order, and can be safely ignored in the asymptotic regime. First we need a regularity result associated with the domain Ω. Let r be the largest number $0 < r \leq 1$ such that, for any $f \in H^{r-1}(\Omega)$,

$$(\omega, v)_1 = (f, v),\ \forall v \in H^1(\Omega) \implies \|\omega\|_{1+r} \leq C\|f\|_{r-1} \leq C\|f\| \ .$$

It is clear that $r = 1$ for convex domains, and it is well-understood how r shrinks as the measure of any interior angles of Ω increase beyond π (cf. [14, 15]).

We now present a result about the superconvergence of enhanced Ritz values to generic, degenerate eigenvalues.

Theorem 4.1. *Let $\hat{S}_m$ be the subspace from which we approximate the single eigenvalue of λ_q of multiplicity m and let the enhanced Ritz values be as in* (4.3)*. Then*

$$\sum_{i=1}^m \frac{\hat{\mu}_i^* - \lambda_q}{\lambda_q} \leq C_{m,B} C_r h_{\mathcal{T}}^{2r} \sum_{i=1}^m \tilde{\eta}_i^2(\hat{S}_m) \ . \tag{4.5}$$

Proof. We now use the results of Appendix B. Let $P_V : \mathcal{H} \to V$ be the L^2 orthogonal projection onto the space of piecewise linear continuous functions on $\mathcal{T}$, $V = V(\mathcal{T})$. We define the operator $\mathcal{W}_V$ so that

$$(\psi, \mathcal{W}_V \phi) = B(\psi, \phi), \qquad \psi, \phi \perp \mathrm{Ran}(P_V) \cap \mathcal{H}$$

holds. Obviously, we can conclude

$$\lim_{h_{\mathcal{T}} \to 0} \|\mathcal{W}_V^{-1/2}\| = 0 \ . \tag{4.6}$$

Let us now apply the trace operator onto the identity (B.8). We obtain

$$\sum_{i=1}^m \frac{\hat{\mu}_i - \lambda_q}{\hat{\mu}_i} = \sum_{i=1}^m \eta_i^2(\hat{S}_m) + F(\eta_1^2(\hat{S}_m), \ldots, \eta_m^2(\hat{S}_m)), \tag{4.7}$$

where $F(\cdot)$ is the real-valued function defined by applying the trace operator on the second term on the right hand side of (B.8). For the function $F(\cdot)$ we have the estimate

$$|F(\eta_1^2(\hat{S}_m), \ldots, \eta_m^2(\hat{S}_m))| \leq \mathrm{RelGap}(\lambda_q, \hat{S}_m) \|\mathcal{W}_V^{-1/2}\|^2 \sum_{i=1}^m \eta_i^2(\hat{S}_m).$$

If we now add and subtract $\frac{\hat{\mu}_i^*}{\hat{\mu}_i}$, $i = 1, \dots, m$ on the left hand side of the equality and note the identity (4.1), we obtain

$$\sum_{i=1}^{m} \frac{|\hat{\mu}_i^* - \lambda_q|}{\hat{\mu}_i} \leq \mathrm{RelGap}(\lambda_q, \hat{S}_m) \|\mathcal{W}_V^{-1/2}\|^2 \sum_{i=1}^{m} \eta_i^2(\hat{S}_m).$$

An application of Theorem 3.2 yields that there is a constant $C(\mathcal{T}, B)$ which solely depends on the shape regularity of $\mathcal{T}$ and the coefficients of B such that

$$\sum_{i=1}^{m} \frac{|\hat{\mu}_i^* - \lambda_q|}{\hat{\mu}_i} \leq C(\mathcal{T}, B)\, \mathrm{RelGap}(\lambda_q, \hat{S}_m) \|\mathcal{W}_V^{-1/2}\|^2 \sum_{i=1}^{m} \tilde{\eta}_i^2(\hat{S}_m). \tag{4.8}$$

To prove superconvergence of $\hat{\mu}_i^*$ it remains to establish an *a priori* estimate for the asymptotic behavior of

$$\|\mathcal{W}_V^{-1/2}\|^2 = \sup_{u \in \mathcal{H}} \frac{\|u - P_V u\|^2}{|\!|\!| u - P_V u |\!|\!|^2} \tag{4.9}$$

as $h_{\mathcal{T}} \to 0$. Let $S_V : \mathcal{H} \to V$ be the H^1 orthogonal projection from $\mathcal{H}$ onto V. For $u \in \mathcal{H}$ it holds that

$$\|u - P_V u\| \leq \|u - S_V u\| \leq C_r h_{\mathcal{T}}^r \|u - S_V u\|_1 \tag{4.10}$$

$$\leq C_r h_{\mathcal{T}}^r \|u - P_V u\|_1 \leq C_r (c_0)^{-1/2} h_{\mathcal{T}}^r |\!|\!| u - P_V u |\!|\!| \ . \tag{4.11}$$

The second inequality in (4.10) uses a standard duality argument (Aubin-Nitsche), where $f = u - S_V u \in \mathcal{H} \subset L^2(\Omega) \subset H^{r-1}$ is used as data for the dual BVP in order to gain the fractional power of $h_{\mathcal{T}}$. We recall here that c_0 is the coercivity constant introduced in Section 2. To define the constant $C_{m,B}$ we switched $\hat{\mu}_i$ in the denominator of (4.8) for λ_q. Q.E.D.

Let us emphasize that the error equation (4.7) contains the measure of the eigenvalue sensitivity — the eigenvalue gap — in the asymptotically higher order term. Subsequently, in the asymptotic regime the estimates are cluster robust, e.g. the asymptotic results are not spoiled by the possibly clustered eigenvalues.

Corollary 4.2. *Let $\lambda_{q-1} < \lambda_q \leq \cdots \leq \lambda_{q+m-1} < \lambda_{q+m}$ and let $\hat{S}_m = \hat{S}_m(\mathcal{T}) = \mathrm{span}(\hat{\phi}_k) \subset V = V(\mathcal{T})$ be the computed approximation of the invariant subspace corresponding to the eigenvalues $\{\lambda_q, \cdots, \lambda_{q+m-1}\}$ from the given mesh $\mathcal{T}$ and let the enhanced Ritz value be as in* (4.3)*. Then*

$$\lim_{h_{\mathcal{T}} \to 0} \frac{\sum_{i=1}^{m} \frac{|\hat{\mu}_i^* - \lambda_{q+i-1}|}{\lambda_{q+i-1}}}{\sum_{i=1}^{m} \frac{|\hat{\mu}_i - \lambda_{q+i-1}|}{\lambda_{q+i-1}}} = 0.$$

Remark 4.3. The claim (4.5) follows from the analysis of $\|\mathcal{W}_V^{-1/2}\|$ as given in (4.9). One can say that (4.9) depends on the regularity property of all eigenvectors. Estimate (4.5) could seriously be improved if we were to obtain more detailed estimates of the quantity $\|\mathcal{W}_V^{-1/2}\big|_{\mathrm{Ran}(\Gamma)}\|^2$ from Theorem B.1. This quantity depends only on the regularity property of the target eigenvectors and could be much smaller than the global quantity (4.9).

Remark 4.4. The condition necessary for the conclusion (4.6) to hold is the non-degeneracy assumption [12, Assumption 6.1]. Namely, we require that there is no open set $U \subset \Omega$ such that the restriction of an eigenfunction $\psi_i\big|_U \in \mathbb{P}_l(U)$ for some $l \in \mathbb{N}$. Here $\mathbb{P}_l$ is the space of polynomials of degree l. This assumption is satisfied by the operators which are defined by a form B for which the matrix-valued function A is continuous and piecewise-$\mathbb{P}_1$ and the potential c is piecewise constant. We can prove the theorem in more general setting by using the fact that we can substitute $\|\mathcal{W}_V^{-1/2}\big|_{\mathrm{Ran}(\Gamma)}\|^2$ for $\|\mathcal{W}_V^{-1/2}\|^2$. The quantity $\|\mathcal{W}_V^{-1/2}\big|_{\mathrm{Ran}(\Gamma)}\|^2$ is local and has robust asymptotic properties, see Appendix B.

4.1. **Eigenvector enhancement.** We now turn to the eigenvector enhancement. The results are somewhat weaker than in the eigenvalue case. For this discussion we assume that we have a cluster of eigenvalues on the lower end of the spectrum so that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m < \lambda_{m+1}$ holds. We further assume that λ_i, $i = 1, \ldots, m$ are approximated by $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \cdots \leq \hat{\lambda}_m$ for which the usual Cauchy inequality holds, e.g $\lambda_i \leq \hat{\lambda}_i$, $i = 1, \ldots, m$.

Associated to $\hat{\lambda}_i$, $i = 1, \ldots, m$ are the Ritz vectors $\hat{\psi}_i$, $i = 1, \ldots, m$. We also use the notation

$$\rho(\phi) = \frac{B(\phi, \phi)}{(\phi, \phi)}, \qquad \phi \in \mathcal{H}$$

for the standard Rayleigh functional. Given the vector $\phi \in \mathcal{H}$ and the eigenvector $\mathcal{A}\psi_i = \lambda_i \psi_i$ the following equality due to Strang holds

$$\frac{\| \phi - \psi_i \|^2}{\| \psi_i \|^2} = \|\phi - \psi_i\|^2 + \frac{\rho(\phi) - \lambda_i}{\lambda_i}$$

which together with the inequality[1]

$$\|\psi_i - \hat{\psi}_i\| \leq \max_{\lambda \in \mathrm{Spec}(\mathcal{A}) \setminus \{\lambda_i\}} \frac{\sqrt{2\lambda\hat{\lambda}_i}}{|\lambda - \hat{\lambda}_i|} \frac{\eta_m(\hat{S}_m)}{\sqrt{1 - \eta_m(\hat{S}_m)}} \tag{4.12}$$

enables us to reduce the eigenvector problem to the eigenvalue problem.

Furthermore, let $\varepsilon_i = \varepsilon(\hat{\psi}_i)$, $i = 1, \ldots, m$ be the associated hierarchical residual approximations. We choose the enhanced Ritz vectors $\hat{\psi}_i^*$, $i = 1, \ldots, m$ as the Ritz vectors associated to the lowermost Ritz values of the form $B(\cdot, \cdot)$ from the subspace

$$\mathrm{span}\{\hat{\psi}_1, \ldots, \hat{\psi}_m, \varepsilon_1, \ldots, \varepsilon_m\}.$$

Using the standard interlacing results we conclude that the inequalities

$$\lambda_i \leq \rho(\hat{\psi}_i^*) \leq \hat{\lambda}_i, \qquad i = 1, \ldots, m$$

hold. If one of the equalities holds, then it must be that $\lambda_i = \rho(\hat{\psi}_i^*) = \hat{\lambda}_i$, and $\hat{\psi}_i = \hat{\psi}_i^* = \psi_i$ is an eigenvector of $\mathcal{A}$. Furthermore, (4.12) together with Theorem 3.8 implies

$$\sum_{i=1}^m \|\psi_i - \hat{\psi}_i^*\|^2 \leq C_{\mathrm{vec}} \sum_{i=1}^m \frac{\rho(\hat{\psi}_i^*) - \lambda_i}{\lambda_i},$$

where constant C_{VEC} depends on the spectral gap—as introduced in (4.12)—and the joint multiplicity m. We can now combine this with Strang's identity to obtain

$$\begin{aligned} \sum_{i=1}^m \frac{\| \hat{\psi}_i^* - \psi_i \|^2}{\| \psi_i \|^2} &\leq C_{\mathrm{VEC}} \sum_{i=1}^m \frac{\rho(\hat{\psi}_i^*) - \lambda_i}{\lambda_i} \\ &\leq C_{\mathrm{VEC}} \Big[\frac{3}{2} \sum_{i=1}^m \frac{\rho(\hat{\psi}_i^*) - \hat{\lambda}_i^*}{\hat{\lambda}_i} + \sum_{i=1}^m \frac{\hat{\lambda}_i^* - \lambda_i}{\lambda_i}. \Big] \end{aligned}$$

where constant C_{VEC} is suitably modified but depends again on the spectral gap and the joint multiplicity m. The constant $\frac{3}{2}$ comes from the equivalence result (B.20). To be able to appreciate the significance of this estimate, note that $\sum_{i=1}^m \frac{\rho(\hat{\psi}_i^*) - \hat{\lambda}_i^*}{\hat{\lambda}_i}$ is a fully *a posteriori* estimate which can be monitored, whereas we know that $\sum_{i=1}^m \frac{\hat{\lambda}_i^* - \lambda_i}{\lambda_i}$ super-converges to zero at a higher rate than—when compared to $\sum_{i=1}^m \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i}$.

[1]Let us note that the inequality (4.12) holds with potentially much sharper residual estimate $\eta_m(\hat{S}_m)$. Namely, if we assume that λ_i has the multiplicity m_i and that $\hat{S}_{m_i}$ are those Ritz vectors from $\hat{S}_m$ which approximate the λ_i of multiplicity m_i, then we can substitute $\eta_{m_i}(\hat{S}_{m_i})$ for $\eta_m(\hat{S}_m)$ in inequality (4.12).

5. Experiments

For the numerical examples given below we have used PLTMG [2] and we have solved the partial eigenvalue problem with ARPACK [23] to compute the approximations, $\hat{s}_m, \hat{S}_m$, of s_m, S_m, and for purposes of error estimation, adaptive refinement, accelerating eigenvalue convergence. Because adaptive refinement is done only in the case of single, simple eigenvalue computations below, the adaptive refinement is driven by local norms of the approximate error function $\varepsilon(\hat{\psi})$ for the computed eigenpair $(\hat{\mu}, \hat{\psi})$. The refinement strategy for clusters of eigenpairs, as described in [18], is not used here. For this section, we use the notation (λ, ϕ), $(\hat{\lambda}, \hat{\phi})$ and $(\hat{\lambda}^*, \hat{\phi}^*)$ for generic eigenpairs and their approximations, instead of the (μ, ψ) notation used in earlier sections to denote eigenpairs at some arbitrary point in the spectrum. In each case, we make explicit which eigenpairs we are considering.

