Numer. Math. (2000) 85: 579–608
Digital Object Identifier (DOI) 10.1007/s002110000135

Numerische Mathematik

Adaptive finite element methods for elliptic equations with non-smooth coefficients

C. Bernardi[1], **R. Verfürth**[2]

[1] Analyse Numérique, C.N.R.S. Université Pierre et Marie Curie, B.C. 187, 4 place Jussieu, F-75252 Paris Cedex 05, France; e-mail: bernardi@ann.jussieu.fr

[2] Fakultät für Mathematik, Ruhr-Universität Bochum, D-44780 Bochum, Germany; e-mail: rv@silly.num1.ruhr-uni-bochum.de

Received February 5, 1999 / Published online March 16, 2000 –

Dedicated to Prof. P.-A. Raviart who showed us the beauty of Numerical Analysis

Summary. We consider a second-order elliptic equation with discontinuous or anisotropic coefficients in a bounded two- or three dimensional domain, and its finite-element discretization. The aim of this paper is to prove some a priori and a posteriori error estimates in an appropriate norm, which are independent of the variation of the coefficients.

Résumé. Nous considérons une équation elliptique du second ordre à coefficients discontinus ou anisotropes dans un domaine borné en dimension 2 ou 3, et sa discrétisation par éléments finis. Le but de cet article est de démontrer des estimations d'erreur a priori et a posteriori dans une norme appropriée qui soient indépendantes de la variation des coefficients.

Mathematics Subject Classification (1991): 65N30

1. Introduction

We consider the Dirichlet problem for second-order elliptic equations

$$\begin{aligned} -\mathrm{div}\,(A\ \mathbf{grad}\,u) &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{1.1}$$

in a bounded two- or three dimensional polyhedral domain with a Lipschitz–continuous boundary. Here, A denotes a function with values in square, symmetric, positive definite matrices of order 2 or 3 according to the space dimension d. We are interested in two rather different situations:

Correspondence to: C. Bernardi

– Either the function A is discontinuous: it is only smooth on a finite number of subdomains and has large jumps across the interfaces between the subdomains. This models for instance several layers of fluids with rather different viscosities which weakly depend on the depth [8, Chap. 3].
– Or the matrix A is constant on the whole domain but has eigenvalues of very different sizes, which results in an anisotropy of equation (1.1). This models for instance elastic materials in thin layers.

We work with a finite element discretization of problem (1.1), relying on possibly anisotropic triangulations of the initial domain. Here "anisotropic" means that the triangulations do not satisfy the standard regularity property which excludes very flat triangles or tetrahedra. It is our goal to prove a priori and a posteriori estimates that are independent of the large parameters linked to equation (1.1), i.e. the size of the jumps in the case of a discontinuous matrix A or the ratio of the eigenvalues in the case of an anisotropic matrix A. Once these estimates are proven, the finite element mesh can be constructed adaptively such that the error is the smallest possible for a fixed number of degrees of freedom.

Let us briefly describe the main ideas which enable us to achieve our goal. If A is discontinuous, the mesh should be aligned with the discontinuities, i.e. jumps of A may only occur across inter-element boundaries. A modification of Clément's quasi-interpolation operator [3] then allows us to obtain estimates for the interpolation error which are independent of the size of the jumps of A. In addition the scaling factors of the error estimator must correctly take into account the local size of A. If A is anisotropic, the mesh should take account of this anisotropy. This means that element geometries should be measured not with respect to the standard Euclidean norm but must be computed using a new metric depending on A.

The outline of this paper is as follows. We analyze the case of a discontinuous, isotropic function A in Sect. 2, the one of a continuous, anisotropic function A in Sect. 3. In each case, we first consider a simple model problem. We give its variational formulation, describe the discrete problem, and prove a priori and a posteriori error estimates. Then we explain how to extend the analysis to more complex situations. In order to clarify the analysis and to avoid unnecessary technicalities, we always work with the simplest finite element space consisting of continuous, piecewise affine functions. All arguments and estimates, however, are formulated in such a way that they immediately carry over to higher order finite element discretizations.

2. Isotropic discontinuous coefficients

We first consider equation (1.1) with $A = \alpha I$ where α is a given, scalar, piecewise constant function on Ω. Accordingly we introduce a disjoint par-

