A posteriori error control for finite element approximations of elliptic eigenvalue problems

Vincent Heuveline and Rolf Rannacher

Institute of Applied Mathematics
University of Heidelberg, INF 294
D-69120 Heidelberg, Germany

Revision : February 2, 2001

Abstract

We develop a new approach to *a posteriori* error estimation for Galerkin finite element approximations of symmetric and non-symmetric elliptic eigenvalue problems. The idea is to embed the eigenvalue approximation into the general framework of Galerkin methods for nonlinear variational equations. In this context residual-based *a posteriori* error representations are available with explicitly given remainder terms. The careful evaluation of these error representations for the concrete situation of an eigenvalue problem results in *a posteriori* error estimates for the approximations of eigenvalues as well as eigenfunctions. These suggest local error indicators that are used in the mesh refinement process.

1 Introduction

We consider the Galerkin finite element approximations of symmetric or non-symmetric elliptic eigenvalue problems. As a prototypical problem, we consider the convection-diffusion problem

$$\mathcal{A}v \;:=\; -\nabla\cdot\{a\nabla v) + b\cdot\nabla v + cv = \lambda v \quad \text{in } \ \Omega, \quad u = 0 \ \text{ on } \partial\Omega. \tag{1}$$

For $b \not\equiv 0$, the governing operator is non-symmetric and may possess non-real eigenvalues. In this case complex analysis is required. The Galerkin finite element approximation of (1) is based on its variational formulation. We seek non-trivial pairs $\{v_h, \lambda_h\} \in H_h \times \mathbb{C}$ satisfying

$$a(v_h, \varphi_h) = \lambda_h(v_h, \varphi_h) \quad \forall \varphi_h \in H_h, \tag{2}$$

where $a(\cdot,\cdot)$ is the sesqui-linear form associated to the differential operator $\mathcal{A}$, and $(\cdot,\cdot)$ is the usual L^2 scalar product. The finite element subspace $H_h \subset H := H_0^1(\Omega)$ consists of piecewise linear or bilinear functions on decompositions $\mathbb{T}_h$ of the domain $\bar{\Omega}$ into cells T (triangles, quadrilaterals, etc.) with diameter $h_T := \text{diam}\{T\}$.

For solutions $\{v_h, \lambda_h\}$ of the discrete eigenvalue problems (2), we define the usual 'cell-residuals' by

$$\rho_T(v_h, \lambda_h) := \|\mathcal{A}v_h - \lambda_h v_h\|_T + \tfrac{1}{2}\|a[\partial_n v_h]\|_{\partial T \setminus \partial\Omega},$$

where $[\partial_n v_h]$ indicates the jump of $\partial_n v_h$ across inter-element boundaries, and $\|\cdot\|_T$ and $\|\cdot\|_{\partial T}$ are the L^2-norms over T and ∂T, respectively. We want to derive *a posteriori* error estimates for the eigenvalues and eigenfunctions in terms of the cell residuals $\rho_T(v_h, \lambda_h)$. For the symmetric eigenvalue problem with $b = 0$, *a posteriori* error estimates of this type have been derived by Larson [14],

$$|\lambda - \lambda_h| \le c\Big(\sum_{T\in\mathbb{T}_h} h_T^4\, \rho_T(v_h, \lambda_h)^2\Big)^{1/2}, \qquad (3)$$

and by Nystedt [15],

$$|\lambda - \lambda_h| \le c\sum_{T\in\mathbb{T}_h} h_T^2\, \rho_T(v_h, \lambda_h)^2. \qquad (4)$$

Both estimates are of optimal order for the present situation. Analogues results have also been given for the H^1- as well the L^2-errors of the eigenfunctions. In the proofs, the H^2-regularity of the eigenfunctions is used which excludes domains with reentrant corners or discontinuous coefficients $a(x)$. Further, *a priori* conditions are required in the sense that the current mesh is fine enough to ensure that the approximate eigenvalue λ_h is closer to λ than any other eigenvalue of (2). The result (3) is obtained by direct analysis largely exploiting the symmetry of $\mathcal{A}$. For deriving (4) the heavy machinery of the resolvent integral calculus from the *a priori* error analysis of Bramble & Osborn [7] (see also Osborn [16], Babuska & Osborn [2]) is used. A result similar to (3), though of only suboptimal order, has been derived by Verfürth [17] by considering the eigenvalue problem as a parameter-dependent nonlinear equation and using general results for the Galerkin approximation of such type of problems. This approach is close to the one used in this paper for treating the general non-symmetric case. Earlier references dealing also with a posteriori error estimates for symmetric eigenvalue problems are Babuska & Osborn [1] and Babuska & Tsuchiya [3].

The central idea of the following analysis is to associate a 'dual' eigenvalue problem $\mathcal{A}^*v^* = \lambda^*v^*$ to the general non-symmetric 'primal' eigenvalue problem (2), together with its discrete counterpart

$$a^*(v_h^*, \psi_h) = \lambda^*(v_h^*, \psi_h) \quad \forall \psi_h \in H_h. \qquad (5)$$

Then, the two problems (2) and (5) are embedded into the general optimal-control framework of Galerkin approximations of nonlinear variational equations developed in Becker & Rannacher [6] (for earlier references see also Johnson [12], Eriksson, Estep, Hansbo

& Johnson [10] and Becker & Rannacher [4]). By simultaneous consideration of the approximation of primal and dual eigenvalue problems, we shall derive *a posteriori* error estimates of the form

$$|\lambda-\lambda_h| \le c \sum_{T\in\mathbb{T}_h} h_T^2 \left\{\rho_T(v_h,\lambda_h)^2 + \rho_T^*(v_h^*\,\lambda_h^*)^2\right\}, \tag{6}$$

provided that λ_h is already sufficiently close to λ and that λ_h and λ are simple. Here, ρ_T and ρ_T^* are the cell residuals of the primal and dual eigenvalue problem, respectively. This result is the natural extension of the error estimate (4) to the non-symmetric case and is derived without requiring the H^2-regularity of the problem. By applying the same general theory also the expected H^1 and L^2 error estimates for the corresponding eigenfunctions are derived. Further, similar estimates are obtained for the error with respect to more general out-put quantities $j(v)$ like point values $j(v) = v(P)$, line integrals $j(v) = \int_\Gamma v\,ds$, or weighted averages $j(v) = \int_\Omega v\omega\,dx$:

$$|j(v-v_h)| \le c \sum_{T\in\mathbb{T}_h} h_T^2\,\rho_T(v_h,\lambda_h)\,\|\nabla^2 z\|_T + R_h, \tag{7}$$

where z is the solution of an auxiliary 'dual' boundary value problem depending on the particular choice of the functional $j(\cdot)$. The remainder term R_h is quadratic in $|\lambda–\lambda|$ and $\|v-v_h\|$. The local quantities $h_T^2\|\nabla^2 z\|_T$ can be viewed as sensitivity factors measuring the effect of the cell-residuals ρ_T on the error $j(v–v_h)$, and are evaluated computationally (see Becker & Rannacher [5] or Becker & Rannacher [6] for the details of the resulting feed-back procedure for mesh adaptation).

We note that the approach of simultaneously considering the primal and dual eigenvalue problems is also crucial for deriving efficient multigrid solvers for the discrete problems. This has been successfully exploited for non-selfadjoint elliptic eigenvalue problems by Heuveline & Bertsch [11].

The further contents of this paper are as follows. In Section 2, we set out the general framework of elliptic eigenvalue problems and their finite element approximation and recall the main results of the *a priori* error analysis. Section 3 introduces an abstract approach to residual-based *a posteriori* error estimation for the Galerkin approximation of variational equations. In Section 4 the eigenvalue problem is embedded into the framework of Section 3 and *a posteriori* error estimates are derived for the eigenvalues and the eigenfunctions in the energy norm as well as with respect to general out-put functionals. In Section 5 the theoretical results are illustrated by some numerical tests.

2 Preliminaries

2.1 Formulation of the eigenvalue problem

Let $\mathcal{A}$ be a uniformly elliptic operator of order $m = 2$ defined on a bounded domain $\Omega \subset \mathbb{R}^d$. The classical formulation of the eigenvalue problem for this operator is

$$\mathcal{A}v = \lambda v \quad \text{in } \Omega, \quad v = 0 \text{ on } \partial\Omega. \tag{8}$$

One can replace (8) by the variational formulation seeking pairs $\{v, \lambda\} \in H \times \mathbb{C}$, satisfying

$$a(v, \psi) = \lambda\,(v, \psi) \quad \forall \psi \in H, \quad \|v\| = 1, \tag{9}$$

where $a(\cdot,\cdot)$ is the sesqui-linear form generated by the operator $\mathcal{A}$, $(\cdot,\cdot)$ is the (complex) L^2-scalar product on Ω with corresponding norm $\|\cdot\|$, and $H := H_0^1(\Omega)$ is the usual first-order (complex) Sobolev space. The sesqui-linear form $a(\cdot,\cdot)$ is assumed to be bounded and H-elliptic, that is

$$|a(\varphi, \psi)| \le \alpha \|\varphi\|_H \, \|\psi\|_H, \quad \gamma \|\varphi\|_H^2 \le |a(\varphi, \varphi)| + \beta \|\varphi\|^2, \quad \varphi,\, \psi \in H, \tag{10}$$

with certain constants $\alpha, \gamma > 0$, and $\beta \ge 0$. For simplicity, we assume in the following that (10) holds with $\beta = 0$, such that $0 \notin \Sigma(\mathcal{A})$. The "adjoint" (or "dual") eigenvalue problem associated to (9) seeks pairs $\{v^*, \lambda^*\} \in H \times \mathbb{C}$, satisfying

$$a(\psi, v^*) = \bar{\lambda}^*\,(\psi, v^*) \quad \forall \psi \in H, \quad (v, v^*) = 1. \tag{11}$$

Since the embedding $H \hookrightarrow L^2(\Omega)$ is compact, the classical Riesz-Schauder theory applies (see, e.g., Kato [13]). Hence, the primal as well as the dual eigenvalue problem possess countable infinite sets $\Sigma(\mathcal{A}) := \{\lambda_i\}_{i=1}^{\infty} \subset \mathbb{C}$ and $\Sigma(\mathcal{A}^*) := \{\lambda_i^*\}_{i=1}^{\infty} \subset \mathbb{C}$ of isolated eigenvalues with finite (algebraic) multiplicities which have no finite accumulation point. Further, one easily sees that $\lambda_i^* = \bar{\lambda}_i$. The 'algebraic' and 'geometric' multiplicities of an eigenvalue λ are denoted by σ_λ and ρ_λ, respectively. Further, its 'ascent' α_λ is the smallest integer such that $Ker\{(\mathcal{A} - \lambda I)^{\alpha+1}\} = Ker\{(\mathcal{A} - \lambda I)^{\alpha}\}$. The case $\alpha_\lambda = 1$ is related to $\sigma_\lambda = \rho_\lambda$.

For an eigenvalue λ with eigenspace $E(\lambda) := Ker\{\mathcal{A} - \lambda I\}$, the form $a_\lambda(\cdot,\cdot) := a(\cdot,\cdot) - \lambda(\cdot,\cdot)$ is regular on the quotient-space $H/E(\lambda)$, that is

$$\|\varphi\|_H \le c_{s,\lambda}^{(1)} \sup_{\psi \in H/E(\lambda)} \Big\{ \frac{a_\lambda(\varphi, \psi)}{\|\psi\|_H} \Big\}, \quad \varphi \in H/E(\lambda), \tag{12}$$

with the 'weak stability constant' $c_{s,\lambda}^{(1)}$ of λ, depending on $\operatorname{dist}(\lambda, \Sigma(\mathcal{A})\backslash\lambda)$. Further, the operator $\mathcal{A}$ is called 'H^2-regular', if any function $w \in H$ satisfying $\mathcal{A}w \in L^2(\Omega)$ is also in $H^2(\Omega)$ and admits the *a priori* bound

$$\|w\|_{H^2} \le c_s^{(2)} \|\mathcal{A}w\|, \tag{13}$$

with the 'strong stability constant' $c_s^{(2)}$. This particularly implies that all eigenfunctions have H^2-regularity. In this case the unique solution $w \in H/E(\lambda)$ of the problem

$$a_\lambda(\psi, w) = l(\psi) \quad \forall \psi \in H, \tag{14}$$

with $l(\cdot) \in E(\lambda)^\perp$ is in $H^2(\Omega)$ and satisfies

$$\|w\|_{H^2} \le c_s^{(2)} \sup_{\psi \in L^2(\Omega)} \Big\{ \frac{|l(\psi)|}{\|\psi\|} \Big\}, \tag{15}$$

with some constant $c_{s,\lambda}^{(2)}$ called the 'strong stability constant' of λ. Clearly, the strong stability constant behaves like $c_s^{(2)} \sim c_s^{(2)} + |\lambda| c_{s,\lambda}^{(1)}$.