In several cases, both tables and convergence graphs describing the same data are given. Whereas the tables provide more detailed information, the convergence graphs more clearly convey general behavior. As a shorthand for the standard scientific notation $y = x \times 10^m$, we use $y = x_m$ in the tables. Letting λ denote the eigenvalue of interest, $\hat{\lambda}$ its approximation computed in V, and $\hat{\lambda}^*$ the enhanced (or accelerated) approximation, we report the relative errors $E = |\lambda - \hat{\lambda}|/\lambda$ and $E^* = |\lambda - \hat{\lambda}^*|/\lambda$ both graphically and numerically, generally for both uniform and adaptive refinement. In the numerical tables, we also give the reduction factor for E and E^* between successive meshes. Because both the uniform and adaptive refinement schemes increase the number of unknowns by roughly a factor of four, an error reduction factor of 4 corresponds to roughly quadratic convergence, a reduction factor of 8 to roughly cubic convergence, and a reduction factor of 16 to roughly quartic convergence. The numerical results given below can be summarized briefly as follows: accelerated/enhanced eigenvalues are noticeably better than their unaccelerated/unenhanced counterpart, whether or not adaptive refinement is used; in cases where adaptivity is not needed ("smooth" eigenfunctions), the best option is to use acceleration on uniformly refined meshes, because the additional structure of the mesh yields better superconvergence properties; for more singular eigenfunctions, adaptivity truly is needed to get the best behavior out of the accelerated eigenvalues.

5.1. **Simple and Degenerate Eigenvalues on the Unit Square.** The eigenvalues and vectors of the Dirichlet Laplacian on the unit square $\Omega = (0,1) \times (0,1)$,

$$-\Delta\phi = \lambda\phi \ , \ \phi \in H_0^1(\Omega) \ , \ \|\phi\|_{L^2(\Omega)} = 1 \ ,$$

are well-known,

$$\lambda_{mn} = (m^2 + n^2)\pi^2 \quad , \quad \phi_{mn} = 2 \sin m\pi x \, \sin n\pi y \quad , \quad m, n \in \mathbb{N} \ .$$

With this problem, we demonstrate the exceptional performance of the acceleration procedure outlined in Section 4, for both simple and degenerate eigenvalues. Convergence histories are given for the first eigenvalue $\lambda = \lambda_{11} = 2\pi^2$, and the degenerate eigenvalue $\lambda = \lambda_{12} = \lambda_{21} = 5\pi^2$ in Figure 1 and Table 1. In order to obtain a single a single, accelerated eigenvalue $\hat{\lambda}^* \approx \lambda = 5\pi^2$ from the two computed approximations μ_1, μ_2 we use (4.4). Based on the form of $\hat{\lambda}^*$, we use the harmonic mean of μ_1, μ_2 to obtain a single, unaccelerated eigenvalue—although the arithmetic, harmonic and geometric means are nearly identical in this case, and all would yield the same results.

We see from these convergence histories that unaccelerated eigenvalues on uniform meshes perform the worst—i.e. they converge quadratically, as the standard theory predicts. Although adaptive refinement is marginally better for the unaccelerated eigenvalues, it too exhibits quadratic convergence. When the eigenvalues are accelerated, however, the convergence rates improve dramatically, with uniform refinement winning out over adaptivity in the case that adaptivity is used, and optimal convergence rates generally between cubic and quartic!

5.2. **The Dirichlet Laplacian on the L-Shaped Domain.** The much-studied L-shaped domain provides an example for which some of the eigenfunctions have singular behavior near the re-entrant corner, and therefore provide a good test of our adaptivity and eigenvalue acceleration approaches. The problem is

$$-\Delta\phi = \lambda\phi \ , \ \phi \in H_0^1(\Omega) \ , \ \|\phi\|_{L^2(\Omega)} = 1 \ , \tag{5.1}$$

where Ω is the union of three unit squares, pictured in Figure 3 with the contours of six of the eigenfunctions associated with the eigenvalues given below—the eigenfunction $\phi_3 = (2/\sqrt{3}) \sin \pi x \, \sin \pi y$ associated with

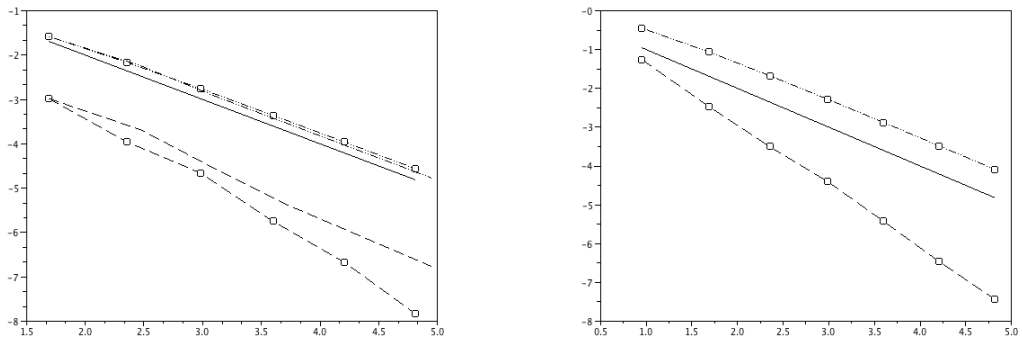

FIGURE 1. Convergence histories for approximations of the smallest eigenvalue and the smallest degenerate eigenvalue for the Square domain. domain (left to right). These are log-log plots of the relative errors $|\hat{\lambda} - \lambda|/\lambda$ (dotted lines) and $|\hat{\lambda}^* - \lambda|/\lambda$ (dashed lines). Lines marked with "$\circ$" are based on uniform refinement. The solid line corresponds to N^{-1}—quadratic convergence.

TABLE 1. Data for the square problem for the smallest two eigenvalues, from top to bottom. For λ_1, the left half of the table corresponds to uniform refinement, and the right half to adaptive refinement. For λ_2, only uniform refinement is done.

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	49	2.62_{-2}		1.05_{-3}		49	2.62_{-2}		1.05_{-3}	
	225	6.89_{-3}	3.80	1.11_{-4}	9.42	296	5.76_{-3}	4.55	2.08_{-4}	5.04
λ_1	961	1.75_{-3}	3.94	2.20_{-5}	5.06	1304	1.16_{-3}	4.97	2.67_{-5}	7.79
	3969	4.39_{-4}	3.98	1.79_{-6}	12.2	5406	2.80_{-4}	4.14	4.07_{-6}	6.55
	16129	1.10_{-4}	4.00	2.13_{-7}	8.43	22039	6.85_{-5}	4.09	8.36_{-7}	4.87
	65025	2.75_{-5}	4.00	1.47_{-8}	14.5	88957	1.70_{-5}	4.02	1.72_{-7}	4.87

	N	RE	red.	RE^*	red.
	49	8.61_{-2}		3.36_{-3}	
	225	2.13_{-2}	4.04	3.16_{-4}	10.6
λ_2	961	5.30_{-3}	4.01	4.05_{-5}	7.79
	3969	1.33_{-3}	4.00	3.77_{-6}	10.7
	16129	3.31_{-4}	4.00	3.44_{-7}	11.0
	65025	8.28_{-5}	4.00	3.74_{-8}	9.20

$\lambda_3 = 2\pi^2$ is not pictured. In [31] several of the eigenvalues in the lower portion of the spectrum are given, accurate to eight digits (up to rounding), and we list them here

i	λ_i	i	λ_i
1	9.6397238	5	31.912636
2	15.197252	6	41.474510
3	19.739209	20	101.60529
4	29.521481		

Convergence histories for the smallest six eigenvalues are given graphically in Figure 4 and numerically in Table 2. Instead of describing the convergence histories for the smallest six eigenvalues in the level of detail provided for the square problem, we provide a few key observations. In cases where the eigenfunctions

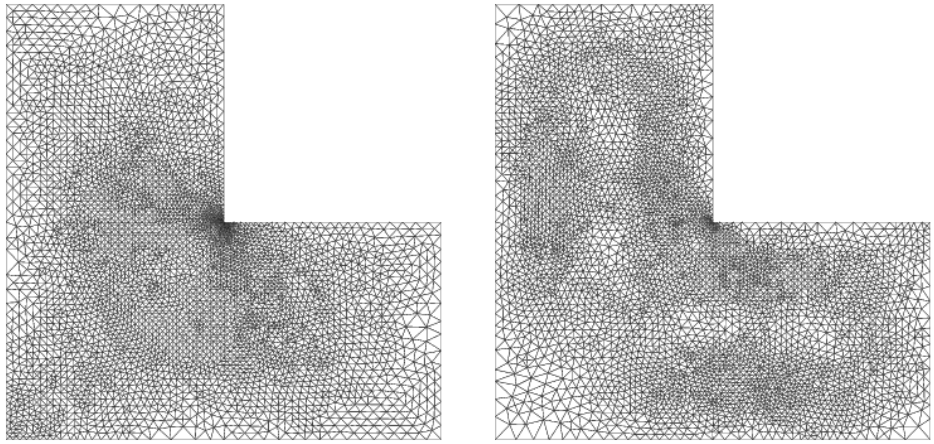

FIGURE 2. Adaptively refined meshes for the first (left) and sixth (right) eigenvalues for the L-shaped; both having roughly 4400 degrees of freedom.

have higher regularity, λ_k for $k = 2, 3, 4$, acceleration on both uniform and adaptive meshes yield better-than-quadratic convergence rates, with uniform refinement being best of all (between cubic and quartic). In cases where the eigenfunctions have lower regularity, λ_k for $k = 1, 5, 6$, it is clear that adaptivity is truly needed to achieve at least quadratic convergence. In these cases, acceleration (with adaptivity) is the only version which yields better-than-quadratic convergence rates, though adaptivity alone recovers the quadratic convergence rate. Since the exact eigenvalues are only given to eight digits—except in the case of $\lambda_3 = 2\pi^2$—reported relative errors below 10^{-8} for λ_2 and λ_4 should be viewed accordingly. This explains the odd behavior below this threshold in the convergence graphs for accelerated eigenvalues under uniform refinement for these eigenvalues. Finally, for λ_{20}, which is clearly singular, the relative errors for $\hat{\lambda}_{20}$ and $\hat{\lambda}^*_{20}$, on an adaptively refined mesh having 72634 degrees of freedom, are 3.27×10^{-4} and 6.33×10^{-6}, respectively.

5.3. **A Schrödinger-Type Operator with Discontinuous Potential.** We shall now consider the eigenvalue problem for the following Schrödinger-type operator, $\mathcal{A}_V w := -\Delta w + cw$,

$$\mathcal{A}_V \phi = \lambda \phi \ , \ \phi \in H^1(\mathbb{R}^2) \ , \ \|\phi\|_{L^2(\mathbb{R}^2)} = 1 \ . \tag{5.2}$$

For properties of such operators and further motivation see [21, 25]. In our example, we have chosen

$$c(x, y) = \begin{cases} 10 + y^2 & |x| > 2 \\ 1 + y^2 & |x| \le 2 \end{cases} .$$

Traditionally, the potential c is denoted by V; but to keep our notation consistent, we have opted for c. Using separation-of-variables, it is determined that the eigenfunctions of $\mathcal{A}_V$ are $C^1(\mathbb{R}^2)$ and decay exponentially away from the origin, and its discrete eigenvalues are $\Sigma = [\Sigma_0 \oplus (2\mathbb{N} - 1)] \cap (1, 10)$, where Σ_0 consists of the four numbers in $(1, 10)$ which satisfy

$$\sqrt{\alpha - 1} \tan(2\sqrt{\alpha - 1}) = \sqrt{10 - \alpha} \quad \text{OR} \quad \sqrt{\alpha - 1} \cot(2\sqrt{\alpha - 1}) = -\sqrt{10 - \alpha}$$

All 12 of these eigenvalues, accurate to 8 digits, are given below.

i	λ_i	i	λ_i
1	2.4520888	7	7.7939697
2	3.7939697	8	7.9717026
3	4.4520888	9	8.4520888
4	5.7939697	10	8.8276737
5	5.9717026	11	9.7939697
6	6.4520888	12	9.9717026

