Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

On Estimators for Eigenvalue/Eigenvector Approximations

by

Luka Grubišić, and Jeffrey Ovall

Preprint no.: 18 2007

MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

ON ESTIMATORS FOR EIGENVALUE/EIGENVECTOR APPROXIMATIONS

LUKA GRUBIŠIĆ AND JEFFREY S. OVALL

Abstract. We consider a large class of residuum based *a posteriori* eigenvalue/eigenvector estimates and present an abstract framework for proving their asymptotic exactness. Equivalence of the estimator and the error is also established. To demonstrate the strength of our abstract approach we present a detailed study of hierarchical error estimators for Laplace eigenvalue problems in planar polygonal regions. To this end we develop new error analysis for the Galerkin approximation which avoids a use of the strengthened Cauchy-Schwarz inequality and the saturation assumption, and gives reasonable and explicitly computable upper bounds on the discretization error. Brief discussion is also given concerning the design of estimators which are in the same spirit, but are based on different *a posteriori* techniques – notably, those of gradient recovery type.

1. Introduction

The purpose of this paper is to analyse *a posteriori* eigenvalue/eigenvector estimators for a class of positive definite symmetric eigenvalue problems. We reduce the study of the eigenvalue/eigenvector estimators to the study of associated boundary value problems and reuse available results on the *a posteriori* error analysis for those auxiliary problems. In particular we consider those estimators for boundary value problems which are asymptotically exact and show that under a natural non-degeneracy assumption on the spectral approximation problems our derived eigenvalue/eigenvector estimators are also asymptotically exact.

Our analysis also yields equivalence of the *a posteriori* estimator and the *relative* eigenvalue/eigenvector error with reasonable and computable equivalence constants. Our results are based on the techniques of the relative perturbation theory from Numerical Linear Algebra—we are particularly influenced by the approach of [12]— which have recently been considered in [17, 18, 15] in the setting of infinite dimensional Hilbert spaces.

Estimators for the adaptive finite element eigenvalue approximations have recently been considered in literature from several viewpoints. One possible approach is that of Heuveline and Rannacher [21] and Verfürth [33] which is based on a general analysis of the nonlinear (single vector) residuum equations. Such approaches make an analysis of the approximations of multiple eigenvalues somewhat more involved. On the other hand, the approaches of Neymeyr [28], Durán, Padra and Rodríguez [13], Larson [23] and Mao, Shen and Zhou [25] analyse the same residual

2000 *Mathematics Subject Classification.* Primary .
Key words and phrases. eigenvalue problem, finite element method, a posteriori error estimates .

equations directly. Eigenvalue estimates for multiple eigenvalues and the associated invariant subspaces are then derived by maximising the residual estimate over the approximate test subspace. The approach of [13] is essentially asymptotic in nature since the equivalence is shown up to the higher order terms. The analysis of when these higher order terms may be neglected is given, but the equivalence results are still not constructive in nature. The analysis of [23] is performed by a combination of *a posteriori* and *a priori* analysis and it unfortunately requires that the associated boundary value problem be H^2 regular.

On the other hand, we start from the abstract block matrix residual equation for the invariant subspace—as presented in [18]—which allows a natural treatment of the eigenvalue multiplicity without incurring unnecessary regularity constraints. This error representation formula is used both to prove the equivalence (with explicit and reasonable constants) of the residuum based estimator as well as its asymptotic exactness. We also indicate that there is a class of $\sin\Theta$-type theorems which use the same type of residual measures to obtain computable bounds on the invariant subspace error, see [19]. In this paper we do not quote those results explicitly (an application of the results from [19] in our context is straightforward), but rather concentrate on obtaining estimates of the norm of the gradient of the eigenvector error. Of all the approaches which we have mentioned the closest in spirit to our considerations are those from [28] and [25] since they both reduce the study of the eigenvalue problem on the study of the associated boundary value problem.

More to the point, we use a similar preconditioned hierarchical error estimator to the one which is used in [28] and prove that our modified estimator is not only rigorous/reliable but also efficient – in other words, equivalent to the error. Furthermore, we show that it is asymptotically exact on the model problem of the Dirichlet Laplacian and provide reasonable and computable equivalence constants. Eigenvector error estimates, which were not considered in [28] are also given. The authors of [25] analyze local averaging type error estimators and prove their asymptotic exactness. To illustrate the generality of our block-matrix (invariant subspace) residual equations we briefly discuss how to obtain similar results for some other gradient recovery type error estimators.

In our detailed analysis of the Dirichlet Laplace eigenvalue problem we wanted to reuse the known results on the error of the Galerkin approximation. However, the available estimates did not suit our needs, since in the eigenvalue problem we wanted to simultaneously consider the associated boundary value problem for a large class of right-hand side vectors. The standard estimates involved constants which were intricately dependent on the right-hand side vector and it was not possible to decouple this dependence easily. Therefore we have developed a new error analysis—which is interesting in its own right—for the Galerkin approximation which avoids a use of the strengthened Cauchy-Schwarz inequality and the saturation assumption. Furthermore, this analysis yields reasonable and explicitly computable upper bounds on the discretisation error.

The theory of [17, 18, 15] has been developed—in the framework of the perturbation theory from [22, Chapters VI–VIII]—for an abstract positive definite symmetric and closed form h in a general Hilbert space $\mathcal{H}$. We use this abstract approach to establish the asymptotic exactness of the *scaled residual* and the relative eigenvalue/eigenvector error in Section 4. It is often the case that one loses information about important specific examples by making general abstract arguments. However, to show the strength of our theory we focus on the Dirichlet Laplacian in polygonal domains with possibly reentrant corners, demonstrating that nothing important is lost in the general arguments. In Section 3 we introduce our measures of the size of the *scaled residual*—which we call approximation defects—and give a detailed constructive (equivalence) analysis of their behaviour. Furthermore, to this end in Section 5 we revisit the class of error estimators for boundary value problems from [11, 26, 27] and obtain new reasonable and computable upper bounds on the discretisation error.

2. Notation and preliminaries

Let $\mathcal{R} \subset \mathbb{R}^2$ be a bounded *polygonal region*, possibly with reentrant corners. By $H^1_0(\mathcal{R})$ we denote the subspace of the first order Sobolev space $H^1(\mathcal{R})$ which consists of all those functions which vanish on the boundary $\partial\mathcal{R}$ (this is meant in the sense of the *trace operator*). The space $H^1_0(\mathcal{R})$ is assumed to be equipped with the norm $\|u\|_{H^1_0} = |u|_{1,2}$. By $\|\cdot\|$ we always denote the norm on $L^2(\mathcal{R})$ and we use $|\cdot|_{k,2}$, $k \in \mathbb{N}$ to denote the standard Sobolev semi-norms. For other real $\alpha \in \langle 0, 1]$ we also use $H^{1+\alpha}(\mathcal{R})$ to denote standard interpolation spaces.

In Section 4 we shall deal with variational eigenvalue problems for a general closed symmetric and positive definite form h in a Hilbert space in the sense of [22, Theorem VI-2.23, pp. 331]. Indeed, this is a natural framework for most of our theory and this is the generality in which the results of [18] have been proved. However, in this paper it is our aim to discuss the finer properties of the construction from [18]. To this end, and also to ease the presentation, we concentrate on the Dirichlet Laplace eigenvalue problem. In the weak form this reads: Find the real eigenvalue λ and a nonzero eigenfunction v such that

$$\begin{cases} \int_{\mathcal{R}} \nabla v \cdot \nabla \psi &= \lambda \int_{\mathcal{R}} v\psi, \qquad \text{for all } \psi \in H^1_0(\mathcal{R}) \\ \qquad \int_{\mathcal{R}} |v|^2 &= 1. \end{cases} \tag{2.1}$$

Problem (2.1) is attained by a sequence of positive eigenvalues λ_i—ordered in the ascending order $\lambda_i \leq \lambda_{i+1}$, $i \in \mathbb{N}$ according to multiplicity—such that $\lambda_i \to \infty$ and a sequence of associated eigenvectors such that $v_j \in H^{1+\alpha}(\mathcal{R})$ (the parameter α depends on the regularity of $\mathcal{R}$). In the operator form this can be written as

$$\begin{cases} -\triangle v_i &= \lambda_i v_i, \qquad \psi \in \mathcal{R} \\ \qquad v_i &= 0 \qquad \text{on } \partial\mathcal{R}. \end{cases} \tag{2.2}$$

The gradient operator ∇ and the Laplace operator $\triangle$ are meant in the distributional sense. We will also use the notation $-\triangle$ to denote the positive definite self-adjoint operator $\mathbf{H}$ which represents the symmetric form

$$h(\psi, \phi) = \int_{\mathcal{R}} \nabla\psi \cdot \nabla\phi, \qquad \psi, \phi \in H^1_0(\mathcal{R}) \tag{2.3}$$

in the sense of [22, Theorem VI-2.23, pp. 331], i.e. $h(\psi, \phi) = (\mathbf{H}^{1/2}\psi, \mathbf{H}^{1/2}\phi)$, $\psi, \phi \in H^1_0(\mathcal{R})$ and the domain of definition of $\mathbf{H}^{1/2}$ equals $H^1_0(\mathcal{R})$. The associated quadratic form is denoted by $h[\psi] = h(\psi, \psi)^{1/2}$. In this paper we use $\mathcal{D}(\mathbf{H})$ to denote the *domain of the operator* $\mathbf{H}$, $\Sigma(\mathbf{H})$ to denote its spectrum and $\mathcal{Q}(h)$ to *denote the domain of the symmetric form h*. Furthermore, for a linear operator A we use $\mathsf{R}(A)$ and $\mathsf{N}(A)$ to denote its *range* an the *null space.*

To compute finite element approximations for the eigenvalues and eigenvectors of (2.1) we define a family of finite element spaces. Let $\mathcal{T}_d$ be a collection of closed triangles such that $\overline{\mathcal{R}} = \bigcup_{\tau\in\mathcal{T}_d}\tau$. The diameter of a triangle $\tau \in \mathcal{T}_d$ is given by d_τ, and the maximal diameter

$$d = \max_{\tau\in\mathcal{T}_d} d_\tau$$

is used to index the triangulation. We will only consider *conforming triangulations* $\mathcal{T}_d$ of $\mathcal{R}$ - triangulations such that the intersection of any two triangles in $\mathcal{T}_d$ is either empty, or consists of a common edge or vertex. For a given triangulation $\mathcal{T}_d$ we define the finite dimensional function spaces:

$$\mathfrak{L}(\mathcal{T}_d) = \{u \in \mathcal{Q} \cap C(\ \overline{\mathcal{R}}\ ) \mid \text{ for } T \in \mathcal{T}_d,\ u|_T \text{ is a linear function}\}, \tag{2.4}$$

$$\mathfrak{Q}(\mathcal{T}_d) = \{u \in \mathcal{Q} \cap C(\ \overline{\mathcal{R}}\ ) \mid \text{ for } T \in \mathcal{T}_d,\ u|_T \text{ is a quadratic function}\}\ . \tag{2.5}$$

We will also make use of the space $\mathfrak{B}(\mathcal{T}_d)$ of edge bubble functions, which are those functions from $\mathfrak{Q}(\mathcal{T}_d)$ which vanish at the vertices of all triangles in $\mathcal{T}_d$. We have the hierarchical decomposition $\mathfrak{Q}(\mathcal{T}_d) = \mathfrak{L}(\mathcal{T}_d) \oplus \mathfrak{B}(\mathcal{T}_d)$, so we use $\mathfrak{B}(\mathcal{T}_d) = \mathfrak{Q}(\mathcal{T}_d) \ominus \mathfrak{L}(\mathcal{T}_d)$ as a compact definition.

We take the standard bases for $\mathfrak{L}(\mathcal{T}_d)$ and $\mathfrak{B}(\mathcal{T}_d)$, which are described as follows. Let $\mathcal{V}_d$ be the set of interior vertices, $\bar{\mathcal{V}}_d$ the set of all vertices and $\mathcal{E}_d$ be the set of the interior edges in the triangulation $\mathcal{T}_d$. Then the bases for $\mathfrak{L}(\mathcal{T}_d)$ and $\mathfrak{B}(\mathcal{T}_d)$ are, respectively,

$$\begin{aligned}&\{\ell_z \in \mathfrak{L}(\mathcal{T}_d) \mid \ell_z(z') = \delta_{zz'} \text{ for } z \in \mathcal{V}_d,\ z' \in \bar{\mathcal{V}}_d\} \quad \text{and}\\ &\{b_e \in \mathfrak{B}(\mathcal{T}_d) \mid b_e = 4\ell_z\ell_{z'} \text{ for } e \in \mathcal{E}_d \text{ with endpoints } z, z'\}.\end{aligned}$$

The factor of 4 in the definition of b_e is chosen so that the coefficients of a function in $\mathfrak{B}(\mathcal{T}_d)$ with respect to this basis coincide with the values of the function at the midpoints of the corresponding edges. The union of these sets forms a (hierarchical) basis for $\mathfrak{Q}(\mathcal{T}_d)$. The cardinalities of the sets $\mathcal{T}_d$, $\mathcal{I}_d$ and $\mathcal{E}_d$ are related by Euler's formula, $|\mathcal{E}_d| = |\mathcal{T}_d| + |\mathcal{I}_d| - 1$, and we generally expect that $\mathcal{E}_d$ has between three and four times the cardinality of $\mathcal{I}_d$. We will use the spaces $\mathfrak{L}(\mathcal{T}_d)$ to compute eigenvalue/eigenvector approximations and the spaces $\mathfrak{B}(\mathcal{T}_d)$ to assess the quality of the approximation.

A discrete variant of (2.1) now reads

$$\begin{cases}\int_{\mathcal{R}} \nabla u \cdot \nabla\psi &= \lambda \int_{\mathcal{R}} u\psi, \qquad \text{for all } \psi \in \mathfrak{L}(\mathcal{T}_d)\\ \qquad \int_{\mathcal{R}} |u|^2 &= 1.\end{cases} \tag{2.6}$$

and it is attained by a finite number of discrete eigenvalues $\lambda_i(\mathcal{T}_d)$ and discrete eigenvectors $u_i(\mathcal{T}_d)$, $i = 1, \ldots, \dim \mathfrak{L}(\mathcal{T}_d)$. The discrete eigenvalues $\lambda_i(\mathcal{T}_d)$ (discrete eigenvectors $u_i(\mathcal{T}_d)$) are often called the Ritz values/vectors. We reserve these terms for those discrete eigenvalues/eigenvectors which approximate a particular eigenvalue of interest and have a joint multiplicity which is equivalent to the multiplicity of the eigenvalue.