The eigenvalue problem (9) and (11) are approximated by a Galerkin finite element methods. Let $H_h \subset H$ be a (finite dimensional) finite element space where $h \in \mathbb{R}_+$ is a discretization parameter. Then, the approximate eigenvalue problems read as follows. Find non-trivial pairs $\{v_h, \lambda_h\}$ and $\{v_h^*, \lambda_h^*\}$ satisfying

$$a(v_h, \varphi_h) = \lambda_h\,(v_h, \varphi_h) \quad \forall \varphi_h \in H_h, \quad \|v_h\| = 1, \tag{16}$$

$$a(\varphi_h, v_h^*) = \bar{\lambda}_h^*\,(\varphi_h, v_h^*) \quad \forall \varphi_h \in H_h, \quad (v_h, v_h^*) = 1. \tag{17}$$

In algebraic notation (16) and (17) take the form of generalized eigenvalue problems

$$\mathbf{A}_h x_h = \lambda_h \mathbf{M}_h x_h, \qquad \mathbf{A}_h^H x_h^* = \lambda_h^* \mathbf{M}_h x_h^*, \tag{18}$$

where $\mathbf{A}_h$ is the "stiffness matrix" (not necessarily symmetric) and $\mathbf{M}_h$ is the symmetric and positive-definite mass matrix. Again, the approximate primal and dual eigenvalues are related to each other by $\lambda_{h,i}^* = \bar{\lambda}_{h,i}$

Remark 1. *Problem (8) represents only a model case. The results of this paper can be carried over to more complicated situations involving higher-order elliptic operators and other boundary conditions.*

2.2 A priori error estimates

The *a posteriori* error analysis to be developed takes advantage of known *a priori* error estimates for the eigenvalue approximation which are available in the literature. The following lemma summarizes some results which can be extracted from the papers of Bramble & Osborn [7], Osborn [16] and Babuska & Osborn [2]. Below, c is a generic constant which may vary with the context, but is independent of the mesh size h.

Lemma 1. *Suppose that the elliptic operator $\mathcal{A}$ is of order $m = 2$ and the finite element approximation of order $r = 2$. Suppose further that the operator $\mathcal{A}$ is H^2-regular. Let λ be an eigenvalue of $\mathcal{A}$ with algebraic multiplicity σ and ascent α. Then, for sufficiently small h, there are exactly σ approximating eigenvalues $\{\lambda_{h,i}\}_{i=1,\dots,\sigma}$, counted according to their algebraic multiplicity, such that*

$$\Big|\lambda - \tfrac{1}{\sigma}\sum_{i=1}^{\sigma} \lambda_{h,i}\Big| \le c_\lambda\, h^2. \tag{19}$$

Further, let $v_h \in Ker\{(\mathcal{A}_h - \lambda_h I_h)^\alpha\}$ be a normalized (algebraic) eigenvector. Then, there is a $v^h \in H$, such that

$$\|v^h - v_h\| \le c_\lambda\, h^{2/\alpha}. \tag{20}$$

If the eigenvalue λ is simple, or if the approximate eigenvalue λ_h has the same index as λ, then (19) and (20) simplify to

$$|\lambda - \lambda_h| \le c_\lambda\, h^2, \qquad \|v^h - v_h\| \le c_\lambda h^2. \tag{21}$$

Remark 2. *In the following sections, we will derive a posteriori error estimates for the eigenvalues as well as the corresponding eigenfunctions. In this context, there arises the fundamental question of how to associate a suitable eigenpair* $\{v, \lambda\}$ *of (9) to a computed eigenpair* $\{v_h, \lambda_h\}$ *of (16), such that the error estimates make sense. First, we note that the 'ascent' of an eigenvalue is not stable under general perturbations. This may be seen by considering the* 2×2*-matrix* A *and its perturbation* A_ϵ *defined by*

$$A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \qquad A_\epsilon = \begin{pmatrix} \lambda & 1 \\ \epsilon & \lambda \end{pmatrix}, \quad \epsilon \in \mathbb{C}.$$

Here, the eigenvalue λ *with ascent two splits into the two eigenvalues* $\lambda_\epsilon^\pm = \lambda \pm \sqrt{\epsilon}$ *each with ascent one, for arbitrarily small* ϵ. *It is common to consider the set of eigenvalues* $\Sigma_\epsilon(A) := \{\lambda_\epsilon^+, \lambda_\epsilon^-\}$ *as approximation to the eigenvalue* λ *being counted according to its (algebraic) multiplicity. Then, in this sense, the total algebraic multiplicity* $\sigma(\lambda) = 2 = \sigma(\lambda^+) + \sigma(\lambda^-)$ *is stable under the perturbation, while the total geometric multiplicity is unstable:* $\rho(\lambda) < 2 = \rho(\lambda^+) + \rho(\lambda^-)$. *This example reflects the generic situation of continuous eigenvalue problems under discretization. In general, any eigenvalue* λ *of (9) with algebraic multiplicity* $\sigma \geq 2$ *will split into a group of* σ *simple eigenvalues* $\{\lambda_h^{(i)},\, i = 1, ..., \sigma\}$ *which may be considered as the approximation to* λ. *Only in special cases, for instance when the discretization preserves certain symmetries of the given problem, some of the discrete eigenvalues* $\lambda_h^{(i)}$ *may have geometric multiplicity* $\rho(\lambda_h^{(i)}) > 1$. *The case that* $\sigma(\lambda_h^{(i)}) > 1$ *will hardly occur in practice exactly because of asymmetries in the discretization and round-off errors. In view of the foregoing discussion, in computing eigenvalues, we have to expect the following scenario (for 'sufficiently accurate' discretization): There are* $\sigma \geq 1$ *discrete eigenvalues* $\{\lambda_h^{(i)},\, i=1, ..., \sigma\}$ *which are usually all simple and seem to converge (for* $h \to 0$*) to some eigenvalue* $\lambda \in \Sigma(\mathcal{A})$, *such that*

(a) the limit eigenvalue λ *has geometric multiplicity* $\rho = \sigma$, *and hence index* $\alpha = 1$, *or*

(b) the limit eigenvalue λ *has geometric multiplicity* $\rho < \sigma$, *and hence index* $\alpha > 1$.

Here, the term 'sufficiently accurate' is left vague since it depends on the characteristics of the given problem (1), on the discretization and on the location of the eigenvalue λ *in* $\Sigma(\mathcal{A})$. *For practical situations involving non-symmetric operators (like for example in hydrodynamic stability) the case (b) is of particular interest. The most interesting question is how to detect a posteriori whether case (a) or (b) is valid. Unfortunately, this type of a posteriori error analysis becomes difficult and has to be postponed to a later paper. In the following, we concentrate on the simpler case (a) and, for further simplification, suppose that* $\rho = 1$. *The latter assumption is not very critical since the arguments presented below can easily be extended to the case* $\sigma = \rho > 1$.

Remark 3. *The general setting laid out above contains elliptic eigenvalue problems of arbitrary order. Common examples are the (non-selfadjoint) second-order operator* $\mathcal{A}$ *in (1) or the fourth-order biharmonic operator* $\mathcal{A} = \Delta^2$ *in plate-bending theory. The boundary conditions may be of Dirichlet or Neumann type. In practical applications,* *especially in plate and shell theory, the eigenvalue problem may be of generalized form*

$$\mathcal{A}v = \lambda\,\mathcal{B}v, \tag{22}$$

with an symmetric, positive-definite operator $\mathcal{B}$ *that is bounded on the space* H *, e.g., the Laplace operator* $\mathcal{B} = -\Delta$ *on* $H \subset H^2(\Omega)$ *. In this case, the variational formulation is*

$$a(v,\varphi) = \lambda\, b(v,\varphi) \quad \forall \varphi \in H, \tag{23}$$

with the bilinear form $b(\cdot,\cdot)$ *generated by* $\mathcal{B}$ *.*

3 A paradigm for a posteriori error control

At first, we recall from Becker & Rannacher [6] an abstract framework for the *a posteriori* error analysis of Galerkin approximation of general nonlinear variational equations. Let $A(\cdot;\cdot)$ be a differentiable semi-linear form and $F(\cdot)$ a linear functional defined on some real or complex function space V. The derivatives of $A(\cdot;\cdot)$ at a point v in direction δv are denoted by $A'(v;\delta v,\cdot)$, $A''(v;\delta v,\cdot,\cdot)$ and $A'''(v;\delta v,\cdot,\cdot,\cdot)$. We assume that the variational equation

$$A(u;\varphi) = F(\varphi) \quad \forall \varphi \in V, \tag{24}$$

has a unique solution $u \in V$. Suppose that the goal is to compute a certain physical quantity extracted from the solution u by applying a differential functional $J(\cdot)$, with derivatives denoted by $J'(v;\delta v)$, $J''(v;\delta v,\cdot)$ and $J'''(v;\delta v,\cdot,\cdot)$.

Problem (24) is approximated by a Galerkin method using finite dimensional subspaces $V_h \subset V$ parametrized by $h > 0$. We assume that the associated discrete problems

$$A(u_h;\varphi_h) = F(\varphi_h) \quad \forall \varphi_h \in V_h, \tag{25}$$

also possess unique solutions $u_h \in V_h$ and take $J(u_h)$ as approximation to the target quantity $J(u)$. The aim is now to derive *a posteriori* estimates for the error $J(u)-J(u_h)$. To this end, we employ the Euler-Lagrange approach of optimal control theory. The problem of computing $J(u)$ from the the solution of (24) can be equivalently formulated as solving the following (trivial) constrained optimization problem

$$J(u) = \min_{v\in V} \big\{ J(v),\;\; A(v;\varphi) = F(\varphi)\;\; \forall \varphi \in V \big\}. \tag{26}$$

Minima u of (26) correspond to stationary points $\{u,z\} \in V\times V$ of the Lagrangian

$$L(u;z) := J(u) + F(z) - A(u;z),$$

with the adjoint variable $z \in V$. Hence, we seek solutions $\{u,z\} \in V\times V$ to the Euler-Lagrange system

$$A(u;\varphi) = F(\varphi) \qquad \forall \varphi \in V, \tag{27}$$

$$A'(u;\varphi,z) = J'(u;\varphi) \quad \forall \varphi \in V. \tag{28}$$

We note that the first equation of this system is just the considered variational equation (24). The Galerkin approximation of system (27), (28) in the subspace $V_h \subset V$ seeks pairs $\{u_h, z_h\} \in V_h \times V_h$ satisfying

$$A(u_h;\varphi_h) = F(\varphi_h) \qquad \forall \varphi_h \in V_h, \tag{29}$$
$$A'(u_h;\varphi_h, z_h) = J'(u_h;\varphi_h) \quad \forall \varphi_h \in V_h. \tag{30}$$

To the approximate solutions $u_h \in V_h$ of (29) and $z_h \in V_h$ of (30) we associate the residuals

$$\rho(u_h;\cdot) := F(\cdot) - A(u_h;\cdot),$$
$$\rho^*(z_h;\cdot) := J'(u_h;\cdot) - A'(u_h;\cdot, z_h),$$

which are defined on all of V. For $\varphi_h \in V_h$, we have $\rho(u_h;\varphi_h) = \rho^*(z_h;\varphi_h) = 0$, by definition. For this situation, we have from Becker & Rannacher [6] the following result for which the proof is supplied for completeness.

Proposition 1. *For the Galerkin approximation (29), (30) of the saddle-point problem (27), (28), we have the a posteriori error representation*

$$J(u) - J(u_h) = \tfrac{1}{2}\,\rho(u_h; z-\psi_h) + \tfrac{1}{2}\,\rho^*(z_h; u-\varphi_h) + R, \tag{31}$$

for arbitrary elements $\varphi_h, \psi_h \in V_h$. *The remainder term* R *is given by*

$$\begin{aligned} R := \tfrac{1}{2}\int_0^1 \Big\{ & J'''(u_h+se_h; e_h, e_h, e_h) - A'''(u_h+se_h; z_h+se_h^*; e_h, e_h, e_h) \\ & - 3A''(u_h+se_h; e_h, e_h, e_h^*)\Big\}\, s(s-1)\, ds\,, \end{aligned} \tag{32}$$

where $e_h := u-u_h$ *and* $e_h^* := z-z_h$.