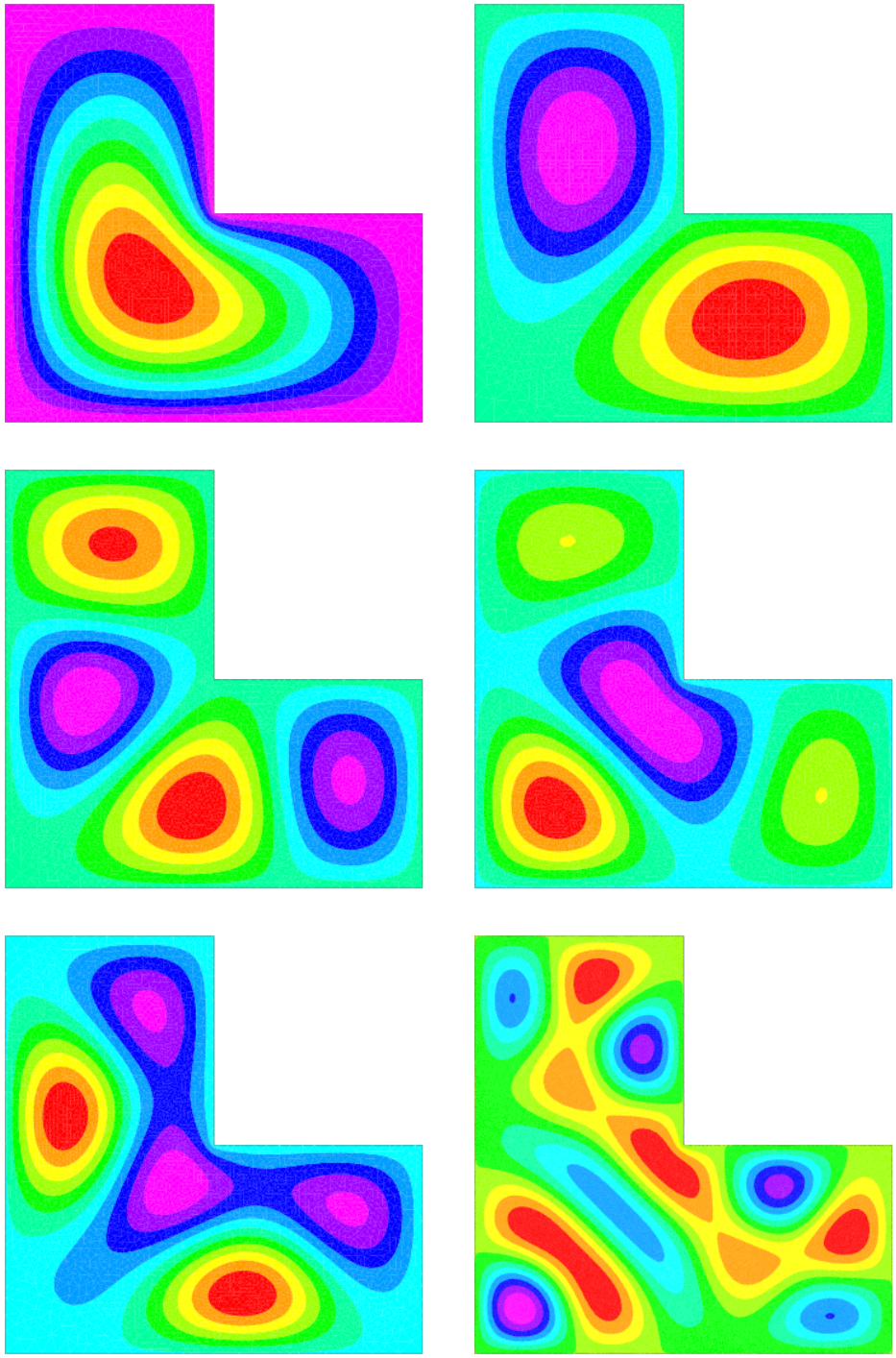

FIGURE 3. Contour plots of the eigenfunctions corresponding to $\lambda_k,\ k = 1, 2, 4, 5, 6, 20$ for the L-Shaped domain (left to right, top to bottom).

TABLE 2. Data for the L-Shape problem for the smallest six eigenvalues, from top to bottom. The left half of each table corresponds to uniform refinement, and the right half to adaptive refinement.

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	33	1.04_{-1}		1.28_{-2}		33	1.04_{-1}		1.28_{-2}	
	161	3.23_{-2}	3.24	3.37_{-3}	3.79	243	1.77_{-2}	5.88	2.10_{-3}	6.08
λ_1	705	1.02_{-2}	3.16	1.38_{-3}	2.44	1054	3.96_{-3}	4.48	2.70_{-4}	7.81
	2945	3.39_{-3}	3.01	5.44_{-4}	2.54	4406	8.30_{-4}	4.77	2.95_{-5}	9.13
	12033	1.17_{-3}	2.88	2.14_{-4}	2.54	17976	1.92_{-4}	4.33	3.56_{-6}	8.29
	48641	4.25_{-4}	2.77	8.48_{-5}	2.53	72638	4.65_{-5}	4.13	6.13_{-7}	5.80

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	33	8.74_{-2}		6.24_{-3}		33	8.74_{-2}		6.24_{-3}	
	161	2.38_{-2}	3.67	7.55_{-4}	8.26	232	1.66_{-2}	5.27	7.76_{-4}	8.04
λ_2	705	6.12_{-3}	3.89	4.32_{-5}	17.5	1050	3.49_{-3}	4.75	1.31_{-4}	5.90
	2945	1.55_{-3}	3.96	8.88_{-6}	4.86	4397	8.15_{-4}	4.28	2.12_{-5}	6.21
	12033	3.88_{-4}	3.98	5.12_{-7}	17.4	17971	1.98_{-4}	4.12	3.93_{-6}	5.38
	48641	9.73_{-5}	3.99	2.35_{-9}	218	72650	4.87_{-5}	4.07	8.25_{-7}	4.77

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	33	1.85_{-1}		2.91_{-1}		33	1.85_{-1}		2.91_{-1}	
	161	3.73_{-2}	4.97	5.06_{-4}	575	232	2.89_{-2}	6.42	1.08_{-3}	27.0
λ_3	705	9.56_{-3}	3.90	1.01_{-5}	50.3	1039	4.60_{-3}	6.29	1.41_{-4}	7.65
	2945	2.40_{-3}	3.98	3.58_{-6}	2.81	4367	1.05_{-3}	4.40	2.02_{-5}	6.97
	12033	6.02_{-4}	3.99	3.86_{-7}	9.26	17918	2.53_{-4}	4.12	3.41_{-6}	5.91
	48641	1.51_{-4}	4.00	9.12_{-8}	4.24	72530	6.28_{-5}	4.03	6.68_{-7}	5.10

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	33	1.57_{-1}		1.99_{-2}		33	1.57_{-1}		1.99_{-2}	
	161	4.48_{-2}	3.50	1.64_{-3}	12.2	232	3.12_{-2}	5.01	1.42_{-3}	14.0
λ_4	705	1.16_{-2}	3.85	1.82_{-4}	8.97	1043	6.94_{-3}	4.50	2.04_{-4}	6.94
	2945	2.94_{-3}	3.96	1.12_{-5}	16.3	4390	1.59_{-3}	4.35	3.25_{-5}	6.28
	12033	7.36_{-4}	3.99	7.60_{-7}	14.7	17976	3.85_{-4}	4.13	5.66_{-6}	5.74
	48641	1.84_{-4}	4.00	1.22_{-7}	6.21	72642	9.52_{-5}	4.05	1.23_{-6}	4.61

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	33	8.83_{-2}		2.91_{-2}		33	8.83_{-2}		5.95_{-2}	
	161	3.26_{-2}	2.71	5.06_{-2}	0.75	232	4.20_{-2}	2.10	3.78_{-3}	15.7
λ_5	705	2.06_{-2}	1.58	1.01_{-3}	6.28	1049	8.79_{-3}	4.78	5.40_{-4}	7.00
	2945	5.81_{-3}	3.55	3.58_{-4}	2.47	4388	1.92_{-3}	4.57	5.97_{-5}	9.06
	12033	1.69_{-3}	3.42	3.86_{-4}	2.72	17956	4.54_{-4}	4.24	9.16_{-6}	6.51
	48641	5.21_{-4}	3.25	9.12_{-5}	2.54	72593	1.10_{-4}	4.11	1.45_{-6}	6.32

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	33	1.80_{-2}		2.42_{-1}		33	1.80_{-2}		2.42_{-1}	
	161	6.15_{-2}	0.29	5.75_{-3}	42.1	236	5.30_{-2}	3.40	4.83_{-3}	50.0
λ_6	705	1.94_{-2}	3.63	9.09_{-4}	6.23	1055	1.02_{-2}	5.17	5.36_{-4}	9.02
	2945	4.64_{-3}	3.65	3.09_{-4}	2.94	4423	2.30_{-3}	4.49	6.22_{-5}	8.63
	12033	1.30_{-3}	3.56	1.11_{-4}	2.77	18031	5.46_{-4}	4.22	9.86_{-6}	6.31
	48641	3.82_{-4}	3.41	4.29_{-5}	2.59	72776	1.33_{-4}	4.10	1.98_{-6}	4.97

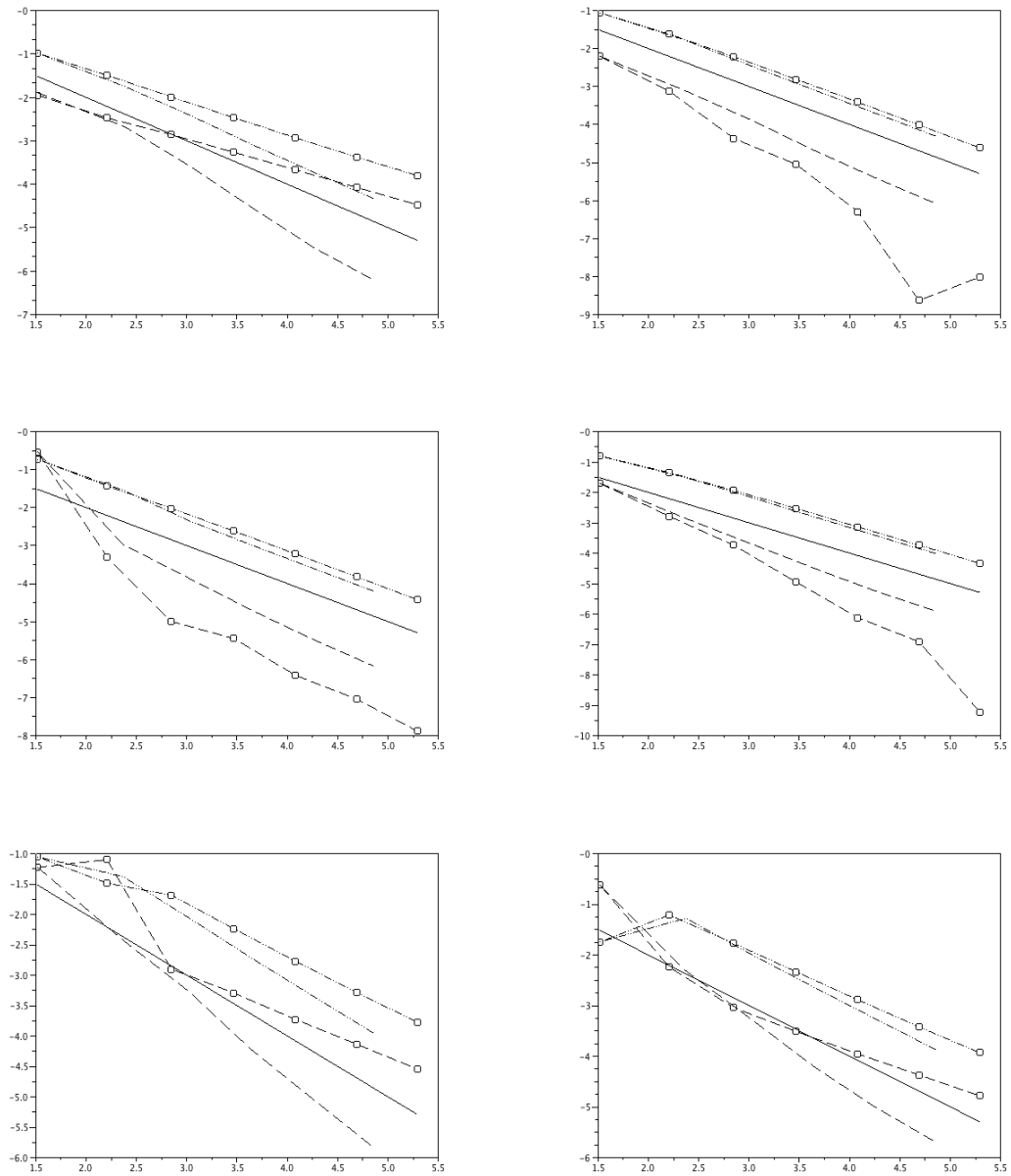

FIGURE 4. Convergence histories for approximations of $\lambda_k,\ k = 1, 2, 3, 4, 5, 6$ for the L-Shaped domain (left to right, top to bottom). These are log-log plots of the relative errors $|\hat{\lambda}_k - \lambda_k|/\lambda_k$ (dotted lines) and $|\hat{\lambda}_k^* - \lambda_k|/\lambda_k$ (dashed lines). Lines marked with "$\circ$" are based on uniform refinement. The solid line corresponds to N^{-1}—quadratic convergence.

We use the strong exponential decay of the eigenfunctions to truncate the original eigenvalue problem to the finite domain $\Omega = (-8, 8) \times (-8, 8)$ *without affecting the accuracy of the eigenvalues reported above.* In particular, we consider the problem

$$\mathcal{A}_V \phi = \lambda \phi \ , \ \phi \in H_0^1(\Omega) \ , \ \|\phi\|_{L^2(\Omega)} = 1 \ . \tag{5.3}$$

We note that any number of (sufficiently large) bounded domains, together with a variety of homogeneous boundary conditions (Dirichlet, Neumann, Robin), would work just as well as the simple ones we have chosen.

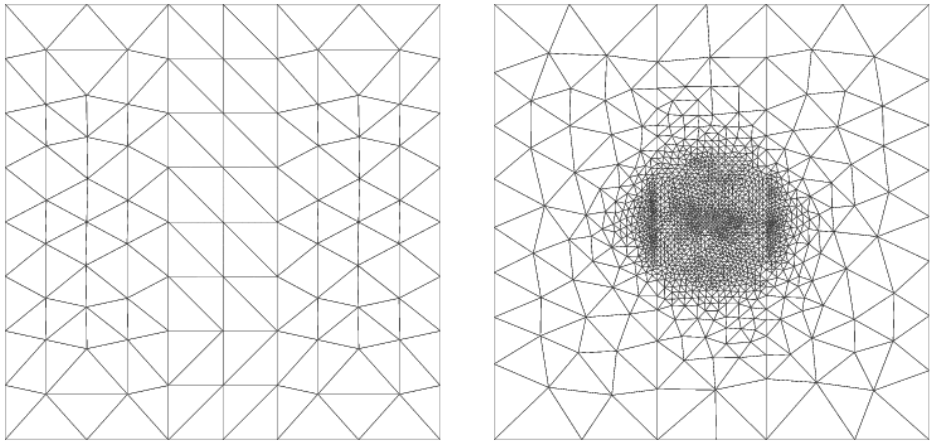

FIGURE 5. The initial mesh for the Schrödinger problem together with the adaptively refined mesh for the smallest eigenvalue, having 1592 degrees of freedom.