Let us assume that we want to approximate the eigenvalue λ_q of multiplicity $m \in \mathbb{N}$ of (2.1). This is to say we assume that λ_i satisfy

$$\lambda_{q-1} < \lambda_q = \lambda_{q+m-1} < \lambda_{q+m}. \tag{2.7}$$

We also assume that $q+m < \dim \mathfrak{L}(\mathcal{T}_d)$. By P_d we denote the orthogonal projection onto the linear span of $\{u_q(\mathcal{T}_d), \ldots, u_{q+m-1}(\mathcal{T}_d)\}$. We use $\mathsf{R}(P_d)$ to denote the range of the projection P_d and we write

$$\mathsf{R}(P_d) = \operatorname{span}\{u_q(\mathcal{T}_d), \ldots, u_{q+m-1}(\mathcal{T}_d)\}. \tag{2.8}$$

Given such a subspace $\mathsf{R}(P_d)$ we set $\mu_i^d = \lambda_{q-1+i}(\mathcal{T}_d)$ and $\psi_i^d = u_{q-1+i}(\mathcal{T}_d)$. We call μ_i^d and ψ_i^d Ritz vectors/vectors from the subspace $\mathsf{R}(P_d)$. In choosing our notation we will suppress the dependence on the parameter d wherever there is no danger of confusion.

Take $\psi \in \mathsf{R}(P) \subset H_0^1(\mathcal{R})$ and consider the solution $u(\psi)$ of the problem

$$-\triangle u = \psi, \qquad u \in H_0^1(\mathcal{R}).$$

Let the functions $u_P(\psi), u_1(\psi, \mathcal{T}_d)$, for $\psi \in \mathsf{R}(P)$, be such that

$$\|\nabla u(\psi) - \nabla u_P(\psi)\| = \min_{v \in \mathsf{R}(P)} \|\nabla u(\psi) - \nabla v\|$$
$$\|\nabla u(\psi) - \nabla u_1(\psi, \mathcal{T}_d)\| = \min_{v \in \mathfrak{L}(\mathcal{T}_d)} \|\nabla u(\psi) - \nabla v\|.$$

We define the approximation defects

$$\eta_i(P) = \max_{\substack{\mathcal{S} \subset \mathsf{R}(P) \\ \dim \mathcal{S} = m-i+1}} \min_{\psi \in \mathcal{S}} \frac{\|\nabla u(\psi) - \nabla u_P(\psi)\|}{\|\nabla u(\psi)\|} \tag{2.9}$$

$$\eta_i(P, \mathcal{T}_d) = \max_{\substack{\mathcal{S} \subset \mathsf{R}(P) \\ \dim \mathcal{S} = m-i+1}} \min_{\psi \in \mathcal{S}} \frac{\|\nabla u(\psi) - \nabla u_1(\psi)\|}{\|\nabla u(\psi)\|}. \tag{2.10}$$

The quantities $\eta_i(P)$ are defined for any projection P such that $\mathsf{R}(P) \subset H_0^1(\mathcal{R})$. They are the main ingredient of the error estimates below and we call them the *approximation defects* of $\mathsf{R}(P)$. Obviously, for P_d from (2.8) we have $\eta_i(P_d) = \eta_i(P_d, \mathcal{T}_d)$ and we can profit from the information on the approximation properties of the spaces $\mathfrak{L}(\mathcal{T}_d)$ in the quest for obtaining computable estimates of $\eta_i(P_d)$. In case in which the projection P, $\mathsf{R}(P) \subset \mathfrak{L}(\mathcal{T}_d)$ does not satisfy the assumption (2.8) we do not have the equality $\eta_i(P) = \eta_i(P, \mathcal{T}_d)$, but a simple perturbation argument can be used to obtain estimates of the approximation defect $\eta_i(P)$. We will comment on this more in Section 3 where we develop practical procedures for the computation of $\eta_i(P)$, cf. (3.10). The reason why we have chosen such test spaces is that we use an adaptation of the standard results on the hierarchical decomposition $\mathfrak{Q}(\mathcal{T}_d) = \mathfrak{L}(\mathcal{T}_d) \oplus \mathfrak{B}(\mathcal{T}_d)$ to obtain practical computational estimates for $\eta_i(P)$. This is the main subject of the following section.

The analysis of [18] yields the conclusion that the test space $\mathsf{R}(P)$ contains sufficiently good approximation for the eigenvalue $\lambda_q(\mathbf{H})$ when $\eta_m(P)$ is smaller than half of the *relative gap*

$$\gamma_q := \min\Big\{\frac{\lambda_{q+m}(\mathbf{H}) - \mu_m}{\lambda_{q+m}(\mathbf{H}) + \mu_m}, \frac{\mu_1 - \lambda_{q-1}(\mathbf{H})}{\mu_1 + \lambda_{q-1}(\mathbf{H})}\Big\}.$$

This is in line with the convergence analysis in Numerical Linear Algebra, see [12, Proposition 2.3].

Theorem 2.1. *Let the discrete eigenvalues of the positive definite operator* $\mathbf{H}$ *be so ordered that* $\lambda_{q-1} < \lambda_q = \lambda_{q+m-1} < \lambda_{q+m}$. *Let* $\mathsf{R}(P)$ *be the test subspace such that* $\dim \mathsf{R}(P) = m$ *and* $\frac{\eta_m(P)}{1-\eta_m(P)} < \gamma_q$. *Then we have*

$$\interleave \operatorname{diag}\Big(\frac{|\lambda_q - \mu_i|}{\mu_i}\Big)_{i=1}^m \interleave \leq \frac{\eta_m(P)}{\mathfrak{g}_{q,\eta_m(P)}} \interleave \operatorname{diag}(\eta_i(P))_{i=1}^m \interleave . \tag{2.11}$$

where $\mathfrak{g}_{q,\zeta} := \max\{\frac{\mu_1(1-\zeta)-(1+\frac{\zeta}{1-\zeta})\lambda_{q-1}}{(1+\frac{\zeta}{1-\zeta})\lambda_{q-1}}, \frac{(1-\frac{\zeta}{1-\zeta})\lambda_{q+m}-(1+\zeta)\mu_m}{(1-\frac{\zeta}{1-\zeta})\lambda_{q+m}}\}$ *for* $q > 1$ *and we set* $\mathfrak{g}_{1,\zeta} := \mathfrak{g}_1 := \frac{\lambda_{m+1}-\mu_m}{\lambda_{m+1}+\mu_m}$. *Here we use* $\operatorname{diag}(\alpha_i)_{i=1}^m$ *to denote the* $m \times m$ *diagonal matrix with scalars* α_i *on its diagonal and* $|\!|\!| \cdot |\!|\!|$ *denotes any unitary invariant matrix norm.*

In the case in which we do not have explicit information on the multiplicity of λ_q we have a weaker upper estimate. There is also an accompanying lower estimate which establishes the equivalence of the estimators η_i and the error.

Theorem 2.2. *Let the discrete eigenvalues of the positive definite operator* $\mathbf{H}$ *be so ordered that* $\lambda_m < \lambda_{m+1}$ *and let* $\lambda_{s_1} < \lambda_{s_2} < \cdots < \lambda_{s_p}$ *be all the elements*[1] *of* $\Sigma(\mathbf{H}) \setminus \{\lambda \in \Sigma(\mathbf{H}) \ : \ \lambda \geq \lambda_{m+1}\}$. *If* $\frac{\eta_m(P)}{1-\eta_m(P)} < \frac{\lambda_{m+1}-\mu_m}{\lambda_m+\mu_m}$ *then*

$$\frac{\mu_1}{2\mu_m}\sum_{i=1}^m \eta_i^2(P) \leq \sum_{i=1}^m \frac{|\lambda_i - \mu_i|}{\mu_i} \leq \frac{1}{\min\limits_{i=1,\ldots,p} \mathfrak{g}_{s_i,\eta_{m_i}(P_{s_i})}} \sum_{i=1}^m \eta_i^2(P). \tag{2.12}$$

Here P_{s_i} *is the orthogonal projection onto the linear span of* $\mathcal{S}_i := \{u_j \ : \ j = \sum_{k=1}^i m_k + 1, \ldots, \sum_{k=1}^{i+1} m_k\}$ *and* m_i *is the multiplicity of the eigenvalue* λ_{s_i}, $i = 1, \ldots, p$. *Obviously the identity* $P_{s_1} \oplus P_{s_2} \oplus \cdots \oplus P_{s_p} = P$ *holds. In the case in which* $\lambda_1 = \lambda_m$ *we can drop the constant* $\frac{\mu_1}{2\mu_m}$ *from the lower estimate.*

Remark 2.3. Note that as $\eta_{s_i}(P_i) \to 0$ we have

$$\mathfrak{g}_{s_i,\eta_{m_i}(P_i)} \to \min\{\frac{\lambda_{s_{i+1}} - \lambda_{s_i}}{\lambda_{s_i}}, \frac{\lambda_{s_i} - \lambda_{s_{i-1}}}{\lambda_{s_{i-1}}}\}$$

and $\min_{i=1,\ldots,p} \mathfrak{g}_{s_i,\eta_{m_i}(P_i)}$ quantifies the minimal *relative* gap among the eigenvalues $\lambda_{s_1} < \lambda_{s_2} < \cdots < \lambda_{s_p}$. Note that the relative gap $\mathfrak{g}_{s_i,\eta_{s_i}(P_i)}$ distinguishes better between the close eigenvalues than the *absolute* gap, eg. $\min\{\lambda_{s_{i+1}} - \lambda_{s_i}, \lambda_{s_i} - \lambda_{s_{i-1}}\}$ is an example of an absolute gap. In Theorem 2.2, equivalently as in [12, Proposition 2.3], we have that when $\eta_{m_i}(P_i) < \frac{1}{3}\min_{k\neq j} \frac{|\lambda_{s_k} - \lambda_{s_j}|}{\lambda_{s_k} + \lambda_{s_j}}$, $i = 1, \ldots, p$ then

$$\frac{1}{\min\limits_{i=1,\ldots,p} \mathfrak{g}_{s_i,\eta_{m_i}(P_i)}} \leq \frac{1}{\min\limits_{k\neq j} \frac{|\lambda_{s_k} - \lambda_{s_j}|}{\lambda_{s_k} + \lambda_{s_j}}}.$$

Remark 2.4. The constant $\frac{\mu_1}{\mu_m}$ is not satisfactory, since it implies that the estimate is not quantitatively useful for higher eigenvalues. Establishing sharper lower estimate for higher eigenvalues is technically involved and does not promise any significant new quantitative information. Any type of estimate is bound to include the minimal relative gap between the computed Ritz values and the unwanted component of the spectrum, and estimating this distance is in practice only asymptotically possible. As an alternative we establish, in Section 4, an asymptotic exactness of the eigenvalue error and the approximation defects. This result holds for all discrete eigenvalues of a positive definite operator. It even holds for the eigenvalues which are in gaps of the essential spectrum (in case we are considering unbounded domains or periodic boundary conditions).

[1] We assume that $1 \leq s_1 < s_2 < \cdots < s_p \leq m$.

2.1. **The operator theoretic perturbation construction.** Let us give some basic information on the perturbation construction which led to the aforementioned theorems. This abstract construction will be used later for a general proof of the asymptotic exactness of the estimators $\eta_i(P)$ in Section 4. For a positive definite form h and some orthogonal projection Y, such that $\mathsf{R}(Y) = \mathcal{Y} \subset \mathcal{Q}(h)$ and $\dim\mathcal{Y} < \infty$, we define the positive definite form, generically using $Y^\perp := \mathbf{I} - Y$,

$$h_Y(\psi, \phi) = h(Y\psi, Y\phi) + h(Y^\perp\psi, Y^\perp\phi), \qquad \psi, \phi \in \mathcal{Q}(h). \tag{2.13}$$

By $\mathbf{H}_Y$ we denote the self-adjoint operator which is defined by h_Y in the sense of Kato. It holds, assuming $N = \dim\mathcal{Y}$, that

$$\max_{\substack{\mathcal{S}\subset\mathcal{Y},\\ \dim\mathcal{S}=N-i+1}} \min_{\psi\in\mathcal{S}} \frac{h(\psi,\psi)}{(\psi,\psi)} = \max_{\substack{\mathcal{S}\subset\mathcal{Y},\\ \dim\mathcal{S}=N-i+1}} \min_{\psi\in\mathcal{S}} \frac{(\psi,\mathbf{H}_Y\psi)}{(\psi,\psi)} = \lambda_i(\mathbf{H}_Y\big|_{\mathcal{Y}}),$$

and so we have constructed the operator $\mathbf{H}_Y$, such that[2] $\mathcal{D}(\mathbf{H}^{1/2}) = \mathcal{D}(\mathbf{H}_{\mathcal{Y}}^{1/2})$ and $\mathbf{H}_Y$ has the Ritz values of $\mathbf{H}$ from the subspace $\mathcal{Y}$, in its discrete spectrum. Furthermore, if we assume that there is a sequence of orthogonal projections Y_d, $\dim\mathsf{R}(Y_d) < \infty$, such that $Y_d \to \mathbf{I}$ strongly then the spectral projections of $\mathbf{H}_{Y_d}$ converge to the spectral projections of $\mathbf{H}$ in norm. This convergence can be estimated if we interpret $\mathbf{H}$ as a perturbation of $\mathbf{H}_{Y_d}$ (in the quadratic form sense) and then use an adapted version of the relative perturbation theory of Kato (from [22, Chapters VI–X]) to obtain both the eigenvalue and eigenvector estimates (for details on this construction and estimates see [15, 16]). By $u_i(\mathcal{T}_d) \in \mathcal{Y}_d$, $i = 1, \dots, N$ we denote the vectors such that

$$\begin{aligned} \mathbf{H}_{Y_d} u_i(\mathcal{T}_d) &= \lambda_i(\mathbf{H}_{Y_d}\big|_{\mathcal{Y}_d}) u_i(\mathcal{T}_d), \\ (u_i(\mathcal{T}_d), u_j(\mathcal{T}_d)) &= \delta_{ij}, \qquad i, j = 1, \dots, N, \end{aligned} \tag{2.14}$$

Furthermore, assuming we want to approximate the eigenvalue λ_q of multiplicity m we give the following abstract definition. Let P be an orthogonal projection such that $\dim\mathsf{R}(P) = m$ and $\mathsf{R}(P) \subset \mathcal{Q}(h)$. We call $\mathsf{R}(P)$ the *test subspace* for (the approximation of) λ_q. The operator

$$\Xi = (\mathbf{H}^{1/2}P)^*\mathbf{H}^{1/2}P\Big|_{\mathsf{R}(P)} \tag{2.15}$$

will be called the (generalized) *Rayleigh quotient.* Its eigenvalues $\mu_1 \leq \cdots \leq \mu_m$ will be called the *Ritz values* from the *test subspace* $\mathsf{R}(P)$ and the vectors $\psi_i \in \mathsf{R}(P)$, $\Xi\psi_i = \mu_i\psi_i$, $\|\psi_i\| = 1$ will be called the *Ritz vectors.* The operator theoretic version of (2.9) now reads

$$\eta_i(P) = \max_{\substack{\mathcal{S}\subset\mathsf{R}(P)\\ \dim(\mathcal{S})=m-i+1}} \min_{\substack{\psi\in\mathcal{S}\\ \|\psi\|=1}} \frac{\|\mathbf{H}\Xi^{-1}\psi - \psi\|_{\mathbf{H}^{-1}}}{\|\psi\|_{\mathbf{H}^{-1}}}\ , \tag{2.16}$$

for $i = 1, \dots, m$. Obviously, for the projection P_d from (2.8) we have $\mu_i^d = \lambda_{q-1+i}(\mathcal{T}_d)$ and $\psi_i^d = u_{q-1+i}(\mathcal{T}_d)$, $i = 1, \dots, m$.

