hp-Finite Elements for Elliptic Eigenvalue Problems: Error estimates which are explicit with respect to λ, h, and p.

S. Sauter*

Abstract

Convergence rates for finite element discretisations of elliptic eigenvalue problems in the literature usually are of the form: If the mesh width h is fine enough then the eigenvalues resp. eigenfunctions converge at some well-defined rate. In this paper, we will determine the maximal mesh width h_0 – more precisely the minimal dimension of a finite element space – so that the asymptotic convergence estimates hold for $h \leq h_0$. This mesh width will depend on the size and spacing of the exact eigenvalues, the spatial dimension and the local polynomial degree of the finite element space.

For example in the one-dimensional case, the condition $\lambda^{3/4} h_0 \lesssim 1$ is sufficient for piecewise linear finite elements to compute an eigenvalue λ with optimal convergence rates as $h_0 \geq h \rightarrow 0$. It will turn out that the condition for eigenfunctions is slightly more restrictive. Furthermore, we will analyse the dependence of the ratio of the errors of the Galerkin approximation and of the best approximation of an eigenfunction on λ and h.

In this paper, the error estimates for the eigenvalue/-function are limited to the selfadjoint case. However, the regularity theory and approximation property cover also the non-selfadjoint case and, hence, pave the way towards the error analysis of non-selfadjoint eigenvalue/-function problems.

1 Eigenvalue problems for second order elliptic problems

In this paper, we shall deal with the numerical approximation of eigenvalue problems for linear second order partial differential equations.

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with boundary Γ and let $H^k(\Omega)$ denote the usual Sobolev space equipped with the scalar product $(\cdot,\cdot)_{H^k(\Omega)}$ and norm $\|\cdot\|_{H^k(\Omega)}$. For simplicity we restrict to the pure Dirichlet problem and denote by $H_0^1(\Omega)$ the subspace of $H^1(\Omega)$ consisting of all functions with vanishing boundary traces. We introduce the usual

*(stas@math.uzh.ch), Institut für Mathematik, Universität Zürich, Winterthurerstr 190, CH-8057 Zürich, Switzerland

seminorms formally by

$$\left\|\nabla^{\ell}u\right\|_{L^2(\Omega)}^2 := \sum_{|\alpha|=\ell}\frac{\ell!}{\alpha!}\left\|D^{\alpha}u\right\|_{L^2(\Omega)}^2 \quad \text{and} \quad \left\|\nabla^{\ell}u\right\|_{L^{\infty}(\Omega)} := \left\|\sqrt{\sum_{|\alpha|=\ell}\frac{\ell!}{\alpha!}\left|D^{\alpha}u\right|^2}\right\|_{L^{\infty}(\Omega)} \tag{1.1}$$

and set $|u|_{\ell} := \left\|\nabla^{\ell}u\right\|_{L^2(\Omega)}$ and $|u|_{\ell,\infty} := \left\|\nabla^{\ell}u\right\|_{L^{\infty}(\Omega)}$.

We shall deal with the problem of seeking eigenpairs $(\lambda, e) \in \mathbb{C} \times H_0^1(\Omega) \setminus \{0\}$ such that

$$a(e,v) = \lambda (e,v)_{L^2(\Omega)} \qquad \forall v \in H_0^1(\Omega), \tag{1.2}$$

where

$$a(u,v) = a_0(u,v) + a_1(u,v) \tag{1.3}$$

with[1]

$$a_0(u,v) := \int_{\Omega} \langle A\nabla u, \nabla \bar{v}\rangle + cu\bar{v} \quad \text{and} \quad a_1(u,v) := \int_{\Omega}\langle b, \nabla u\rangle \bar{v}. \tag{1.4}$$

The set of all eigenvalues is the spectrum and denoted by $\sigma(a)$. In this paper, we will consider the case of real analytic coefficients A, b, c and domains with analytic boundary.

Assumption 1.1 *The coefficients in (1.4) satisfy*

1. $A \in C^{\infty}\left(\Omega, \mathbb{R}^{d\times d}\right)$ *is symmetric and uniformly positive definite, i.e., there exists constants* $0 < a_{\min}$, $a_{\max} < \infty$ *such that*

$$a_{\min} \leq \inf_{x\in\Omega}\inf_{v\in\mathbb{C}^d\setminus\{0\}}\frac{\langle A(x)v, \bar{v}\rangle}{\|v\|^2} \leq \sup_{x\in\Omega}\sup_{v\in\mathbb{C}^d\setminus\{0\}}\frac{\langle A(x)v, \bar{v}\rangle}{\|v\|^2} \leq a_{\max},$$

where $\langle\cdot,\cdot\rangle$ *is the Euclidean scalar product in* $\mathbb{R}^d$ *and* $\|\cdot\|$ *denotes the Euclidean norm.*

2. $b \in C^{\infty}\left(\Omega, \mathbb{R}^d\right)$,
3. $c \in C^{\infty}\left(\Omega, \mathbb{R}_{\geq 0}\right)$,
4. $-\frac{1}{2}\operatorname{div} b + c \geq 0$.
5. *There exist constants* $C_{\mathcal{A}}$, C_b, C_c, $\gamma_{\mathcal{A}}$, γ_b, γ_c *such that, for all* $n \in \mathbb{N}_0$,

$$|A|_{n,\infty} \leq C_{\mathcal{A}} n! \gamma_{\mathcal{A}}^n, \quad |b|_{n,\infty} \leq C_b n! \gamma_b^n, \quad |c|_{n,\infty} \leq C_c n! \gamma_c^n.$$

The assumption on the domain are as follows.