TABLE 3. Data for the Schrödinger problem, first eigenvalue (top) and second eigenvalue (bottom). The left half of each table corresponds to uniform refinement, and the right half to adaptive refinement.

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	67	1.60_{-1}		6.62_{-2}		67	1.60_{-1}		6.62_{-2}	
	293	5.96_{-2}	2.69	9.33_{-3}	7.10	372	9.59_{-3}	16.7	8.01_{-4}	82.7
λ_1	1225	1.60_{-2}	3.73	8.77_{-4}	10.6	1592	1.18_{-3}	8.15	3.96_{-5}	20.2
	5009	4.08_{-3}	3.91	5.72_{-5}	15.3	6472	2.60_{-4}	4.52	4.73_{-6}	8.36
	20257	1.02_{-3}	3.98	4.78_{-6}	12.0	25992	6.20_{-5}	4.20	8.07_{-7}	5.86
	81473	2.57_{-4}	3.99	4.08_{-7}	11.7	104072	1.57_{-5}	3.93	1.84_{-7}	4.35

	N	RE	red.	RE^*	red.	N	RE	red.	RE^*	red.
	67	2.99_{-1}		4.45_{-2}		67	2.99_{-1}		4.45_{-1}	
	293	1.31_{-1}	2.28	3.52_{-2}	12.6	372	2.07_{-2}	14.5	1.38_{-3}	323
λ_2	1225	3.86_{-2}	3.41	3.04_{-3}	11.6	1592	2.73_{-3}	7.57	8.98_{-5}	15.3
	5009	1.01_{-2}	3.82	2.52_{-4}	12.1	6472	6.09_{-4}	4.49	1.00_{-5}	8.96
	20257	2.56_{-3}	3.95	1.19_{-5}	21.2	25992	1.46_{-4}	4.17	1.80_{-6}	5.56
	81473	6.42_{-4}	3.99	1.16_{-6}	10.2	104072	3.61_{-5}	4.04	3.99_{-7}	4.52

The domain, divided into the three regions in which c is continuous, the initial mesh and an adaptively refined mesh for λ_1—both of which align with these regions—are given in Figure 5. Contour plots of four of the eigenfunctions are given in Figure 6. Convergence histories for the first two eigenvalues are given in Figure 7. The numerical data on which these graphs are based is given in Table 3. Concerning these convergence histories, we see that both unaccelerated versions (uniform and adaptive refinement) have asymptotic rates of convergence with agree with standard theory, and both accelerated versions converge at more rapidly, with adaptive-accelerated winning out over uniform-accelerated in terms of error on the final mesh, but the reverse holding (asymptotically) in terms of convergence rate. In the computation for the second eigenvalue on the coarsest mesh, it is actually the lowest eigenpair which is approximated—which accounts for the massive error reduction in the first refinement, and the fact that the accelerated eigenvalue is actually worse than the original one on this mesh.

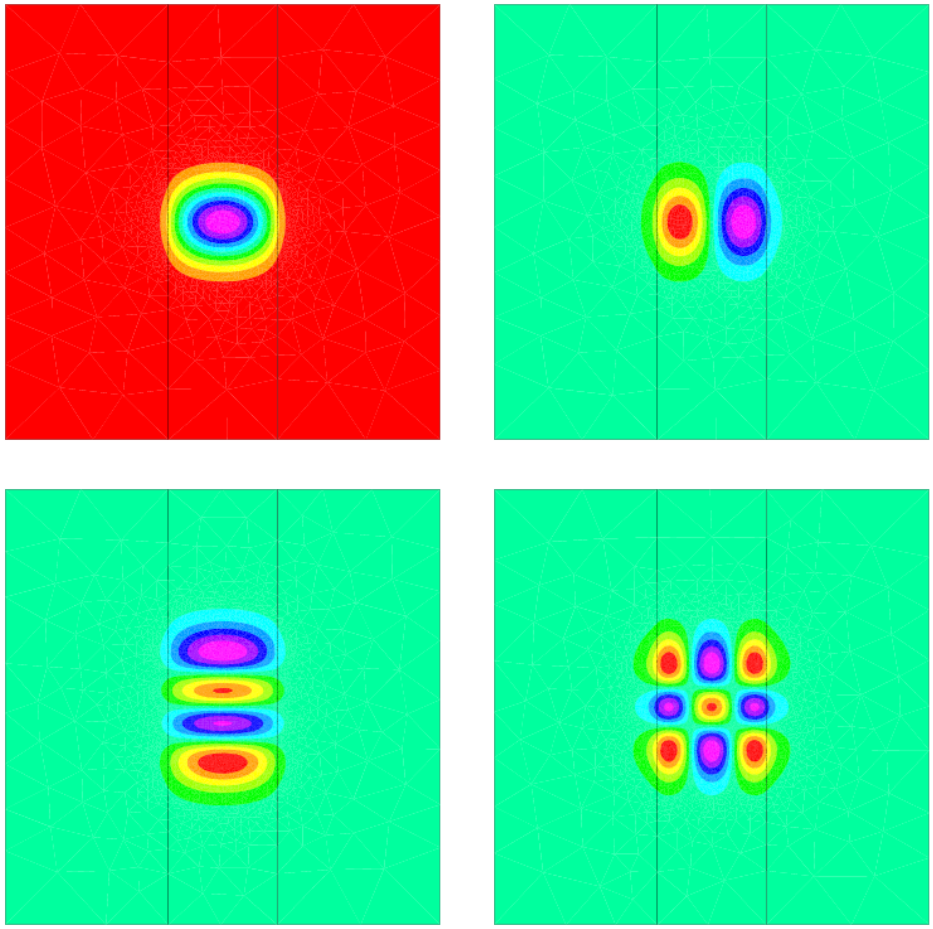

FIGURE 6. Contour plots of the first, second, ninth and twelfth eigenfunctions (left to right, top to bottom) for the Schrödinger problem.

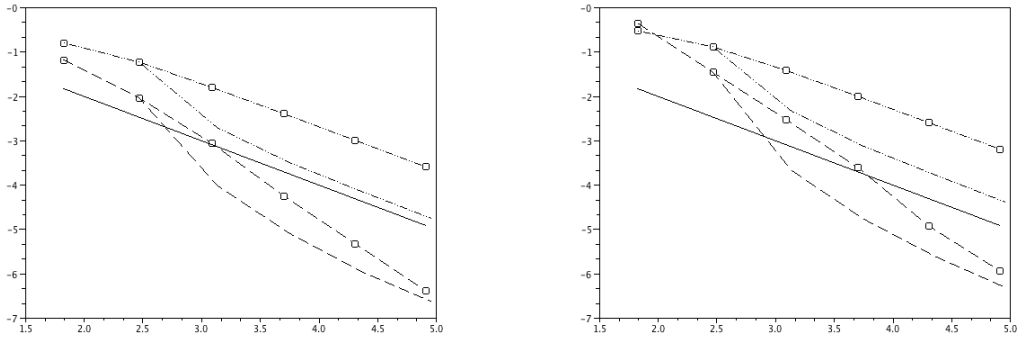

FIGURE 7. Convergence histories for approximations of $\lambda_k,\ k = 1, 2$ for the Schrödinger problem (left to right). These are log-log plots of the relative errors $|\hat{\lambda}_k - \lambda_k|/\lambda_k$ (dotted lines) and $|\hat{\lambda}_k^* - \lambda_k|/\lambda_k$ (dashed lines). Lines marked with "$\circ$" are based on uniform refinement. The solid line corresponds to N^{-1}—quadratic convergence.

Remark 5.1. Although the operator on the unbounded domain has only twelve *discrete* eigenvalues, the operator on the truncated domain has infinitely many. It is only the first twelve which are physically relevant, as they coincide with high accuracy to those of the unbounded domain.

Acknowledgement

R.B. was supported by the U. S. National Science Foundation under contract DMS-0915220, and the Alexander Humboldt Foundation through a Humboldt Research Award.

L.G. was supported by the grant: "Spectral decompositions – numerical methods and applications", Grant Nr. 037-0372783-2750 of the Croatian MZOS.

J.O. thanks the Max Planck Institute for Mathematics, in Leipzig, Germany, for graciously hosting him while this manuscript was being completed. Thanks are also due to the University of Kentucky Mathematics Department, and the College of Arts and Sciences, for additional funding during the final stages of this project.

References

[1] R. E. Bank. Hierarchical bases and the finite element method. In *Acta numerica, 1996*, volume 5 of *Acta Numer.*, pages 1–43. Cambridge Univ. Press, Cambridge, 1996.

[2] R. E. Bank. PLTMG: A software package for solving elliptic partial differential equations. users' guide 10.0. Technical report, University of California at San Diego, Department of Mathematics, 2007.

[3] R. E. Bank and R. K. Smith. A posteriori error estimates based on hierarchical bases. *SIAM J. Numer. Anal.*, 30(4):921–935, 1993.

[4] F. A. Bornemann, B. Erdmann, and R. Kornhuber. A posteriori error estimates for elliptic problems in two and three space dimensions. *SIAM J. Numer. Anal.*, 33(3):1188–1204, 1996.

[5] P. G. Ciarlet and J.-L. Lions, editors. *Handbook of numerical analysis. Vol. II.* Handbook of Numerical Analysis, II. North-Holland, Amsterdam, 1991. Finite element methods. Part 1.

[6] W. Dahmen. Multiskalen-Methoden und Wavelets—Konzepte und Anwendungen. *Jahresber. Deutsch. Math.-Verein.*, 97(3):97–114, 1995.

[7] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[8] P. Deuflhard, P. Leinen, and H. Yserentant. Concepts of an adaptive hierarchical finite element code. *IMPACT Comput. Sci. Eng.*, 1(1):3–35, 1989.

[9] R. A. DeVore and B. J. Lucier. Wavelets. In *Acta numerica, 1992*, Acta Numer., pages 1–56. Cambridge Univ. Press, Cambridge, 1992.

[10] W. Dörfler and R. H. Nochetto. Small data oscillation implies the saturation assumption. *Numer. Math.*, 91(1):1–12, 2002.

[11] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[12] E. M. Garau, P. Morin, and C. Zuppa. Convergence of adaptive finite element methods for eigenvalue problems. *Math. Models Methods Appl. Sci.*, 19(5):721–747, 2009.

[13] S. Giani and I. G. Graham. A convergent adaptive method for elliptic eigenvalue problems. *SIAM J. Numer. Anal.*, 47(2):1067–1091, 2009.

[14] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[15] P. Grisvard. *Singularities in boundary value problems*, volume 22 of *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*. Masson, Paris, 1992.

[16] L. Grubišić. On eigenvalue and eigenvector estimates for nonnegative definite operators. *SIAM J. Matrix Anal. Appl.*, 28(4):1097–1125 (electronic), 2006.

[17] L. Grubišić. On Temple–Kato like inequalities and applications. *arXiv:math/0511408v2*, preprint:1–22, 2009.

[18] L. Grubišić and J. S. Ovall. On estimators for eigenvalue/eigenvector approximations. *Math. Comp.*, 78:739–770, 2009.

[19] V. Heuveline and R. Rannacher. A posteriori error control for finite approximations of elliptic eigenvalue problems. *Adv. Comput. Math.*, 15(1-4):107–138 (2002), 2001. A posteriori error estimation and adaptive computational methods.

[20] T. Kato. *Perturbation theory for linear operators*. Springer-Verlag, Berlin, second edition, 1976. Grundlehren der Mathematischen Wissenschaften, Band 132.

[21] T. Koprucki, R. Eymard, and J. Fuhrmann. Convergence of a finite volume scheme to the eigenvalues of a schroedinger operator. Technical report, WIAS Preprint No. 1260, 2007.

[22] M. G. Larson. A posteriori and a priori error analysis for finite element approximations of self-adjoint elliptic eigenvalue problems. *SIAM J. Numer. Anal.*, 38(2):608–625 (electronic), 2000.

[23] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK users' guide*, volume 6 of *Software, Environments, and Tools*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods.

[24] D. Mao, L. Shen, and A. Zhou. Adaptive finite element algorithms for eigenvalue problems based on local averaging type a posteriori error estimates. *Adv. Comput. Math.*, 25(1-3):135–160, 2006.
[25] A. Messiah. *Quantum mechanics (Two volumes bound as one)*. Dover Publications, 1999.
[26] A. Naga, Z. Zhang, and A. Zhou. Enhancing eigenvalue approximation by gradient recovery. *SIAM J. Sci. Comput.*, 28(4):1289–1300 (electronic), 2006.
[27] K. Neymeyr. A posteriori error estimation for elliptic eigenproblems. *Numer. Linear Algebra Appl.*, 9(4):263–279, 2002.
[28] J. S. Ovall. Function, gradient, and Hessian recovery using quadratic edge-bump functions. *SIAM J. Numer. Anal.*, 45(3):1064–1080 (electronic), 2007.
[29] B. N. Parlett. *The symmetric eigenvalue problem*, volume 20 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Corrected reprint of the 1980 original.
[30] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.
[31] L. N. Trefethen and T. Betcke. Computed eigenmodes of planar regions. In *Recent advances in differential equations and mathematical physics*, volume 412 of *Contemp. Math.*, pages 297–314. Amer. Math. Soc., Providence, RI, 2006.
[32] C. Tretter. *Spectral theory of block operator matrices and applications*. Imperial College Press, London, 2008.
[33] H. Wu and Z. Zhang. Enhancing eigenvalue approximation by gradient recovery on adaptive meshes. *IMA J. Numer. Anal.*, 29(4):1008–1022, 2009.

Appendix A. Technical Finite Element Results

A.1. **Basic Geometric Results.** We implicitly assume that any vertex in $\mathcal{V}_D$ will have at least one adjacent vertex in $\mathcal{V}$. Such an assumption is very natural, and simple to enforce in practice. Let $T \in \mathcal{T}$ be given, having vertices z_k, angles θ_k, and opposite edges e_k, $k = 1, 2, 3$ We further take $\ell_k = \ell_{z_k}$, $b_k = b_{e_k}$, and $\mathbf{n}_k$ to be the outward unit normal to e_k (with respect to T). Throughout this manuscript, the notation $|X|$ will be used to denote the length of a curve, the area of a region, the cardinality of a (finite) set, or the Euclidean norm of a vector, and the appropriate interpretation should be clear from the context. We collect several well-known results which will be used liberally, often without explicit reference, in later arguments.