Let us note that Theorem 2.2 essentially solves the eigenvector approximation problem, too. By this we mean that we have both upper as well as lower estimates for the eigenvector error. This is made explicit in the following proposition.

[2] In the case when $\mathbf{H}$ is as in (2.3), then $\mathcal{D}(\mathbf{H}^{1/2}) = \mathcal{D}(\mathbf{H}_Y^{1/2}) = H_0^1(\mathcal{R})$.

Proposition 2.5. *For eigenvectors $-\triangle v_i = \lambda_i v_i$ and Ritz vectors $\Xi\psi_i = \mu_i\psi_i$, $\psi_i \in \mathsf{R}(P)$—assuming $\lambda_{q-1} < \lambda_q \leq \cdots \leq \lambda_{q+m-1} < \lambda_{q+m}$ and $2\eta_m < \gamma_q$— we have*

$$\|v_i - \psi_i\| \leq \max_{\lambda\in\Sigma(\mathbf{H})\setminus\{\lambda_i\}} \frac{\sqrt{2\lambda\mu_i}}{|\lambda-\mu_i|}\frac{\eta_m(P)}{\sqrt{1-\eta_m(P)}}, \tag{2.17}$$

$$\frac{\|\nabla\psi_i - \nabla v_i\|^2}{\|\nabla v_i\|^2} = \|v_i - \psi_i\|^2 + \frac{\mu_i - \lambda_i}{\lambda_i}, \qquad i = q, ..., q+m-1. \tag{2.18}$$

The proof of (2.17) can be found in [18] and identity (2.18) is well-known. We can now combine (2.18) with (2.17) and Theorem 2.2 to obtain equivalent estimators for the eigenvector error $\frac{\|\nabla\psi_i-\nabla v_i\|^2}{\|\nabla v_i\|^2}$. For a general form h we estimate the quotient $\frac{h[\psi_i-v_i]}{h[v_i]}$ and the same formula holds.

3. Equivalence of the eigenvalue/eigenvector estimator

In this section we consider computable estimates of the approximation defects $\eta_i(P)$ of $\mathsf{R}(P)$. In particular, we show that these estimates are equivalent to the approximation defects. We assume that the test subspace $\mathsf{R}(P)$ satisfies the same conditions as in (2.8), and for $\psi \in \mathsf{R}(P)$ we consider the functions $u_1(\psi, \mathcal{T}_d) \in \mathfrak{L}(\mathcal{T}_d)$, $u_2(\psi, \mathcal{T}_d) \in \mathfrak{Q}(\mathcal{T}_d)$ and $\varepsilon(\psi, \mathcal{T}_d) \in \mathfrak{B}(\mathcal{T}_d)$ defined by:

$$\int_{\mathcal{R}} \nabla u_1(\psi, \mathcal{T}_d)\cdot\nabla v = \int_{\mathcal{R}} \psi v \quad \text{ for all } v \in \mathfrak{L}(\mathcal{T}_d) \tag{3.1}$$

$$\int_{\mathcal{R}} \nabla u_2(\psi, \mathcal{T}_d)\cdot\nabla v = \int_{\mathcal{R}} \psi v \quad \text{ for all } v \in \mathfrak{Q}(\mathcal{T}_d) \tag{3.2}$$

$$\int_{\mathcal{R}} \nabla \varepsilon(\psi, \mathcal{T}_d)\cdot\nabla v = \int_{\mathcal{R}} \psi v - \nabla u_1(\psi, \mathcal{T}_d)\cdot\nabla v \quad \text{ for all } v \in \mathfrak{B}(\mathcal{T}_d)\ . \tag{3.3}$$

This definition for $u_1(\psi, \mathcal{T}_d)$ coincides with the minimization formulation given in the Section 2, and $u_2(\psi, \mathcal{T}_d)$ satisfies the analagous minimization problem. The function $\varepsilon(\psi, \mathcal{T}_d)$ is the projection of the linear residual error onto the space of edge bump functions, and is an example of a hierarchical basis error estimator, which are well-known in the literature (see, for example, [2, ch.5] and [3]). The quantity $\|\nabla\varepsilon(\psi, \mathcal{T}_d)\|$ is much cheaper to compute than $\|\nabla u_2(\psi, \mathcal{T}_d) - \nabla u_1(\psi, \mathcal{T}_d)\|$, and provides a reliable estimate of the actual error $\|\nabla u(\psi) - \nabla u_1(\psi, \mathcal{T}_d)\|$.

Using arguments similar to those of Dörfler and Nochetto in [11], we will prove in Corollary 5.10 of Section 5 that there exists a constant $C_1(\mathcal{T}_d)$ depending solely on the shape regularity of $\mathcal{T}_d$ such that

$$\|\nabla u(\psi) - \nabla u_1(\psi, \mathcal{T}_d)\| \leq C_1(\mathcal{T}_d)\|\nabla\varepsilon(\psi, \mathcal{T}_d)\| + \operatorname{osc}(\psi, \mathcal{T}_d)\ . \tag{3.4}$$

The term $\operatorname{osc}(\psi, \mathcal{T}_d)$ is a measure of the oscillation in the data ψ. There are various ways of describing data oscillation to be found in the literature (see, for example, [11, 26]), and our definition will be similar in spirit to those. In Section 5, we will define $\operatorname{osc}(\psi, \mathcal{T}_d)$ explicitly and give a computable bound on $C_1(\mathcal{T}_d)$. For now, we merely state that, for $\psi \in \mathcal{H}_0^1(\mathcal{R})$,

$$\operatorname{osc}(\psi, \mathcal{T}_d) \leq C_2(\mathcal{T}_d)d^2\|\nabla\psi\|\ , \tag{3.5}$$

where $C_2(\mathcal{T}_d)$ depends solely on the shape regularity of $\mathcal{T}_d$. Our estimate of $\eta_i(P)$ based on $\varepsilon(\psi, \mathcal{T}_d)$, for some $\mathsf{R}(P) \subset \mathfrak{L}(\mathcal{T}_d)$, is given by

$$\eta_i(\mathfrak{B}_d, P) = \max_{\substack{\mathcal{S} \subset \mathsf{R}(P) \\ \dim \mathcal{S} = m-i+1}} \min_{\psi \in \mathcal{S}} \frac{\|\nabla \varepsilon(\psi, \mathcal{T}_d)\|}{\sqrt{\|\nabla u_1(\psi, \mathcal{T}_d)\|^2 + \|\nabla \varepsilon(\psi, \mathcal{T}_d)\|^2}} \ , \tag{3.6}$$

and we have the following theorem.

Theorem 3.1. *Let* $\mathbf{H} = -\triangle$ *be the Dirichlet Laplacian in* $\mathcal{R}$, *which we triangulate with* $\mathcal{T}_d$. *If we take* P_d *to be the orthogonal projection onto the linear span of* ψ_i^d, $i = 1, \dots m$ *from (2.8), then*

$$1 \le \frac{\eta_i(P_d)}{\eta_i(\mathfrak{B}_d, P_d)} \le C_1(\mathcal{T}_d) + \max_{\psi \in \mathsf{R}(P_d), \|\psi\|=1} \frac{\operatorname{osc}(\psi, \mathcal{T}_d)}{\|\nabla \varepsilon(\psi, \mathcal{T}_d)\|} \ ,$$

where $C_1(\mathcal{T}_d)$ *is the optimal constant in (3.4).*

Proof. Take an arbitrary $\psi \in \mathsf{R}(P_d)$. To establish the left-hand inequality, we first note that

$$\begin{aligned} \|\nabla u(\psi)\|^2 &= \|\nabla(u(\psi) - u_1(\psi, \mathcal{T}_d) - \varepsilon(\psi, \mathcal{T}_d))\|^2 + \|\nabla u_1(\psi, \mathcal{T}_d)\|^2 + \|\nabla \varepsilon(\psi, \mathcal{T}_d)\|^2 \\ &\ge \|\nabla u_1(\psi, \mathcal{T}_d)\|^2 + \|\nabla \varepsilon(\psi, \mathcal{T}_d)\|^2 \ . \end{aligned}$$

Therefore, we have

$$\begin{aligned} \frac{\|\nabla(u(\psi) - u_1(\psi, \mathcal{T}_d))\|^2}{\|\nabla u(\psi)\|^2} &= 1 - \frac{\|\nabla u_1(\psi, \mathcal{T}_d)\|^2}{\|\nabla u(\psi)\|^2} \\ &\ge 1 - \frac{\|\nabla u_1(\psi, \mathcal{T}_d)\|^2}{\|\nabla u_1(\psi, \mathcal{T}_d)\|^2 + \|\nabla \varepsilon(\psi, \mathcal{T}_d)\|^2} \\ &= \frac{\|\nabla \varepsilon(\psi, \mathcal{T}_d)\|^2}{\|\nabla u_1(\psi, \mathcal{T}_d)\|^2 + \|\nabla \varepsilon(\psi, \mathcal{T}_d)\|^2} \ . \end{aligned}$$

To prove the right-hand estimate, we note that

$$\frac{\|\nabla u(\psi) - \nabla u_1(\psi, \mathcal{T}_d)\|}{\|\nabla u(\psi)\|} \le C_1(\mathcal{T}_d) \frac{\|\nabla \varepsilon(\psi, \mathcal{T}_d)\|}{\|\nabla u(\psi)\|} + \frac{\operatorname{osc}(\psi, \mathcal{T}_d)}{\|\nabla u(\psi)\|} \ .$$

Replacing $\|\nabla u(\psi)\|$ by $\sqrt{\|\nabla u_1(\psi, \mathcal{T}_d)\|^2 + \|\nabla \varepsilon(\psi, \mathcal{T}_d)\|^2}$ in the denominator only increases the righthand side. The conclusions of the theorem now follow readily from the definitions. Q.E.D.

Under the standard non-degeneracy assumption,

$$\|\nabla u(\psi) - \nabla u_1(\psi, \mathcal{T}_d)\| \sim d\|\psi\| \text{ for sufficiently small } d \ , \tag{3.7}$$

Theorem 3.1 establishes the equivalence of $\eta_i(P_d)$ and the computable $\eta_i(\mathfrak{B}_d, P_d)$. In particular, the non-degeneracy assumption together with (3.4) and (3.5), imply that

$$\operatorname{osc}(\psi, \mathcal{T}_d) \le C_2(\mathcal{T}_d)\, \mu_{\max}(\mathsf{R}(P_d))\, d^2 \|\psi\| \ \text{ and} \tag{3.8}$$

$$\|\varepsilon(\psi, \mathcal{T}_d)\| \sim \|\nabla u(\psi) - \nabla u_1(\psi, \mathcal{T}_d)\| \sim d\|\psi\| \ , \tag{3.9}$$

where $\mu_{\max}(\mathsf{R}(P_d))$ is the largest of the Ritz values associated with the orthonormal Ritz basis of $\mathsf{R}(P_d)$.

The actual computation of the $\eta_i(\mathfrak{B}_d, P_d)$ involves solving a small, $m \times m$, generalized eigenvalue problem. Given a target range in which to look for eigenvalues of $-\Delta$ and a target number m of Ritz values/vectors of the discretized operator to compute, the initial stage of computation returns Ritz values $\{\mu_1^d, \ldots, \mu_m^d\}$ and corresponding Ritz vectors $\{\psi_1^d, \ldots, \psi_m^d\}$ – recall (2.14) from Section 2. It is these Ritz vectors that form the orthonormal basis for the space $\mathsf{R}(P_d)$ onto which P_d projects. If we take $\varepsilon_i^d = \varepsilon(\psi_i^d, \mathcal{T}_d)$, it is clear that computing (3.6) for every i is equivalent to solving the generalized eigenvalue problem

$$E\mathbf{v} = \eta^2\,(E + D)\mathbf{v} \quad , \quad D = \mathrm{diag}(\mu_1^d, \ldots, \mu_m^d) \quad , \quad E_{ij} = \int_{\mathcal{R}} \nabla\varepsilon_i^d \cdot \nabla\varepsilon_j^d \ . \tag{3.10}$$

Here we have relied the fact that both $u_1(\psi, \mathcal{T}_d)$ and $\varepsilon(\psi, \mathcal{T}_d)$ depend on ψ linearly.

We make a few final remarks before returning to consider the perturbation construction from Section 2 and what it tells us about the asymptotic behavior of the approximation defects. Our first remark is that, although we deliberately phrased Theorem 3.1 in such a way that the constant $C_1(\mathcal{T}_d)$ depends on nothing other than the shape regularity of the mesh – in particular, $C_1(\mathcal{T}_d) \not\to 1$ as $d \to 0$ – it is actually not unreasonable to expect that

$$\frac{\eta_i(P_d)}{\eta_i(\mathfrak{B}_d, P_d)} \to 1 \text{ as } d \to 0 \tag{3.11}$$

in practice, certainly for convex domains – where we have smoothness of all eigenvectors up to the boundary. Justification of this expectation is based on recent work of the second author on the asymptotic exactness of $\|\nabla\varepsilon(\psi, \mathcal{T}_d)\|$ as an approximation of $\|\nabla(u(\psi) - u_1(\psi, \mathcal{T}_d))\|$ for **fixed** ψ [29], and the fact that the invariant subspaces $\mathsf{R}(P_d)$ are converging.