Assumption 1.2 Ω *is a bounded Lipschitz domain with analytic boundary, i.e., there is a finite family* $\mathcal{U}$ *of open subset in* $\mathbb{R}^d$ *along a family of bijective maps*[2] $\left\{\chi_U : \overline{B_1} \to \overline{U}\right\}_{U\in\mathcal{U}}$ *such that*

$$\begin{array}{llll} \forall U \in \mathcal{U}: & \chi_U \in C^{0,1}\left(\overline{B_1}, \overline{U}\right), & \chi_U^{-1} \in C^{0,1}\left(\overline{U}, \overline{B_1}\right), & \\ \forall U \in \mathcal{U}: & \chi_U\left(B_1^0\right) = U \cap \partial\Omega, & \chi_U\left(B_1^{+}\right) = U \cap \Omega, & \chi_U\left(B_1^{-}\right) = U \cap \mathbb{R}^d\backslash\overline{\Omega}, \\ \exists C_{\Gamma}, \gamma_{\Gamma} \quad \forall U \in \mathcal{U}: & |\chi_U|_{n,\infty} \leq C_{\Gamma}\gamma_{\Gamma}^n n! \quad \forall n \in \mathbb{N}_0. & & \end{array} \tag{1.5}$$

[1] For vectors $a, b \in \mathbb{C}^d$, we set $\langle a, b\rangle = \sum_{i=1}^d a_i b_i$ (without complex conjugation).

[2] B_1 denotes the unit ball in $\mathbb{R}^d$ and $B_1^0 := \{x \in B_1 \mid x_d = 0\}$. For $\sigma \in \{+,-\}$, we set $B_1^{\sigma} := \{x \in B_1 \mid \sigma x_d > 0\}$.

The standard example for an elliptic problem is given by the Laplace operator.

Example 1.3 *a. Let* $\Omega \subset \mathbb{R}^d$ *be a bounded domain. The bilinear form* $a : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ *corresponding to the weak formulation of the Laplace operator is given by*

$$a(u,v) = \int_\Omega \langle \nabla u, \nabla v \rangle$$

and the eigenvalue problem reads: Find $(\lambda, e) \in \mathbb{C} \times H_0^1(\Omega) \setminus \{0\}$ *such that*

$$a(e,v) = \lambda (e,v)_0 \qquad \forall v \in H_0^1(\Omega).$$

For $d = 1$ *and* $\Omega = (0,1)$*, the eigenpairs are given explicitly by*

$$e_n = c_n \sin(n\pi x), \qquad \lambda_n = (n\pi)^2 \qquad n = 1, 2, \ldots,$$

where the normalization factor $c_n \in \mathbb{R}$ *is chosen such that* $\|e_n\|_{L^2(\Omega)} = 1$*. A simple calculation shows that the isolation distance between the eigenvalues satisfies*

$$3\pi\sqrt{\lambda} \geq \operatorname{dist}(\lambda, \sigma(a) \setminus \{\lambda\}) \geq \pi\sqrt{\lambda} \qquad \forall \lambda \in \sigma(a). \tag{1.6}$$

b. For general $d > 1$*, the isolation distance can be arbitrary small: Consider the Laplace eigenvalue problem with Dirichlet boundary conditions on the rectangle* $(0,1) \times (0, a_\varepsilon)$ *with* $a_\varepsilon = \frac{3}{4}\sqrt{2(1+\varepsilon)}$ *and some* $\varepsilon > 0$*. Then, the following values*

$$\lambda = 9 + \frac{128}{9(1+\varepsilon)} \quad \text{and} \quad \lambda' = 1 + \frac{200}{9(1+\varepsilon)}$$

belong to the spectrum of the Laplacian with homogeneous Dirichlet boundary conditions and satisfy

$$\lambda - \lambda' = 8\frac{\varepsilon}{\varepsilon + 1} \to 0 \text{ as } \varepsilon \to 0.$$

c. The study of the asymptotic distribution of eigenvalues of selfadjoint elliptic operators goes back to H. Weyl [26] and was refined e.g., in [8, Sec. VI, § 4, Satz 17 and 19], [4], [3], [20, Theorem 13.1]. The main result reads

$$\lim_{t\to\infty} \frac{N(t)}{t^{d/2}} = C_d, \tag{1.7}$$

where $N : \mathbb{R} \to \mathbb{R}$ *is a smooth, strictly monotonously increasing function which satisfies* $N(\lambda) := \operatorname{card}\{\lambda' \in \sigma(a) : \lambda' \leq \lambda\}$ *for all* $\lambda \in \sigma(a)$*. and* C_d *is a positive constant which only depends on the space dimension* d*. If we assume in this light – for the selfadjoint case – that* $\lambda_- < \lambda < \lambda_+$ *is a triple of consecutive eigenvalues such that there exists a (slowly varying) function* $g : [\lambda_-, \lambda_+] \to \mathbb{R}$ *with*

$$N(t) = \tilde{C}_d t^{d/2}(1 + g(t)) \quad \text{and} \quad g(\lambda) = 0 \tag{1.8a}$$

and, for all $t \in [\lambda_-, \lambda_+]$*,*

$$c_g^{\mathrm{I}} \leq 1 + g(t) \wedge -\alpha\frac{d}{2} \leq \frac{t g'(t)}{1 + g(t)} \leq C_g^{\mathrm{II}} \wedge \frac{t^2 g''(t)}{1 + g(t)} \leq C_g^{\mathrm{III}}, \tag{1.8b}$$

where c_g^{I}, C_g^{II}, C_g^{III} are positive constants and $0 < \alpha < 1$, then, we can derive an estimate for the spectral gap as follows. To reduce technicalities we assume that λ has multiplicity 1. For all $t \in [\lambda_-, \lambda_+]$, (1.8) implies $N'(t) \geq \tilde{C}_d \frac{d}{2} t^{d/2-1} c_g^{\mathrm{I}} (1-\alpha)$ and for the relative spectral gap, we obtain

$$\begin{aligned}\frac{|\lambda_\pm - \lambda|}{\lambda} &= \frac{|N^{-1}(N(\lambda) \pm 1) - \lambda|}{\lambda} = \frac{\left|\pm (N'(\lambda))^{-1} + \frac{(N^{-1})''(\xi)}{2}\right|}{\lambda} \\ &= \left|\pm \frac{1}{\lambda N'(\lambda)} - \frac{(N'' \circ N^{-1})(\xi)}{2\lambda (N' \circ N^{-1})^3 (\xi)}\right| \\ &\geq \frac{2}{d\tilde{C}_d} \frac{\lambda^{-d/2}}{1 + \frac{2}{d}\lambda g'(\lambda)} \left| 1 - \frac{d\tilde{C}_d \left(1 + \frac{2}{d} C_g^{\mathrm{II}}\right) (N'' \circ N^{-1})(\xi)}{4\lambda^{1-d/2} (N' \circ N^{-1})^3 (\xi)} \right|,\end{aligned}$$