Lemma A.1. *On T, for $j \neq k$ we have*

$$\ell_k = 1 - \frac{1}{d_k}(x - z_k) \cdot \mathbf{n}_k = -\frac{1}{d_k}(x - z_j) \cdot \mathbf{n}_k = (x - z_j) \cdot \nabla \ell_k \ . \tag{A.1}$$

The ratio $d_k := \frac{2|T|}{|e_k|}$ is the altitude of the triangle with respect to the base e_k. Let $p, q, r \in \mathbb{Z}_{\geq 0}$. The following hold:

$$\int_T \ell_1^p \ell_2^q \ell_3^r = \frac{2|T| p! q! r!}{(p+q+r+2)!} \qquad \int_{e_k} \ell_{k-1}^p \ell_{k+1}^q = \frac{|e_k| p! q!}{(p+q+1)!} \tag{A.2}$$

$$\nabla \ell_k \cdot \nabla \ell_k = \frac{\cot\theta_{k-1} + \cot\theta_{k+1}}{2|T|} \qquad \nabla \ell_{k-1} \cdot \nabla \ell_{k+1} = -\frac{\cot\theta_k}{2|T|} \tag{A.3}$$

$$\int_T \nabla b_k \cdot \nabla b_k = \frac{4}{3}(\cot\theta_1 + \cot\theta_2 + \cot\theta_3) \qquad \int_T \nabla b_{k-1} \cdot \nabla b_{k+1} = -\frac{4}{3}\cot\theta_k \tag{A.4}$$

For any $v \in H^1(T)$, $d_k \int_{e_k} v = \int_T 2v + (x - z_k) \cdot \nabla v$.

Proof. Using an affine change of variables $[0, 1] \mapsto e_k$, we see that the second result in (A.2) is just the classical beta function. A similar affine change of variables and induction yields the first result in (A.2). The other explicitly geometric results are simple consequences of (A.1). The final result is an application of the Divergence Theorem. Q.E.D.

A.2. **Quasi-Interpolant Results.**

Lemma A.2. *Let $v \in \mathcal{H}$. There is a quasi-interpolant $\mathcal{I}v = \hat{v} + \hat{w} \in V(\mathcal{T}) \oplus W(\mathcal{T})$, with $\hat{v} = \sum_{z \in \bar{\mathcal{V}}} v_z$ and $\hat{w} = \sum_{z \in \bar{\mathcal{V}}} w_z$, where $v_z \in V(\mathcal{T})$ and $w_z \in W(\mathcal{T})$, satisfying the zero-mean properties:*

$$\int_{\omega_z} (v\ell_z - v_z - w_z) = 0 \text{ for each } z \in \bar{\mathcal{V}} \ ,$$

$$\int_e (v - \hat{v} - \hat{w}) = 0 \text{ for each } e \in \mathcal{E} \ .$$

We can (and will) choose v_z, w_z to be supported in ω_z for $z \in \mathcal{V}$; and choose v_z, w_z to be supported in $\omega_{z'}$ for $z \in \mathcal{V}_D$, where $z' \in \mathcal{V}$ is shares an edge $e \in \mathcal{E}$ with z.

Proof. We first note that, if $\int_e (v\ell_z - v_z - w_z) = 0$ for each $e \in \mathcal{E}_z$ and each $z \in \bar{\mathcal{V}}$, then $\int_e (v - \hat{v} - \hat{w}) = 0$ for each $e \in \mathcal{E}$ (in fact, for each $e \in \bar{\mathcal{E}}$). This is so because, if z, z' are the endpoints of e, then

$$v|_e = (v\ell_z + v\ell_{z'})|_e \quad , \quad \hat{v}|_e = (v_z + v_{z'})|_e \quad , \quad \hat{w}|_e = (w_z + w_{z'})|_e \ .$$

Let $z \in \mathcal{V}$ be given. We have $v_z = \alpha_z \ell_z$ and $w_z = \sum_{e \in \mathcal{E}_z} \beta_{ez} b_e$, so the equations which must be satisfied for the zero-mean condition to hold are

$$\frac{|\omega_z|}{3}\,\alpha_z + \sum_{e \in \mathcal{E}_z} \frac{|\omega_e|}{3}\,\beta_{ez} = \int_{\omega_z} v\ell_z \tag{A.5}$$

$$\frac{|e|}{2}\,\alpha_z + \frac{2|e|}{3}\,\beta_{ez} = \int_e v\ell_z \text{ for all } e \in \mathcal{E}_z \ . \tag{A.6}$$

These can be solved explicitly, and yield

$$\alpha_z = \frac{6}{|\omega_z|}\left(\sum_{e \in \mathcal{E}_z} \frac{|\omega_e|}{2|e|}\int_e v\ell_z - \int_{\omega_z} v\ell_z\right) \quad , \quad \beta_{ez} = \frac{3}{2|e|}\int_e v\ell_z - \frac{3}{4}\,\alpha_z \ . \tag{A.7}$$

We also note that, if $v \in V(\mathcal{T})$, then $v\ell_z \in V(\mathcal{T}) \oplus W(\mathcal{T})$ is supported in ω_z, and the conditions on $e \in \mathcal{E}_z$ and ω_z force $v\ell_z = v_z + w_z$. Here and elsewhere, ω_e denotes the support of b_e—the one or two triangles adjacent to e.

For $z \in \mathcal{V}_D$, we can maintain the zero-mean condition $\int_{\omega_z} (v\ell_z - v_z - w_z) = 0$, provided z has at least one adjacent vertex $z' \in \mathcal{V}$. This is easy to enforce in practice, and a very natural assumption. Let $z' \in \mathcal{V}$ be adjacent to z, with common edge e, and let $v_z = \alpha_z \ell_{z'}$ and $w_z = \sum_{e' \in \mathcal{E}_{z'}} \beta_{e'z} b_{e'}$. The equations which must be satisfied for the zero-mean conditions are

$$\frac{|\omega_{z'}|}{3}\,\alpha_z + \sum_{e' \in \mathcal{E}_{z'}} \frac{|\omega_{e'}|}{3}\,\beta_{e'z} = \int_{\omega_z} v\ell_z \tag{A.8}$$

$$\frac{|e'|}{2}\,\alpha_z + \frac{2|e'|}{3}\,\beta_{e'z'} = \int_{e'} v\ell_z = \delta_{ee'}\int_e v\ell_z \text{ for all } e' \in \mathcal{E}_{z'} \ . \tag{A.9}$$

Just as before, we solve this system to obtain

$$\alpha_z = \frac{6}{|\omega_{z'}|}\left(\frac{|\omega_e|}{2|e|}\int_e v\ell_z - \int_{\omega_z} v\ell_z\right) \quad , \quad \beta_{e'z} = -\frac{3}{4}\,\alpha_z + \frac{3\delta_{ee'}}{2|e|}\int_e v\ell_z \ . \tag{A.10}$$

We emphasize in (A.10) that we integrate $v\ell_z$ over ω_z, not $\omega_{z'}$. Q.E.D.

Theorem A.3. *Let $v \in \mathcal{H}$, and $\mathcal{I}v$ be the quasi-interpolant described in Lemma A.2. There are scale-invariant constants c_{1z}, c_{2z}, c_{3z}, c_{4z} and C_1 such that*

(1) $|w_z|_1 \le c_{1z}|v|_{1,\omega_z}$ *and* $\|\hat{w}\|_1 \le C_1|v|_1$
(2) $|v\ell_z - v_z - w_z|_1 \le c_{2z}|v|_{1,\omega_z}$
(3) $\|v\ell_z - v_z - w_z\|_0 \le c_{3z}d_z|v|_{1,\omega_z}$
(4) $\|v - \hat{v} - \hat{w}\|_{0,e} \le c_{4e}|e|^{1/2}|v|_{1,\omega_z \cup \omega_{z'}}$, *where z and z' are the endpoints of e.*

Proof. (of (1)). We first consider $z \in \mathcal{V}$. We begin by re-expressing the coefficients α_z and β_{ez} in a form which is more convenient for analysis. Given $e \in \mathcal{E}_z$, ω_e consists of the one or two triangles adjacent to e. Suppose that there are two such triangles T_e and $\hat{T}_e$, and that the vertices opposite e in these triangles are z_e and $\hat{z}_e$, respectively. We define the function $\mathbf{d}_e$ on ω_e by

$$\mathbf{d}_e = \begin{cases} x - z_e & x \in T_e \\ x - \hat{z}_e & x \in \hat{T}_e \end{cases} \ .$$

The version when ω_e consists of one triangle is defined analogously. Using this definition, we obtain

$$\alpha_z = \frac{3}{2|\omega_z|}\sum_{e \in \mathcal{E}_z}\int_{\omega_e} \mathbf{d}_e \cdot \nabla(v\ell_z) = \kappa_z + \frac{3}{2|\omega_z|}\sum_{e \in \mathcal{E}_z}\int_{\omega_e} \ell_z \mathbf{d}_e \cdot \nabla v \ ,$$

where $\kappa_z = \frac{3}{|\omega_z|}\int_{\omega_z} v\ell_z$ is the weighted-average of v on ω_z. Therefore,

$$\begin{aligned}\beta_{ez} &= \frac{3}{2|e|}\int_e (v-\kappa_z)\ell_z - \frac{9}{8|\omega_z|}\sum_{\hat{e}\in\mathcal{E}_z}\int_{\omega_{\hat{e}}} \ell_z \mathbf{d}_{\hat{e}}\cdot\nabla v \\ &= \frac{9}{4|\omega_e|}\int_{\omega_e}(v-\kappa_z)\ell_z + \frac{3}{4|\omega_e|}\int_{\omega_e}\ell_z\mathbf{d}_e\cdot\nabla v - \frac{9}{8|\omega_z|}\sum_{\hat{e}\in\mathcal{E}_z}\int_{\omega_{\hat{e}}}\ell_z\mathbf{d}_{\hat{e}}\cdot\nabla v\ .\end{aligned}$$

We note that $|w_z|^2_{1,\omega_z} = \beta_z^T A_z \beta_z$, where β_z is the coefficient vector of w_z with respect to $\{b_e : e\in\mathcal{E}_z\}$, and $(A_z)_{ee'} = \int_{\omega_z}\nabla b_e\cdot\nabla b_{e'}$. The eigenvalues of the $|\mathcal{E}_z|\times|\mathcal{E}_z|$ matrix A_z are bounded above by a scale-invariant constant, independent of the mesh (assuming shape-regularity of the family), so we bound $|w_z|_{1,\omega_z}$ by bounding the sizes of the coefficients β_{ez}. We have

$$\begin{aligned}\beta_{ez} \le{}& \frac{9\|\ell_z^{1/2}\|_{0,\omega_e}}{4|\omega_e|}\|(v-\kappa_z)\ell_z^{1/2}\|_{0,\omega_e} + \left|\frac{3}{4|\omega_e|}-\frac{9}{8|\omega_z|}\right|\|\ell_z\mathbf{d}_e\|_{0,\omega_e}|v|_{1,\omega_e} \\ &+ \frac{9}{8|\omega_z|}\sum_{\hat{e}\in[\mathcal{E}_z\setminus\{e\}]}\|\ell_z\mathbf{d}_e\|_{0,\omega_{\hat{e}}}|v|_{1,\omega_{\hat{e}}}\ .\end{aligned}$$

We have

$$\begin{aligned}\|(v-\kappa_z)\ell_z^{1/2}\|_{0,\omega_e} \le \|(v-\kappa_z)\ell_z^{1/2}\|_{0,\omega_z} &= \inf_{a\in\mathbb{R}}\|(v-a)\ell_z^{1/2}\|_{0,\omega_z} \\ &\le \inf_{a\in\mathbb{R}}\|v-a\|_{0,\omega_z} \le K_z d_z |v|_{1,\omega_z}\ ,\end{aligned}$$

where we have used the Poincaré inequality in the last bound. Therefore, we deduce that β_{ez} can be bounded by a scale-invariant constant times $|v|_{1,\omega_z}$, which completes the proof of the first claim for $z\in\mathcal{V}$.

The argument for $z\in\mathcal{V}_D$ is actually slightly simpler because of the fact that v vanishes on some portion of $\partial\omega_z$ having non-zero length. We have

$$|v\ell_z - v_z - w_z|_1 \le |v\ell_z|_{1,\omega_z} + |v_z + w_z|_{1,\omega_{z'}}\ .$$

A Poincaré-Friedrichs' inequality gives $|v\ell_z|^2_{1,\omega_z} = |\ell_z|^2_{1,\omega_z}\|v\|^2_{0,\omega_z} + \|(\nabla v)\ell_z\|_{0,\omega_z} \le k_z|v|^2_{1,\omega_z}$. As before, we know that bounds on α_z and $\beta_{e'z}$ involving scale-invariant constants times $|v|_{1,\omega_z}$ will yield a bound on $|v_z+w_z|_{1,\omega_{z'}}$ involving a scale-invariant constant times $|v|_{1,\omega_z}$. In particular, we have

$$\begin{aligned}|\alpha_z| &\le \frac{6}{|\omega_{z'}|}\left|\frac{|\omega_e|}{2|e|}\int_e v\ell_z\right| + \frac{6\|v\|_{0,\omega_z}\|\ell_z\|_{0,\omega_z}}{|\omega_{z'}|} \\ &\le \frac{6}{|\omega_{z'}|}\left|\int_{\omega_e} 2v\ell_z + \mathbf{d}_e\cdot\nabla(v\ell_z)\right| + \frac{\sqrt{6}c_z d_z|\omega_z|^{1/2}}{|\omega_{z'}|}|v|_{1,\omega_z} \\ &\le \frac{6}{|\omega_{z'}|}\left|\int_{\omega_e} 3v\ell_z + \ell_z\mathbf{d}_e\cdot\nabla v\right| + \frac{\sqrt{6}c_z d_z|\omega_z|^{1/2}}{|\omega_{z'}|}|v|_{1,\omega_z} \le \hat{c}_z|v|_{1,\omega_z}\ .\end{aligned}$$

In these estimates, we freely used Poincaré-Friedrichs' inequalities to bound $\|v\|_{0,\omega}$ in terms of d_z and $|v|_{1,\omega_z}$. The coefficients $\beta_{e'z}$ are also clearly bounded by scale-invariant constants times $|v|_{1,\omega_z}$, so we have proved the first part of (1) for all $z\in\bar{\mathcal{V}}$.