This same sort of reasoning suggests that error estimates based on gradient recovery might also work very well in this context - many such estimators have also been proven to yield asymptotically exact approximations of error under certain assumptions, and are seen to do so in practice even when these assumptions do not hold (or cannot be verified). For concreteness, we briefly mention the recovery scheme of Bank and Xu [5, 6] and how it can be used in our context. Given the piecewise constant $\nabla u_1(\psi, \mathcal{T}_d)$, each component is projected into $\mathfrak{L}(\mathcal{T}_d)$ using the L^2-projection, and this is potentially followed by a few sweeps of multigrid-like smoothing process in order to force superconvergence of the estimator if adaptive refinement has negatively affected approximate mesh symmetries. Written briefly[3],

$$\nabla u_1(\psi, \mathcal{T}_d) \mapsto S^k(\mathcal{T}_d)Q(\mathcal{T}_d)\nabla u_1(\psi, \mathcal{T}_d) \in \mathfrak{L}(\mathcal{T}_d) \times \mathfrak{L}(\mathcal{T}_d).$$

The corresponding analoge of our $\eta_i(\mathfrak{B}_d, P_d)$ is

$$\eta_i(\mathfrak{L}_d \times \mathfrak{L}_d, P_d) = \max_{\substack{\mathcal{S} \subset \mathsf{R}(P_d) \\ \dim \mathcal{S} = m-i+1}} \min_{\psi \in \mathcal{S}} \frac{\|S^k(\mathcal{T}_d)Q(\mathcal{T}_d)\nabla u_1(\psi, \mathcal{T}_d) - \nabla u_1(\psi, \mathcal{T}_d)\|}{\|S^k(\mathcal{T}_d)Q(\mathcal{T}_d)\nabla u_1(\psi, \mathcal{T}_d)\|} \ .$$

Although we have only explicitly mentioned the Bank/Xu recovery scheme, others might also be used – but one should take care that the recovery scheme is linear with respect to $u_1(\psi, \mathcal{T}_d)$ so that the $\eta_i(\mathfrak{L}_d \times \mathfrak{L}_d, P_d)$ can be computed by solving a small generalized eigenvalue problem analagous to the one described above for the $\eta_i(\mathfrak{B}_d, P_d)$. In fact, many *a posteriori* error estimators could feasibly be used in this context – the main theoretical differences being with what we might be able to prove similar to Theorem 3.1.

[3] Here we have deliberately (for optical reasons) used the notation from [5, 6]. In our notation is $Y_d = Q(\mathcal{T}_d)$ and $S^k(\mathcal{T}_d)$ represents a couple of smoothing multigrid-steps.

Finally we remark briefly that, whatever procedure is used to compute ψ_i^d, we will in fact get a perturbation $\tilde{\psi}_i^d$ of it. However, the well-conditioning of the system associated with the computation of $\varepsilon_i(\psi)$ – which we establish explicitly in Section 5 – guarantees that this approximation error is not unduly magnified. In symbols, $\tilde{\varepsilon}_i(\tilde{\psi}_i^d) \sim \varepsilon_i(\tilde{\psi}_i^d) \sim \varepsilon_i(\psi_i^d)$. In other words, the approximation of E which we actually compute is of good quality and (3.10) is a well-behaved positive definite $m \times m$ generalized eigenvalue problem.

4. On the asymptotic behavior of the estimators $\eta_i(P_d)$

We now reuse the general perturbation construction from [18, 15], which was outlined in Section 2. To make our analysis we go back to [18, Remark 3.2]. Let Y_d be a sequence of orthogonal projections which converges strongly to the identity operator $\mathbf{I}$. In order to be concrete one could think of Y_d as an orthogonal projections onto the space of piecewise linear functions $\mathfrak{L}(\mathcal{T}_d)$. Let us assume that we want to approximate the eigenvalue $\lambda_q(\mathbf{H})$ of finite multiplicity m and that we are given (by some finite element procedure) a sequence of projections P_d, such that P_d and $\mathbf{H}_{Y_d}$ commute. We further assume that $\dim \mathsf{R}(P_d) = m$ and that P_d satisfy the assumptions of Theorem 2.1. For such a P_d the formula (2.10) holds. In this abstract formulation (2.10) reads

$$\eta_i(P_d) = \max_{\substack{\mathcal{S}\subset \mathsf{R}(P_d)\\ \dim\mathcal{S}=m-i+1}} \min_{\psi\in\mathcal{S}} \frac{\|\mathbf{H}^{-1}\psi - \mathbf{H}_{Y_d}^{-1}\psi\|_{\mathbf{H}}}{\|\mathbf{H}^{-1}\psi\|_{\mathbf{H}}}.$$

In the rest of this section we freely use both notations $\mathbf{H}^{-1}\psi = u(\psi)$ and $\mathbf{H}_{Y_d}^{-1}\psi = u_{Y_d}(\psi)$ as is technically at a given instance more convenient.

To study the asymptotic behavior we make the following standard assumption, cf. [20, Assumption (2.14)],

$$\|u(\psi) - u_{Y_d}(\psi)\|_{\mathbf{H}} \leq \mathfrak{C}_1 \; d^{\alpha_1} \; \|\mathbf{H}^{1/2}\psi\|, \qquad \psi \in \mathcal{Q}(h) \tag{4.1}$$

and $0 < \alpha_1, \mathfrak{C}_1$. We also make an abstract non-degeneracy assumption (similar in spirit to (3.7))

$$\|u(\psi) - u_{P_d}(\psi)\|_{\mathbf{H}} \geq \mathfrak{c} \; d^{\alpha_1}\|\psi\|_{\mathbf{H}} = \mathfrak{c} \; d^{\alpha_1}\|\mathbf{H}^{1/2}\psi\|, \qquad \psi \in \mathsf{R}(P_d). \tag{4.2}$$

The constants $\mathfrak{C}$ and $\mathfrak{c}$ are independent from d and ψ. In the case in which $\mathbf{H} = -\triangle$ in $H_0^1(\mathcal{R})$ and there exists a disc $\mathfrak{D} \subset \mathcal{R}$ and a constant $k > 0$ such that

$$\min\{\mathrm{diam}(T) \;:\; T \in \mathcal{T}_d \text{ and } T \subset \mathfrak{D}\} \geq kd$$

then according to [13, Remark 4.1] the assumption (4.2) holds.

For the projections P_d and Y_d we construct the forms h_{P_d}, h_{Y_d} and the operators $\mathbf{H}_{P_d}$ and $\mathbf{H}_{Y_d}$ as in (2.14). If we define the operators $\Xi_d : \mathsf{R}(P_d) \to \mathsf{R}(P_d)$—as in (2.15)—and $\mathbf{W}_d : \mathsf{R}(P_d)^\perp \to \mathsf{R}(P_d)^\perp$ —as the operator which is defined by the form h_{P_d} in the space $\mathsf{R}(P_d)^\perp$—then the form

$$h(\mathbf{H}_{P_d}^{-1/2}\cdot, \mathbf{H}_{P_d}^{-1/2}\cdot) - \lambda_q(\mathbf{H}_{P_d}^{-1/2}\cdot, \mathbf{H}_{P_d}^{-1/2}\cdot)$$

can be represented—with respect to $P \oplus P^\perp = \mathbf{I}$—by the bounded operator matrix

$$H_s(\lambda_q) = \begin{bmatrix} \mathbf{I} - \lambda_q \Xi_d^{-1} & \Gamma_d^* \\ \Gamma_d & \mathbf{I} - \lambda_q \mathbf{W}_d^{-1} \end{bmatrix}. \tag{4.3}$$

Let us now outline some further properties of this construction. The proofs can be found in [16, 17, 18] and the references therein. It holds that

$$\begin{aligned} h(\psi,\phi)-h_{P_d}(\psi,\phi) &= h(P_d^{\perp}\psi,P_d\phi)+h(P_d\psi,P_d^{\perp}\phi), \qquad \psi\in\mathcal{Q}(h)\\ h(\psi,\phi)-h_{Y_d}(\psi,\phi) &= h(Y_d^{\perp}\psi,Y_d\phi)+h(Y_d\psi,Y_d^{\perp}\phi), \qquad \psi\in\mathcal{Q}(h)\\ h_{Y_d}(\psi,\phi) &= h_{P_d}(\psi,\phi), \qquad \phi\in\mathsf{R}(P_d),\ \psi\in\mathcal{Q}(h)\\ h(Y_d^{\perp}\psi,Y_d\phi) &= h(P_d^{\perp}\psi,P_d\phi), \qquad \phi\in\mathsf{R}(P_d),\ \psi\in\mathcal{Q}(h),\\ \mathbf{H}_{P_d}\psi &= \mathbf{H}_{Y_d}\psi, \qquad \psi\in\mathsf{R}(Y_d) \text{ or } \psi\in\mathsf{R}(Y_d^{\perp})\cap\mathcal{D}(\mathbf{H}_{Y_d}). \end{aligned}$$

With this in hand we can now prove a theorem about the asymptotic behavior of $\eta_i(P_d)$. Without reducing the level of generality we may assume that we are given an eigenvalue λ_q of multiplicity m and a sequence of projections P_d, $\dim\mathsf{R}(P_d)=m$ such that the associated Ritz values μ_i^d, $i=1,\ldots,m$ converge to λ_q when $d\to 0$. For Ritz vectors ψ_i^d we assume that they are always paired with eigenvectors v_i, $\mathbf{H}v_i=\lambda_q v_i$, $i=1,...,m$ in the sense of Proposition 2.5, cf. [18, Theorem 6.2]. The general case of a cluster of eigenvalues $\lambda_q\le\lambda_{q+1}\le\cdots\le\lambda_{q+m-1}$ can be reduced to the case of a single multiple eigenvalue and the conclusions, identities (4.4) and (4.5) below, remain unchanged.

Theorem 4.1. *Let Y_d be a sequence of projections such that the assumptions (4.1)–(4.2) hold. Let us assume we have a sequence of projections P_d such that all P_d satisfy the assumptions of Theorem 2.1 and for each d, P_d commutes with $\mathbf{H}_{Y_d}$. Then; assuming $\mathbf{H}_{P_d}\psi_i^d=\mu_i^d\psi_i^d$, $\psi_i^d\in\mathsf{R}(P_d)$, $\|\psi_i^d\|=1$; we have*

$$\lim_{d\to 0}\frac{\sum_{i=1}^m\frac{|\mu_i^d-\lambda_q|}{\mu_i^d}}{\eta_1^2(P_d)+\cdots+\eta_m^2(P_d)}=1 \tag{4.4}$$

$$\lim_{d\to 0}\frac{\sum_{i=1}^m\frac{\|\nabla\psi_i^d-\nabla v_i\|^2}{\|\nabla v_i\|^2}}{\eta_1^2(P_d)+\cdots+\eta_m^2(P_d)}=1. \tag{4.5}$$

Analogous asymptotic properties are shared by other measures of the relative error from (2.11).

Proof. Using standard Schur complement Wilkinson's tricks (from Numerical Linear Algebra) on the identity (4.3) we can conclude—as has been done in [18, Equation (3.8)]—that

$$\mathbf{I}-\lambda_q\Xi_d^{-1}=\Gamma_d^*(\mathbf{I}-\lambda_q\mathbf{W}_d^{-1})^{-1}\Gamma_d. \tag{4.6}$$

This error representation formula is the basis for this argument. Before we proceed, note that $\mathrm{tr}(\mathbf{I}-\lambda_q\Xi_d^{-1})=\sum_{i=1}^m\frac{\mu_i^d-\lambda_i}{\mu_i^d}$ and that $\mathrm{tr}(\Gamma_d^*\Gamma_d)=\eta_1(P_d)^2+\cdots\eta_m^2(P_d)$. In particular, we have the following characterization

$$\begin{aligned} \|\Gamma_d\|=\eta_m(P_d) &= \max_{\substack{\psi,\phi\in\mathcal{Q}(h)\\ \psi,\phi\neq 0}}\frac{|h(\psi,\phi)-h_{P_d}(\psi,\phi)|}{\sqrt{h_{P_d}(\psi,\psi)h_{P_d}(\phi,\phi)}}\\ &= \max_{\substack{\psi,\phi\in\mathcal{Q}(h)\\ \psi,\phi\neq 0}}\frac{|h(\psi,\phi)-h_{P_d}(\psi,\phi)|}{\|\mathbf{H}_{P_d}^{1/2}\psi\|\|\mathbf{H}_{P_d}^{1/2}\psi\|} = \max_{\substack{\psi\in\mathcal{Q}(h),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}}\frac{|h(\psi,\phi)-h_{P_d}(\psi,\phi)|}{\|\mathbf{H}_{P_d}^{1/2}\psi\|\|\mathbf{H}_{P_d}^{1/2}\psi\|}\\ &= \max_{\substack{\psi\in\mathcal{Q}(h),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}}\frac{|h(\psi,\phi)-h_{Y_d}(\psi,\phi)|}{\|\mathbf{H}_{P_d}^{1/2}\psi\|\|\mathbf{H}_{P_d}^{1/2}\phi\|} = \max_{\substack{\psi\in\mathcal{Q}(h),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}}\frac{|h(Y_d^{\perp}\psi,Y_d\phi)|}{\|\mathbf{H}_{P_d}^{1/2}\psi\|\|\mathbf{H}_{P_d}^{1/2}\phi\|} \end{aligned}$$

$$
\begin{aligned}
&= \max_{\substack{\psi\in\mathcal{Q}(h),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}} \frac{\sqrt{h[Y_d^{\perp}\psi]h[Y_d\phi]}}{\|\mathbf{H}_{P_d}^{1/2}\psi\|\|\mathbf{H}_{P_d}^{1/2}\phi\|} \leq \max_{\substack{\psi\in\mathcal{Q}(h),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}} \frac{\sqrt{h[Y_d^{\perp}\psi]h[Y_d\phi]}}{\|\mathbf{H}_{P_d}^{1/2}\psi\|\|\mathbf{H}_{P_d}^{1/2}\phi\|}\\
&\leq \frac{\mathfrak{C}_1}{\sqrt{1-\eta_m(P_d)}}\, d^{\alpha_1}. \qquad (4.7)
\end{aligned}
$$

We can write (4.6) as

$$
\mathbf{I}-\lambda_q\Xi_d^{-1} = \Gamma_d^*\Gamma_d + \lambda_q\Gamma_d^*\mathbf{W}_d^{-1/2}(\mathbf{I}-\lambda_q\mathbf{W}_d^{-1})^{-1}\mathbf{W}_d^{-1/2}\Gamma_d. \tag{4.8}
$$