Proof. The proof is given in two steps.
i) At first, we derive a general error representation for the Galerkin approximation of stationary points of functionals. Let $K(\cdot)$ a differentiable functional on a (real or complex) function space X and $x \in X$ a stationary point, that is

$$K'(x;y) = 0 \quad \forall y \in X. \tag{33}$$

Further, let X_h be a finite dimensional subspace, indicated by a discretization parameter $h \in \mathbb{R}_+$, and let $x_h \in X_h$ be stationary points of $K(\cdot)$ on X_h, that is

$$K'(x_h;y_h) = 0 \quad \forall y_h \in X_h. \tag{34}$$

For the error $e_h := x - x_h$, we have the identity

$$K(x) - K(x_h) = \int_0^1 K'(x_h+se_h; e_h)\, ds \;=:\; K'(\overline{xx_h}; e_h)\,.$$

Adding zero on the right-hand results in

$$K(x) - K(x_h) = K'(\overline{xx_h}; e_h) + \tfrac{1}{2}K'(x_h; e_h) - \tfrac{1}{2}K'(x_h; e_h) - \tfrac{1}{2}K'(x; e_h)\,.$$

We observe that the last two terms on the right are just the approximation of the second by the trapezoidal rule. Recalling the corresponding remainder term

$$R = \tfrac{1}{2}\int_0^1 K'''(x_h + se_h; e_h, e_h, e_h)\, s(s-1)\, ds,$$

and noting that $K'(x_h; y_h) = 0$ for all $y_h \in X_h$, we obtain the error representation

$$K(x) - K(x_h) = \tfrac{1}{2}\, K'(x_h; x - y_h) + R, \quad y_h \in X_h. \tag{35}$$

ii) Next, we give the proof of the representation (32). To this end, we embed problems (27-28) and (29-30) into the framework laid out in step (i). To this end, we define the spaces $X := V \times V$ and $X_h := V_h \times V_h$, such that the Lagrangian $L(\cdot;\cdot)$ defines a functional $K(\cdot) := L(\cdot;\cdot)$ on X. Correspondingly the solutions $x := \{u, z\} \in X$ of (27-28) and $x_h := \{u_h, z_h\} \in X_h$ of (29-30) are stationary points of $L(\cdot)$ on X and X_h, respectively. At these solutions, by definition,

$$\begin{aligned} K(x) - K(x_h) &= L(u; z) - L(u_h; z_h) \\ &= J(u) + F(z) - A(u; z) - J(u_h) + F(z_h) - A(u_h; z_h) \\ &= J(u) - J(u_h). \end{aligned}$$

Therefore, from the general error representation (35) we obtain that

$$\begin{aligned} J(u) - J(u_h) &= \tfrac{1}{2}\, K'(x_h; x - y_h) + R \\ &= \tfrac{1}{2}\, \rho(u_h, z - \psi_h) + \tfrac{1}{2}\, \rho^*(z_h; u - \varphi_h) + R, \end{aligned}$$

with arbitrary $y_h = \{\varphi_h, \psi_h\} \in V_h \times V_h$. Since $L(\cdot;\cdot)$ is linear in its second argument, the third derivative of $K(\cdot)$ consists of only three terms, namely,

$$\begin{aligned} K'''(x_h + s\hat{e}_h; \hat{e}_h, \hat{e}_h, \hat{e}_h) = J'''(u_h + se_h; e_h, e_h, e_h) - A'''(u_h + se; e_h, e_h, e_h, z_h + se_h^*) \\ - 3A''(u_h + se_h; e_h, e_h, e_h^*). \end{aligned}$$

Therefore, the remainder term R has the asserted form (32). This completes the proof of the lemma. □

Remark 4. *We note that the evaluation of the error representation (31) requires us to determine approximations to the primal as well as the dual solution u and z, respectively. Strategies for achieving those with sufficient accuracy will be discussed below.*

In the following the abstract result of Proposition 1 will be applied to the special situation of the Galerkin approximation of an eigenvalue problem.

4 Application to eigenvalue problems

4.1 General a posteriori error representations

Our error analysis for the approximation of the eigenvalue problems (9), (11) by the discrete problems (16), (17) relies on their interpretation as nonlinear equations and their embedding into the framework laid out in the previous section. In order to keep the presentation as simple as possible, we assume from now on that the eigenvalue λ to be computed is simple, that is its multiplicity is $\sigma = \rho = 1$. Accordingly, for sufficiently fine mesh, the approximate eigenvalues λ_h are likewise simple.

We introduce the spaces $V := H \times \mathbb{C}$ and $V_h := H_h \times \mathbb{C}$ and the eigenpairs $u := \{v, \lambda\}$ and $u_h := \{v_h, \lambda_h\}$, respectively. Further, for $\varphi = \{\psi, \chi\} \in V$, let the nonlinear form $A(\cdot;\cdot)$ be defined by

$$A(u;\varphi) := -a(v,\psi) + \lambda(v,\psi)\} + \overline{\chi}\{\|v\|^2 - 1\}.$$

Then, the primal eigenvalue problems (9) and (16) are equivalent to the nonlinear variational equations

$$A(u;\varphi) = 0 \quad \forall \varphi \in V, \tag{36}$$

and

$$A(u_h;\varphi_h) = 0 \quad \forall \varphi_h \in V_h, \tag{37}$$

respectively. The corresponding residual of the discrete solution $u_h = \{v_h, \lambda_h\}$, satisfying $\|v_h\| = 1$, is defined by $\rho(u_h;\cdot) := a(v_h,\cdot) - \lambda_h(v_h,\cdot)$. For controlling the error of this approximation, we shall use again functionals $J(\cdot)$ defined on V, the associated dual problem

$$A'(u;\varphi,z) = J'(u;\varphi) \quad \forall \varphi \in V, \tag{38}$$

as well as its discrete analogue

$$A'(u_h;\varphi_h,z_h) = J'(u_h;\varphi_h) \quad \forall \varphi_h \in V_h. \tag{39}$$

The solvability of (38) and (39) will have to be discussed for each particular choice of $J(\cdot)$ separately. For later purposes, we note that, for $z = \{w, \mu\} \in V$ and $\varphi = \{\psi, \chi\} \in V$,

$$A'(u;\varphi,z) = -a(\psi,w) + \lambda(\psi,w) + \chi\,(v,w) + 2\,\bar{\mu}\,\mathrm{Re}\{(\psi,v)\} + \bar{\mu}\{\|v\|^2 - 1\}. \tag{40}$$

Note that the last term on the right vanishes in view of the normalization condition $\|v\| = 1$. After this preparation, we give a general representation for the eigenvalue error.

Proposition 2. *Let $\{v_h, \lambda_h\}$, $\{v_h^*, \lambda_h^*\}$ be computed eigenpairs of (16), (17), and $\{v, \lambda\}$, $\{v^*, \lambda^*\}$ any associated eigenpairs of (9), (11). Then, we have*

$$\begin{aligned}(\lambda - \lambda_h)(1 - \sigma_h) = {} & \tfrac{1}{2}\big\{a(v_h, v^* - \psi_h) - \lambda_h\,(v_h, v^* - \psi_h)\big\} \\ & + \tfrac{1}{2}\big\{a(v - \varphi_h, v_h^*) - \bar{\lambda}_h^*\,(v - \varphi_h, v_h^*)\big\},\end{aligned} \tag{41}$$

for arbitrary $\varphi_h, \psi_h \in H$, where $\sigma_h := \frac{1}{2}(v - v_h, v^* - v_h^*)$.

Proof. We choose the functional

$$J(\varphi) := \chi\|\psi\|^2, \quad \varphi = \{\psi,\chi\} \in V,$$

noting that $J(u) := \lambda$, since $\|v\| = 1$. Then, we have

$$J'(u;\varphi) = \chi\,\|v\|^2 + 2\lambda\,\mathrm{Re}\{(\psi,v)\}.$$

Hence, observing that $(v,v^*) = 1 = \|v\|^2$, and $\lambda = \bar\lambda^*$, in the present case the abstract dual problem (38) has the form

$$a(\psi,v^*) = \bar\lambda^*(\psi,v^*) \quad \forall\psi\in H, \tag{42}$$

that is the associated dual solution is given as $z = u^*$. Accordingly, the abstract discrete dual problem (39) has the solution $z_h = u_h^*$, satisfying

$$a(\psi_h,v_h^*) = \bar\lambda_h^*(\psi_h,v_h^*) \quad \forall\psi_h\in H_h. \tag{43}$$

The corresponding residual is $\rho(u_h^*;\cdot) = a(\cdot,v_h^*) - \bar\lambda_h^*(\cdot,v_h^*)$. Finally, we compute the remainder R. In the present situation, by a simple calculation, we have ,

$$\begin{aligned} J'''(u_h{+}se;e,e,e) &= 6(\lambda{-}\lambda_h)\|v{-}v_h\|^2,\\ A'''(u_h{+}se;e,e,e,z_h{+}se^*) &= 0,\\ -3A''(u_h{+}se;e,e,e^*) &= -6(\lambda{-}\lambda_h)\,(v{-}v_h,v^*{-}v_h^*) - 6(\overline{\lambda^*{-}\lambda_h^*})\|v{-}v_h\|^2, \end{aligned}$$

and consequently, noting that $\lambda{-}\lambda_h = \overline{\lambda^*{-}\lambda_h^*}$,

$$R = -3\int_0^1 (\lambda{-}\lambda_h)\,(v{-}v_h,v^*{-}v_h^*)s(s{-}1)\,ds = \tfrac12(\lambda{-}\lambda_h)\,(v{-}v_h,v^*{-}v_h^*).$$

Collection all the derived relations, we obtain the desired result (41). □

Remark 5. *We note that the error representation (41) holds true without any assumption on the multiplicity of the eigenvalue λ. Such a restriction will become necessary only in dealing with the error of the eigenfunctions. Moreover, once one knows the result, the identity (41) can be easily derived also by direct manipulation using the defining equations for $\{v,\lambda\}$, $\{v_h,\lambda_h\}$ and their adjoint counterparts together with the normalization relations $\|v\| = \|v_h\| = 1$ and $(v,v^*) = (v_h,v_h^*) = 1$. For just deriving (41) the general theory of Section 3 would not have been needed. The power of this abstract approach will be used below in deriving error estimates for the eigenfunctions. Furthermore, for the adjoint residual $\rho(u_h;\cdot)$, we have*

$$\rho^*(u_h^*;v{-}v_h) = (\lambda{-}\lambda_h)(v,v_h^*) = (\lambda{-}\lambda_h)\big\{(v{-}v_h,v^*)+1\big\}. \tag{44}$$

Hence, the adjoint residual $\rho^*(u_h^*;v{-}\varphi_h)$ could be eliminated from the representation (41) in favor of an additional second-order remainder term. However, this does not mean that the influence of the dual problem can be avoided, since for strong transport the normalization $(v,v^*) = 1$ forces v^* to be large, such that the term $(v{-}v_h,v^*)$ may not be small enough on meshes constructed only by using $\rho(u_h;v{-}v_h)$. This claim is confirmed by the numerical results presented in Section 5, below.

Remark 6. *In the case of a symmetric eigenvalue problem (1), with* $b \equiv 0$ *, the general error identity (41) reduces to the following simpler form*

$$(\lambda - \lambda_h)(1 - \sigma_h) = a(v_h, v - \varphi_h) - \lambda_h \, (v_h, v - \varphi_h), \tag{45}$$

for arbitrary $\varphi_h \in H_h$ *, where* $\sigma_h := \frac{1}{2} \|v - v_h\|^2$.

Remark 7. *In view of the a priori error estimate (19) for an eigenvalue* λ *with ascent* $\alpha_\lambda > 1$ *, one would have to consider the whole group of approximating eigenvalues* $\{\lambda_h^{(i)}\}$ *and corresponding (algebraic) eigenfunctions, in order to achieve optimal-order accuracy. The appropriate extension of the representation (41) to this situation is the subject of current research.*

Now, we turn to the estimation of the error in the eigenfunctions. To this end, let $j(\cdot) : H \to \mathbb{C}$ be an arbitrary (linear) functional which is used for controlling the error in the eigenfunctions, that is, we want to estimate the error $j(v) - j(v_h)$. Let $\{v, \lambda\}$ be an eigenpair of the eigenvalue problem (9), and suppose again that λ is simple. For applying the general theory of Section 3, we now set for $\varphi = \{\psi, \mu\} \in V$:

$$J(\varphi) := j(\psi).$$

Using this notation and recalling (40), the associated dual problem (38) seeks to determine $z = \{w, \mu\} \in V$, such that

$$-a(\psi, w) + \lambda(\psi, w) + \chi \, (v, w) + 2 \, \bar{\mu} \, \mathrm{Re}\{(\psi, v)\} = j(\psi) \quad \forall \varphi = \{\psi, \chi\} \in V,$$

which reduces to the equation

$$a(\psi, w) - \lambda(\psi, w) = 2 \, \bar{\mu} \, \mathrm{Re}\{(\psi, v)\} - j(\psi) \quad \forall \psi \in H, \tag{46}$$

together with the normalization condition $(v, w) = 0$. Since λ is a simple eigenvalue, this system can be solved if its right-hand side vanishes on the eigenvector v, that is

$$2 \, \bar{\mu} \, \mathrm{Re}\{(v, v)\} - j(v) = 0 \qquad \Leftrightarrow \qquad \bar{\mu} = \tfrac{1}{2} \, j(v).$$

Consequently, for the given eigenpair $\{v, \lambda\}$ of (9) the reduced dual problem

$$a(\psi, w) - \lambda(\psi, w) = j(v) \, \mathrm{Re}\{(\psi, v)\} - j(\psi) \quad \forall \psi \in H, \tag{47}$$

has a solution $w \in H$, which is uniquely determined by the condition $(v, w) = 0$. By an analogous argument, the discrete dual problem (39) is seen to have the form

$$a(\psi_h, w_h) - \lambda_h(\psi_h, w_h) = j(v_h) \, \mathrm{Re}\{(\psi_h, v_h)\} - j(\psi_h) \quad \forall \psi_h \in H_h, \tag{48}$$

where $\{v_h, \lambda_h\}$ is the eigenpair of the approximate eigenvalue problem (16). This problem also has a solution $w_h \in H_h$ which is uniquely determined by the condition $(v_h, w_h) = 0$. This equation determines the dual residual

$$\rho^*(z_h; \cdot) : \, = a(\cdot, w_h) - \lambda_h(\cdot, w_h) - j(v_h) \, \mathrm{Re}\{(\cdot, v_h)\} + j(\cdot). \tag{49}$$

After these preparations, we state the following intermediate result for the error of the eigenfunction approximation.