Standard inverse estimates guarantee the existence of a scale-invariant constant k_{1z} such that $\|w_z\|_{1,\omega_z} \le \sqrt{k_{1z}^2|\omega_z| + c_{1z}^2}|v|_{1,\omega_z}$. Using the discrete Cauchy-Schwarz inequality, we can take $C_1^2 = 3\max_{z\in\mathcal{V}}(k_{1z}^2|\omega_z| + c_{1z}^2)$, which completes the proof of the second claim. Q.E.D.

Proof. (of (2)). Let $\mathbb{P}_0^2$ consist of componentwise piecewise-constant functions. We first note that, for any $\mathbf{F}\in\mathbb{P}_0^2$,

$$\int_{\omega_z}\mathbf{F}\cdot\nabla(v\ell_z - v_z - w_z) = \sum_{T\subset\omega_z}\int_{\partial T}\mathbf{F}\cdot\mathbf{n}_T(v\ell_z - v_z - w_z) = 0.$$

Therefore, we have

$$\begin{aligned}|v\ell_z - v_z - w_z|_{1,\omega_z} &\le \inf_{\mathbf{F}\in\mathbb{P}_0^2} \|\nabla(v\ell_z - w_z) - \mathbf{F}\|_{0,\omega_z} \\ &\le |w_z|_{1,\omega_z} + \inf_{\mathbf{F}\in\mathbb{P}_0^2} \|\nabla(v\ell_z) - \mathbf{F}|_{0,\omega_z} \\ &\le |w_z|_{1,\omega_z} + \left(\sum_{T\subset\omega_z} |(v - v_T)\ell_z|_{1,T}^2\right)^{1/2} ,\end{aligned}$$

where v_T is the average value of v on T. Using (1), we bound $|w_z|_{1,\omega_z}$ in terms of $|v|_{1,\omega_z}$, so we need only consider $|(v - v_T)\ell_z|_{1,T}^2$. We have

$$|(v - v_T)\ell_z|_{1,T}^2 \le 2|v|_{1,T}^2 + 2|(\nabla\ell_z)_{|T}|^2 \, \|v - v_T\|_{0,T}^2 \le 2\left(1 + |(\nabla\ell_z)_{|T}|^2 \, \frac{h_T^2}{\pi^2}\right) |v|_{1,T}^2 \ .$$

The bound $\|v - v_T\|_{0,T}^2 \le \frac{h_T^2}{\pi^2} \, |v|_{1,T}^2$, where h_T is the longest edge of T, is due to a Poincaré inequality [30]. Q.E.D.

Proof. (of (3)). Noting that $v\ell_z - v_z - w_z$ has zero-mean on ω_z, we use a Poincaré inequality to establish that $\|v\ell_z - v_z - w_z\|_{0,\omega_z} \le c_z d_z |v\ell_z - v_z - w_z|_{1,\omega_z}$. Combining this with (2) yields (3). Q.E.D.

Remark A.4. The optimal constant $c_z d_z$ is related to the Dirichlet eigenvalue problem $-\Delta\phi = \lambda\phi$ on $\tilde{\omega}_z = \omega_z$ (if $z \in \mathcal{V}$) or on $\tilde{\omega}_z = \omega_z \cup \omega_{z'}$ (if $z \in \mathcal{V}_D$ and z' is the vertex adjacent z which was chosen for the quasi-interpolant). In particular, $c_z d_z$ is bounded by the reciprocal of the smallest eigenvalue. The bound is generally strict because of the additional zero-mean conditions.

Proof. (of (4)). Using the final claim of Lemma A.1 and the notation $\mathbf{d}_e$ from the proof of (1) we have

$$\begin{aligned}\|v\ell_z - v_z - w_z\|_{0,e}^2 &= \frac{|e|}{2|\omega_e|} \int_{\omega_e} 2(v\ell_z - v_z - w_z)^2 + \mathbf{d}_e \cdot \nabla[(v\ell_z - v_z - w_z)^2] \\ &\le \frac{|e|}{2|\omega_e|} \int_{\omega_e} 3(v\ell_z - v_z - w_z)^2 + [\mathbf{d}_e \cdot \nabla(v\ell_z - v_z - w_z)]^2 \\ &\le \frac{|e|}{2|\omega_e|} \left(3c_{3z}^2 d_z^2 + c_{2z}^2 \|\mathbf{d}_e\|_{0,\omega_e}^2\right) |v|_{1,\omega_z}^2\end{aligned}$$

The fact that $\|v - \hat{v} - \hat{w}\|_{0,e} \le \|v\ell_z - v_z - w_z\|_{0,e} + \|v\ell_{z'} - v_{z'} - w_{z'}\|_{0,e}$, where z and z' are the endpoints of e finishes the proof. Q.E.D.

A.3. **Key BVP Error Estimation Result.**

Lemma A.5. *There are scale-invariant constants $K_1 = K_1(\mathcal{T}, B)$ and $K_2 = K_2(\mathcal{T})$ such that*

$$\begin{aligned}B(u - \hat{u}, v) &\le K_1 |\!|\!|\varepsilon|\!|\!| \, |\!|\!|v|\!|\!| + K_2 \operatorname{osc}(R, r)|v|_1 \ , \\ [\operatorname{osc}(R, r)]^2 &= \sum_{z\in\bar{\mathcal{V}}} d_z^2 \inf_{R_z\in\mathbb{R}} \|R - R_z\|_{0,\omega_z}^2 + \sum_{e\in\mathcal{E}} |e| \inf_{r_e\in\mathbb{R}} \|r - r_e\|_{0,e}^2 \ ,\end{aligned}$$

where d_z is the diameter of ω_z.

Proof. Recalling the key error equation, we have

$$B(u - \hat{u}, v) = B(\varepsilon, \hat{w}) + \sum_{z\in\bar{\mathcal{V}}} \int_{\omega_z} R(v\ell_z - v_z - w_z) \, dV + \sum_{e\in\mathcal{E}} \int_e r(v - \hat{v} - \hat{w}) \, dS \ .$$

The term $|B(\varepsilon, \hat{w})|$ is easily bounded by $|\!|\!|\varepsilon|\!|\!| \, |\!|\!|\hat{w}|\!|\!| \le K_1 |\!|\!|\varepsilon|\!|\!| \, |\!|\!|v|\!|\!|$, where K_1 is related to both the constant in $\|\hat{w}\|_1 \le C_1 \|v\|_1$ and the coercivity and boundedness constants in $m\|v\|_1^2 \le |\!|\!|v|\!|\!|^2 \le M\|v\|_1^2$. We also have the bounds

$$\begin{aligned}\left|\int_{\omega_z} R(v\ell_z - v_z - w_z) \, dV\right| &= \left|\int_{\omega_z} (R - R_z)(v\ell_z - v_z - w_z) \, dV\right| \le c_{3z} d_z \|R - R_z\|_{0,\omega_z} \\ \left|\int_e r(v - \hat{v} - \hat{w}) \, dS\right| &= \left|\int_e (r - r_e)(v - \hat{v} - \hat{w}) \, dS\right| \le c_{4e} |e|^{1/2} \|r - r_e\|_{0,e} |v|_{1,\omega_z\cup\omega_{z'}}\end{aligned}$$

Using the discrete Cauchy-Schwarz inequality and the finite-overlap of patches completes the proof. Q.E.D.

Remark A.6. It is clear from the argument that we did not require symmetry of the bilinear form—the argument and results do not change when $B(u,v) = \int_\Omega A\nabla u \cdot \nabla v + (\mathbf{b}\cdot\nabla u + cu)v\,dV$. We need not have homogeneous Neumann condition either—$F(v) = \int_\Omega fv\,dV + \int_{\partial\Omega_N} gv\,dS$ is fine for the right-hand side.

Remark A.7. It is natural at this point to ask what the natural analogue of the "error space" $W(\mathcal{T})$, in which we compute the approximate error function $\varepsilon_\mathcal{T}$, should be for problems in $\mathbb{R}^3$. Should it be spanned by quadratic edge-bubbles, as the traditional analysis (saturation assumption) might seem to suggest, or by cubic face-bubbles? The most natural extension of the analysis presented here leads us to opt for the latter, in light of the fact that we wish to have zero-mean properties analogous to those in Lemma A.2, namely

$$\int_{\omega_z} (v\ell_z - v_z - w_z) = 0 \quad \text{and} \quad \int_F (v - \hat{v} - \hat{w}) = 0 \text{ for each } F \in \mathcal{F}\ . \tag{A.11}$$

Here, ω_z is the union of the tetrahedra having z as a vertex, and $\mathcal{F}$ is the collection of (non-Dirichlet) tetrahedral faces. The corresponding error estimator, and its extension to eigenvalue applications, is a subject of current investigation.

A.4. **The Conditioning of the System Associated with Computing $\varepsilon(f)$.** We argue here that the matrix associated with computing ε is spectrally equivalent to its diagonal *independent of the mesh scaling* for shape-regular families of meshes. Therefore, the computation of ε will require few iterations of a Krylov solver (CG, GMRES) to sufficiently converge—either with no preconditioning at all, or with (symmetric) diagonal preconditioning.

Given a global ordering b_i of the basis functions for W (and hence of the edges in $\mathcal{E}$), let $B_{ij} = B(b_j, b_i)$ and $\hat{B}_{ij} = (b_j, b_i)_{H^1(\Omega)}$. The first of these is the matrix associated with computing ε. In our setting, both B and $\hat{B}$ are symmetric and positive definite, with

$$m\,\mathbf{v}^t\hat{B}\mathbf{v} = m\|v\|_1^2 \le \mathbf{v}^t B\mathbf{v} = B(v,v) \le M\|v\|_1^2 = M\,\mathbf{v}^t\hat{B}\mathbf{v}\ .$$

So it is clear that B and $\hat{B}$ are spectrally equivalent, independent of any mesh. We also define $D = \mathrm{diag}(B)$ and $\hat{D} = \mathrm{diag}(\hat{B})$. The same argument shows that D and $\hat{D}$ are spectrally equivalent. What remains is to show that $\hat{B}$ and $\hat{D}$ are spectrally equivalent.

In order to estimate the spectra of $\hat{B}$ and $\hat{D}$, we consider their "element-matrices" for each triangle T. Let $v \in W$ have coefficient vector $\mathbf{v}$ and let $\mathcal{E}_T$ denote the set of non-Dirichlet edges touching T. If $\{k_j : 1 \le j \le |\mathcal{E}_T|\}$ are the indices associated with the non-Dirichlet edges of T, then $\hat{B}_T$, $\hat{D}_T$ and $\mathbf{v}_T$ are

$$(\hat{B}_T)_{ij} = (b_{k_j}, b_{k_i})_{H^1(T)} \quad , \quad \hat{D}_T = \mathrm{diag}(\hat{B}_T) \quad , \quad (\mathbf{v}_T)_i = \mathbf{v}_{k_i}\ . \tag{A.12}$$

It is clear from the definitions that $\mathbf{v}^t\hat{B}\mathbf{v} = \sum_{T\in\mathcal{T}} \mathbf{v}_T^t\hat{B}_T\mathbf{v}_T$, and that the analogous result holds for the diagonal matrices. We will show that there are scale-invariant constants $k_0, k_1 > 0$, depending only on the angles in $\mathcal{T}$ such that $k_0\mathbf{v}_T^t\hat{D}_T\mathbf{v}_T \le \mathbf{v}_T^t\hat{B}_T\mathbf{v}_T \le k_1\mathbf{v}_T^t\hat{D}_T\mathbf{v}_T$ for all triangles $T \in \mathcal{T}$, and hence that $k_0\mathbf{v}^t\hat{D}\mathbf{v} \le \mathbf{v}^t\hat{B}\mathbf{v} \le k_1\mathbf{v}^t\hat{D}\mathbf{v}$—i.e. these matrices are spectrally equivalent. For these purposes, we lose no generality by assuming that $\hat{B}_T, \hat{D}_T \in \mathbb{R}^{3\times 3}$.

Let $0 < \theta_k < \pi$ be the measures of the three angles of T, with $c_k = \cot\theta_k$ and $c = c_1 + c_2 + c_3$. We have

$$\hat{B}_T = \frac{4}{3}\begin{pmatrix} c & -c_3 & -c_2 \\ -c_3 & c & -c_1 \\ -c_2 & -c_1 & c \end{pmatrix} + \frac{4|T|}{45}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = A_T + M_T\ . \tag{A.13}$$

One way of obtaining reasonable constants k_0, k_1 is by computing (or estimating) the eigenvalues of $X = \mathrm{diag}(M_T)^{-1}M_T$ and $Y = \mathrm{diag}(A_T)^{-1}A_T$. The eigenvalues of X are readily seen to be $[1/2, 1/2, 2]$. The characteristic polynomial of Y is $p(t) = (1-t)^3 - \frac{1}{c^2}(c_1^2 + c_2^2 + c_3^2)(1-t) + \frac{2c_1c_2c_3}{c^3}$; so we deduce that its eigenvalues are

$$\sigma_k = 1 - \frac{2}{c}\sqrt{\frac{c_1^2 + c_2^2 + c_3^2}{3}}\cos\left(\frac{\theta + 2k\pi}{3}\right)\ ,\ \cos\theta = c_1c_2c_3\left(\frac{3}{c_1^2 + c_2^2 + c_3^2}\right)^{3/2}\ , \tag{A.14}$$

where $k = 0, 1, 2$ and $\theta \in [0, \pi]$. It is clear from the $\cos\left(\frac{\theta + 2k\pi}{3}\right)$ terms that $\sigma_0 \le \sigma_2 \le \sigma_1$. The optimal case, $\theta_k = \frac{\pi}{3}$, yields $\sigma_0 = \frac{1}{3}$ and $\sigma_1 = \sigma_2 = \frac{4}{3}$. Shape-regularity is equivalent to a minimal angle condition,

$\theta_k \geq \alpha\pi > 0$, which bounds σ_0 away from 0. Choosing $k_0 = \min\{\sigma_0(T) : T \in \mathcal{T}\ ,\ \mathcal{T} \in \mathcal{F}\}$ and $k_1 = \max\{2, \max\{\sigma_1(T) : T \in \mathcal{T}\ ,\ \mathcal{T} \in \mathcal{F}\}\}$ completes the argument.