Note that $\min\left\{\frac{\lambda_1}{\lambda_q-\lambda_1},1\right\} \leq \nu \leq \mathfrak{g}_{q,\eta_m(P_i)}$, for all $\nu \in \Sigma(|(\mathbf{I}-\lambda_q\mathbf{W}_d^{-1})^{-1}|)$, and so asymptotically it is sufficient to analyze $s_i(\mathbf{W}^{-1/2}\Gamma_d)$, $i=1,\ldots,m$, i.e. the singular values of $\mathbf{W}^{-1/2}\Gamma_d$. As in (4.7), the estimate

$$
\begin{aligned}
\|\mathbf{W}_d^{-1/2}\Gamma_d\| &= \max_{\substack{\psi\in\mathcal{D}(\mathbf{H}_{P_d}),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}} \frac{|h(\psi,\phi)-h_{P_d}(\psi,\phi)|}{\|\mathbf{H}_{P_d}\psi\|\|\mathbf{H}_{P_d}^{1/2}\phi\|}\\
&= \max_{\substack{\psi\in\mathcal{D}(\mathbf{H}_{P_d}),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}} \frac{|h(Y_d^{\perp}\psi,Y_d\phi)|}{\|\mathbf{H}_{P_d}\psi\|\|\mathbf{H}_{P_d}^{1/2}\phi\|}\\
&\leq \max_{\substack{\psi\in\mathsf{R}(Y_d)^{\perp},\psi\in\mathcal{D}(\mathbf{H}_{P_d}),\phi\in\mathsf{R}(P_d)\\ \psi,\phi\neq 0}} \frac{\sqrt{h[Y_d^{\perp}\psi]h[Y_d\phi]}}{\|\mathbf{H}_{Y_d}\psi\|\|\mathbf{H}_{Y_d}^{1/2}\phi\|}\\
&\leq \eta_m(P_d)\ \|\mathbf{H}_{Y_d}^{-1/2}Y_d^{\perp}\| \qquad (4.9)
\end{aligned}
$$

holds. By an analogous computation we obtain

$$
\|\mathbf{W}_d^{-1/2}\Gamma_d\phi\| \leq \|\mathbf{H}_{Y_d}^{-1/2}Y_d^{\perp}\|\|\Gamma_d\phi\|, \qquad \phi\in\mathsf{R}(P_d),
$$

which yields the estimate

$$
s_i(\mathbf{W}_d^{-1/2}\Gamma_d) \leq \eta_i(P_d)\ \|\mathbf{H}_{Y_d}^{-1/2}Y_d^{\perp}\|, \qquad i=1,\ldots,m.
$$

Let us now combine the assumptions (4.1) and (4.2) with (4.7) and (4.9). Assumptions (4.1) and (4.2) and the characterization(4.7) imply $\mathrm{tr}(\Gamma_d^*\Gamma_d)=O(d^{2\alpha_1})$. Furthermore, since $\lim_{d\to 0}\|\mathbf{H}_{Y_d}^{-1/2}Y_d^{\perp}\|\to 0$, we conclude that

$$
\lim_{d\to 0}\frac{\mathrm{tr}(\Gamma_d^*\mathbf{W}_d^{-1/2}(\mathbf{I}-\lambda_q\mathbf{W}_d^{-1})^{-1}\mathbf{W}_d^{-1/2}\Gamma_d)}{\mathrm{tr}(\Gamma_d^*\Gamma_d)} = 0. \tag{4.10}
$$

If we now apply the trace operator $\mathrm{tr}(\cdot)$ on the equation (4.8) and utilize (4.9)–(4.10) we obtain the conclusion (4.4). The eigenvector estimate follows with the help of Proposition 2.5, which also holds for a general closed symmetric and positive definite form h. Q.E.D.

Remark 4.2. The asymptotic assumptions (4.1)–(4.2) are only made for technical convenience. The conclusion (4.4) follows under much milder assumptions. In fact, we only need to assume that $\eta_m(P_d)\to 0$, $d\to 0$ and $\eta_1(P_d)>0$, for $d>0$. In the case in which $\eta_1(P_d)=0$ for some $d>0$ we have that $v\in\mathsf{R}(P_d)$, for v an eigenvector of the operator $\mathbf{H}$. This case can be considered as trivial and excluded without reducing the level of generality. The assumptions (4.1)–(4.2) are needed to make the proof of (4.5) more convenient, see [18, Section 6.1] and cf. [25, Section 4.]. Without such an assumption we can still prove convergence, but only to some constant between 1 and a quantity dependent on the relative gap. The assumptions, (4.1)–(4.2) are however typically satisfied by most approximation methods, cf. [13, Remark 4.1]

5. Concerning Constants and Computational Cost

In this section we discuss what we might reasonably expect from the constant $C_{\mathcal{T}_d}$ appearing in Theorem 3.1 , and give a sense of the cost of computing the bump function error estimators by providing some bounds on the condition number and spectral radius of the corresponding linear system. All of these quantities of interest depend only on the underlying triangulation $\mathcal{T}_d$, so it is natural to discuss them together. We will make use of the following identities.

Lemma 5.1. *Let $\tau \in \mathcal{T}_d$ be given and let $z_k, x_k, \theta_k, e_k, L_k, h_k, \ell_k$ and b_k, $k = 1, 2, 3$, denote, respectively: vertex k, the coordinates of z_k, the measure of the associated interior angle, the edge opposite z_k, the length of e_k, the distance from z_k to the line containing e_k, the linear basis function associated with z_k and the quadratic bubble basis function associated with e_k. Furthermore, let $p, q, r \in \mathbb{Z}_{\geq 0}$. The following hold:*

$$\int_\tau \ell_1^p \ell_2^q \ell_3^r = \frac{p!q!r!}{(p+q+r+2)!}\, 2|\tau| \quad , \quad \int_{e_k} \ell_{k-1}^p \ell_{k+1}^q = \frac{p!q!}{(p+q+1)!}\, L_k \tag{5.1}$$

$$\|\nabla \ell_k\|_\tau^2 = \frac{1}{2}(\cot\theta_{k-1} + \cot\theta_{k+1}) \quad , \quad \int_\tau \nabla\ell_{k-1}\cdot\nabla\ell_{k+1} = -\frac{1}{2}\cot\theta_k \tag{5.2}$$

$$\|\nabla b_k\|_\tau^2 = \frac{4}{3}(\cot\theta_1 + \cot\theta_2 + \cot\theta_3) \quad , \quad \int_\tau \nabla b_{k-1}\cdot\nabla b_{k+1} = -\frac{4}{3}\cot\theta_k \tag{5.3}$$

$$\nabla\ell_k\cdot(x - x_{k-1}) = \nabla\ell_k\cdot(x - x_{k+1}) = \ell_k \tag{5.4}$$

$$\|x - x_k\|_\tau^2 = \frac{\cot\theta_{k-1} + 3\cot\theta_k + \cot\theta_{k+1}}{3}\,|\tau|^2 \tag{5.5}$$

$$\textit{For } f \in H^1(\tau),\ h_k \int_{e_k} f = \int_\tau 2f + (x - x_k)\cdot\nabla f\ . \tag{5.6}$$

Most, if not all, of these identities are well-known, and (5.1)-(5.5) can be verified by direct computation; integration by parts yields (5.6). We will also use

$$g(z_k, \tau) \doteq \|x - x_k\|_\tau / |\tau|$$

in what follows.

5.1. **A More Careful Look at (3.4).**

A derivation of (3.4) which will give us fairly detailed information on the constants involved and the data oscillation will require a careful look at some Clément-like quasi-interpolation estimates and the key arguments of Dörfler and Nochetto in [11]. For the arguments that follow, we consider a fixed $\psi \in L^2(\mathcal{R})$ – this is certainly more general than we need for the results in Section 3, but the arguments given below do apply in more general circumstances. To make the notation less cumbersome, we will generally suppress explicit dependencies on ψ and $\mathcal{T}_d$. For any $v \in H_0^1(\mathcal{R})$ and any $\mathcal{I}v \in \mathfrak{L}(\mathcal{T}_d)$, we have the well-known identity

$$\int_{\mathcal{R}} \nabla(u - u_1)\cdot\nabla v = \int_{\mathcal{R}} \psi(v - \mathcal{I}v) - \nabla u_1\cdot\nabla(v - \mathcal{I}v)\ . \tag{5.7}$$

We aim to bound this in terms of $\|\nabla\varepsilon\|$ and $\|\nabla v\|$, the data oscillation $\mathrm{osc}(\psi)$ - which will be defined later - and constants which depend only on the shape regularity of the mesh.

For the analysis below, we take $\mathcal{I}v$ to be the modified version of the Clément interpolant which is identical to that introduced by Carstensen [9], except near the boundary. Our analysis is partially motivated by that in [9, 34, 11, 27], and is quite similar at some points to that in [27]. It will be convenient for us to consider the set $\bar{\mathcal{V}}$ of **all** vertices in $\mathcal{T}_d$, including those on $\partial\mathcal{R}$. Recall that ℓ_z is the continuous, piecewise linear function such that $\ell_z(z) = 1$ and $\ell_z(z') = 0$ for $z' \in \bar{\mathcal{V}} \setminus \{z\}$. We take ω_z to be the support of ℓ_z. In what follows, for $f \in L^2(\mathcal{R})$, we define

$$f_z = \left(\int_{\omega_z} f\,\ell_z\right) / \left(\int_{\omega_z} \ell_z\right) \quad , \quad \mathcal{I}f = \sum_{z\in\mathcal{V}} f_z\ell_z \ . \tag{5.8}$$

The key properties of this interpolant are that,

$$\int_{\mathcal{R}} \psi(v - \mathcal{I}v) = \sum_{z\in\mathcal{V}} \int_{\omega_z} (\psi - \psi_z)(v - v_z)\ell_z + \sum_{z\in\bar{\mathcal{V}}\setminus\mathcal{V}} \int_{\omega_z} \psi v\ell_z \ , \tag{5.9}$$

(5.10)
$$\int_{\mathcal{R}} \nabla u_1 \cdot \nabla(v - \mathcal{I}v) = \sum_{z\in\mathcal{V}} \int_{\omega_z} \nabla u_1 \cdot \nabla[(v - v_z)\ell_z] + \sum_{z\in\bar{\mathcal{V}}\setminus\mathcal{V}} \int_{\omega_z} \nabla u_1 \cdot \nabla(v\ell_z) \ .$$

These identities follow directly from

$$\sum_{z\in\bar{\mathcal{V}}} \ell_z = 1 \text{ on } \mathcal{R} \quad , \quad \int_{\omega_z} (v - v_z)\ell_z = 0 \ .$$

We first treat (5.9), and to do so we will use the following Lemma.

Lemma 5.2. *Let $f \in H_0^1(\mathcal{R})$, $D_z = \mathrm{diam}(\omega_z)$ and $r_z = \max_{e\in\mathcal{E}_z} L_e$, where $\mathcal{E}_z \subset \mathcal{E}$ is the set of all interior edges having z as a vertex. The following hold:*

$$\text{For } z \in \bar{\mathcal{V}},\ \|(f - f_z)\ell_z^{1/2}\|_{\omega_z}^2 \le \frac{D_z^2}{\pi^2} \|\nabla f\|_{\omega_z}^2 \ , \tag{5.11}$$

$$\text{For } z \in \bar{\mathcal{V}} \setminus \mathcal{V},\ \|f\|_{\omega_z} \le r_z \|\nabla f\|_{\omega_z} \ . \tag{5.12}$$

Proof. The first inequality is a direct application of recent work by Chua and Wheeden [10, Thms. 1.1 and 1.2] on weighted Poincaré inequalities. Their work extends previous contributions by Payne and Weinberger, Acosta and Durán, and Bebendorf [30, 1, 8]. Although the results in [10] are stated for convex domains, we need not be concerned here with whether or not ω_z is convex, because the weight functions ℓ_z are supported in ω_z.

For the second inequality, we will assume, without loss of generality, that $f \in C_0^1(\mathcal{R})$. In particular it holds that $f(z) = 0$, and we will also assume, without loss of generality, that $z = 0$. We have $|f(x)| \le |\nabla f(x)||x|$, which yields

$$\|f\|_{\omega_z}^2 \le \left(\sup_{\omega_z} |x|^2\right) \|\nabla f\|_{\omega_z}^2 = r_z^2 \|\nabla f\|_{\omega_z}^2 \ .$$

Q.E.D.

We are now ready to provide a bound for $|\int_{\mathcal{R}} \psi(v - \mathcal{I}v)|$ in terms of $\|\nabla v\|$ and our first definition of oscillation of ψ.

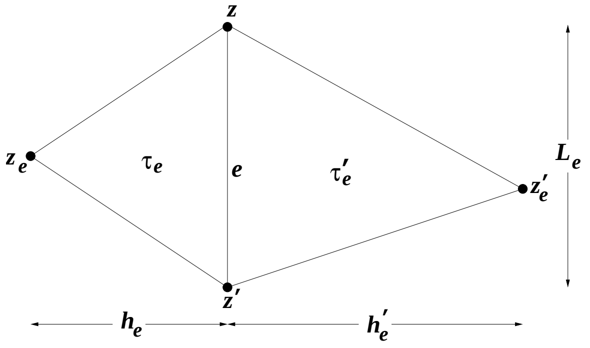

Figure 1. The edge $e \in \mathcal{E}$ and $\omega_e = \mathrm{supp}(b_e)$.

Theorem 5.3. *For $v \in H_0^1(\mathcal{R})$, it holds that*

$$\left|\int_{\mathcal{R}} \psi(v - \mathcal{I}v)\right| \leq \mathrm{osc}_1(\psi)\|\nabla v\| \quad \textit{where}$$

$$\mathrm{osc}_1^2(\psi) = \frac{3}{\pi^2}\sum_{z\in\mathcal{V}} D_z^2\|(\psi - \psi_z)\ell_z^{1/2}\|_{\omega_z}^2 + 3\sum_{z\in\bar{\mathcal{V}}\setminus\mathcal{V}} r_z^2\|\psi\ell_z\|_{\omega_z}^2 \ .$$

Proof. This follows directly from using the continuous Cauchy-Schwarz inequality, followed by the results of Lemma 5.2 with $f = v$, and then the discrete Cauchy-Schwarz inequality. The "extra" factor of 3 in $\mathrm{osc}_1(\psi)$ is due the fact that $\sum_{z\in\bar{\mathcal{V}}}\|\nabla v\|_{\omega_z}^2 = 3\|\nabla v\|^2$. Q.E.D.

We now consider the gradient terms (5.10). To analyze them we will we collect in the following lemma identities from [11] which are most useful to that end.