Proposition 3. *Let $\{v_h, \lambda_h\}$ be a computed eigenpair of (16) and $\{v, \lambda\}$ an associated eigenpair of (9). Then, for the given functional $j(\cdot) : H \to \mathbb{C}$ and the associated solution $w \in H$ of the dual problem (47), we have the identity*

$$
\begin{aligned}
j(v - v_h) = a(v_h, w - \psi_h) - \lambda_h(v_h, w - \psi_h) + (\lambda - \lambda_h)(v - v_h, w) \\
+ \tfrac{1}{2} j(v) \|v - v_h\|^2,
\end{aligned} \tag{50}
$$

for arbitrary $\psi_h \in H_h$.

Proof. We derive the assertion from the general error identity (31) of Section 3. First, we note that with the above settings the primal residual is given by $\rho(u_h; \cdot) = a(v_h, \cdot) - \lambda(v_h, \cdot)$ and the dual residual $\rho^*(z_h; \cdot)$ by (49). Therefore, we have

$$
\begin{aligned}
j(v - v_h) = \tfrac{1}{2} \{a(v_h, w - \psi_h) - \lambda_h(v_h, w - \psi_h)\} + \tfrac{1}{2} \{a(v - \varphi_h, w_h) - \lambda_h(v - \varphi_h, w_h) \\
- j(v_h)\mathrm{Re}\{(v - \varphi_h, v_h)\} + j(v - \varphi_h)\} + R,
\end{aligned}
$$

and consequently, taking $\varphi_h = v_h$,

$$
\begin{aligned}
j(v - v_h) = a(v_h, w - \psi_h) - \lambda_h(v_h, w - \psi_h) + a(v - v_h, w_h) - \lambda_h(v - v_h, w_h) \\
- j(v_h)\mathrm{Re}\{(v - v_h, v_h)\} + 2R.
\end{aligned}
$$

To identify the remainder R , we note that

$$
\begin{aligned}
J'''(u_h + s e_h; e_h, e_h, e_h) &= 0, \\
A'''(u_h + s e_h; z_h + s e^*)(e_h, e_h, e_h) &= 0, \\
-3A''(u_h + s e_h; e_h, e_h, e_h^*) &= -6(\lambda - \lambda_h)(v - v_h, w - w_h) - 6(\bar\mu - \bar\mu_h)\|v - v_h\|^2,
\end{aligned}
$$

which yields

$$
R = \tfrac{1}{2}(\lambda - \lambda_h)(v - v_h, w - w_h) + \tfrac{1}{2}(\bar\mu - \bar\mu_h)\|v - v_h\|^2.
$$

We recall that $\bar\mu = \frac{1}{2} j(v)$ and $\bar\mu_h = \frac{1}{2} j(v_h)$ and obtain

$$
R = \tfrac{1}{2}(\lambda - \lambda_h)(v - v_h, w - w_h) + \tfrac{1}{4} j(v - v_h)\|v - v_h\|^2.
$$

From this, we infer as an intermediate result that

$$
\begin{aligned}
j(v - v_h) = a(v_h, w - \psi_h) - \lambda_h(v_h, w - \psi_h) + a(v - v_h, w_h) - \lambda_h(v - v_h, w_h) \\
- j(v_h)\mathrm{Re}\{(v - v_h, v_h)\} + \tfrac{1}{2} j(v - v_h)\|v - v_h\|^2 + (\lambda - \lambda_h)(v - v_h, w - w_h).
\end{aligned}
$$

Next, by definition and since $(v_h, w_h) = 0$, we have

$$
a(v - v_h, w_h) - \lambda_h(v - v_h, w_h) = (\lambda - \lambda_h)(v, w_h) = (\lambda - \lambda_h)(v - v_h, w_h).
$$

Further, noting that $\|v\| = \|v_h\| = 1$, there holds

$$
\begin{aligned}
\|v - v_h\|^2 &= \|v\|^2 + \|v_h\|^2 - 2\mathrm{Re}\{(v, v_h)\} \\
&= \|v\|^2 - \|v_h\|^2 - 2\mathrm{Re}\{(v - v_h, v_h)\} = -2\mathrm{Re}\{(v - v_h, v_h)\}.
\end{aligned} \tag{51}
$$

Then, combining the last three relations gives us

$$\begin{aligned} j(v-v_h) &= a(v_h, w-\psi_h) - \lambda_h(v_h, w-\psi_h) \\ &\quad + (\lambda-\lambda_h)(v-v_h, w_h) + \tfrac{1}{2}j(v)\|v-v_h\|^2 + (\lambda-\lambda_h)(v-v_h, w-w_h). \end{aligned}$$

From this, we finally obtain the desired identity (50). □

Remark 8. *Proposition 3 requires* λ *to be simple. In the case of multiplicity* $\rho > 1$, *we have to simultaneously consider a whole basis* $\{v^{(i)}, i = 1, ..., \rho\}$ *of the eigenspace kern*$(\mathcal{A} - \lambda I)$ *in setting up the dual problem (47). The tedious details are omitted.*

4.2 Practical a posteriori error estimates

In the following, we want to convert the general *a posteriori* error representation of Proposition 3 into an *a posteriori* error estimate which can be used in practice. For simplicity, we restrict our analysis to the special case of the convection-diffusion equation (1) with Dirichlet boundary conditions. Further, we assume the domain Ω to be polygonal (in 2D) or polyhedral (in 3D), in order to avoid the complications caused by the approximation of curved boundaries.

Let $H_h \subset H$ be finite element subspaces of order $r = 2$, using piecewise linear or d-linear trial and test functions in d dimensions on regular decompositions $\mathbb{T}_h = \{T\}$ of $\bar{\Omega}$. The 'cells' T have width $h_T := \mathrm{diam}(T)$, while $h := \max_{T\in\mathbb{T}_h} h_T$ is the global mesh-size. For more details on the terminology of finite element discretization, we refer to the text books of Ciarlet [9] or Brenner & Scott [8]. For later purposes, we recall the basic facts on finite element interpolation theory. Let $i_h : H \to H_h$ denote the usual operator of (generalized) nodal interpolation satisfying

$$\|v-i_hv\|_T + h_T^{1/2}\|v-i_hv\|_{\partial T} \le c_i h_T \|\nabla v\|_{\tilde{T}}, \quad v \in H^1(\tilde{T}), \tag{52}$$

$$\|v-i_hv\|_T + h_T^{1/2}\|v-i_hv\|_{\partial T} \le c_i h_T^2 \|\nabla^2 v\|_T, \quad v \in H^2(T), \tag{53}$$

with certain 'interpolation constants' c_i, where $\tilde{T} := \cup\{T'\in\mathbb{T}_h : T\cap T' \neq \emptyset\}$.

For the following, we suppose that the computed eigenpairs $\{v_h, \lambda_h\}$, $\{v_h^*, \lambda_h^*\}$ of (16), (17), and the associated eigenpairs $\{v, \lambda\}$, $\{v^*, \lambda^*\}$ of (9), (11) are chosen, such that

$$\Delta_h := \max\{|\lambda-\lambda_h|, \|v-v_h\|, \|v^*-v_h^*\|\} \le 1. \tag{54}$$

This 'saturation' assumption essentially means that the approximate eigenvalue λ_h is already closer to λ than any other eigenvalue of the operator $\mathcal{A}$. Further, according to Proposition 1, for a computed eigenvector v_h of λ_h the corresponding eigenvector v of λ has to be properly chosen. This can be accomplished in general, provided that the initial mesh $\mathbb{T}_h$ is sufficiently fine. As before, c will be a generic constant that varies with the context but is always independent of h and λ. Dependence on λ will mainly be introduced through the 'stability constants' $c_{s,\lambda}^{(1)}$ and $c_s^{(2)}$ specified by (12) and (13), respectively. All these dependencies will be collected in another generic constant C_λ.

To the approximate eigenpairs $\{v_h, \lambda_h\}$ and $\{v_h^*, \lambda_h^*\}$, we associate the primal and dual residuals

$$\begin{aligned} \rho_T &:= \big(\|\mathcal{A}v_h - \lambda_h v_h\|_T^2 + \tfrac{1}{2}h_T^{-1}\|a[\partial_n v_h]\|_{\partial T}^2\big)^{1/2}, \\ \rho_T^* &:= \big(\|\mathcal{A}^* v_h^* - \lambda_h^* v_h^*\|_T^2 + \tfrac{1}{2}h_T^{-1}\|a[\partial_n v_h^*]\|_{\partial T\setminus\partial\Omega}^2\big)^{1/2}, \end{aligned}$$

where $[\partial_n v_h]_\Gamma := \partial_n v_{h|\Gamma} + \partial_{n'} v_{h|\Gamma}$ is related to the jump of $\partial_n v_h$ across the common edge of two cells, $\Gamma = T \cap T'$, with the convention that $[\partial_n v_h] := \partial_n v_h$ along the boundary $\partial\Omega$. With this notation, we have the following first theoretical result.

Proposition 4. *Let $\{v_h, \lambda_h\}$, $\{v_h^*, \lambda_h^*\}$ be computed eigenpairs of (16), (17), and $\{v, \lambda\}$, $\{v^*, \lambda^*\}$ associated eigenpairs of (9), (11), such that the saturation condition (54) is satisfied. Then, we have the a posteriori error estimates*

$$|\lambda - \lambda_h| \le C_\lambda \sum_{T\in\mathbb{T}_h} h_T^2\{\rho_T^2 + \rho_T^{*2}\}, \tag{55}$$

$$\|\nabla(v - v_h)\| + \|\nabla(v^* - v_h^*)\| \le C_\lambda \Big(\sum_{T\in\mathbb{T}_h} h_T^2\{\rho_T^2 + \rho_T^{*2}\}\Big)^{1/2}. \tag{56}$$

with constants C_λ explicitly depending on $|\lambda|$ and the weak stability constant $c_s^{(1)}$ defined in Section 2.

Proof. (i) We begin with the estimate for the eigenvalue error. To this end, we recall the general error representation (41):

$$\begin{aligned} (\lambda - \lambda_h)(1 - \sigma_h) &= \tfrac{1}{2}\{a(v_h, v^* - \psi_h) - \lambda_h\,(v_h, v^* - \psi_h)\} \\ &\quad + \tfrac{1}{2}\{a(v - \varphi_h, v_h^*) - \bar\lambda_h^*\,(v - \varphi_h, v_h^*)\}, \end{aligned} \tag{57}$$

for arbitrary $\varphi_h, \psi_h \in H$, where $|\sigma_h| := \frac{1}{2}|(v - v_h, v^* - v_h^*)| \le \frac{1}{2}$. To evaluate the first residual on the right-hand side, we set $\psi := v^* - \psi_h$. Recalling the structure of the differential operator $\mathcal{A}$, we have by cellwise integration by parts

$$a(v_h, \psi) - \lambda_h\,(v_h, \psi) = \sum_{T\in\mathbb{T}_h} \Big\{(\mathcal{A}v_h - \lambda_h v_h, \psi)_T - (a\partial_n v_h, \psi)_{\partial T}\Big\},$$

and observing that each cell edge occurs twice but with opposite orientation of the outward normal vector n,

$$a(v_h, \psi) - \lambda_h\,(v_h, \psi) = \sum_{T\in\mathbb{T}_h} \Big\{(\mathcal{A}v_h - \lambda_h v_h, \psi)_T - \tfrac{1}{2}(a[\partial_n v_h], \psi)_{\partial T}\Big\}, \tag{58}$$

using the above notation $[\partial_n v_h]$ for the jump across the cell boundary. Analogously, with $\varphi := v - \varphi_h$, we have

$$a(\varphi, v_h^*) - \lambda_h\,(\varphi, v_h^*) = \sum_{T\in\mathbb{T}_h} \Big\{(\varphi, \mathcal{A}^* v_h^* - \lambda_h^* v_h^*)_T - \tfrac{1}{2}(\varphi, a[\partial_n v_h^*])_{\partial T}\Big\}. \tag{59}$$

Using the identities (58) and (59) in (57), we obtain

$$(\lambda-\lambda_h)(1-\sigma_h) = \tfrac{1}{2}\sum_{T\in\mathbb{T}_h}\Big\{(\mathcal{A}v_h-\lambda_h v_h,\psi)_T - \tfrac{1}{2}(a[\partial_n v_h],\psi)_{\partial T} + (\varphi,\mathcal{A}^*v_h^*-\lambda_h^* v_h^*)_T - \tfrac{1}{2}(\varphi,a[\partial_n v_h^*])_{\partial T}\Big\}.$$