Remark A.8. It is clear from the argument that, as the mesh is refined, the relative contribution of M_T to these bounds becomes insignificant.

Remark A.9. Although we are here only concerned with symmetric matrices B, the above arguments are readily generalized to non-symmetric matrices. In the non-symmetric case, if μ is an eigenvalue of B, then $m\lambda_{\min}(\hat{B}) \leq \mathrm{Re}\,(\mu) \leq M\lambda_{\max}(\hat{B})$, and $|\mathrm{Im}\,(\mu)| \leq M\lambda_{\max}(\hat{B})$. Therefore, the spectrum of B is controlled by the spectrum of $\hat{B}$, which does not deteriorate as the mesh is refined provided the meshes satisfy some fixed minimal angle condition.

Appendix B. Enhancing Eigenvalue Convergence Using Approximation Defects–the Proofs

In this section we review the approximation theory for the eigenvalues of the self-adjoint operator $\mathcal{A}$ from (2.6) as presented in [18]. Herein, "orthogonal projection" will be taken to mean "L^2-orthogonal projection"; in fact, all projections will be L^2-orthogonal onto various subspaces.

Let λ_q be a discrete eigenvalue of the operator $\mathcal{A}$ of multiplicity m and let E_{λ_q} be the projection onto the eigenspace of λ_q. Let $m \in \mathbb{N}$ denote the multiplicity of λ_q. Let us now assume that we have two sequences of projections P^h and Y^h, parameterized by a positive parameter h, with the properties

(A1) For each h we have $\mathsf{R}(P^h) \subset \mathsf{R}(Y^h) \subset \mathcal{H}$.
(A2) For each h we have $\dim \mathsf{R}(Y^h) \leq \infty$ and there exist $r > 0$ and $C > 0$ such that
- $Y^h \to \mathbf{I}$ strongly as $h \to 0$ and
- $\displaystyle\sup_{u \in \mathcal{H}\setminus\{0\}} \frac{\|u - Y^h u\|^2}{|\!|\!| u - Y^h u |\!|\!|^2} \leq Ch^{2r}$.

(A3) For each h we have $\dim \mathsf{R}(P^h) = m$, the multiplicity of the eigenvalue λ_q.

In the notation of earlier sections we have $s_m = \{\lambda_q\}$, and $S_m = \mathsf{R}(E_{\lambda_q})$, $\dim S_m = m$, $\hat{S}_m = \mathsf{R}(P^h)$ and $V = \mathsf{R}(Y^h)$, but we emphasize that the theory presented below applies to more general projections satisfying these assumptions. The operator definition of $\hat{s}_m$ will be given below.

We will study the asymptotic behavior of the approximations which can be derived from (A1)–(A3) by the Rayleigh–Ritz method as $h \to 0$. Our method will involve rigorous efficiency and reliability bounds on the approximation errors. The bounds will hold for all h and will be sharp and thus will allow a precise and reliable asymptotic analysis of the convergence of the approximations.

B.1. Block matrix representation.

To this end let h be temporarily frozen and so we drop it from the notation and write only P. We also further simplify the notation by temporarily denoting the one element set s_m by its single element λ_q.

Our theory of eigenvalue estimation is based on the Frobenius-Schur factorization of the resolvent of the operator $\mathcal{A}$. Let P be the orthogonal projection in $\mathcal{H}$, such that $\mathsf{R}(P) \subset \mathcal{H}$ and $\dim \mathsf{R}(P) = m < \infty$. We will represent the form B as the product of block operator matrices in the product space $\mathsf{R}(P) \oplus \mathsf{R}(P_\perp)$. We will use $\langle \cdot, \cdot \rangle$ to denote the scalar product on the product space $\mathsf{R}(P) \oplus \mathsf{R}(P_\perp)$ and we tacitly assume that the product space $\mathsf{R}(P) \oplus \mathsf{R}(P_\perp)$ is equivalent to $L^2(\Omega)$ and so we do not notationally distinguish them, apart from the distinction in the notation for the scalar product. More to the point, let $\phi, \psi \in L^2(\Omega)$ be given then there exist unique $x_1, y_1 \in \mathsf{R}(P)$ and $x_2, y_2 \in \mathsf{R}(P_\perp)$ such that

$$\phi = x_1 \oplus x_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \psi = y_1 \oplus y_2 = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

We freely use the above notation in the manner outlined above and also write

$$\langle x_1 \oplus x_2, y_1 \oplus y_2 \rangle = \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\rangle = (\phi, \psi),$$

where $(\cdot, \cdot)$ is the scalar product on $L^2(\Omega)$ from (2.3).

We shall now precisely derive the block matrix representation of the resolvent of $\mathcal{A}$ in the product space $\mathsf{R}(P) \oplus \mathsf{R}(P_\perp) = L^2$. Define the operators $\mathcal{M} : \mathsf{R}(P) \to \mathsf{R}(P)$ and $\mathcal{W} : \mathrm{Dom}(\mathcal{W}) \subset \mathsf{R}(P_\perp) \to \mathsf{R}(P_\perp)$

in the sense of [20, Theorem VI-2.23, pp. 331] by the bilinear forms $B(P\cdot, P\cdot) : \mathsf{R}(P) \times \mathsf{R}(P) \to \mathbb{R}$ and $B(P_\perp\cdot, P_\perp\cdot) : \mathsf{R}(P) \cap \mathcal{H} \times \mathsf{R}(P) \cap \mathcal{H} \to \mathbb{R}$ respectively. Let also the operator $\Gamma : \mathsf{R}(P) \to \mathsf{R}(P_\perp)$ be such that

$$(\psi, \Gamma\phi) = B(\mathcal{W}^{-1/2}\psi, \mathcal{M}^{-1/2}\phi),$$

$\psi \in \mathsf{R}(P_\perp)$, $\phi \in \mathsf{R}(P)$ holds. The explicit operator formula for Γ can be extracted from [16, equation (4.37)]. At this point we note that the singular values of Γ are precisely the approximation defects introduced in Section 2. More to the point, we have

$$\sigma_{i-m+1}(\Gamma) = \eta_i.$$

where η_i are defined in (2.11) and we have dropped the reference to h from the notation.

Let $\zeta \in \mathbb{C} \setminus \mathrm{Spec}(\mathcal{W})$, we now use the formal matrix decomposition[2]

$$\mathcal{A} - \zeta\mathbf{I} = \begin{bmatrix} \mathcal{M}^{1/2} & \\ & \mathcal{W}^{1/2} \end{bmatrix} \cdot_\omega \begin{bmatrix} \mathbf{I} - \zeta\mathcal{M}^{-1} & \Gamma^* \\ \Gamma & \mathbf{I} - \zeta\mathcal{W}^{-1} \end{bmatrix} \begin{bmatrix} \mathcal{M}^{1/2} & \\ & \mathcal{W}^{1/2} \end{bmatrix}$$

as a shorthand for

$$B(\psi, \phi) - \zeta(\phi, \psi) = \left\langle \begin{bmatrix} \mathbf{I} - \zeta\mathcal{M}^{-1} & \Gamma^* \\ \Gamma & \mathbf{I} - \zeta\mathcal{W}^{-1} \end{bmatrix} \begin{bmatrix} \mathcal{M}^{1/2} & \\ & \mathcal{W}^{1/2} \end{bmatrix} \phi, \begin{bmatrix} \mathcal{M}^{1/2} & \\ & \mathcal{W}^{1/2} \end{bmatrix} \psi \right\rangle,$$

$\phi, \psi \in \mathcal{H}$. We call the block matrix valued function

$$\zeta \longmapsto S(\zeta) := \begin{bmatrix} \mathbf{I} - \zeta\mathcal{M}^{-1} & \Gamma^* \\ \Gamma & \mathbf{I} - \zeta\mathcal{W}^{-1} \end{bmatrix}^{-1}$$

the scaled pseudo-resolvent of the operator $\mathcal{A}$. This name is justified by the fact that the resolvent of $\mathcal{A}$ has the product representation

$$(\mathcal{A} - \zeta\mathbf{I})^{-1} = \begin{bmatrix} \mathcal{M}^{-1/2} & \\ & \mathcal{W}^{-1/2} \end{bmatrix} S(\zeta) \begin{bmatrix} \mathcal{M}^{-1/2} & \\ & \mathcal{W}^{-1/2} \end{bmatrix}. \tag{B.1}$$

The entries in the matrix $S(\zeta)$, $\zeta \in \mathbb{C} \setminus \mathrm{Spec}(\mathcal{A})$ are bounded operators, so we can use the Frobenius-Schur decomposition of bounded block operator matrices to study the components of $S(\zeta)$, see [32, Chaper 2].

One possible representation of the matrix components of $S(\zeta)$ can be achieved by the help of the operator valued function

$$\zeta \longmapsto S_u(\zeta) := \mathbf{I} - \zeta\mathcal{M}^{-1} - \Gamma^*(\mathbf{I} - \zeta\mathcal{W}^{-1})^{-1}\Gamma, \qquad \zeta \notin \mathrm{Spec}(\mathcal{A}) \cup \mathrm{Spec}(\mathcal{W}), \tag{B.2}$$

which is called the upper Schur complement of $S(\zeta)$. The term "upper" signifies that the Frobenius-Schur decomposition of $S(\zeta)$ is performed by starting from the upper left-hand corner of $S(\zeta)$, see [32, Proposition 1.6.2]. More accurately formulated, $S(\zeta)$ can be expressed as a function of $\mathcal{M}$, $\mathcal{W}$, Γ and the inverse $S_u(\zeta)^{-1}$. We omit the technical details and refer a reader to the monograph [32, Definition 1.6.1 and particularly formula (1.7.4)]. Let us also note that there is a lower Schur complement associated with $S(\zeta)$. This Schur complement is obtained from the Frobenius-Schur factorization which starts from the lower right hand side of $S(\zeta)$. For our considerations we shall only need the Schur complement $S_u(\zeta)$, see [32, Chaper 2].

Now, let us consider what happens when $\zeta = \lambda_q$. In this case $\|(\mathcal{A} - \zeta\mathbf{I})^{-1}\| = \infty$ and the precise analysis—given in [17, Theorem 3.3 and formula (3.8)] and [18, Theorem 4.1]—of the Schur complement yields that the zero which gets inverted to ∞ is precisely and solely—with multiplicity—restricted to

$$S_u(\lambda_q) = 0. \tag{B.3}$$

This technique is a generalization of the standard Wilkinson's trick from matrix analysis, see [29, p. 183].

A closer look at (B.2) shows that $S_u(\lambda_q) = 0$ is the representation result for the eigenvalue error. To see this note that the singular values of $\mathbf{I} - \lambda_q\mathcal{M}^{-1}$, together with their multiplicities, are precisely the relative errors

$$\frac{|\hat{\mu}_i - \lambda_q|}{\hat{\mu}_i}, \qquad i = 1, \ldots, m \text{ and } m \text{ is the multiplicity of } \lambda_q. \tag{B.4}$$

Here $\hat{\mu}_1 \le \hat{\mu}_2 \le \cdots \le \hat{\mu}_m$ are all the eigenvalues of $\mathcal{M}$ with their multiplicities and we assume—as is customary—that $\hat{\mu}_i$ approximate λ_q. Identity (B.2) now reads

$$I - \lambda_q\mathcal{M}^{-1} = \Gamma^*(I - \lambda_q\mathcal{W}^{-1})^{-1}\Gamma \tag{B.5}$$

[2]The symbol $\cdot_\omega$ signifies that this is a "weak" block matrix product.

and a direct application of the statement (B.4) together with the application of the trace operator on the equation (B.5) yields

$$\sum_{i=1}^{m} \frac{|\hat{\mu}_i - \lambda_q|}{\hat{\mu}_i} \leq \mathrm{RelGap}(\lambda_q, P) \sum_{i=1}^{m} \sigma_i^2(\Gamma).$$

Here

$$\mathrm{RelGap}(\lambda_q, P) = \|(\mathbf{I} - \lambda_q \mathcal{W}^{-1})^{-1}\| \tag{B.6}$$

is a real number and $\sigma_1(\Gamma) \geq \sigma_2(\Gamma) \geq \cdots \geq \sigma_m(\Gamma)$ are the singular values of Γ. For ways to efficiently compute $\sigma_i(\Gamma)$ and $\mathrm{RelGap}(\lambda_q, P)$ see [17, 18].

Intuitively, the singular values $\sigma_i(\Gamma)$ measure the size of the residual—that is the reason why we have called them in [18] the approximation defects. The quantity $\mathrm{RelGap}(\lambda_q, P)$ measures the sensitivity of the eigenvalue, and is typically related to the quantity

$$\max_{\zeta \in \mathrm{Spec}(\mathcal{A}) \setminus \{\lambda_q\}} \frac{\zeta + \lambda_q}{|\zeta - \lambda_q|} ,$$

which is why we call it the relative gap (in the spectrum).