where $\xi \in [\lambda, \lambda_+]$ for the "+" sign and $\xi \in [\lambda_-, \lambda]$ for "-". The last quotient on the right-hand side can be estimated from above by $C\lambda_-^{-d/2}$, where C only depends on the constants in (1.8b). According to the middle inequality in (1.8b), the quantity $\lambda g'(\lambda) = \tilde{C}_g^{\mathrm{II}}$ is bounded from below and above independent of the size of λ. Hence, we have derived under the hypotheses (1.8) and sufficiently large λ_- the estimate

$$\frac{\operatorname{dist}(\lambda, \sigma(a) \setminus \{\lambda\})}{\lambda} = \lambda^{-d/2} \left(\check{C}_d \pm \mathcal{O}\left(\lambda_-^{-d/2}\right)\right). \tag{1.9}$$

d. *If we consider the eigenvalues of the Laplacian with homogenous Dirichlet boundary conditions on $\Omega = (0,\pi)^2$, then, e.g., the values $\lambda_{379} = 509$, $\lambda = \lambda_{380} = 512$, $\lambda_{381} = 514$ are three consecutive eigenvalues. Some tedious calculations yield that $\tilde{C}_d$ and $g(t)$ in the formula (1.8a) for $N(t)$ can be chosen as*

$$\tilde{C}_d = \frac{95}{128} \quad \textit{and} \quad g(t) = -\frac{303407(t-512)}{372817050} + \frac{33361(t-512)^2}{372817050}$$

and the constants c_g^{I}, C_g^{II}, C_g^{III}, and $0 < \alpha < 1$ according to

$$c_g^{\mathrm{I}} = 0.998\ldots, \quad C_g^{\mathrm{II}} = 0, \quad C_g^{\mathrm{III}} = 47.34\ldots, \quad \alpha = 0.685\ldots.$$

2 Galerkin Finite Element Method

The *Galerkin discretisation* of (1.2) is based on the definition of a finite dimensional subspace $S \subset H_0^1(\Omega)$ and given by seeking pairs $(\lambda_S, e_S) \in \mathbb{C} \times S \setminus \{0\}$ such that

$$a(e_S, v) = \lambda_S (e_S, v)_0 \qquad \forall v \in S. \tag{2.1}$$

The space S is chosen as a conforming finite element space $S_{\mathcal{G}}^p \subset H_0^1(\Omega)$ being defined in the usual way via a finite element mesh $\mathcal{G}$ of maximal mesh width h which consists locally of polynomials of degree p. The conformity condition implies that the functions in $S_{\mathcal{G}}^p$ are continuous, i.e., $S_{\mathcal{G}}^p \subset C^0(\Omega)$, and thus $p \geq 1$.

Since domains with (curved) boundary are relevant geometries for our theory, we consider triangulations with possibly curved elements: The triangulation $\mathcal{G}$ consists of elements which are the image of a *reference* simplex (i.e., the unit simplex in $\mathbb{R}^d$). We do not allow hanging nodes and assume – as is standard – that the element maps of elements sharing an edge or a face induce the same parametrisation on that edge or face. The maximal mesh width is denoted by $h := \max_{\tau \in \mathcal{G}} h_\tau$, where $h_\tau := \operatorname{diam} \tau$. Additionally, we make the following assumption on the element maps $F_\tau : \widehat{\tau} \to \tau$.

Definition 2.1 (quasi-uniform regular triangulation) *Each element map F_τ can be written as $F_\tau = R_\tau \circ F_\tau^{\text{affine}}$, where F_τ^{affine} is an affine map and the maps R_τ and F_τ^{affine} satisfy for constants C_{affine}, $C_{\text{metric}} > 0$ independent of h and $\tau \in \mathcal{G}$*

$$
\begin{array}{ll}
\|DF_\tau^{\text{affine}}\|_\infty \leq C_{\text{affine}} h_\tau, & \|(DF_\tau^{\text{affine}})^{-1}\|_\infty \leq C_{\text{affine}} h_\tau^{-1} \\
\|(DR_\tau)^{-1}\|_\infty \leq C_{\text{metric}}, & \|\nabla^n R_\tau\|_\infty \leq C_{\text{metric}} \gamma^n n! \qquad \forall n \in \mathbb{N}_0,
\end{array}
$$

where D denotes the (multidimensional) derivative.

Remark 2.2 *Triangulations satisfying Definition 2.1 can be obtained by patchwise construction of the mesh: Let $\mathcal{G}^{\text{macro}}$ be a* fixed *triangulation (with curved elements) with analytic element maps that resolves the geometry. If the triangulation $\mathcal{G}$ is obtained by quasi-uniform refinements of the reference element $\widehat{\tau}$ and the final mesh is obtained by mapping the subdivisions of the reference element with the macro element maps, then the resulting element maps satisfy the assumptions of Definition 2.1.*

For meshes $\mathcal{G}$ satisfying Def. 2.1 with element maps F_τ we denote the usual space of piecewise (mapped) polynomials by

$$
S_{\mathcal{G}}^p := \{u \in H_0^1(\Omega) \mid \forall \tau \in \mathcal{G} : u|_\tau \circ F_\tau \in \mathbb{P}_p\}. \tag{2.2}
$$

3 Regularity and Approximability of Eigenfunctions

The a priori error analysis for eigenvalue problems requires subtle regularity properties of eigenfunctions and corresponding approximation properties of finite element spaces. Their derivation is the topic of this section.