Noting that $\sigma_h \le \frac{1}{2}$, we conclude

$$\begin{aligned}|\lambda-\lambda_h| &\le \sum_{T\in\mathbb{T}_h}\Big\{\|\mathcal{A}v_h-\lambda_h v_h\|_T\,\|\psi\|_T + \tfrac{1}{2}\|a[\partial_n v_h]\|_{\partial T}\,\|\psi\|_{\partial T} \\ &\quad + \|\varphi\|_T\,\|\mathcal{A}^*v_h^*-\lambda_h^* v_h^*\| + \tfrac{1}{2}\|\varphi\|_{\partial T}\,\|a[\partial_n v_h^*]\|_{\partial T}\Big\} \\ &\le \sum_{T\in\mathbb{T}_h}\Big\{\rho_T\big(\|\psi\|_T^2+\tfrac{1}{2}h_T\|\psi\|_{\partial T}^2\big)^{1/2} + \rho_T^*\big(\|\varphi\|_T^2+\tfrac{1}{2}h_T\|\varphi\|_{\partial T}^2\big)^{1/2}\Big\}.\end{aligned}$$

Consequently, for any $m\in\mathbb{N}$,

$$|\lambda-\lambda_h| \le \rho_h^{(m)}\omega_h^{(m)*} + \rho_h^{(m)*}\omega_h^{(m)}, \tag{60}$$

with the abbreviations

$$\rho_h^{(m)} := \Big(\sum_{T\in\mathbb{T}_h} h_T^{2m}\rho_T^2\Big)^{1/2}, \quad \rho_h^{(m)*} := \Big(\sum_{T\in\mathbb{T}_h} h_T^{2m}\rho_T^{*2}\Big)^{1/2}.$$

and

$$\begin{aligned}\omega_h^{(m)} &:= \Big(\sum_{T\in\mathbb{T}_h} h_T^{-2m}\big\{\|\varphi\|_T^2+\tfrac{1}{2}h_T\|\varphi\|_{\partial T}^2\Big)^{1/2}, \\ \omega_h^{(m)*} &:= \Big(\sum_{T\in\mathbb{T}_h} h_T^{-2m}\big\{\|\psi\|_T^2+\tfrac{1}{2}h_T\|\psi\|_{\partial T}^2\Big)^{1/2}.\end{aligned}$$

First, we set $m=1$. Then, taking $\varphi_h = i_h v = i_h(v-v_h)+v_h \in H_h$ and $\psi_h = i_h v^* = i_h(v^*-v_h^*)+v_h^* \in H_h$, we obtain by the interpolation estimate (52):

$$\omega_h^{(1)} \le c\,\|\nabla(v^*-v_h^*)\|, \quad \omega_h^{(1)*} \le c\,\|\nabla(v-v_h)\|.$$

This gives us the first intermediate result

$$|\lambda-\lambda_h| \le c\,\rho_h^{(1)}\|\nabla(v^*-v_h^*)\| + c\,\rho_h^{(1)*}\|\nabla(v-v_h)\|. \tag{61}$$

(ii) Next, we turn to the energy-norm error estimate. To this end, we choose the functional

$$j(\psi) := \mathrm{Re}\{(\nabla(v-v_h),\nabla\psi)\}, \quad \psi\in H,$$

and note that $j(v-v_h) = \|\nabla(v-v_h)\|^2$. The corresponding dual solution $w \in H$ is determined by the equations $(v, w) = 0$ and

$$a(\psi, w) - \lambda(\psi, w) = \mathrm{Re}\{(\nabla(v-v_h), \nabla v)\}\, \mathrm{Re}\{(\psi, v)\} - \mathrm{Re}\{(\nabla(v-v_h), \nabla\psi)\},$$

for all $\psi \in H$. By the regularity of form $a_\lambda(\cdot,\cdot)$ on $H/\mathrm{span}\{v\}$, and observing that $\|\nabla v\|^2 \le c|a(v,v)| \le c|\lambda|\|v\|^2 = c|\lambda|$, we have

$$\|w\|_{H^1} \le c c_{s,\lambda}^{(1)} |\lambda|^{1/2} \|\nabla(v-v_h)\|, \tag{62}$$

with the weak stability constant $c_{s,\lambda}^{(1)}$. Next, we recall the general error identity (50) which in the present case reads as follows:

$$\begin{aligned} \|\nabla(v-v_h)\|^2 &= a(v_h, w-\psi_h) - \lambda_h(v_h, w-\psi_h) + (\lambda-\lambda_h)(v-v_h, w) \\ &\quad + \tfrac{1}{2}\mathrm{Re}\{(\nabla(v-v_h), \nabla v)\}\|v-v_h\|^2, \end{aligned} \tag{63}$$

for arbitrary $\psi_h \in H_h$. Using the bound (62), we obtain for the second and third term on the right of (63) that

$$|(\lambda-\lambda_h)(v-v_h, w)| \le c c_{s,\lambda}^{(1)} |\lambda|^{1/2} |\lambda-\lambda_h| \|v-v_h\| \|\nabla(v-v_h)\|,$$

and,

$$\tfrac{1}{2}\mathrm{Re}\{(\nabla(v-v_h), \nabla v)\}\|v-v_h\|^2 \le c\,|\lambda|^{1/2} \|\nabla(v-v_h)\| \|v-v_h\|^2.$$

Using the previous two estimates in (63) yields

$$\begin{aligned} \|\nabla(v-v_h)\|^2 &\le |a(v_h, w-\psi_h) - \lambda_h(v_h, w-\psi_h)| \\ &\quad + C_\lambda \|\nabla(v-v_h)\| \big\{ |\lambda-\lambda_h| \|v-v_h\| + \|v-v_h\|^2 \big\}. \end{aligned} \tag{64}$$

For the residual term on the right in (64), we obtain in a similar way as above taking $\psi_h := i_h w$ and using (52):

$$\begin{aligned} |a(v_h, w-i_h w) - \lambda_h(v_h, w-i_h w)| &\le \sum_{T\in\mathbb{T}_h} \big| (\mathcal{A}v_h - \lambda_h v_h, w-i_h w)_T - (a[\partial_n v_h], w-i_h w)_{\partial T} \big| \\ &\le c \Big(\sum_{T\in\mathbb{T}_h} h_T^2 \rho_T^2 \Big)^{1/2} \Big(\sum_{T\in\mathbb{T}_h} \|\nabla w\|_{\tilde{T}}^2 \Big)^{1/2} \le c\,\rho_h^{(1)} \|\nabla w\|. \end{aligned}$$

In view of the bound (62), it follows that

$$|a(v_h, w-i_h w) - \lambda_h(v_h, w-i_h w)| \le c c_{s,\lambda}^{(1)} \rho_h^{(1)} \|\nabla(v-v_h)\|. \tag{65}$$

Using this estimate in (64) gives us

$$\begin{aligned} \|\nabla(v-v_h)\|^2 &\le c c_{s,\lambda}^{(1)} \rho_h^{(1)} \|\nabla(v-v_h)\| \\ &\quad + C_\lambda \|\nabla(v-v_h)\| \big\{ |\lambda-\lambda_h| \|v-v_h\| + \|v-v_h\|^2 \big\}. \end{aligned}$$

and, consequently,

$$\|\nabla(v-v_h)\| \le c c_{s,\lambda}^{(1)}\, \rho_h^{(1)} + C_\lambda \|v-v_h\| \big\{ |\lambda-\lambda_h| + \|v-v_h\| \big\}. \tag{66}$$

In a completely analogous way, we may also derive the following dual estimate:

$$\|\nabla(v^*-v_h^*)\| \le c c_{s,\lambda}^{(1)} \rho_h^{(1)*} + C_\lambda \|v^*-v_h^*\| \big\{ |\lambda-\lambda_h| + \|v^*-v_h^*\| \big\}. \tag{67}$$

(iii) In the next step, we supply sub-optimal L^2-norm error estimates. To this end, we choose the functional

$$j(\psi) := \mathrm{Re}\{(v-v_h,\psi)\}, \quad \psi \in H,$$

and note that $j(v-v_h) = \|v-v_h\|^2$. In this case, the dual solution $w \in H$ is determined by the equations $(v,w)=0$ and

$$a(\psi,w) - \lambda(\psi,w) = \mathrm{Re}\{(v-v_h,v)\}\,\mathrm{Re}\{(\psi,v)\} - \mathrm{Re}\{(v-v_h,\psi)\},$$

for all $\psi \in H$. Hence, we have the *a priori* bound

$$\|w\|_{H^1} \le c c_{s,\lambda}^{(1)} \|v-v_h\|. \tag{68}$$

Next, we recall the general error identity (50) which in the present case reads as follows:

$$\begin{aligned} \|v-v_h\|^2 &= a(v_h, w-\psi_h) - \lambda_h(v_h, w-\psi_h) \\ &\quad + \tfrac{1}{2}\mathrm{Re}\{(v-v_h,v)\}\|v-v_h\|^2 + (\lambda-\lambda_h)(v-v_h,w), \end{aligned}$$

for arbitrary $\psi_h \in H_h$. This results in

$$\|v-v_h\|^2 \le |a(v_h,w-\psi_h) - \lambda_h(v_h,w-\psi_h)| + \tfrac{1}{2}\|v-v_h\|^3 + |\lambda-\lambda_h|\|v-v_h\|\,\|w\|. \tag{69}$$

Setting $\psi_h := i_h w$ in the residual term on the right, we conclude analogously as above:

$$|a(v_h,w-\psi_h) - \lambda_h(v_h,w-\psi_h)| \le c\rho_h^{(1)}\|\nabla w\|.$$

Then, collecting the last two estimates and using the *a priori* bound (68) and the assumption $\|v-v_h\| \le 1$, we arrive at

$$\|v-v_h\| \le c c_{s,\lambda}^{(1)} \rho_h^{(1)} + C_\lambda |\lambda-\lambda_h|. \tag{70}$$

In an analogous way, we may also derive the corresponding dual L^2-error estimate

$$\|v^*-v_h^*\| \le c c_{s,\lambda}^{(1)}\, \rho_h^{(1)*} + C_\lambda |\lambda-\lambda_h|. \tag{71}$$

(iv) Finally, we combine the preliminary results (61), (66), (67), (70) and (71), to obtain

$$\begin{aligned}
|\lambda-\lambda_h| &\le c\,\rho_h\|\nabla(v^*-v_h^*)\| + c\,\rho_h^{(1)*}\|\nabla(v-v_h)\| \\
&\le c\,\rho_h\Big\{c_{s,\lambda}^{(1)}\rho_h^{(1)*} + C_\lambda\big\{\|v^*-v_h^*\|^2 + |\lambda-\lambda_h|^2\big\}\Big\} \\
&\quad + c\,\rho_h^{(1)*}\Big\{c_{s,\lambda}^{(1)}\,\rho_h^{(1)} + C_\lambda\big\{\|v-v_h\|^2 + |\lambda-\lambda_h|^2\big\}\Big\} \\
&\le c\,\rho_h^{(1)}\Big\{c_{s,\lambda}^{(1)}\rho_h^{(1)*} + C_\lambda\big\{\rho_h^{(1)*} + |\lambda-\lambda_h|\big\}\|v^*-v_h^*\| + C_\lambda|\lambda-\lambda_h|^2\big\}\Big\} \\
&\quad + c\,\rho_h^{(1)*}\Big\{c_{s,\lambda}^{(1)}\rho_h^{(1)} + C_\lambda\big\{\rho_h^{(1)} + |\lambda-\lambda_h|\big\}\|v-v_h\| + C_\lambda|\lambda-\lambda_h|^2\big\}\Big\}.
\end{aligned}$$

From this we conclude the first result, observing that $|\lambda-\lambda_h| \le 1$,

$$|\lambda-\lambda_h| \le C_\lambda\,\{\rho_h^{(1)2} + \rho_h^{(1)*2}\}.$$

Combining this with (70) and (71) yields

$$\|v-v_h\| + \|v^*-v_h^*\| \le C_\lambda\,\{\rho_h^{(1)2} + \rho_h^{(1)*2}\}^{1/2},$$

and then, by (66) and (67),

$$\|\nabla(v^*-v_h^*)\| + \|\nabla(v^*-v_h^*)\| \le C_\lambda\,\{\rho_h^{(1)2} + \rho_h^{(1)*2}\}^{1/2}.$$

This completes the proof. □

Remark 9. *We note that the a posteriori error estimates (55) and (56) are not only reliable but also (asymptotically) efficient. By arguments as described in Verfürth [17], we can derive a bound of the form*

$$\sum_{T\in\mathbb{T}_h} h_T^2\{\rho_T^2 + \rho_T^{*2}\} \le c\big\{|\lambda-\lambda_h| + \|\nabla(v-v_h)\|^2 + \|\nabla(v^*-v_h^*)\|^2\big\}.$$

Since this result is of only little importance in practice, the details of the proof are omitted.

Remark 10. *The constants C_λ in the estimates (55) and (56) can be traced to be of the more precise form $C_\lambda = c\,\{c_{s,\lambda}^{(1)} + \mathcal{O}(\Delta_h)\}$. Using the stronger saturation assumption that Δ_h is sufficiently small (depending on $c_{s,\lambda}^{(1)}$), we could for instance state the eigenvalue error estimate (55) in the sharper form*

$$|\lambda-\lambda_h| \le c\,c_{s,\lambda}^{(1)}\Big(\sum_{T\in\mathbb{T}_h} h_T^2\rho_T^2\Big)^{1/2}\Big(\sum_{T\in\mathbb{T}_h} h_T^2\rho_T^{*2}\Big)^{1/2}. \tag{72}$$

Proposition 4 has been proven under the minimal regularity assumption $v,\, v^* \in H$. If the eigenvalue problem (1) is more regular, that is H^2-regular, then we can derive further higher-order *a posteriori* error bounds.