We have shown how to obtain reliable estimates of the unitary invariant matrix norms of the error by the use of representation formula (B.5). We will now present a reason why this approach yields accurate estimations of the error. Starting from (B.5) and by using the simple Neumann series argument on the operator $(\mathbf{I} - \lambda_q \mathcal{W}^{-1})^{-1}$ we obtain

$$I - \lambda_q \mathcal{M}^{-1} = \Gamma^* (I - \lambda_q \mathcal{W}^{-1})^{-1} \Gamma \tag{B.7}$$

$$= \Gamma^* \Gamma + \lambda_q \Gamma^* \mathcal{W}^{-1/2} (\mathbf{I} - \lambda_q \mathcal{W}^{-1})^{-1} \mathcal{W}^{-1/2} \Gamma \tag{B.8}$$

$$= \Gamma^* \Gamma + \lambda_q \Gamma^* \mathcal{W}^{-1} \Gamma + \lambda_q^2 \Gamma^* \mathcal{W}^{-1} (\mathbf{I} - \lambda_q \mathcal{W}^{-1})^{-1} \mathcal{W}^{-1} \Gamma. \tag{B.9}$$

The analysis from [18, Section 4] started from the equation (B.8), and yielded that the singular values of Γ are asymptotically exact estimators of the eigenvalue approximation errors (the singular values of $I - \lambda_q \mathcal{M}^{-1}$). The key ingredient of this analysis was the Rayleigh-Ritz orthogonality property of the "residual" Γ. We now repeat this argument—in Theorem B.1 below—to show that, starting instead from

$$I - \lambda_q \mathcal{M}^{-1} - \Gamma^* \Gamma = \lambda_q \Gamma^* \mathcal{W}^{-1} \Gamma + \lambda_q^2 \Gamma^* \mathcal{W}^{-1} (\mathbf{I} - \lambda_q \mathcal{W}^{-1})^{-1} \mathcal{W}^{-1} \Gamma, \tag{B.10}$$

we can prove that the eigenvalues $\hat{\mu}_i^{\#}$, $i = 1, \ldots, m$ of of the operator $\mathcal{M} - \mathcal{M}^{1/2} \Gamma^* \Gamma \mathcal{M}^{1/2}$ — actually representable by an $m \times m$ matrix — are superior approximations when compared to the standard Ritz values $\hat{\mu}_i \in \mathrm{Spec}(\mathcal{M})$, $i = 1, \ldots, m$. This is the reason why we call $\hat{\mu}_i^{\#}$ the enhanced Ritz values. Here we have assumed that $\hat{\mu}_1^{\#} \leq \cdots \leq \hat{\mu}_m^{\#}$ are counted according to their multiplicity.

B.2. **Estimates for the enhanced approximations.** Let us go back to the assumptions (A1)–(A3). Since we suppress the notational dependence on h by assuming it is fixed, let us assume that we have the orthogonal projection P and Y such that

$$\mathsf{R}(Y) \subset \mathcal{H} \quad \text{and} \quad \dim \mathsf{R}(Y) < \infty \tag{B.11}$$

$$\mathsf{R}(P) \subset \mathsf{R}(Y) \quad \text{and} \quad \mathcal{M}_Y P = P \mathcal{M}_Y \tag{B.12}$$

$$\dim \mathsf{R}(P) = m \quad \text{the multiplicity of } \lambda_q \tag{B.13}$$

$$\sigma_1(\Gamma) \leq \mathrm{RelGap}(\lambda_q, P). \tag{B.14}$$

Here we have used that $\mathcal{M}_Y : \mathsf{R}(Y) \to \mathsf{R}(Y)$ is the operator which is defined by the sesquilinear form $B(Y\cdot, Y\cdot) : \mathsf{R}(Y) \times \mathsf{R}(Y) \to \mathbb{R}$. Let us also define the operator $\mathcal{W}_Y : \mathsf{R}(Y_\perp) \to \mathsf{R}(Y_\perp)$ as the self-adjoint operator which is defined in $\mathsf{R}(Y_\perp)$ by the positive definite form $B(Y_\perp\cdot, Y_\perp\cdot) : (\mathsf{R}(Y_\perp) \cap \mathcal{H}) \times (\mathsf{R}(Y_\perp) \cap \mathcal{H}) \to \mathbb{R}$ in the sense of [20, Theorem VI-2.23, pp. 331]. These operators, $\mathcal{M}_Y$ and $\mathcal{W}_Y$, as well as those defined earlier, $\mathcal{M} = \mathcal{M}_P$, $\mathcal{W} = \mathcal{W}_P$ $\Gamma = \Gamma_P$, are related to the relative errors in the enhanced Ritz values via the following theorem.

Theorem B.1. *Let the assumptions (B.11)–(B.14) hold. Then we have*

$$\begin{aligned}\sum_{i=1}^{m} \frac{|\hat{\mu}_i^{\#} - \lambda_q|}{\lambda_q} &\leq \mathrm{RelGap}(\lambda_q, P) \ \sum_{i=1}^{m} \sigma_i^2(\mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2}) \\ &\leq \mathrm{RelGap}(\lambda_q, P) \ \|\mathcal{W}_Y^{-1/2}\Big|_{\mathrm{Ran}(\Gamma)}\|^2 \ \sum_{i=1}^{m} \sigma_i^2(\Gamma\mathcal{M}^{1/2}) \\ &\leq \sqrt{\hat{\mu}_m}\, \mathrm{RelGap}(\lambda_q, P) \ \|\mathcal{W}_Y^{-1/2}\|^2 \ \sum_{i=1}^{m} (\eta_i)^2\end{aligned}$$

where $\hat{\mu}_1^{\#} \leq \cdots \leq \hat{\mu}_m^{\#}$ are the all the eigenvalues of the matrix $\mathcal{M} - \mathcal{M}^{1/2}\Gamma^*\Gamma\mathcal{M}^{1/2}$ (counting according to multiplicity), and $\mathrm{RelGap}(\lambda_q, P)$ *is defined in (B.6).*

Proof. We can write (B.8) as

$$\begin{aligned}\frac{1}{\lambda_q}\Big(\mathcal{M}^{1/2}(\mathbf{I}-\Gamma^*\Gamma)\mathcal{M}^{1/2} - \lambda_q\Big) &= \mathcal{M}^{1/2}\Gamma^*\mathcal{W}^{-1/2}(\mathbf{I}-\lambda_q\mathcal{W}^{-1})^{-1}\mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2} \\ &= \mathcal{M}^{1/2}\Gamma^*\mathcal{W}^{-1}\Gamma\mathcal{M}^{1/2} \\ &\quad + \lambda_q\mathcal{M}^{1/2}\Gamma^*\mathcal{W}^{-1}(\mathbf{I}-\lambda_q\mathcal{W}^{-1})^{-1}\mathcal{W}^{-1}\Gamma\mathcal{M}^{1/2}\end{aligned} \tag{B.15}$$

Note that the operator $(\mathbf{I} - \lambda_q\mathcal{W}^{-1})$ is both bounded and has a bounded inverse and so asymptotically it is sufficient to analyze $\sigma_i(\mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2})$, $i = 1, \ldots, m$, i.e. the singular values of $\mathcal{W}^{-1/2}\Gamma$. We estimate $\sigma_1(\mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2})$ and argue that the estimate for other singular values can be obtained from the standard min-max characterization of singular values. In particular we have for the bounded operator $\mathcal{W}^{-1/2}\Gamma$ the "componentwise" identity

$$(\psi, \mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2}\phi) = \frac{B(\psi,\phi)}{\|\mathcal{W}\psi\|\|\phi\|}, \qquad \psi \in \mathrm{Dom}(\mathcal{W}),\ \phi \in \mathsf{R}(P), \tag{B.16}$$

so we may compute, essentially using (B.12), the estimate

$$\begin{aligned}\sigma_1(\mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2}) &= \max_{\substack{\psi\in\mathrm{Dom}(\mathcal{W}),\phi\in\mathsf{R}(P)\\ \psi,\phi\neq 0}} \frac{|B(\psi,\phi)|}{\|\mathcal{W}\psi\|\|\phi\|} \\ &= \max_{\substack{\psi\in\mathrm{Dom}(\mathcal{W}_Y),\phi\in\mathsf{R}(P)\\ \psi,\phi\neq 0}} \frac{|B(Y_\perp\psi, Y\phi)|}{\|\mathcal{W}_Y\psi\|\|\phi\|} \\ &\leq \sigma_1(\Gamma\mathcal{M}^{1/2})\ \|\mathcal{W}_Y^{-1/2}\|.\end{aligned} \tag{B.17}$$

An analogous argument yields

$$\sigma_i(\mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2}) \leq \sigma_i(\Gamma\mathcal{M}^{1/2})\ \|\mathcal{W}_Y^{-1/2}\|, \qquad i = 1, \ldots, m.$$

The conclusion follows by applying the trace operator on (B.17) and the properties of the operator norm. Q.E.D.

Remark B.2. Were we to assume (A2) in the above theorem, which does hold for our finite element applications, it yields the upper bound $\|\mathcal{W}_Y^{-1/2}\| \leq Ch^{2r}$ which is key for the superconvergence or our enhanced Ritz values.

Remark B.3. Let us note that we may apply other unitary invariant norms on the equation (B.15) and thus obtain estimates for other symmetric gauge functions—in the sense of the von Neumann theory of unitary invariant operator norms—of the error, not just the trace. This is the reason why we call this enhanced Ritz value optimal.

Remark B.4. A natural extension of our methodology would be to practically estimate the singular values $\sigma_i(\mathcal{W}^{-1/2}\Gamma\mathcal{M}^{1/2})$ and thus obtain even higher order corrections by the repetition of the same enhancement argument. At present, we do not have an algorithm for doing so, but this may be revisited at a future point.

The approach outlined in this paper uses *a priori* estimates of this quantity in order to count these singular values as a higher-order terms which are ignored in enhancement procedure.

B.3. **Corollaries of the main theorem.** Let us now repeat the singular value identity for the approximation defects

$$\sigma_{i-m+1}(\Gamma^h) = \eta_i^h.$$

Here we have made explicit the dependence of the quantities on P^h. This is a variational characterization of the singular values of Γ. A similar formula could be derived for the singular values of the operator $\mathcal{W}_{P^h}^{-1/2}\Gamma^h\mathcal{M}_{P^h}^{-1/2}$. However, unlike for the approximation defects η_i, an algorithm for the efficient computation of $\sigma_i(\mathcal{W}_{P^h}^{-1/2}\Gamma^h\mathcal{M}_{P^h}^{-1/2})$ does not appear to be feasible in the same generality, cf. Remark B.4. In the following we also use the notation $\hat{\mu}^h$ for the Ritz values from the subspace $\mathrm{Ran}(P^h)$ and $\hat{\mu}_i^{\#(h)}$ for the enhanced Ritz values. Let us now state that the analysis of Theorem B.1 is sharp.

Corollary B.5. *Assume that we have a collection of projections P^h and Y^h which satisfy the assumptions (A1)–(A3). Then*

$$\lim_{h\to 0}\frac{\sum_{i=1}^m \frac{|\hat{\mu}_i^{\#(h)}-\lambda_q|}{\lambda_q}}{\sum_{i=1}^m \frac{|\hat{\mu}_i^{h}-\lambda_q|}{\lambda_q}} = 0 \quad , \quad \lim_{h\to 0}\frac{\sum_{i=1}^m \frac{|\hat{\mu}_i^{\#(h)}-\lambda_q|}{\lambda_q}}{\sum_{i=1}^m \sigma_i^2(\mathcal{W}_{P^h}^{-1/2}\Gamma^h\mathcal{M}_{P^h}^{1/2})} = 1. \tag{B.18}$$

Here we have explicitly stated the dependence of the quantities from Section B.2 on the pair of spaces P^h and Y^h by appending a subscript or superscript to the corresponding object in an obvious way.

Remark B.6 (Practical estimates in Sobolev spaces). We have indicated that we will use both operator as well as variational realization of the eigenvalue problem (2.1) as is more appropriate to the context. Let us go back to the formula (2.7) for the operator theoretic representation of the energy norm. We will now emphasize the notational dependence on h.

We now define the block diagonal energy norm $|||\cdot|||$ by the formula

$$|||\psi|||_{\hat{S}_m}^2 = |||P^h\psi|||^2 + |||P_\perp^h\psi|||^2, \qquad \psi\in\mathcal{H}$$

and recall the definition for the approximation defects (2.11). The basic result of [16] is the equivalence of the norms $|||\cdot|||_{\hat{S}_m}$ and $|||\cdot|||$. Precisely, we have

$$(1-\eta_m^h)|||\psi|||_{\hat{S}_m} \le |||\psi||| \le (1+\eta_m^h)|||\psi|||_{\hat{S}_m}, \qquad \psi\in\mathcal{H}. \tag{B.19}$$

The discussion from [18, Section 5] implies that we can assume, without reducing the level of generality, that we have h such that $\eta_m^h < \frac{1}{2}$. Note that even for a very coarse mesh, this upper estimate is quite crude since $\eta_m^h = O(h^2)$. With this we write (B.19) as

$$\frac{1}{2}|||\psi|||_{\hat{S}_m} \le |||\psi||| \le \frac{3}{2}|||\psi|||_{\hat{S}_m}, \qquad \psi\in\mathcal{H} \tag{B.20}$$

and note that the actual equivalence constants, which can be reliably and efficiently estimated by computable quantities $\tilde{\eta}_m^h$ from (2.14) below, are much closer to 1. According to [16] the norm equivalence (B.20) implies that the relative approximation error for the Ritz values is smaller than $\frac{1}{2}$.

Department of Mathematics University of California, San Diego La Jolla, California 92093-0112, USA
E-mail address: rbank@ucsd.edu

University of Zagreb, Department of Mathematics, Bijenička 30, 10000 Zagreb, Croatia
E-mail address: luka.grubisic@math.hr

University of Kentucky, Department of Mathematics, Patterson Office Tower 761, Lexington, KY 40506-0027, USA
E-mail address: jovall@ms.uky.edu