We choose an increasing numbering of the eigenvalues according to their modulus and their multiplicities

$$
|\lambda_1| \leq |\lambda_2| \leq \ldots
$$

and define, for $1 \leq j \leq N$, the space

$$
U_{1,j} := \operatorname{span}\{e_i : 1 \leq i \leq j\}. \tag{3.1}
$$

As a measure for the approximation quality of the finite element space $S = S_{\mathcal{G}}^p$ (cf. (2.2)) we introduce

$$
\tilde{d}^2(U_{1,j}, S) := \sum_{i=1}^{j} \left(\frac{\|(I - Q_S) e_i\|_1}{\|e_i\|_1} \right)^2, \tag{3.2}
$$

where $Q_S : H_0^1(\Omega) \to S$ is the H^1-orthogonal projection. In order to estimate $\tilde{d}(U_{1,j}, S)$, regularity properties for eigenfunctions of elliptic operators are needed and subtle approximation properties for finite element spaces.

Theorem 3.1 *Let Ω be an analytic, bounded Lipschitz domain which satisfies (1.5). Let the coefficients A, b, c satisfy Assumptions 1.1 and 1.2. Then, any eigenfunction u (normalized to $\|u\|_{L^2(\Omega)} = 1$) is analytic. There exist constants $C, K > 0$ depending only on the constants in Assumption 1.1, (1.5), on $a_{\min}$, and the spatial dimension d such that*

$$\left\|\nabla^{n+2} u\right\|_{L^2(\Omega)} \leq C K^{n+2} \max\left\{n, \sqrt{|\lambda|}\right\}^{n+2} \qquad \forall n \in \mathbb{N}_0, \tag{3.3}$$

where λ is the eigenvalue corresponding to u.

Proof. The statement can be derived from [21, Theorem 5.3.10] as follows. First, let $|\lambda| \geq 1$ and consider (1.2) in the strong form, written as

$$-\varepsilon^2 \nabla \cdot (A \nabla u) + \left\langle \tilde{b}, \nabla u \right\rangle + (\tilde{c} - 1)\, u = f \qquad \text{in } \Omega \quad \text{with} \quad u|_{\partial\Omega} = 0$$

where $\varepsilon^2 = \lambda_j^{-1}$, $\tilde{b} = \lambda_j^{-1} b$, $\tilde{c} = c/\lambda_j$, and $f \equiv 0$. For the quantity $\mathcal{E}$ in [21, Theorem 5.3.10] we obtain the estimate

$$\mathcal{E}^{-1} := C_b + \frac{\sqrt{1 + C_c/|\lambda|}}{|\lambda|^{-1/2}} + 1 \leq 1 + C_b + \sqrt{|\lambda| + C_c} \leq C_1 \sqrt{|\lambda|},$$

where $C_1 := 1 + C_b + \sqrt{1 + C_c}$. The other quantities which appear in [21, Theorem 5.3.10] have to be substituted therein by

$$C_f \leftarrow 0, \quad C_c \leftarrow C_c + 1, \quad \mathcal{E} \leftarrow C_2 |\lambda|^{-1/2}, \quad \left(\frac{\mathcal{E}}{\varepsilon}\right)^2 \leftarrow C_2^2$$

with $C_2 := \left(\sqrt{1 + C_c} + C_b\right)^{-1}$. From Assumption 1.1(4) we conclude that

$$\begin{aligned} \operatorname{Re} a(u, u) &= \int_\Omega \langle A \nabla u, \nabla \overline{u} \rangle + \operatorname{Re}\left(\langle b, \nabla u \rangle \overline{u}\right) + c\,|u|^2 \\ &= \int_\Omega \langle A \nabla u, \nabla \overline{u} \rangle + \frac{1}{2} \left\langle b, \nabla\left(|u|^2\right)\right\rangle + c\,|u|^2 \\ &= \int_\Omega \langle A \nabla u, \nabla \overline{u} \rangle + \left(-\frac{1}{2} \operatorname{div} b + c\right) |u|^2 \\ &\overset{\text{Assumpt. 1.1(4)}}{\geq} a_{\min} \|\nabla u\|^2_{L^2(\Omega)} \end{aligned} \tag{3.4}$$

holds. Since u is an eigenfunction corresponding to λ and $\|u\|_{L^2(\Omega)} = 1$ we obtain

$$\|\nabla u\|_{L^2(\Omega)} \leq a_{\min}^{-1/2} \sqrt{\operatorname{Re} a(u, u)} = \sqrt{(\operatorname{Re} \lambda)/a_{\min}}. \tag{3.5}$$

Plugging these quantities into the estimate in [21, Theorem 5.3.10] we get

$$\left\|\nabla^{n+2} u\right\|_{L^2(\Omega)} \leq C K^{n+2} \max\left\{n, \sqrt{|\lambda|}\right\}^{n+2},$$

where C only depends on the constants $C_{\mathcal{A}}$, C_b, C_c, $\gamma_{\mathcal{A}}$, γ_b, γ_c, C_Γ, γ_Γ, $a_{\min}$. As explained in [21, Remark 5.3.11] the coercivity assumption which is imposed in [21, Theorem 5.3.10] is not required for this estimate. The proof of [21, Theorem 5.3.10] covers only the case $d = 2$.

However, the only part therein, where $d = 2$ (instead of general d) is used explicitly, is the mapping lemma [21, Lemma 4.3.1]. Inspection of the proof shows that the case $d = 3$ can be handled analogously (see [22, Lemma C.1]).[3]

The case $|\lambda| < 1$ is even simpler because we consider directly

$$-\varepsilon^2 \nabla \cdot (A\nabla u) + \langle b, \nabla u \rangle + (c - \lambda)\, u = f \qquad \text{in } \Omega \quad \text{with} \quad u|_{\partial\Omega} = 0$$

where $f = 0$ and $\varepsilon^2 = 1$. By repeating the steps in the first part of the proof with coefficients $\tilde{b} = b$, $\tilde{c} = c - \lambda$ with $|\lambda| < 1$ we obtain

$$\left\| \nabla^{n+2} u \right\|_{L^2(\Omega)} \leq C\,(nK)^{n+2}.$$

■

Theorem 3.2 *Let the assumptions of Theorem 3.1 be satisfied. Let u be an eigenfunction of (1.2) corresponding to an eigenvalue λ (normalized to $\|u\|_{L^2(\Omega)} = 1$). Then, there exist positive constants C_2, C_3, σ, and sufficiently small $C_1 > 0$ independent of h, λ, p such that for all h, p which satisfy*

$$kh/p \leq C_1 \qquad \text{with } k := \sqrt{|\lambda|}$$

there exists a finite element function $u_S \in S$ such that

$$\|u - u_S\|_{1,k} \leq C_3 \left[\left(\frac{C_2 h}{h + \sigma} \right)^p + k \left(\frac{kh}{\sigma p} \right)^p \right],$$

where

$$\|v\|_{1,k} := \sqrt{\|\nabla v\|^2_{L^2(\Omega)} + k^2 \|v\|^2_{L^2(\Omega)}}.$$