Proposition 5. *Let the assumptions of Proposition 4 be satisfied, and let additionally the eigenvalue problem (1) be H^2-regular. Then, we have the a posteriori error estimates*

$$|\lambda-\lambda_h| \le C_\lambda\Big(\sum_{T\in\mathbb{T}_h} h_T^4\{\rho_T^2+\rho_T^{*2}\}\Big)^{1/2}, \tag{73}$$

$$\|v-v_h\|+\|v^*-v_h^*\| \le C_\lambda\Big(\sum_{T\in\mathbb{T}_h} h_T^4\{\rho_T^2+\rho_T^{*2}\}\Big)^{1/2}, \tag{74}$$

with constants C_λ explicitly depending on $|\lambda|$, $\|v^*\|$, and the stability constants $c_{s,\lambda}^{(1)}$ and $c_s^{(2)}$ defined in Section 2.

Proof. (i) We begin with the proof of the eigenvalue error estimate. To this end, we recall the error estimate (60) with $m=2$:

$$|\lambda-\lambda_h| \le \rho_h^{(2)}\omega_h^{(2)*}+\rho_h^{(2)*}\omega_h^{(2)},$$

with $\varphi:=v-\varphi_h$, $\psi:=v^*-\psi_h$, and the abbreviations

$$\rho_h^{(2)} := \Big(\sum_{T\in\mathbb{T}_h} h_T^4\rho_T^2\Big)^{1/2}, \quad \rho_h^{(2)*} := \Big(\sum_{T\in\mathbb{T}_h} h_T^4\rho_T^{*2}\Big)^{1/2}.$$

and

$$\omega_h^{(2)} := \Big(\sum_{T\in\mathbb{T}_h} h_T^{-4}\{\|\varphi\|_T^2+\tfrac12 h_T\|\varphi\|_{\partial T}^2\Big)^{1/2},$$

$$\omega_h^{(2)*} := \Big(\sum_{T\in\mathbb{T}_h} h_T^{-4}\{\|\psi\|_T^2+\tfrac12 h_T\|\psi\|_{\partial T}^2\Big)^{1/2}.$$

Taking $\varphi_h:=i_hv$ and $\psi_h:=i_hv^*$, we obtain by the interpolation estimate (53):

$$\omega_h^{(2)} \le c\,\|\nabla^2 v\|, \qquad \omega_h^{(2)*} \le c\,\|\nabla^2 v^*\|.$$

This gives us the intermediate result

$$|\lambda-\lambda_h| \le c\,\big\{\rho_h^{(2)}\|\nabla^2 v^*\|+\rho_h^{(2)*}\|\nabla^2 v\|\big\}.$$

Since v and v^* are eigenfunctions, there holds

$$\|\nabla^2 v\|+\|\nabla^2 v^*\| \le c_s^{(2)}\,\big\{\|\mathcal{A}v\|+\|\mathcal{A}^*v^*\|\big\} \le c_s^{(2)}|\lambda|\,\{\|v\|+\|v^*\|\},$$

with the constant $c_s^{(2)}$ in the *a priori* estimate (13). This completes the proof of the error estimate (73).

(ii) To derive the L^2-norm error estimate (74), we proceed as in step (iii) in the proof of Proposition 4. Using the functional

$$j(\psi) := \mathrm{Re}\{(v-v_h,\psi)\}, \quad \psi\in H,$$

the corresponding dual solution $w\in H$ is determined by the equations $(v,w)=0$ and

$$a(\psi,w)-\lambda(\psi,w)=\mathrm{Re}\{(v-v_h,v)\}\,\mathrm{Re}\{(\psi,v)\}-\mathrm{Re}\{(v-v_h,\psi)\},$$

for all $\psi\in H$. By the assumed H^2-regularity of the problem, we have the *a priori* bound

$$\|w\|_{H^2}\le c_{s,\lambda}^{(2)}\|v-v_h\|. \tag{75}$$

Next, we recall the general error identity (50) which in the present case reads as follows:

$$\begin{aligned}\|v-v_h\|^2&=a(v_h,w-\psi_h)-\lambda_h(v_h,w-\psi_h)+(\lambda-\lambda_h)(v-v_h,w)\\&\quad+\tfrac{1}{2}\mathrm{Re}\{(v-v_h,v)\}\|v-v_h\|^2,\end{aligned} \tag{76}$$

for arbitrary $\psi_h\in H_h$. Using the saturation assumption (54), this results in

$$\|v-v_h\|^2\le 2|a(v_h,w-\psi_h)-\lambda_h(v_h,w-\psi_h)|+2|\lambda-\lambda_h|\,\|w\|.$$

Setting $\psi_h:=i_hw$ in the residual term on the right, we conclude analogously as above, this time using the interpolation error estimate (53):

$$\begin{aligned}|a(v_h,w-i_hw)-\lambda_h(v_h,w-i_hw)|&\le\sum_{T\in\mathbb{T}_h}\big|(\mathcal{A}v_h-\lambda_hv_h,w-i_hw)_T-(a[\partial_nv_h],w-i_hw)_{\partial T}\big|\\&\le c\Big(\sum_{T\in\mathbb{T}_h}h_T^4\rho_T^2\Big)^{1/2}\|\nabla^2w\|=c\,\rho_h^{(2)}\|\nabla^2w\|.\end{aligned}$$

Combining the last two estimates and using the *a priori* H^2-bound (75) and (54), we arrive at

$$\|v-v_h\|\le C_\lambda\rho_h^{(2)}+C_\lambda|\lambda-\lambda_h|.$$

In an analogous way, we may also derive the corresponding dual L^2-error estimate

$$\|v^*-v_h^*\|\le C_\lambda\rho_h^{(2)*}+C_\lambda|\lambda-\lambda_h|.$$

Hence, in view of the eigenvalue error estimate (73), the proof is complete. □

Remark 11. *The a posteriori error estimates in Proposition 4 as well as in Proposition 5 involve the stability constants $c_{s,\lambda}^{(1)}$ and $c_s^{(2)}$. These constants may deteriorate if the (nonsymmetric) transport $b\cdot\nabla v$ increases what leads to strong boundary layers. Then, the regularity of the eigenfunctions varies strongly over the domain and is not sufficiently well described by a global stability constant. In this case the estimates derived so far strongly overestimate the true errors and, what is more critical, lead to inefficient mesh refinement. This effect is confirmed by the numerical examples presented in Section 5, below. To overcome this deficiency, we propose to step back in the analysis and to base the mesh adaptation on the following error estimate derived in the proof of Proposition 4:*

$$|\lambda-\lambda_h|\le\sum_{T\in\mathbb{T}_h}h_T^2\big\{\rho_T\,\omega_T^*+\rho_T^*\,\omega_T\big\}. \tag{77}$$

with the residuals ρ_T, ρ_T^* *as defined above and the weights*

$$\omega_T := h_T^{-2}\big(\|v-i_hv\|_T^2 + \tfrac{1}{2}h_T\|v-i_hv\|_{\partial T}^2\big)^{1/2},$$
$$\omega_T^* := h_T^{-2}\big(\|v^*-i_hv^*\|_T^2 + \tfrac{1}{2}h_T\|v^*-i_hv^*\|_{\partial T}^2\big)^{1/2}.$$

By the interpolation error estimate (53), we have

$$\omega_T \le c_i\|\nabla^2 v\|_T, \quad \omega_T^* \le c_i\|\nabla^2 v^*\|_T.$$

In order to use the local information contained in the weights, we have to evaluate them locally, rather than collecting them into a global stability constant as in the above estimates. At this point, the further development becomes largely heuristic. Since the eigenfunctions v *and* v^* *are not known, we attempt to estimate their second-order derivatives from second-order difference quotients of the computed approximations* v_h *and* v_h^* *,*

$$\omega_T \approx \tilde\omega_T := \|\nabla_h^2 v_h\|_T, \quad \omega_T^* \approx \tilde\omega_T^* := \|\nabla_h^2 v_h^*\|_T.$$

For more details on this approach, we refer to Becker & Rannacher [6]. On the basis of the relation (77) the local error indicators

$$\eta_T := h_T^2\big\{\rho_T\,\tilde\omega_T^* + \rho_h^*\,\tilde\omega_T\big\}$$

are used for controlling the mesh construction. The effectivity of this approach will be demonstrated by the numerical examples in Section 5, below.

Finally, we consider the case that the eigenfunction error is to be controlled with respect to an arbitrary functional $j(\cdot)$ on H .

Proposition 6. *Let the assumptions of Proposition 4 be satisfied, and let* $w \in H$ *be the dual solution associated to the linear error-control functional* $j(\cdot) : H \to \mathbb{C}$ *. Then, we have the a posteriori error estimate*

$$|j(v-v_h)| \le \sum_{T\in\mathbb{T}_h} h_T^2\rho_T\,\omega_T + R_h, \tag{78}$$

where the weights ω_T *are defined by*

$$\omega_T := h_T^{-2}\Big(\|w-\psi_h\|_T^2 + \tfrac{1}{2}h_T^{1/2}\|w-\psi_h\|_{\partial T}^2\Big)^{1/2},$$

for arbitrary $\psi_h \in H_h$ *, and the remainder* R_h *by*

$$R_h \;:=\; \max\{\|w\|, |j(v)|\}\big\{|\lambda-\lambda_h|^2 + \|v-v_h\|^2\big\}.$$

Proof. We recall the general error representation (50) of Proposition 3:

$$j(v-v_h) = a(v_h, w-\psi_h) - \lambda_h(v_h, w-\psi_h) + (\lambda-\lambda_h)(v-v_h, w) + \tfrac{1}{2}j(v)\|v-v_h\|^2,$$

with arbitrary $\psi_h \in H_h$. From this, we easily conclude (78). □

Remark 12. *The remainder term* R_h *in Proposition 6 is of order* $\mathcal{O}(h^4)$ *with respect to the global mesh size* h *, assuming that the problem is* H^2*-regular. Hence, in most practical cases it may be neglected in comparison to the leading residual term.*

4.3 Refinement strategies

Finally, we want to discuss strategies for automatic mesh adaptation based on the *a posteriori* error estimates derived above. Starting point is an estimate of the form

$$|j(v-v_h)| \approx \eta_h(v_h,\lambda_h) := \sum_{T\in\mathbb{T}_h} \eta_T, \tag{79}$$

for an error quantity $j(v-v_h)$, with certain cell-error indicators η_T. Let TOL be a prescribed error tolerance and $N_{\max}$ the maximum number of mesh cells that can be used.

In the following, we consider quadrilateral meshes of a two-dimensional polygonal domain Ω. Grid refinement is realized by edge-bisection, that is by cutting a cell T on mesh level L into 4 regular cells on mesh level $L+1$. Correspondingly, mesh coarsening is realized by combining 4 cells on mesh level L to one cell on mesh level $L-1$. This process may create cells with 'hanging nodes' such that the mesh is not compatible. The resulting nonconformity of the trial functions can be avoided by eliminating the unknowns corresponding to any irregular node P by linear interpolation of the values at the neighboring nodes P' and P'': $v_h(P) := \frac{1}{2}\{v_h(P')+v_h(P'')\}$. Then, the resulting finite element space is again 'conforming', that is $H_h \subset H$.

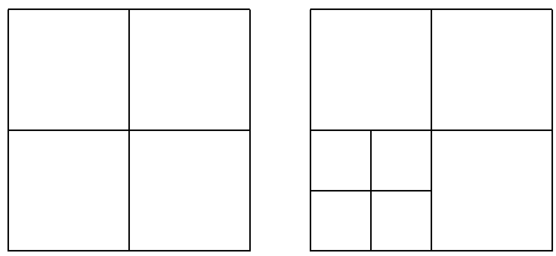

Figure 1: Locally refined mesh patch with 'hanging nodes'.

Starting from a coarse mesh $\mathbb{T}_0 := \mathbb{T}_{h_0}$ with mesh size distribution h_0, let after L refinement cycles the mesh-level $\mathbb{T}_l := \mathbb{T}_{h_l}$ be reached. Let $N_l \approx \dim(H_L)$ be the number of cells of the mesh $\mathbb{T}_l$. On this mesh the approximate solution is $\{v_l,\lambda_l\}$. According to the chosen error-control functional $j(\cdot)$, the corresponding approximate dual solution $w_l \in H_l$ is computed. Then, the associated error estimator $\eta_l(v_l,\lambda_l)$ can be evaluated.

On the basis of an *a posteriori* error estimate of the form (79) the mesh may be generated by a iterative feed-back process consisting of the following components:

Stopping criterion: If the criterion $\eta(v_l,\lambda_l) \le \frac{1}{2}TOL$ is satisfied on the mesh $\mathbb{T}_l$ is satisfied, then the refinement process is stopped and $\{v_l,\lambda_l\}$ is accepted as approximation to $\{v,\lambda\}$ that represents the target quantity $j(v)$ by $j(v_h)$ within the desired tolerance TOL. Otherwise, the the next refinement cycle is started.