Lemma 5.4. *Let ω_e be the support of the quadratic bump basis function b_e associated with edge $e \in \mathcal{E}_z$, and k_z be constant. Let $J_e = J_e(u_1)$ denote the jump in the normal derivative of u_1 accross edge e. For $z \in \mathcal{V}$, (5.13)-(5.16) hold, and for $z \in \bar{\mathcal{V}} \setminus \mathcal{V}$, (5.13)-(5.14) hold.*

$$L_eJ_e = \frac{3}{2}\int_{\omega_e}\nabla u_1\cdot\nabla b_e = \frac{1}{2}k_z|\omega_e| + \frac{3}{2}\int_{\omega_e}(\psi - k_z)b_e - \nabla\varepsilon\cdot\nabla b_e \ , \tag{5.13}$$

$$\sum_{e\in\mathcal{E}_z} L_eJ_e = k_z|\omega_z| + \frac{3}{2}\sum_{e\in\mathcal{E}_z}\int_{\omega_e}(\psi - k_z)b_e - \nabla\varepsilon\cdot\nabla b_e \ , \tag{5.14}$$

$$\sum_{e\in\mathcal{E}_z} L_eJ_e = 2\int_{\omega_z}\nabla u_1\cdot\nabla\ell_z = \frac{2}{3}k_z|\omega_z| + 2\int_{\omega_z}(\psi - k_z)\ell_z \ , \tag{5.15}$$

$$k_z|\omega_z| = 6\int_{\omega_z}(\psi - k_z)\ell_z + \frac{9}{2}\sum_{e\in\mathcal{E}_z}\int_{\omega_e}\nabla\varepsilon\cdot\nabla b_e - (\psi - k_z)b_e \ . \tag{5.16}$$

Proof. These identities are what appear in [11], apart from our choice of sign for J_e, our notation, and the fact that we have replaced $u_2 - u_1$ in each of corresponding identities from [11] with ε - which follows from the definition of ε. Identity (5.16) follows from combining (5.14) and (5.15) in such a way as to eliminate the jump terms. Q.E.D.

We note that for $z \in \bar{\mathcal{V}}$,

$$\int_{\omega_z} \nabla u_1 \cdot \nabla[(v - v_z)\ell_z] = \sum_{e\in\mathcal{E}_z} \left(\frac{3}{2L_e}\int_e (v-v_z)\ell_z\right)\left(\int_{\omega_e} \nabla u_1 \cdot \nabla b_e\right) \ , \tag{5.17}$$

and for $z \in \bar{\mathcal{V}} \setminus \mathcal{V}$

$$\int_{\omega_z} \nabla u_1 \cdot \nabla(v\ell_z) = \sum_{e\in\mathcal{E}_z} \left(\frac{3}{2L_e}\int_e v\ell_z\right)\left(\int_{\omega_e} \nabla u_1 \cdot \nabla b_e\right) \ . \tag{5.18}$$

These identities follow from integration by parts and the first equality in (5.13). We will later use Lemma 5.4 to bound the integrals over ω_e, but now we consider the contribution of the edge integrals. Using (5.4)-(5.6) and the continuous Cauchy-Scwarz inequality, we deduce

Lemma 5.5. *Let τ_e be a triangle having $e \in \mathcal{E}$ as an edge, z as a vertex, and z_e as the vertex opposite e. Let $v \in H^1(\mathcal{R})$. We have*

$$\frac{3}{2L_e}\int_e (v-v_z)\ell_z \le \sqrt{\frac{27}{16|\tau_e|}}\|(v-v_z)\ell_z^{1/2}\|_{\tau_e} + \frac{3g(z_e,\tau_e)}{4}\|\nabla v\|_{\tau_e} \ ,$$

$$\frac{3}{2L_e}\int_e v\ell_z \le \sqrt{\frac{27}{32|\tau_e|}}\|v\|_{\tau_e} + \frac{3g(z_e,\tau_e)}{4}\|\nabla v\|_{\tau_e} \ .$$

Turning now to the gradient integrals over ω_e, we see that inserting (5.16) into (5.13) and regrouping terms yields, for $e \in \mathcal{E}$ and $z \in \mathcal{V}$ an endpoint of e,

(5.19)

$$\begin{aligned}\int_{\omega_e} \nabla u_1 \cdot \nabla b_e = \frac{2|\omega_e|}{|\omega_z|}\int_{\omega_z} (\psi - k_z)\ell_z + \left(\frac{3|\omega_e|}{2|\omega_z|} - 1\right)\int_{\omega_e} \nabla\varepsilon\cdot\nabla b_e - (\psi - k_z)b_e \\ + \frac{3|\omega_e|}{2|\omega_z|}\sum_{\hat{e}\in\mathcal{E}_z\setminus\{e\}}\int_{\omega_{\hat{e}}} \nabla\varepsilon\cdot\nabla b_{\hat{e}} - (\psi - k_z)b_{\hat{e}} \ .\end{aligned}$$

We make the choice $k_z = \psi_z$, which eliminates the first term in the previous identity. We have implicitly assumed above that all interior edges will have at least one interior vertex as an endpoint. We make this natural and easy-to-enforce assumption throughout. Using the continuous and discrete Cauchy-Schwarz inequalities on (5.19) and (5.13), we obtain

Lemma 5.6. *For $e \in \mathcal{E}$ and $z \in \mathcal{V}$ an endpoint of e, we have*

$$\left|\int_{\omega_e} \nabla u_1 \cdot \nabla b_e\right| \le c_1(z,e)\|\nabla\varepsilon\|_{\omega_z} + c_2(\psi,z,e) \quad \textit{where}$$

$$c_1^2(z,e) = 2\left(\frac{3|\omega_e|}{2|\omega_z|} - 1\right)^2\|\nabla b_e\|_{\omega_e}^2 + 2\left(\frac{3|\omega_e|}{2|\omega_z|}\right)^2\sum_{\hat{e}\in\mathcal{E}_z\setminus\{e\}}\|\nabla b_{\hat{e}}\|_{\omega_{\hat{e}}}^2$$

$$c_2(\psi,z,e) = \left|\frac{3|\omega_e|}{2|\omega_z|} - 1\right|\|(\psi-\psi_z)b_e\|_{\omega_e} + \frac{4|\omega_e|}{2|\omega_z|}\sum_{\hat{e}\in\mathcal{E}_z\setminus\{e\}}\|(\psi-\psi_z)b_{\hat{e}}\|_{\omega_{\hat{e}}} \ ,$$

and for $z \in \bar{\mathcal{V}} \setminus \mathcal{V}$ and $e \in \mathcal{E}_z$, we have

$$\left|\int_{\omega_e} \nabla u_1 \cdot \nabla b_e\right| \le \|\nabla b_e\|_{\omega_e}\|\nabla\varepsilon\|_{\omega_e} + \sqrt{\frac{8|\omega_e|}{45}}\|\psi\|_{\omega_e} \ .$$

We make two remarks before moving on to our main theorem concerning the gradient term (5.10). The first is to make a connection between terms of the form $\|fb_e\|_{\omega_e}$, for $f = \psi - \psi_z$, in Lemma 5.6 and the analogous terms $\|f\ell_z^{1/2}\|_{\omega_z}$ in Theorem 5.3. Namely, it holds that

$$\int_{\omega_e} fb_e = 4\int_{\omega_e} f\ell_z\ell_{z'} \leq 4\|\ell_z^{1/2}\ell_{z'}\|_{\omega_e}\|f\ell_z^{1/2}\|_{\omega_e} = 4\sqrt{\frac{|\omega_e|}{30}}\|f\ell_z^{1/2}\|_{\omega_e}\ .$$

Here, z' is the other endpoint of e. Our second remark concerns our treatment of boundary terms $\|\psi\ell_z\|_{\omega_z}$ and $\|\psi\|_{\omega_e}$ for $z \in \bar{\mathcal{V}} \setminus \mathcal{V}$ and $e \in \mathcal{E}_z$ in Theorem 5.3 and Lemma 5.6, and it deserves special notice.

Remark 5.7. Because we are particularly interested in eigenvalue problems in this paper, we have $\psi \in H_0^1(\mathcal{R})$ and can therefore use (5.12) to obtain bounds on the boundary terms $\|\psi\ell_z\|_{\omega_z}$ and $\|\psi\|_{\omega_e}$ of the form $r_z\|\nabla\psi\|_S$ for $S = \omega_z$ or ω_e. However, for more general $\psi \in H^1$, we can **still** obtain similar bounds for these boundary terms. For example, the first part of Lemma 5.6 can be used for $z \in \bar{\mathcal{V}}\setminus\mathcal{V}$ and $e \in \mathcal{E}_z$ by taking the bound $c_1(z', e)\|\nabla\varepsilon\|_{\omega_{z'}} + c_2(\psi, z', e)$, where $z' \in \mathcal{V}$ is the other endpoint of e.

Using Lemmas 5.5 and 5.6, and the discrete Cauchy-Schwarz inequality gives us a theorem concerning the gradient term (5.10).

Theorem 5.8. *For $v \in H_0^1(\mathcal{R})$, there exists a scale invariant constant C_1, depending only on the mesh $\mathcal{T}_d$, such that*

$$\int_{\mathcal{R}} \nabla u_1 \cdot \nabla(v - \mathcal{I}v) \leq C_1\|\nabla\varepsilon\|\|\nabla v\| + \mathrm{osc}_2(\psi)\|\nabla v\|,$$

where $\mathrm{osc}_2(\psi)$ is defined in the proof below.

Proof. It holds that, for $z \in \mathcal{V}$,

$$\int_{\omega_z} \nabla u_1 \cdot \nabla[(v - v_z)\ell_z] \leq \underbrace{\sqrt{\sum_{e\in\mathcal{E}_z} c_1^2(z,e)\left(\frac{27D_z^2}{16\pi^2|\tau_e|} + \frac{9g^2(z_e,\tau_e)}{16}\right)}}_{c_1(z)}\|\nabla\varepsilon\|_{\omega_z}\|\nabla v\|_{\omega_z}$$

$$+ \underbrace{\sqrt{\sum_{e\in\mathcal{E}_z} c_2^2(\psi,z,e)\left(\frac{27D_z^2}{16\pi^2|\tau_e|} + \frac{9g^2(z_e,\tau_e)}{16}\right)}}_{c_2(\psi,z)}\|\nabla v\|_{\omega_z}\ ,$$

and for $z \in \bar{\mathcal{V}} \setminus \mathcal{V}$,

$$\int_{\omega_z} \nabla u_1 \cdot \nabla(v\ell_z) \leq \underbrace{\max_{e\in\mathcal{E}_z}\sqrt{2\|\nabla b_e\|_{\omega_e}^2\left(\frac{27r_z^2}{32|\tau_e|} + \frac{9g^2(z_e,\tau_e)}{16}\right)}}_{\tilde{c}_1(z)}\sqrt{\sum_{e\in\mathcal{E}_z}\frac{\|\nabla\varepsilon\|_{\omega_e}^2}{2}}\|\nabla v\|_{\omega_z}$$

$$+ \underbrace{\sqrt{\sum_{e\in\mathcal{E}_z}\frac{8|\omega_e|}{45}\|\psi\|_{\omega_e}^2\left(\frac{27r_z^2}{32|\tau_e|} + \frac{9g^2(z_e,\tau_e)}{16}\right)}}_{\tilde{c}_2(\psi,z)}\|\nabla v\|_{\omega_z}\ .$$

Therefore, we have

$$\begin{aligned}\int_{\mathcal{R}} \nabla u_1 \cdot \nabla(v - \mathcal{I}v) \le{}& \underbrace{3\max\left\{\max_{z\in\mathcal{V}} c_1(z)\,,\ \max_{z\in\bar{\mathcal{V}}\setminus\mathcal{V}} \tilde{c}_1(z)\right\}}_{C_1} \|\nabla\varepsilon\|\|\nabla v\| \\ &+ \underbrace{\sqrt{3\sum_{z\in\mathcal{V}} c_2^2(\psi,z) + 3\sum_{z\in\bar{\mathcal{V}}\setminus\mathcal{V}} \tilde{c}_2^2(\psi,z)}}_{\mathrm{osc}_2(\psi)}\, \|\nabla v\|\ ,\end{aligned}$$

which completes the proof. Q.E.D.

Combining Theorems 5.3 and 5.8 the overall main theorem of this section, and its immediate corollaries.

Theorem 5.9. *For* $v \in H_0^1(\mathcal{R})$,

$$\int_{\mathcal{R}} \nabla(u - u_1)\cdot\nabla v \le (C_1\|\nabla\varepsilon\| + \mathrm{osc}(\psi))\|\nabla v\|,$$

where $\mathrm{osc}(\psi) = \mathrm{osc}_1(\psi) + \mathrm{osc}_2(\psi)$.

Corollary 5.10. *It holds that* $\|\nabla\varepsilon\| \le \|\nabla(u-u_1)\| \le C_1\|\nabla\varepsilon\| + \mathrm{osc}(\psi)$.

Corollary 5.11. *For* $\psi \in H_0^1(\mathcal{R})$*, there exists a scale invariant constant* C_2*, depending only on the mesh* $\mathcal{T}_d$*, such that* $\mathrm{osc}(\psi) \le C_2\|\nabla\psi\| d^2$.

This last corollary follows from the definition of $\mathrm{osc}(\psi)$, and the use of (5.11) and Lemma 5.2 with $f = \psi$. For $\psi \in H^1(\mathcal{R})$, if we still wanted a bound on the oscillation of the sort above, we would need to handle the boundary terms $z \in \bar{\mathcal{V}}\setminus\mathcal{V}$ slightly differently, in line with Remark 5.7.

The nearest we have found in the literature to bounds of the sort given in Corollary 5.10 is the discussion of Ern and Guermond given on pages 444–445 of [14], but the constants there are not given explicitly. We end this subsection with a few remarks on the above discussion. The first is that Corollary 5.10 replaces both the saturation assumption and the strengthened Cauchy-Schwarz inequality from the traditional analysis of hierarchical basis estimators. For the sake of clarity, we briefly state what the traditional analysis yields for our example. Suppose that there are constants $0 < \beta_1, \beta_2 < 1$ such that

$$\|\nabla(u-u_2)\| \le \beta_1\|\nabla(u-u_1)\|\ , \tag{5.20}$$

$$\int_{\mathcal{R}} \nabla v\cdot\nabla w \le \beta_2\|\nabla v\|\|\nabla w\| \text{ for } v \in \mathfrak{L}(\mathcal{T}_d),\ w \in \mathfrak{B}(\mathcal{T}_d)\ , \tag{5.21}$$

then we have the bounds

$$\|\nabla\varepsilon\| \le \|\nabla(u-u_1)\| \le \frac{1}{\sqrt{(1-\beta_1^2)(1-\beta_2^2)}}\,\|\nabla\varepsilon\|\ . \tag{5.22}$$

The assumption (5.20) is referred to as the saturation assumption, and its removal from the analysis of error estimators was the motivation of [11] as well as other work. Even though this assumption often holds asymptotically in practice, with $\beta_1 \to 0$ as $d \to 0$, the rate of convergence and the constants involved depend on u, and are not readily accessible. At any rate, one cannot disentangle this dependence upon u from $\|\nabla\varepsilon\|$ in (5.22). Maitre and Musy [24] have estimated β_2 solely in terms of the triangulation. In fact, they show that

$$\beta_2^2 \leq \max_{\tau\in\mathcal{T}_d} \frac{1}{2} + \frac{1}{3}\sqrt{\cos^2\theta_1 + \cos^2\theta_2 + \cos^2\theta_3 - \frac{3}{4}}\ .$$

In contrast, we have completely eliminated dependencies on any unknown quantity from our constant C_1. In fact, we even removed dependence upon the known quantity ε in order to obtain a constant which depends solely on the mesh, by taking the pessimistic bound

$$\int_{\omega_e} \nabla\varepsilon\cdot\nabla b_e \leq \|\nabla\varepsilon\|_{\omega_e}\|\nabla b_e\|_{\omega_e}\ .$$

If we were happy to have something more in the spirit of the traditional analysis – but computable! – we could use

$$\left|\int_{\omega_e} \nabla\varepsilon\cdot\nabla b_e\right| = \beta(\varepsilon,e)\|\nabla\varepsilon\|_{\omega_e}\|\nabla b_e\|_{\omega_e}\ ,$$

and modify the above analysis accordingly. A potential benefit of this is that we generally expect that $\beta(\varepsilon, e) < 1$ – perhaps much smaller than 1.