Proof. In [22, Proof of Theorem 5.4, formula (5.9)] the approximation of some analytic functions is investigated. However, the conditions on the functions there differ slightly from (3.3) and, hence, we have to adapt the analysis accordingly. In view of (3.3) we will consider functions which satisfy

$$\|\nabla^n v\|_{L^2(\Omega)} \leq C K^n \max\{n, k\}^n \qquad \forall n \in \mathbb{N}_0 \tag{3.6}$$

for some constants $C, K, k > 0$. Let

$$C_\tau^2 := \sum_{n \in \mathbb{N}_0} \frac{\|\nabla^n v\|^2_{L^2(\tau)}}{(2K \max\{k, n\})^{2n}}. \tag{3.7}$$

The combination of (3.6) and (3.7) yields

$$\sum_{\tau \in \mathcal{G}} C_\tau^2 \leq \frac{4}{3} C^2 \quad \text{and (trivially) it holds} \quad \|\nabla^n v\|_{L^2(\tau)} \leq C_\tau \left(2K \max\{k, n\}\right)^n. \tag{3.8}$$

We employ [22, Lemma C.1] for functions which satisfy the second estimate in (3.8). We conclude that the pullback $\tilde{v}_\tau := v|_\tau \circ R_\tau$ on $\tilde{\tau} := R_\tau^{-1}(\tau)$ satisfies (for suitable constants

[3] Methods for discretising eigenvalue problems in cases where the spatial dimension d is large are presented in [14].

$\tilde{C}$, $\hat{C}$ which depend additionally on the constants in Assumption 1.2 and Definition 2.1) the estimate

$$\left\|\nabla^n \tilde{v}_\tau\right\|_{L^2(\tilde{\tau})} \leq \hat{C} C_\tau \left(\tilde{C} \max\{k,n\}\right)^n.$$

Since F_τ^{affine} is affine, the function $\hat{v}_\tau := v|_\tau \circ F_\tau = \tilde{v}_\tau \circ F_\tau^{\text{affine}}$ satisfies

$$\left\|\nabla^n \hat{v}_\tau\right\|_{L^2(\hat{\tau})} \leq C C_\tau h^{-d/2} \check{C}^n h^n \max\{n,k\}^n \qquad \forall n \in \mathbb{N}_0.$$

Hence, the assumptions of [22, Lemma C.3] are satisfied and we get from [22, (5.9)] and the first estimate in (3.8) the existence of a finite element function $v_S \in S$ such that

$$\|v - v_S\|_{1,k}^2 \leq \frac{4}{3} C^2 \left[\left(1 + \left(\frac{kh}{h+\sigma}\right)^2\right)\left(\frac{h}{h+\sigma}\right)^{2p} + k^2 \left(\frac{kh}{\sigma p}\right)^{2p} \left(\frac{1}{p^2} + \left(\frac{kh}{\sigma p}\right)^2\right)\right].$$

By choosing h, p such that

$$kh/p \lesssim 1$$

we obtain

$$\|v - v_S\|_{1,k} \leq C_3 \left[\left(\frac{C_2 h}{h+\sigma}\right)^p + k\left(\frac{kh}{\sigma p}\right)^p\right].$$

■

4 Convergence Analysis for the Selfadjoint Case. Error Estimate for the Eigenvalues.

The a priori analysis for elliptic eigenvalue problems is classical (see, e.g., [25], [6], [7], [1], [13]) and convergence rates for the finite element method are proved provided the mesh width h is fine enough. In this section we will consider the selfadjoint case, i.e., the sesquilinear form a_1 in (1.3) is zero so that $a = a_0$. For this case, sharp error estimates for Ritz values and Ritz vectors are proved in [16], [11], [18], [23]. We briefly recall and combine them with the regularity and approximation properties derived in the previous section.

The eigenvectors are denoted by e_i and the normalization is chosen so that $(e_n, e_m)_0 = \delta_{n,m}$. Note that this implies

$$a(e_n, e_m) = \lambda_n (e_n, e_m)_0 = \lambda_n \delta_{n,m}. \tag{4.1}$$

The finite element discretisation (2.1) has eigenvalues

$$\lambda_{S,1} \leq \lambda_{S,2} \leq \ldots \leq \lambda_{S,N},$$

where $N := \dim S$ and the corresponding eigenvectors are denoted by $e_{S,n}$, $1 \leq n \leq N$.

Theorem 4.1 *Let Assumption 1.1 be satisfied with $b = 0$. Let $U_{1,j}$ and $\tilde{d}^2$ be defined by (3.1) and (3.2). Then*

$$0 \leq \frac{\lambda_j^S - \lambda_j}{\lambda_j^S} \leq \tilde{d}^2(U_{1,j}, S). \tag{4.2}$$

If $\tilde{d}^2(U_{1,j}, S) < 1$ for some $1 \leq j \leq \dim S$ then

$$0 \leq \frac{\lambda_j^S - \lambda_j}{\lambda_j} \leq \frac{\tilde{d}^2(U_{1,j}, S)}{1 - \tilde{d}^2(U_{1,j}, S)}. \tag{4.3}$$

The proof is a combination of [11, Chap. 9, § 2, Theorem 1] with [18, Corollary 2.2] (see also [25], [16]).