Adaptation step: The transition from mesh $\mathbb{T}_l$ to the next mesh $\mathbb{T}_{l+1}$ follows one of the following refinement strategies. At first, the cells $T \in \mathbb{T}_L$ are ordered according to the size of the indicator values η_T ,

$$\{T_i, i = 1, ..., N_L\}: \quad \eta_{T,1} \leq ... \leq \eta_{T,i} \leq \eta_{T,i+1} \leq ... \leq \eta_{T,N_L}.$$

I) 'Error-balancing' strategy: The goal is to balance the error indicators, such that

$$\eta_l(v_l, \lambda_l) \approx \frac{TOL}{N_l}, \quad T \in \mathbb{T}_l\,. \tag{80}$$

Then, we have

$$|j(v-v_l)| \approx \eta(v_l, \lambda_l) \approx \sum_{T\in\mathbb{T}_l} \frac{TOL}{N_l} = TOL.$$

The problem of this strategy is first that the number of cells N_l changes during the adaptation process and second that the balancing criterion (80) requires a delicate choice of control parameters. On the basis of η_{T,N_l} , we check from $j = 0$ downwards whether

$$\eta_{T,i} \leq \frac{TOL}{N_L + 3j}.$$

If this is not satisfied, then the cell T_i is refined, the counters j and i are increased by one, and one proceeds to the next smaller $\eta_{T,i}$. But, if the condition is satisfied, then the new mesh $\mathbb{T}_{l+1}$ is reached.

The 'error balancing' strategy is potentially optimal but involves many expensive checking operations. Alternative simpler but less efficient strategies are the following 'fixed-rate' strategies.

II) 'Fixed-rate' strategies: The goal is in each refinement cycle to increase the number of mesh cells N_L by a fixed rate or to reduce the error estimator $\eta(u_L)$ by a fixed rate. Again, the starting point is an ordering of the cells of $\mathbb{T}_L$ by the size of the indicator values η_T . For prescribed rates $X\%$ and $Y\%$ the cells are grouped according to

$$\#\{T_{N_L-i}, i = 1, ..., N^*\} \approx X\, N_L\,, \qquad \#\{T_i, i = 1, ..., N_*\} \approx Y\, N_L\,.$$

or

$$\sum_{i=N_L-N^*+1}^{N_L} \eta_{T,i} \;\approx\; X\,\eta(u_L), \qquad \sum_{i=1}^{N_*} \eta_{T,i} \;\approx\; Y\,\eta(u_L).$$

Then the cells $T_{N_L-i}\,(i=1, ..., N^*)$ are refined and the cells $T_i\,(i=1, ..., N_*)$ coarsened.

5 Numerical results

For our computational tests, we consider the model problem

$$-\nabla\cdot\{\nabla v\} + b\cdot\nabla v = \lambda v \quad \text{in } \Omega, \quad v = 0 \ \text{ on } \partial\Omega, \tag{81}$$

with $b = (0, b_y)^T$, defined on the rectangular domain $\Omega = (-1,1) \times (-1,3)$ with a slit with tip at $(0,0)$ (see Fig. 2). In the presence of a reentrant corner at $(0,0)$ with angle $\omega = 2\pi$, the solution contains a 'corner singularity', that is it can be written in the form $v = Ar^{1/2}\sin(\theta/2) + \tilde{v}$, with $\tilde{v} \in H^2(\Omega)$ and $\{r,\theta\}$ being polar coordinates. Further, for nontrivial transport, say $b_y > 0$, a boundary layer occurs at the upper boundary $\{y = 3\}$. Note that in this case, the corresponding dual eigenfunction v^* has a boundary layer at the lower boundary $\{y = -1\}$

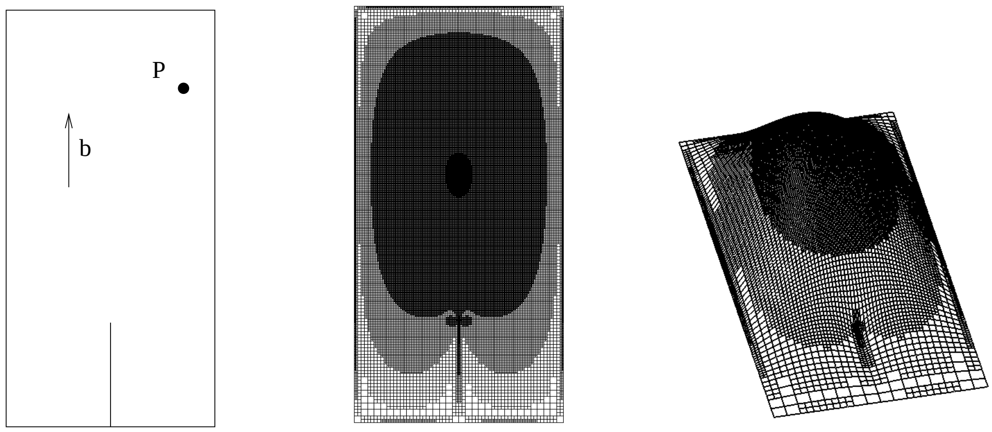

Figure 2: Configuration of the test problem (left), a mesh with about 12,000 cells constructed by the eigenvalue-error estimator $\eta_\lambda^{(1)}$ defined in (82) for the symmetrical case $b_y = 0$ (middle), and the corresponding normalized eigenfunction (right).

We will compare the performance of four different error indicators for controlling the mesh construction in computing the eigenvalue that are derived from the *a posteriori* error estimator (55), a reduced version of this estimator, from the H^2-based estimator (73), and from the 'weighted' estimator (77):

$$\eta_\lambda^{(1)} := \sum_{T\in\mathbb{T}_h} h_T^2\{\rho_T^2 + \rho_T^{*2}\}, \quad \eta_T := h_T^2\{\rho_T^2 + \rho_T^{*2}\}, \tag{82}$$

$$\eta_\lambda^{red} := \sum_{T\in\mathbb{T}_h} h_T^2\rho_T^2, \quad \eta_T := h_T^2\rho_T^2, \tag{83}$$

$$\tilde{\eta}_\lambda^{(2)} := \Big(\sum_{T\in\mathbb{T}_h} h_T^4\{\rho_T^2 + \rho_T^{*2}\}\Big)^{1/2}, \quad \eta_T := h_T^4\{\rho_T^2 + \rho_T^{*2}\}, \tag{84}$$

$$\eta_\lambda^{weight} := \sum_{T\in\mathbb{T}_h} h_T^2\{\rho_T\tilde{\omega}_T^* + \rho_T^*\tilde{\omega}_T\}, \quad \eta_T := h_T^2\{\rho_T\tilde{\omega}_T^* + \rho_T^*\tilde{\omega}_T\}. \tag{85}$$

Testcase 1: At first, we consider the approximation of the dominant eigenvalue λ for the symmetrical case $b_y = 0$. In Fig. 3, we compare the efficiency, that is the observed accuracy TOL versus the number of cells N, of the meshes obtained by using $\eta_\lambda^{(1)}$, $\eta_\lambda^{(2)}$, η_λ^{weight}, and global (uniform) refinement. We see that estimator-driven mesh refinement yields more economical meshes than simple uniform refinement, particularly when higher accuracy is required. Further, due to the lacking H^2 regularity of the present problem, the estimator $\eta_\lambda^{(2)}$ shows weaker performance than the estimators $\eta_\lambda^{(1)}$ and η_λ^{weight}.

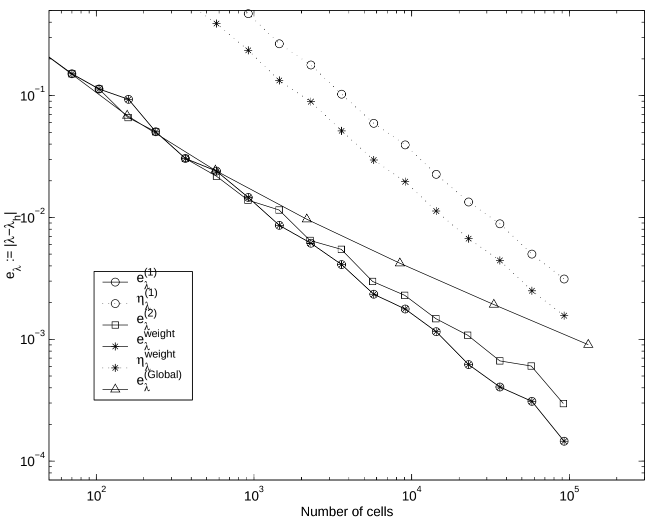

Figure 3: Mesh-efficiencies of the error estimators $\eta_\lambda^{(1)}$ (symbol 'O'), $\eta_\lambda^{(2)}$ (symbol '□'), and η_λ^{weight} (symbol '$*$'), and comparison with uniform refinement (symbol '△') for the symmetrical case.

Testcase 2: Next, we consider the approximation of the dominant eigenvalue λ for the nonsymmetrical case $b_y = 3$. Fig. 4 contains plots of the (normalized) primal and dual eigenfunction corresponding to λ.

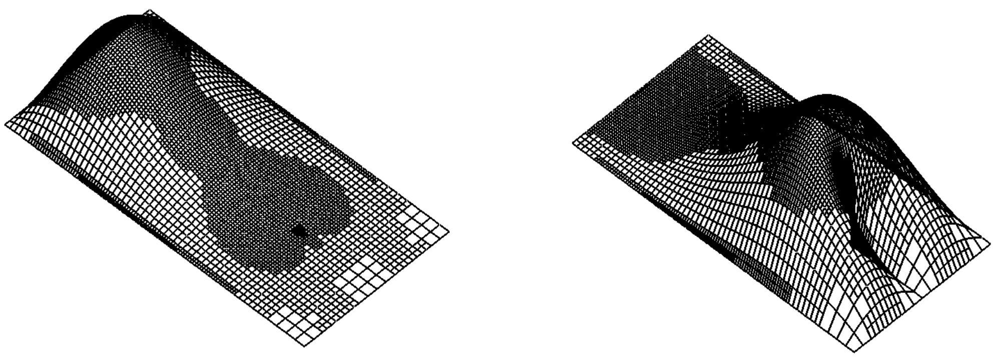

Figure 4: Primal (left) and dual (right) eigenfunctions for the non-symmetrical case.

The 'best' meshes obtained by the error estimators $\eta_\lambda^{(1)}$, η_λ^{red} and η_λ^{weight} are shown in Fig. 5. The 'fixed-rate strategy' with $X = 20\%$ and $Y = 0\%$ has been used in the refinement process. We see that the meshes are quite different, though all three indicators are based on more or less rigorous grounds. In Fig. 6, we compare the efficiency of the error estimators against global (uniform) refinement. We see that the reduced estimator η_λ^{red} shows significantly weaker performance than the other estimators what is due to not resolving the boundary layer of the dual eigenfunction v^* . As predicted, the (heuristical) weighted estimator η_λ^{weight} seems superior over the (rigorous) estimators $\eta_\lambda^{(1)}$.

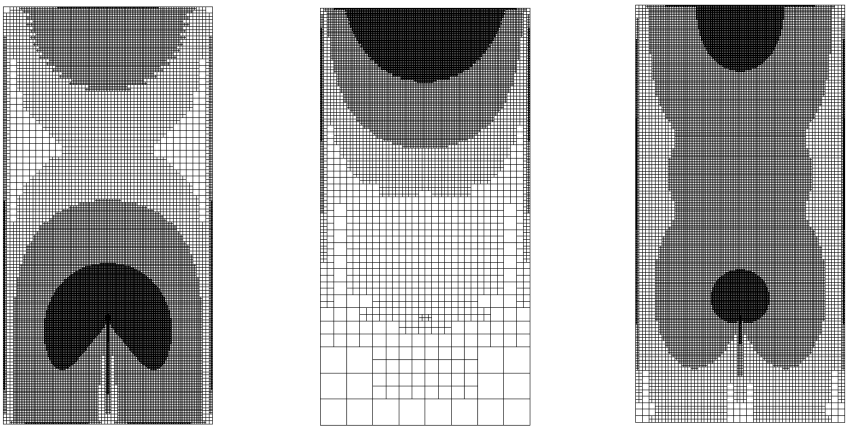

Figure 5: Adapted meshes with about $10,000$ cells constructed by the error estimator $\eta_\lambda^{(1)}$ (left), by its reduced version η_λ^{red} (middle) and by the 'weighted' estimator η_λ^{weight} (right) for the nonsymmetrical case.

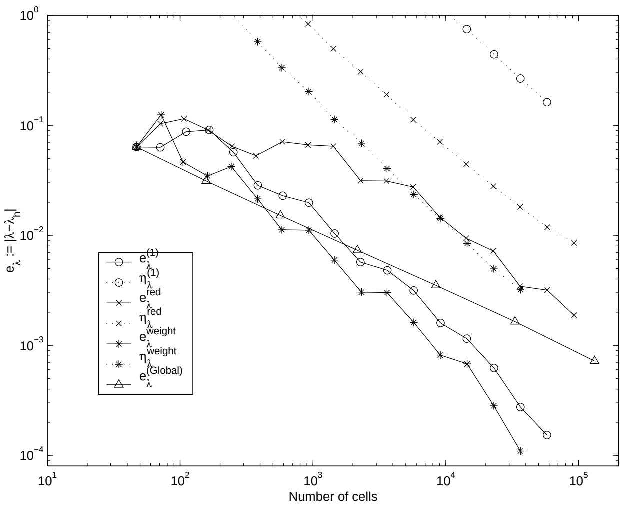

Figure 6: Mesh-efficiencies of the error estimators $\eta_\lambda^{(1)}$ (symbol 'O'), η_λ^{red} (symbol '$\times$'), η_λ^{weight} (symbol '$*$'), compared to uniform refinement (symbol '$\triangle$') for the nonsymmetrical case.