Remark 5.12. Our second remark is that the analysis of nearly all *a posteriori* error estimates are derived from the fundamental identity (5.7) – notable exceptions being those of gradient recovery type – so the analysis provided here has a good chance of improving many known results, certainly in the sense of making all involved constants explicitly computable. A topic of future work of the second author is to extend the analysis given here to more general elliptic operators and boundary conditions, and other types of error estimates – including some of gradient recovery type.

5.2. **The Conditioning of the Bump Stiffness Matrix.** Finally, we consider the cost of computing the error estimator $\varepsilon(\psi,\mathcal{T}_d) \approx u(\psi) - u_1(\psi,\mathcal{T}_d)$. The system matrix for the computation is given by $B_{ij} = (\nabla b_j, \nabla b_i)$, where we recall that $b_k \in \mathfrak{B}(\mathcal{T}_d)$ is the basis function associated with the interior edge e_k. The matrix B is certainly larger than the original stiffness matrix A used for computing $u_1(\psi,\mathcal{T}_d)$ - it can easily have three or four times the number of rows and columns, but never more than five nonzeros per row. However, in contrast to the behavior of A, the condition number of B does not deteriorate as the mesh parameter d decreases, and the diagonal of B is such an effective preconditioner that many opt to solve the diagonal system instead. The resulting error estimator $\hat{\varepsilon}(\psi,\mathcal{T}_d)$ does not lose much of its quality (see, for example [2, Ch. 5]), and many consider the compromise for this further speed-up to be worth it. In fact, this is precisely what is done in [28]. In either case, the cost of computing these sorts of error estimates is comparable to other commonly used methods.

The remainder of this subsection is devoted to a more detailed look at the eigenvalues of B and its diagonal $D = \mathrm{diag}(B)$. In particular, we show that they are both spectrally equivalent to the identity (and hence to each other), with reasonable constants of equivalence under reasonable assumptions on the angles in the mesh. We first consider the element stiffness matrices B_τ, which in our case are given explicitly as

$$B_\tau = \frac{4}{3}\begin{pmatrix} \rho & -\cot\theta_3 & -\cot\theta_2 \\ -\cot\theta_3 & \rho & -\cot\theta_1 \\ -\cot\theta_2 & -\cot\theta_1 & \rho \end{pmatrix}\ ,\ \rho = \cot\theta_1 + \cot\theta_2 + \cot\theta_3\ . \tag{5.23}$$

The eigenvalues of B_τ are

$$\sigma_k = \frac{4}{3}\left(\rho - 2\sqrt{\frac{\cot^2\theta_1 + \cot^2\theta_2 + \cot^2\theta_3}{3}}\cos\left(\frac{\theta + 2(k-1)\pi}{3}\right)\right)\ , \tag{5.24}$$

$$\theta = \arccos\left(\frac{3\sqrt{3}\cot\theta_1\cot\theta_2\cot\theta_3}{(\cot^2\theta_1 + \cot^2\theta_2 + \cot^2\theta_3)^{3/2}}\right)\ ,\ k = 1,2,3\ . \tag{5.25}$$

It holds that $0 < \sigma_1 \leq \sigma_3 \leq \sigma_2$, and

$$0.744\,I \leq B_\tau \leq 4.553\,I\ ,\ 2.309\,I \leq D_\tau \leq 2.667\,I \quad \text{if } \pi/4 \leq \theta_k \leq \pi/2\ , \tag{5.26}$$

$$0.487\,I \leq B_\tau \leq 10.373\,I\ ,\ 2.309\,I \leq D_\tau \leq 5.105\,I \quad \text{if } \pi/8 \leq \theta_k \leq 3\pi/4\ . \tag{5.27}$$

Combining these elementwise estimates into estimates for the full matrices B and D, we obtain

$$1.488\,I \leq B \leq 9.106\,I\ ,\ 4.618\,I \leq D \leq 5.334\,I \tag{5.28}$$
$$\text{if all angles are between } \pi/4 \text{ and } \pi/2\ ,$$

$$0.974\,I \leq B \leq 20.746\,I\ ,\ 4.618\,I \leq D \leq 10.210\,I \tag{5.29}$$
$$\text{if all angles are between } \pi/8 \text{ and } 3\pi/4\ .$$

These estimates give a rough sense of what can be expected of the bump stiffness matrix B. In particular, we see that B is already pretty well-conditioned, and that D provides a very good preconditioner.

6. Experiments

In this section we provide experiments for the model problem on three different domains which illustrate the effectivity of our estimates $\eta_i(\mathfrak{B}_d, P_d)$ of the approximation defects $\eta_i(P_d)$ – which, in turn, are estimates of the relative eigenvalue and eigenvector errors. In particular, we focus on the quality of our trace-type estimates

$$EFF = \frac{\sum_{i=1}^m \frac{|\mu_i^d - \lambda_q|}{\mu_i^d}}{\sum_{i=1}^m \eta_i^2(\mathfrak{B}_d, P_d)} \tag{6.1}$$

for both single (possibly) degenerate eigenvalues, such as λ_q in the above equation, or clusters of eigenvalues which may include degenerate members. We recall that our model problem is the Dirichlet Laplace eigenvalue problem: $-\Delta u = \lambda u$ in Ω, $u = 0$ on $\partial\Omega$. The three domains under consideration are the unit square, the L-shaped domain consisting of a concatenation of three unit squares, and the dumbbell domain consisting of two $\pi \times \pi$ squares connected by a $\pi/4 \times \pi/4$ square (see Figure 2). The unit square was chosen because of the exact knowledge of its eigenvalues and vectors and the fact that it has many degenerate eigenvalues in the lower part of its spectrum. The L-shaped domain appears frequently in the literature as one of the simplest domains for which analytic solutions of the eigenvalue problem are not generally known. The dumbbell domain provides examples of many pairs (and larger collections) of eigenvalues which just barely miss being degenerate, because the small bridge between the two larger squares has a

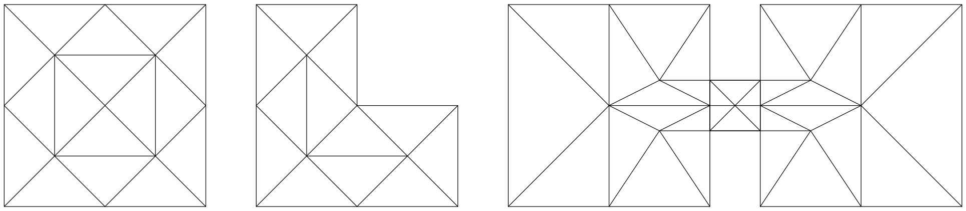

FIGURE 2. The unit square, L-shaped, and dumbbell domains together with their initial triangulations.

symmetry-breaking effect. Although exact eigenvalues are not generally known for the latter two domains, Trefethen and Betcke [32] have computed several of them to a high degree of accuracy, and we will use their computed values as the "exact" values in our effectivity tests.

The code used for the experiments was written by the second author in MATLAB, and makes use of its linear solver and eigenvalue solvers – EIGS for the large sparse generalized eigenvalue problems coming from the finite element discretizations, the "backslash" operation[4] for computing the approximate error functions $\varepsilon(\mu_i^d\psi_i^d, \mathcal{T}_d)$, and EIG for the small generalized eigenvalue problems needed to compute the our estimates of the approximation defects, $\eta_i^2(\mathfrak{B}_d, P_d)$. We also use MATLAB's sparse matrix format. The data structures for the triangulation are triangle-based, and are modeled after those found in PLTMG [4]. For adaptive refinement, we use Rivara's backward-longest-edge bisection algorithm [31] for marked triangles. The marking strategy is based on the local indicators

$$\sum_{i=1}^{m} \frac{\|\nabla\varepsilon(\bar{\psi}_i^d, \mathcal{T}_d)\|_\tau^2}{\|\nabla\varepsilon(\bar{\psi}_i^d, \mathcal{T}_d)\|^2 + \|\nabla u_1(\bar{\psi}_i^d, \mathcal{T}_d)\|^2} \ ,$$

where $\bar{\psi}_i^d \in \mathsf{R}(P_d)$ is an argument satisfying

$$\frac{\|\nabla\varepsilon(\bar{\psi}_i^d, \mathcal{T}_d)\|^2}{\|\nabla\varepsilon(\bar{\psi}_i^d, \mathcal{T}_d)\|^2 + \|\nabla u_1(\bar{\psi}_i^d, \mathcal{T}_d)\|^2} = \eta_i^2(\mathfrak{B}_d, P_d) \ .$$

The triangles whose indicators are larger than the median are marked. Certainly other marking strategies and indicators could be used – if one wished to use a weighting which favored certain approximation defects (relative errors) more heavily than others, then the approach described above could be modified accordingly. Additionally, one could also include local indicators based on the data oscillations (or some suitable simplification of them) for the marking strategy. This might be particularly useful in the early stages of adaptive refinement if larger eigenvalues (with highly oscillatory eigenvectors) are to be approximated.

Remark 6.1. We emphasize that, although the $\eta_i^2(\mathfrak{B}_d, P_d)$ and the corresponding local indicators are computed using the basis for $\mathsf{R}(P_d)$ which is given by the eigensolver, the actual computed quantities are **basis-independent**. This is a very useful quality to have for error estimation and adaptive refinement, especially in

[4] Because our main emphasis in these experiments is to illustrate the effectivity of our estimator (not necessarily to do things in the fastest possible way), and because the asymptotic behavior of our estimators is observed even for relatively small problems, we have not bothered to use a fast iterative solver for the computation of the $\varepsilon(\mu_i^d\psi_i^d, \mathcal{T}_d)$.

the case of degenerate eigenvalues, because one may have little control over the basis computed by the eigensolver. In particular, the computed bases may "drift" as the mesh is refined either uniformly or adaptively – meaning that, although the bases for two different meshes both span spaces which approximate the true invariant subspace, the bases themselves are only approximately equal up to an orthogonal transformation. Moreover, for solvers such as EIGS, the computed basis may be different for consecutive calls on the same mesh! At any rate, an ideal mesh for a given problem should be well-suited for the entire invariant subspace, and not just to a given basis.

6.1. **The Unit Square.** As mentioned above, the eigenvalues and eigenvectors are explicitly known in this case. Namely, we have

$$(k^2+n^2)\pi^2 \quad \text{paired with} \quad 2\sin(k\pi x)\sin(n\pi y),$$

with the eigenfunctions as given above forming an orthonormal basis for the complete eigenspace. For our first experiment, we approximate the smallest six eigenvalues

$$10\pi^2\ ,\ 10\pi^2\ ,\ 8\pi^2\ ,\ 5\pi^2\ ,\ 5\pi^2\ ,\ 2\pi^2$$

having two simple eigenvalues and two degenerate pairs. In Table 1 we see the computed Ritz values in descending order, together with the effectivity indices (6.1) for our estimates of the relative error in approximating these eigenvalues at various levels of adaptive refinement – N indicates the number of degrees of freedom for the problem. We remark that the asymptotic behavior indicated by (4.4) and (3.11) is observed immediately, with nearly perfect effectivity observed throughout the refinement process. For this experiment and the analogous ones for the other domains, we observe quadratic convergence of the absolute and relative eigenvalue errors, which is predicted from *a priori* analysis for convex problems. Although we do not express these computations in table form, they can be deduced from the exact eigenvalues and the information given in Tables 1, 3 and 4.

Table 1. Eigenvalue estimates and effectivity indices for the unit square.

N	μ_6	μ_5	μ_4	μ_3	μ_2	μ_1	EFF
25	125.1204	122.8963	94.9338	54.5014	54.5014	20.7157	1.1471
45	113.1352	112.4708	91.0246	53.3258	53.3074	20.4309	1.1270
68	110.7839	110.2532	87.2628	52.4884	52.2481	20.3112	1.1157
111	106.5602	106.1753	82.5317	50.9809	50.9774	19.9609	1.0852
185	102.6784	102.4893	81.9499	50.4675	50.4675	19.9191	1.0810
295	101.6302	101.3446	80.7847	50.1042	50.0528	19.8777	1.0742
463	100.5934	100.5600	79.8253	49.7767	49.7680	19.7994	1.0652
731	99.7023	99.6333	79.7179	49.6299	49.6294	19.7851	1.0634
1131	99.4619	99.4031	79.4876	49.5507	49.5501	19.7770	1.0593
1765	99.2275	99.1885	79.2009	49.4685	49.4685	19.7592	1.0619
2769	98.9593	98.9461	79.1554	49.4236	49.4228	19.7509	1.0624
4248	98.8957	98.8811	79.1084	49.4033	49.4031	19.7494	1.0546

To illustrate the implications of (4.5) and (3.11), we first reconsider the statement of (4.5). The evaluation of the formula (4.5)—for multiple eigenvalues—is not easy within a practical numerical procedure. The problem is that the formula (4.5) assumes that the finite element eigenvalue procedure has returned the Ritz vectors which are matched to the eigenvectors from the invariant subspace in the sense of Proposition 2.5. For practical tests we will study the quotients $\min_{\psi\in\mathsf{R}(P_d)}\|\nabla\psi-\nabla v_i\|^2/\|\nabla v_i\|^2$ rather then the quotients $\|\nabla\psi_i^d-\nabla v_i\|^2/\|\nabla v_i\|^2$, which appeared in (4.5). According to the analysis of Beattie [7] these quotients have essentially the same asymptotic behavior. Statement (4.5) can now be expressed as

$$\lim_{d\to 0}\frac{\sum_{i=1}^m\frac{\min_{\psi\in\mathsf{R}(P_d)}\|\nabla\psi-\nabla v_i\|^2}{\|\nabla v_i\|^2}}{\sum_{i=1}^m\eta_i^2(P_d)}=1\ , \tag{6.2}$$

where we recall that the v_i form an orthonormal eigen-basis for the invariant subspace which we are trying to approximate with $\mathsf{R}(P_d)$. A direct computation shows that

$$\begin{aligned}\frac{\min_{\psi\in\mathsf{R}(P_d)}\|\nabla\psi-\nabla v_i\|^2}{\|\nabla v_i\|^2}&=1-\sum_{j=1}^m\frac{\lambda_i}{\mu_j}\left(\int_{\mathcal{R}}v_i\psi_j^d\right)^2\\ &=\sin^2\angle(v_i,\mathsf{R}(P_d))+\sum_{j=1}^m\frac{\mu_j-\lambda_j}{\mu_j}\left(\int_{\mathcal{R}}v_i\psi_j^d\right)^2\ ,\end{aligned}\tag{6.3}$$

so it is clear, at least theoretically, how to compute the numerator in (6.2). Furthermore, the result (4.4), together with the $\sin\Theta$ theorem from [19] proves (6.2), cf. [25, Section 4.]. A second natural measure of the effectivity of our computed $\eta_i(\mathfrak{B}_d,P_d)$ is

$$EFF_2=\frac{\sum_{i=1}^m\frac{\min_{\psi\in\mathsf{R}(P_d)}\|\nabla\psi-\nabla v_i\|^2}{\|\nabla v_i\|^2}}{\sum_{i=1}^m\eta_i^2(\mathcal{B}_d,P_d)}\ , \tag{6.4}$$

and we investigate it in the following experiment.