Corollary 4.2 *Let the Assumptions of Theorems 4.1, 3.1, and 3.2 be satisfied.*

1. *Then,*
$$\tilde{d}^2(U_{1,j}, S) \leq C \sum_{i=1}^{j} \left(\left(\frac{C_2 h}{h+\sigma} \right)^{2p} \lambda_i^{-1} + \left(\frac{\sqrt{\lambda_i} h}{\sigma p} \right)^{2p} \right). \tag{4.4}$$

2. *If we assume (in view of (1.7)) that there is a constant C_d independent of j such that*
$$j \leq C_d \lambda_j^{d/2} \tag{4.5}$$
then,
$$\tilde{d}^2(U_{1,j}, S) \leq C_4 \lambda_j^{d/2} \left(\left(\frac{C_2 h}{h+\sigma} \right)^{2p} + \left(\frac{\sqrt{\lambda_j} h}{\sigma p} \right)^{2p} \right) \quad \text{with} \quad C_4 := \frac{C C_d}{\min\{\lambda_1, 1\}}. \tag{4.6}$$

3. *Let (4.5) be satisfied. By choosing the discretisation parameters h and p according to*
$$\frac{\sqrt{\lambda_j} h}{\sigma p} \leq \mathrm{e}^{-d/4}, \quad \frac{C_2 h}{h+\sigma} \leq \mathrm{e}^{-d/4}, \quad p \geq \log\left((4C_4)^{\frac{2}{d}} \lambda_j \right),$$
then, for sufficiently large λ_j, the upper bound in (4.6) is bounded by $1/2$ and the error estimate
$$0 \leq \frac{\lambda_j^S - \lambda_j}{\lambda_j} \leq 2 \frac{C C_d}{\min\{\lambda_1, 1\}} \lambda_j^{d/2} \left(\left(\frac{C_2 h}{h+\sigma} \right)^{2p} + \left(\frac{\sqrt{\lambda_j} h}{\sigma p} \right)^{2p} \right)$$
holds.

4. *Let (4.5) be satisfied. For $p = 1$, the condition on h such that $\tilde{d}^2(U_{1,j}, S) \leq 1/2$ is given by*
$$\lambda_j^{d/4} \left(\sqrt{\lambda_j} h \right) \leq \frac{1}{\sqrt{2} C_5} \quad \text{with} \quad C_5^2 := \frac{C C_d}{\min\{\lambda_1, 1\} \sigma^2} \left(1 + \frac{C_2^2}{\lambda_1} \right) \tag{4.7}$$

Proof. ad 1) Note that Assumption 1.1 implies that $0 < \lambda_1 \leq \lambda_j$ for all $j \in \mathbb{N}$. Hence,
$$c_0 \|v\|_1 \leq \|v\|_{1,k} \qquad \forall v \in H^1(\Omega) \qquad \text{with } c_0 := \min\{1, \lambda_1\}.$$

In view of (4.3) we have to estimate the quantity $\tilde{d}^2(U_{1,j}, S)$ (cf. (3.2)). Let $(e_i)_{i=1}^j$ denote the eigenvectors as in (3.1) which are orthonormal in $L^2(\Omega)$. Hence, by using the previous Theorem, we get (with $k_i := \sqrt{|\lambda_i|}$ and $\|e_i\|_1 \geq c\sqrt{\lambda_i}$, where $c := (\max\{a_{\max}, C_c\})^{-1/2}$)
$$\frac{\|(I - Q_S) e_i\|_1}{\|e_i\|_1} \leq \frac{C_3}{c_0 c} \frac{\left(\frac{C_2 h}{h+\sigma}\right)^p + k_i \left(\frac{k_i h}{\sigma p}\right)^p}{k_i} \leq \frac{C_3}{c_0 c} \left\{ k_i^{-1} \left(\frac{C_2 h}{h+\sigma} \right)^p + \left(\frac{k_i h}{\sigma p} \right)^p \right\}.$$

The quantity $\tilde{d}^2(U_{1,j}, S)$ can therefore be estimated by the right-hand side in (4.4).

ad 2) This is a trivial consequence of part 1 and (4.5).

ad 3) The assumptions on h and p imply

$$C_4\lambda_j^{d/2}\left(C_3\frac{\sqrt{\lambda_j}h}{p}\right)^{2p} \leq C_1\lambda_j^{d/2}\,\mathrm{e}^{-\frac{d}{2}p} \leq C_4\lambda_j^{d/2}\,\mathrm{e}^{-\frac{d}{2}\log\left((4C_4)^{\frac{2}{d}}\lambda_j\right)} = 1/4.$$

The estimate of the first term in (4.6) is just a repetition of the previous arguments.

ad 4) For $p = 1$, we get from (4.6)

$$\tilde{d}^2(U_{1,j}, S) \leq C_5^2\lambda_j^{1+d/2}h^2.$$

and (4.7) follows. ■

5 Convergence Analysis for the Selfadjoint Case. Error Estimate for the Eigenfunctions

In this section, the error of the eigenfunction approximation will be estimated. We assume throughout this section that Assumption 1.1 (with $b = 0$) and Assumption 1.2 are satisfied so that – as a consequence of the Riesz-Schauder theory – the compact solution operator $T : H_0^1(\Omega) \rightarrow H_0^1(\Omega)$ is well defined via

$$a(Tu, v) = (u, v)_0 \qquad \forall v \in H_0^1(\Omega). \tag{5.1}$$

Note that the eigenfunctions of T and (1.2) are the same and the eigenvalues μ of T and λ of (1.2) are reciprocal to each other.

Assumption 1.1 implies the Riesz-Schauder theory: The spectrum $\sigma(a)$ of (1.2) is countable with infinity as the only possible accumulation point. All elements $\lambda \in \sigma(a)$ are eigenvalues. The dimensions of the corresponding eigenspaces

$$E(\lambda, a) := \operatorname{span}\{u : (\lambda, u) \text{ is an eigenpair of } (1.2)\} \tag{5.2}$$

are finite.

For the error analysis, we consider the continuous problem in the form $Tu = \mu u$ The discrete version is formulated in an analogous way by introducing the operator $T_S : S \rightarrow S$ by

$$a(T_S u, v) = (u, v)_0 \qquad \forall u, v \in S$$

and by considering the eigenvalue problem $T_S u^S = \mu^S u^S$. Note that Assumption 1.1 (with $b = 0$) implies that all continuous and discrete eigenvalues are positive. The eigenspace corresponding to a continuous eigenvalue μ is $E(\mu, T) \subset H_0^1(\Omega)$ and $E_S(\mu^S, T_S) \subset S$ is the eigenspace corresponding to a discrete eigenvalue μ^S.