Testcase 3: Finally, we consider the computation of the normalized eigenfunction v corresponding to the dominant eigenvalue λ in the nonsymmetrical case $b_y = 3$. Let the derivative value $\partial_1 v(P)$ to be computed at the point $P = (0.75, 2.25)$. Since the corresponding error control functional $j(v) = \partial_1 v(P)$ is not defined on all of $H = H_0^1(\Omega)$, it has to be properly regularized. For $\epsilon = TOL$, we define

$$j_\epsilon(v) := |B_\epsilon(P)|^{-1} \int_{B_\epsilon(P)} \partial_1 v(x)\, dx,$$

with the ball $B_\epsilon(P) := \{x \in \mathbb{R}^2 : |x-P| < \epsilon\}$. Clearly, this functional is defined on H, such that the corresponding dual solution $w_\epsilon \in H$ exists. It can be argued that w_ϵ behaves like a regularized derivative Green's function, that is

$$|w_\epsilon(x)| \approx d_\epsilon(x)^{-1} + r(x)^{1/2}, \qquad |\nabla^2 w_\epsilon(x)| \approx d_\epsilon(x)^{-3} + r(x)^{-3/2},$$

also containing the corner singularity, where $d_\epsilon(x) := |x - P| + \epsilon$ and $r(x) := |x|$. Recalling the result of Proposition 6, this implies that

$$|\partial_1 e(P)| \;\approx\; \eta_v^{weight} := c \sum_{T \in \mathbb{T}} h_T^3 \rho_T \Big\{ d_{\epsilon,T}^{-3} + r_T^{-3/2} \Big\}, \tag{86}$$

where $d_{\epsilon,T} := \max_{x\in T} d_\epsilon(x)$ and $r_T := \max_{x \in T} r(x)$. Now, suppose that the mesh refinement is organized by systematically balancing the local error indicators according to the 'error-balancing strategy':

$$\eta_T := h_T^3 \rho_T \Big\{ d_{\epsilon,T}^{-3} + r_T^{-3/2} \Big\} \approx \frac{TOL}{N}.$$

Clearly, the resulting mesh will contain cell concentrations at the evaluation point $x = P$ as well at the corner singularity at $x = 0$. However, the influence of the singularity at $x = P$ will dominate that at $x = 0$. We want to estimate the mesh complexity N needed for achieving a tolerance TOL. To this end, we assume that the cell residuals behave like

$$\rho_T \;\approx\; h_T. \tag{87}$$

This can be proven to be satisfied on quasi-uniform meshes by exploiting super-approximation effects but is justified on general (locally refined) meshes only by numerical evidence. Then, neglecting the contribution of the corner singularity at $x = 0$, balancing of the local error indicators yields

$$\eta_T \approx \frac{h_T^4}{d_{\epsilon,T}^3} \approx \frac{TOL}{N} \quad\Rightarrow\quad h_T \approx d_{\epsilon,T}^{3/4} \Big(\frac{TOL}{N}\Big)^{1/4},$$

and, consequently,

$$N = \sum_{T \in \mathbb{T}} h_T^2 h_T^{-2} \approx \Big(\frac{N}{TOL}\Big)^{1/2} \sum_{T \in \mathbb{T}} h_T^2 d_{\epsilon,T}^{-3/2} \approx \Big(\frac{N}{TOL}\Big)^{1/2}.$$

This implies that $N_{\rm opt} \approx TOL^{-1}$. Further, we see that the minimal and maximal cell-width in the optimal mesh $\mathbb{T}_h^{opt}$ are $h_{min} \approx TOL^{5/4}$ and $h_{max} \approx TOL^{1/2}$, respectively.

The mesh $\mathbb{T}_h^{opt}$ can be expected to look quite different from the mesh that would be obtained on the basis of the energy-norm error estimate stated in Proposition 4,

$$\|\nabla(v - v_h)\| \approx \eta_h^E := c\Big(\sum_{T\in\mathbb{T}_h} h_T^2 \{\rho_T^2 + \rho_T^{*2}\}\Big)^{1/2}.$$

Assuming that (87) holds true for ρ_T as well as for ρ_T^*, in this case indicator balancing

$$\eta_T \approx h_T^4 \approx \frac{TOL^2}{N} \quad \Rightarrow \quad h_T \approx \Big(\frac{TOL^2}{N}\Big)^{1/4},$$

leads to

$$N \approx \sum_{T\in\mathbb{T}_h} h_T^2 h_T^{-2} \approx \sum_{T\in\mathbb{T}_h} h_T^2 \Big(\frac{N}{TOL^2}\Big)^{1/2} \approx \Big(\frac{N}{TOL^2}\Big)^{1/2},$$

and consequently to the complexity $N \approx TOL^{-2}$.

This asymptotic behavior is well confirmed by the results obtained for the model problem (81) in the nonsymmetrical case ($b_y = 3$). In Fig. 7, we compare the efficiency of the weighted error estimator η_v^{weight} with that achieved by the eigenvalue-error estimator $\eta_\lambda^{(1)}$, and with uniform refinement. We see that, in computing localized quantities such as the point-value $\partial_1 v$, the use of weights in the error estimator is essential for achieving maximal efficiency. We also see that the estimated error η_h^{weight} reflects the predicted behavior $N \sim TOL^{-1}$, though it would require careful calibration for being used as stopping criterion.

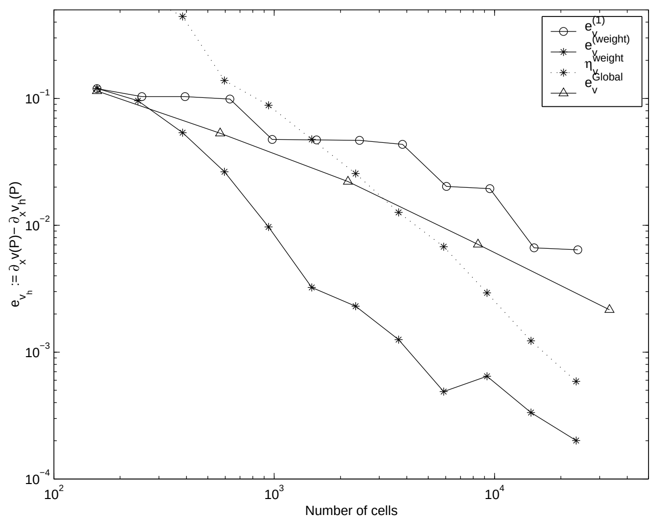

Figure 7: Mesh efficiency of the weighted error estimator η_v^{weight} (symbol '$*$') compared to that achieved by the energy-error estimators $\eta_v^{(1)}$ (symbol 'O') and by uniform refinement (symbol 'Δ') for the approximation of $\partial_1 v(P)$ in the non-symmetrical case.

In Fig. 8, we show an optimized mesh with about $12,000$ cells constructed by the weighted error estimator η_v^{weight} and the corresponding discrete dual solution w_h. Comparing the optimized mesh for computing the derivative-pointvalue of the eigenfunction with that obtained for computing the same quantity for the corresponding boundary value problem (with prescribed right-hand side $f := \lambda v$), we see that the normalization condition $(w, v) = 0$ causes a larger refinement zone. This indicates that in this case the estimator η_v^{weight} has to take care that also the globally determined eigenvalue λ is well approximated besides the local refinement at the point P.

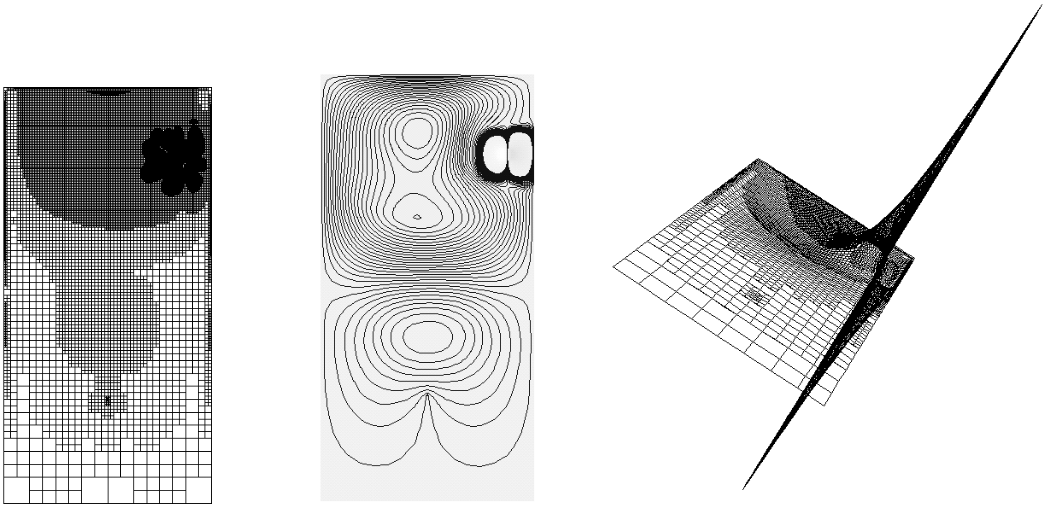

Figure 8: Optimized mesh with about $12,000$ cells for the approximation of $\partial_1 v(P)$ constructed by the weighted error estimator η_v^{weight} (left), and the corresponding discrete dual solution w_h (middle and right).

6 Conclusion

We have derived *a posteriori* error estimates for the Galerkin finite element approximation of second-order, symmetric or nonsymmetric elliptic eigenvalue problems. The estimates for the error in the eigenvalue as well those for the H^1- and L^2-error of the eigenfunctions are asymptotically reliable and efficient. Further, we have obtained 'weighted' error estimates for arbitrary out-put functionals of the eigenfunctions. The efficiency of the resulting estimators in generating optimized meshes has been demonstrated by an example. Here, we have assumed that all eigenvalues have ascent $\alpha = 1$ which excludes the generically interesting case for *nonsymmetric* eigenvalue problems. Also geometric multiplicity $\rho = 1$ is assumed, but this only for technical simplicity. The *a posteriori* error analysis for the 'hard' case $\alpha > 1$ will be the subject of a forthcoming paper.

References

[1] I. Babuška and J. Osborn. A posteriori error estimates for the finite element method. *Int. J. Numer. Methods Engrg*, 12:1597–1615, 1978.

[2] I. Babuska and J. E. Osborn. Eigenvalue problems. In (P.G. Ciarlet and eds.) J.L. Lions, editors, *Handbook of Numerical Analysis*, Vol. 2, Chapter Finite Element Methods (Part 1), pp. 641–792, Elsevier, 1991.

[3] I. Babuska and T. Tsuchiya. A posteriori error estimates of finite element solutions of parametrized nonlinear equations. Technical Report, University of Maryland, 1992.

[4] R. Becker and R. Rannacher. Weighted a posteriori error control in FE methods. In *Proc. ENUMATH'97* (H. G. Bock, et al., eds.), pp. 609–637, World Sci. Publ., Singapore, 1995.

[5] R. Becker and R. Rannacher. A feed-back approach to error control in finite element methods: Basic analysis and examples. *East-West J. Numer. Math.*, 4:237–264, 1996.

[6] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. In *Acta Numerica 2001* (A. Iserles, ed.), Cambridge University Press, Cambridge, 2001, to appear.

[7] J.H. Bramble and J.E. Osborn. Rate of convergence estimates for nonselfadjoint eigenvalue approximations. *Math. Comp.*, 27:525–545, 1973.

[8] S.C. Brenner and R.L. Scott. *The Mathematical Theory of Finite Element Methods*, Springer, Berlin-Heidelberg-New York, 1994.

[9] P.G. Ciarlet. *Finite Element Methods for Elliptic Problems.* North-Holland, Amsterdam, 1978.

[10] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. Introduction to adaptive methods for differential equations. In A. Iserles, editor, *Acta Numerica 1995*, pp. 105–158. Cambridge University Press, 1995.

[11] V. Heuveline and C. Bertsch. On multigrid methods for the eigenvalue computation of non-selfadjoint elliptic operators. *East-West J. Numer. Math.*, 8:275–297, 2000.

[12] C. Johnson. A new paradigm for adaptive finite element methods. In *Proc. MAFELAP 93* (J. Whiteman, ed.), pp. 105-120, Wiley, 1994.

[13] T. Kato. *Perturbation Theory for linear Operators.* Springer, Berlin-Heidelberg-New York, 1966.

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

[15] C. Nystedt. A priori and a posteriori error estimates and adaptive finite element methods for a model eigenvalue problem. Technical Report N0 1995-05, Department of Mathematics, Chalmers University of Technology, 1995.

[16] J. E. Osborn. Spectral approximation for compact operators. *Math. Comp.*, 29:712–725, July 1975.

[17] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques.* John Wiley/Teubner, New York-Stuttgart, 1996.