For this experiment, we consider the degenerate eigenpairs

$$\lambda_2=\lambda_3=5\pi^2\ ,\ v_2=2\sin\pi x\,\sin 2\pi y\ ,\ v_3=2\sin 2\pi x\,\sin\pi y\ ,$$

and see how closely (6.4) comes to the predicted asymptotic behavior (6.2). The results of this experiment are given in Table 2. We see that the asymptotically optimal behavior of the $\eta_i(\mathcal{B}_d,P_d)$ is realized even for coarse meshes. The quantities in the numerator of (6.4) are computed using a twelve-point quadrature rule – which is exact for polynomials of degree six – so the initial effectivity estimates are slightly inflated due to quadrature error. Pictures of the final two adapted meshes for this experiment are given in Figure 3.

Table 2. Eigenvector estimates and effectivity indices for the unit square and the degenerate eigenvalue $\lambda_2=\lambda_3=5\pi^2$.

N	5	9	15	29	45	73	129
EFF_2	1.8844	1.7255	1.4649	1.1744	1.1408	1.1119	1.0940
N	215	353	589	891	1385	2193	3369
EFF_2	1.0721	1.0754	1.0637	1.0551	1.0633	1.0608	1.0534

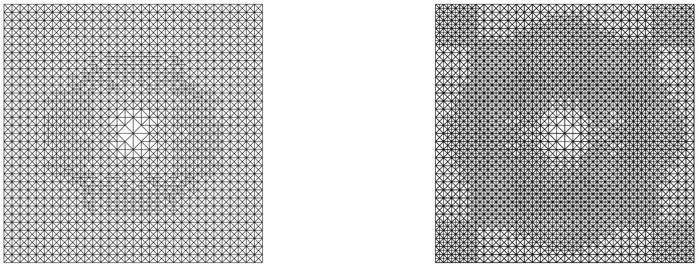

FIGURE 3. Adaptively refined meshes for the unit square and the degenerate eigenvalue $\lambda_2 = \lambda_3 = 5\pi^2$, having $N = 2193$ and $N = 3369$ degrees of freedom.

6.2. **The L-shaped Domain.** Although some of the eigenvalues and eigenvectors for the L-shaped domain are known explicitly, most are not. We take the highly accurate values computed by Trefethen and Betcke for the six smallest eigenvalues

$$41.474510\ ,\ 31.912636\ ,\ 29.521481\ ,\ 19.739209\ (2\pi^2)\ ,\ 15.197252\ ,\ 9.6397238$$

and perform the analogous experiment as was done for the square domain. The results of this experiment are given in Table 3. As before, we see very good effectivity immediately.

TABLE 3. Eigenvalue estimates and effectivity indices for the L-shaped domain.

N	μ_6	μ_5	μ_4	μ_3	μ_2	μ_1	EFF
25	53.1134	42.1280	37.1134	22.8935	16.8847	10.8142	1.2091
36	51.9270	38.9125	35.7211	21.9756	16.6162	10.5431	1.1617
64	47.6832	35.6148	32.5767	21.0477	16.1449	10.2087	1.1499
110	45.4205	34.2515	31.2875	20.5442	15.6511	9.9374	1.1107
169	43.9883	33.5960	30.9005	20.3856	15.5313	9.8560	1.0888
272	43.1556	32.8951	30.2754	20.0585	15.4193	9.7913	1.0970
440	42.3933	32.5259	29.9869	19.9485	15.3198	9.7247	1.0838
680	42.1248	32.3557	29.8686	19.9048	15.2820	9.6989	1.0700
1077	41.9302	32.1737	29.7306	19.8259	15.2555	9.6810	1.0812
1689	41.7377	32.0800	29.6446	19.7916	15.2294	9.6630	1.0747
2588	41.6457	32.0306	29.6128	19.7829	15.2201	9.6559	1.0647
3998	41.5982	31.9877	29.5823	19.7655	15.2136	9.6513	1.0718

6.3. **The Dumbbell Domain.** Again we take the smallest six highly accurate eigenvalues computed by Trefethen and Betcke,

$$4.9968508\ ,\ 4.9968371\ ,\ 4.8298953\ ,\ 4.8007611\ ,\ 1.9606830\ ,\ 1.9557938$$

and again we perform the analogous experiment. As noted in [32], the effect of the small "bridge" between the two $\pi \times \pi$ squares is to take the smallest three eigenvalues of a single $\pi \times \pi$ square – namely 5, 5 and 2 – and split them into nearly degenerate pairs. In Table 4, we again see excellent effectivity even for coarse triangulations.

Table 4. Eigenvalue estimates and effectivity indices for the dumbbell domain.

N	μ_6	μ_5	μ_4	μ_3	μ_2	μ_1	EFF
49	6.4262	6.3624	6.1456	6.1456	2.1777	2.1733	1.2062
85	5.6138	5.6138	5.5491	5.5184	2.0846	2.0809	1.1725
141	5.3668	5.3668	5.1913	5.1651	2.0256	2.0216	1.1348
243	5.2183	5.2128	5.0333	5.0083	1.9967	1.9929	1.1130
397	5.1219	5.1197	4.9619	4.9362	1.9867	1.9827	1.1246
650	5.0778	5.0765	4.9095	4.8847	1.9759	1.9718	1.1065
1060	5.0461	5.0461	4.8798	4.8528	1.9696	1.9652	1.1024
1693	5.0279	5.0278	4.8630	4.8345	1.9673	1.9626	1.1058
2691	5.0175	5.0174	4.8502	4.8221	1.9647	1.9600	1.0967
4274	5.0093	5.0093	4.8428	4.8143	1.9630	1.9583	1.0944

7. Conclusion

The primary aims of this paper were two-fold:

(1) To establish the equivalence of the approximation defects $\eta_i(P_d)$ and the corresponding relative eigenvalue and eigenvector errors.
(2) To provide a practical means of estimating these approximation defects which is provably effective and reliable.

With regard to the first aim, asymptotic exactness was proven in a very general setting, and detailed bounds were also given which always hold. The definition of the approximation defects is such that it is natural to derive estimates for them using the well-developed theory of *a posteriori* error estimation for elliptic boundary value problems. In principle, one could incorporate a number of different *a posteriori* techniques in our framework, and we mentioned the use of gradient recovery techniques explicitly, but our focus was on estimates of hierarchical type, $\eta_i(\mathfrak{B}_d, P_d) \approx \eta_i(P_d)$. For this type of estimator we asserted asymptotic exactness for the model problem on convex domains, and also gave detailed bounds which always hold – complementing the results mentioned above. Experiments verified the effectivity of our estimates of the approximation defects as trustworthy indicators of relative eigenvalue/eigenvector errors. In addition to the strengths of our approach mentioned above, we highlight three more. The case of degenerate eigenvalues is treated very naturally in our framework, requiring no special modification, and we need no assumptions concerning the convexity/non-convexity of the domain. Additionally, the approximation defects and their estimates truly are basis-independent, so one truly obtains information about how well the subspace $\mathsf{R}(P_d)$ – given in terms of some basis by whatever eigensolver is used – approximates the true invariant subspace of interest.

We finish with a brief outlook for future work in this area. All of our analysis was done in the context of piecewise linear finite elements, and the analysis of our hierarchical basis estimates was carried out only for the Laplacian with zero Dirichlet conditions. One clear direction in which our results can be extended is to consider more general elliptic operators and boundary conditions. Item (1) above is already dealt with in principle by the arguments given in this paper, so further work in this direction is really to prove something analogous to Theorem 3.1 in the more general setting. Some of the necessary modifications to our arguments are obvious, but others will require a more detailed look. Another area of future work is to provide similar analysis for other eigenvalue approximation methods, such as those arising from *hp*-finite element discrectizations. The *hp*-approach is in particular well-suited for eigenvalue problems, because of the higher order of smoothness of the eigenfunctions away from (non-convex) boundaries and regions of discontinuity of the coefficients of the differential operator.

Acknowledgement

The authors would like to thank Prof. Dr. Volker Enss, Aachen and Prof. Dr. Wolfgang Hackbusch, Leipzig for helpful discussions and support during the research and the preparation of this manuscript. L. Grubišić was partially supported by the grant :"Spectral decompositions – numerical methods and applications" of the Ministry for Science, Education and Sport of the Republic of Croatia. Support of the Graduate School: "Hierarchy and Symmetry in Mathematical Models", RWTH Aachen University is also kindly acknowledged.

References

[1] G. Acosta and R. G. Durán. An optimal Poincaré inequality in L^1 for convex domains. *Proc. Amer. Math. Soc.*, 132(1):195–202 (electronic), 2004.

[2] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2000.

[3] 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.

[4] R. E. Bank. Pltmg: A software package for solving elliptic partial differential equations, users' guide 9.0. Technical report, University of California, San Diego, 2004.

[5] R. E. Bank and J. Xu. Asymptotically exact a posteriori error estimators. I. Grids with superconvergence. *SIAM J. Numer. Anal.*, 41(6):2294–2312 (electronic), 2003.

[6] R. E. Bank and J. Xu. Asymptotically exact a posteriori error estimators. II. General unstructured grids. *SIAM J. Numer. Anal.*, 41(6):2313–2332 (electronic), 2003.

[7] C. Beattie. Galerkin eigenvector approximations. *Math. Comp.*, 69(232):1409–1434, 2000.

[8] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwendungen*, 22(4):751–756, 2003.

[9] C. Carstensen. Quasi-interpolation and a posteriori error analysis in finite element methods. *M2AN Math. Model. Numer. Anal.*, 33(6):1187–1202, 1999.

[10] S.-K. Chua and R. L. Wheeden. Estimates of best constants for weighted Poincaré inequalities on convex domains. *Proc. London Math. Soc. (3)*, 93(1):197–226, 2006.

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

[12] Z. Drmač and K. Veselić. New fast and accurate Jacobi SVD algorithm: II. *SIAM J. Matrix Anal. Appl.*, to appear. Preprint LAPACK Working Note 170.

[13] R. G. Durán, C. Padra, and R. Rodríguez. A posteriori error estimates for the finite element approximation of eigenvalue problems. *Math. Models Methods Appl. Sci.*, 13(8):1219–1229, 2003.

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

[15] L. Grubišić. *Ritz value estimates and applications in Mathematical Physics*. PhD Thesis, Fernuniversität in Hagen, 2005. Available through dissertation.de Verlag im Internet.

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

[17] L. Grubišić. A posteriori estimates for eigenvalue/vector approximations. *PAMM Proc. Appl. Math. Mech.*, 6(1):59–62, 2006.

[18] L. Grubišić. On Temple–Kato like inequalities and applications. *SIAM J. Numer. Anal.*, submitted. 2005–Preprint available from `http://arxiv.org/abs/math/0511408`.

[19] L. Grubišić and K. Veselić. On weakly formulated sylvester equation and applications. *Integral Equations Operator Theory*, to appear. 2005–Preprint available from `http://arxiv.org/abs/math/0507532`.

[20] W. Hackbusch. On the computation of approximate eigenvalues and eigenfunctions of elliptic operators by means of a multi-grid method. *SIAM J. Numer. Anal.*, 16(2):201–215, 1979.

[21] 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.

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

[23] 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.

[24] J.-F. Maitre and F. Musy. The contraction number of a class of two-level methods; an exact evaluation for some finite element subspaces and model problems. In *Multigrid methods (Cologne, 1981)*, volume 960 of *Lecture Notes in Math.*, pages 535–544. Springer, Berlin, 1982.

[25] 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.

[26] P. Morin, R. H. Nochetto, and K. G. Siebert. Convergence of adaptive finite element methods. *SIAM Rev.*, 44(4):631–658 (electronic) (2003), 2002. Revised reprint of "Data oscillation and convergence of adaptive FEM" [SIAM J. Numer. Anal. **38** (2000), no. 2, 466–488 (electronic); MR1770058 (2001g:65157)].

[27] P. Morin, R. H. Nochetto, and K. G. Siebert. Local problems on stars: a posteriori error estimators, convergence, and performance. *Math. Comp.*, 72(243):1067–1097 (electronic), 2003.

[28] K. Neymeyr. A posteriori error estimation for elliptic eigenproblems. *Numer. Linear Algebra Appl.*, 9(4):263–279, 2002.

[29] J. S. Ovall. Function, gradient and Hessian recovery using quadratic edge bump functions. *SIAM J. Numer. Anal.*, to appear.

[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] M.-C. Rivara. New longest-edge algorithms for the refinement and/or improvement of unstructured triangulations. *Internat. J. Numer. Methods Engrg.*, 40(18):3313–3324, 1997.

[32] 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.

[33] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh refinement techniques*. Wiley-Teubner Series Advances in Numerical Mathematics. John Wiley & Sons Ltd., Chichester, 1996.

[34] R. Verfürth. Error estimates for some quasi-interpolation operators. *M2AN Math. Model. Numer. Anal.*, 33(4):695–713, 1999.

Institut für reine und angewandte Mathematik, RWTH-Aachen, Templergraben 52, D-52062 Aachen, Germany (On leave from Department of Mathematics, University of Zagreb, Croatia)

E-mail address: `luka.grubisic@iram.rwth-aachen.de`

Max-Planck-Institut für Mathematik in den Naturwissenschaften, Inselstr. 22-26, D-04103 Leipzig, Germany

E-mail address: `ovall@mis.mpg.de`