For simplicity, the following convergence theorem covers only the case that all eigenvalues of T have multiplicity 1, i.e.,

$$\mu_1 > \mu_2 > \ldots > 0. \tag{5.3}$$

Theorem 5.1 (Saad) *Let (5.3) and Assumption 1.1 (with $b = 0$) be satisfied. Let (μ_j, u_j), $1 \leq j \leq \dim S$ be the j-th eigenpair of $Tu_j = \mu_j u_j$ with normalization $\|u_j\|_1 = 1$. Let $d_{j,S} := \min\left\{\left|\mu_j - \mu^S\right| : \mu^S \in \sigma(T_S) \setminus \left\{\mu_j^S\right\}\right\}$. Then, there exists some $u_j^S \in E_S\left(\mu_j^S, T_S\right)$ such that*

$$\left\|u_j - u_j^S\right\|_1 \leq \left(1 + \frac{\|(I - P_S)\, T P_S\|_{1\leftarrow 1}^2}{d_{j,S}^2}\right)^{1/2} \inf_{v \in S} \|u_j - v\|_1, \tag{5.4}$$

where P_S denotes the $a(\cdot,\cdot)$-orthogonal projection onto S.

For a proof, we refer to [24, Theorem 3]. The restriction to simple eigenvalues for the eigenvector error estimates is quite strong. The error estimates have been generalized in [17] and [23] to the case of multiple and also clustered eigenvalues.

Estimate (5.4) only makes sense if $d_{j,S} > 0$. This condition can be replaced by a stronger condition which employs the error estimate for the eigenvalue approximation. From the max/min principle we conclude that $\mu_j^S \leq \mu_j$.

Corollary 5.2 *Let the Assumptions of Theorem 5.1 be satisfied and let $1 \leq j \leq \dim S$. If $j > 1$, let the finite-dimensional space S be chosen such that*

$$\tilde{d}^2\left(U_{1,j-1}, S\right) \leq \frac{1}{2} \frac{\lambda_j - \lambda_{j-1}}{\lambda_j} \tag{5.5}$$

1. *Then,*

$$\left\|u_j - u_j^S\right\|_1 \leq \left(1 + \frac{\|(I - P_S)\, T P_S\|_{1\leftarrow 1}^2}{\delta_j^2}\right)^{1/2} \inf_{v \in S} \|u_j - v\|_1, \tag{5.6}$$

where[4]

$$\delta_j := \min_{i \in \{j_+, j+1\}} \frac{\lambda_i - \lambda_{i-1}}{2\lambda_i \lambda_{i-1}} \quad \text{with} \quad j_+ := \max\{j, 2\}.$$

2. *Let Assumptions 1.1 (with $b = 0$) and Assumption 1.2 be satisfied. Further let the assumptions of Theorem 3.2 be valid. Then,*

$$\frac{\left\|u_j - u_j^S\right\|_1}{\|u_j\|_1} \leq C_3 \left(1 + C \frac{h^{\min\{p,2\}}}{\delta_j}\right) \left[\lambda_j^{-1/2} \left(\frac{C_2 h}{h + \sigma}\right)^p + \left(\frac{\sqrt{\lambda_j} h}{\sigma p}\right)^p\right].$$

3. *Assume in addition (cf. (1.9)) that*

$$\delta_j \geq c_6 \lambda_j^{-1-d/2}. \tag{5.7}$$

Then,

$$\frac{\left\|u_j - u_j^S\right\|_1}{\|u_j\|_1} \leq C_3 \left(1 + C \lambda_j^{1+d/2} h^{\min\{p,2\}}\right) \left[\lambda_j^{-1/2} \left(\frac{C_2 h}{h + \sigma}\right)^p + \left(\frac{\sqrt{\lambda_j} h}{\sigma p}\right)^p\right]. \tag{5.8}$$

[4]Note that δ_j is independent of the discretization.

Proof. We give the proof only in the case $j > 1$, because the case $j = 1$ is a simplified version thereof.

ad 1) Part 1: Condition (5.5) implies $\mu_{j-1}^S > \mu_j$.

This follows from (5.5) via

$$\mu_{j-1}^S - \mu_j = \mu_{j-1} - \mu_j - \left(\mu_{j-1} - \mu_{j-1}^S\right) = \mu_{j-1} - \mu_j - \lambda_{j-1}^{-1}\left(\frac{\lambda_{j-1}^S - \lambda_{j-1}}{\lambda_{j-1}^S}\right)$$

$$\overset{(4.2)}{\geq} \lambda_{j-1}^{-1}\left(\frac{\lambda_j - \lambda_{j-1}}{\lambda_j} - \tilde{d}^2\left(U_{1,j-1}, S\right)\right) \geq \frac{\lambda_j - \lambda_{j-1}}{2\lambda_j\lambda_{j-1}} > 0.$$

Part 2) Proof of (5.6).

The min/max principle implies $\mu_j - \mu_{j+1}^S \geq \mu_j - \mu_{j+1}$ and, hence, part 1 yields

$$d_{j,S} \geq \min\left\{\mu_{j-1}^S - \mu_j, \mu_j - \mu_{j+1}\right\} \geq \min\left\{\frac{\lambda_j - \lambda_{j-1}}{2\lambda_j\lambda_{j-1}}, \frac{\lambda_{j+1} - \lambda_j}{\lambda_j\lambda_{j+1}}\right\}.$$

ad 2) Standard approximation properties for finite element spaces imply

$$\|(I - P_S)\, T P_S u\|_{1\leftarrow 1} \leq C h^{\min\{p,2\}}$$

from which the estimate of the first factor of the right-hand side in (5.6) follows. The combination with Theorem 3.2 yields the assertion.

ad 3) The third part follows from part 2 and (5.7). ∎

6 Conclusions

In the following we will discuss different choices of λ, h, p and their implication. For simplicity we assume that $N_S := \dim S$ satisfies $N_S = O\left((p/h)^d\right)$.

1) Possible choices of h_0 and p so that the right-hand side in (5.8) is $\lesssim 1$ for $h \leq h_0$.

(h_0, p)	$\left(\lambda_j^{-\frac{3+d}{4}}, 1\right)$	$\left(\lambda_j^{-\frac{4+d}{8}}, 2\right)$	$\left(\lambda_j^{-\frac{5+d}{10}}, 3\right)$	$\left(\alpha\lambda_j^{-1/2}\log\lambda_j, \log\lambda_j\right)$
N_S	$O\left(\lambda_j^{\frac{3+d}{4}d}\right)$	$O\left(\lambda_j^{\frac{4+d}{8}d}\right)$	$O\left(\lambda_j^{\frac{5+d}{10}d}\right)$	$O\left(C_d\lambda_j^{d/2}\right)$

where $\alpha > 0$ depends only on σ and d. We see that a minimal finite element space S which has the property that the relative eigenfunction error is (starting to be) below 100% is characterized by the choices $p \sim \log\lambda_j$ and $h \sim \alpha\lambda_j^{-1/2}\log\lambda_j$. Afterwards, any reasonable strategy for enriching the finite element space (h version, p version, adaptive hp version – depending on the elliptic regularity) will exhibit its textbook convergence rate.

2) In the analytic case (Assumptions 1.1, 1.2) it is most preferable to employ a p-version of the finite element method (for sufficiently small mesh width h) because, then, the error is converging exponentially.

Remark 6.1 *Nowadays, a posteriori error estimates for finite element discretisations are popular (see, e.g., [10], [19], [15], [12]), [9], [5]). Also here, the question of the minimal dimension of a finite element space plays a role (and is still open) especially when computing higher eigenvalues and corresponding eigenvectors.*

Systematic numerical experiments have been performed and published in [2]. They clearly show that the dependence of the minimal dimension of the finite element space S on λ, δ, and h (such that the relative eigenfunction error is smaller than 100%) is visible also in practical computations.

Acknowledgements. The author is thankful to an anonymous referree whose suggestions lead to an improvement of the paper.

References

[1] I. Babuška and J. Osborn. Eigenvalue Problems. In P. Ciarlet and J. Lions, editors, *Handbook of Numerical Analysis, Vol. II: Finite Element Methods (Part 1)*, pages 641–788. Elsevier Science Publishers, Amsterdam, 1991.

[2] L. Banjai, S. Börm, and S. Sauter. FEM for Elliptic Eigenvalue Problems: How Coarse Can the Coarsest Mesh be Chosen? An Experimental Study. *Comput. Vis. Sci*, pages 363–372, 2008.

[3] F. Brownell. Extended asymptotic eigenvalue distributions for bounded domains in n-space. *Pacific Journ. of Math*, 5:483–499, 1955.

[4] F. Brownell. An extension of Weyl's asymptotic law for eigenvalues. *Pacific Journ. of Math*, 5:483–499, 1955.

[5] C. Carstensen and J. Gedicke. An Oscillation-Free Adaptive FEM For Symmetric Eigenvalue Problems. Technical report, Preprint Series of the DFG Research Center MATHEON, Berlin, 2008. Preprint no. 489.

[6] F. Chatelin. La méthode de Galerkin. Ordre de convergence des éléments propres. *C.R. Acad. Sci. Pairs Sér. A*, 278:1213–1215, 1974.

[7] F. Chatelin. *Spectral Approximation of Linear Operators.* Academic Press, New York, 1983.

[8] R. Courant and D. Hilbert. *Methoden der Mathematischen Physik. Bd. 1.* Springer, Berlin, 1924.

[9] 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.*, 25:135–160, 2006.

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

[11] E. Dyakonov. *Optimization in solving elliptic problems.* CRC Press, Boca Raton, 1996.

[12] S. Giani and I. Graham. A convergent adaptive method for elliptic eigenvalue problems. Technical Report 06/07/2007, Univerity of Bath, Bath Institute for Complex Systems., 2007.

[13] W. Hackbusch. *Elliptic Differential Equations.* Springer Verlag, 1992.

[14] W. Hackbusch, B. Khoromskij, S. Sauter, and E. Tyrtyshnikov. Use of Tensor Formats in Elliptic Eigenvalue Problems. Technical Report 13-2008, Universität Zürich, 2008.

[15] V. Heuveline and R. Rannacher. A posteriori error control for finite element approximations of elliptic eigenvalue problems. *J. Comput. Math. Appl.*, 15:107–138, 2001.

[16] A. Knyazev. Sharp a priori error estimates of the Raleigh-Ritz method without assumptions of fixed sign or compactness. *Mathematical Notes*, 38(5-6):998–1002, 1986.

[17] A. Knyazev. New Estimates for Ritz Vectors. *Math. Comp.*, 66:985–995, 1997.

[18] A. Knyazev and J. Osborn. New a priori FEM error estimates for eigenvalues. *SIAM J. Numer. Anal.*, 48(6):2647–2667, 2006.

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

[20] S. Levendorskii. *Asymptotic Distribution of Eigenvalues of Differential Operators.* Springer, Berlin, 1990.

[21] J. Melenk. *hp-Finite Element Methods for Singular Perturbations.* Springer, 2002.

[22] M. Melenk and S. Sauter. Convergence Analysis for Finite Element Discretizations of the Helmholtz Equation. Part I: The Full Space Problem. Technical Report 09-2008, Universität Zürich, 2008.

[23] E. Ovtchinnikov. Cluster robust error estimates for the Raleigh-Ritz approximation I: Estimates for invariant subspaces. *LAA*, 415(1):167–187, 2006.

[24] Y. Saad. On the Rates of Convergence of the Lanczos and the Block-Lanczos Methods. *SIAM J. Numer. Anal.*, 17:687–706, 1980.

[25] G. Strang and G. Fix. *An Analysis of the Finite Element Method.* Prentice-Hall, 1973.

[26] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.*, 71:441–479, 1912.