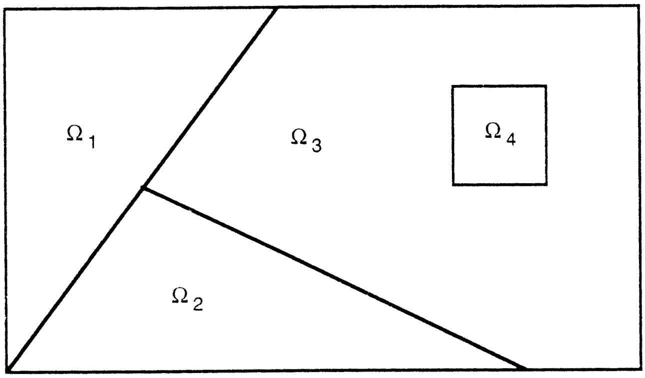

Fig. 1. Partition of the domain Ω

tition of Ω into a finite number of open subdomains Ω_ℓ, $1 \leq \ell \leq L$, such that the function α is equal to the constant α_ℓ on each Ω_ℓ, as illustrated in Fig. 1. We define the two parameters

$$\alpha_{\min} = \min_{1\leq\ell\leq L} \alpha_\ell, \qquad \alpha_{\max} = \max_{1\leq\ell\leq L} \alpha_\ell,$$

and we assume that $\alpha_{\min}$ is positive. We are particularly interested in the critical case where the ratio $\alpha_{\max}/\alpha_{\min}$ is large. Our goal is to establish estimates which are independent of this ratio. In Sects. 2.e and 2.f we will treat the cases of non scalar and piecewise smooth coefficients.

2.a Variational formulation and regularity

We assume that the data f belong to $H^{-1}(\Omega)$. For simplicity, we use the same notation for the scalar product in $L^2(\Omega)$ and for its extension to a duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$. Then, problem (1.1) admits the following equivalent variational formulation: *find* u *in* $H_0^1(\Omega)$ *such that*

$$(2.1)\quad \forall v \in H_0^1(\Omega), \quad \int_\Omega \alpha(\boldsymbol{x})\ \mathbf{grad}\, u \cdot\ \mathbf{grad}\, v \, d\boldsymbol{x} = \int_\Omega f(\boldsymbol{x}) v(\boldsymbol{x})\, d\boldsymbol{x}.$$

Due to the boundedness of α, the bilinear form on the left-hand side of (2.1) is continuous; thanks to the positivity of $\alpha_{\min}$, it is coercive. Hence, the Lax-Milgram lemma leads to the following well-posedness result.

Proposition 2.1 *For any data* f *in* $H^{-1}(\Omega)$, *problem* (2.1) *has a unique solution* u *in* $H_0^1(\Omega)$.

Additional regularity of the solution can be proven thanks to the arguments of Meyers [10]. We refer to [2, Lemma 3.1] for the regularity result in $W^{1,p}(\Omega)$ spaces. In view of the discretization, however, we prefer to work with Hilbertian Sobolev spaces.

Proposition 2.2 *There exists a real number s_0 with $0 < s_0 < \frac{1}{2}$, depending on the geometry of Ω and on the ratio $\alpha_{\max}/\alpha_{\min}$, such that the mapping, which associates with any right-hand side f the unique solution u of problem* (2.1)*, is continuous from $H^{s-1}(\Omega)$ into $H^{s+1}(\Omega) \cap H_0^1(\Omega)$ for all s with $0 \le s \le s_0$.*

Proof. Denote by $\mathcal{L}$ the Laplace operator which associates with any right-hand side f in $H^{-1}(\Omega)$ the unique weak solution u of the Laplace equation

$$\begin{aligned} -\Delta u = f \quad & \text{in } \Omega \\ u = 0 \quad & \text{on } \partial\Omega. \end{aligned}$$

This operator is continuous from $H^{-1}(\Omega)$ into $H_0^1(\Omega)$ with norm 1. Also, there exists a real number s_1 such that $\mathcal{L}$ is continuous from $H^{s_1-1}(\Omega)$ into $H^{s_1+1}(\Omega) \cap H_0^1(\Omega)$. This number is at least $\frac{1}{2}$ in the case of a general polygon or polyhedron [4] and 1 in the case of a convex domain [5, Theorem 3.2.1.2]. Denoting by χ the corresponding norm, an interpolation argument [9, Chap. 1, Théorème 5.1] yields that $\mathcal{L}$ is continuous from $H^{s-1}(\Omega)$ into $H^{s+1}(\Omega) \cap H_0^1(\Omega)$ with norm at most χ^{s/s_1} for all s with $0 \le s \le s_1$.
When dividing equation (2.1) by $\alpha_{\max}$, adding and subtracting Δu, and applying the operator $\mathcal{L}$, we observe that problem (2.1) can equivalently be written as

$$u + \mathcal{L}G(u) = \frac{1}{\alpha_{\max}} \mathcal{L}f, \quad \text{with} \quad G(u) = \Delta u - \operatorname{div} \left(\frac{\alpha}{\alpha_{\max}} \, \mathbf{grad}\, u \right).$$

Hence, the desired regularity result holds for all s such that the norm of $\mathcal{L}G$ from $H^{s+1}(\Omega) \cap H_0^1(\Omega)$ into itself is less than 1. Fix an $s < \frac{1}{2}$ and evaluate $\|G(u)\|_{H^{s-1}(\Omega)}$ for any u in $H^{s+1}(\Omega) \cap H_0^1(\Omega)$. Since the divergence operator is continuous from $L^2(\Omega)^d$ into $H^{-1}(\Omega)$ with norm 1 and from $H^1(\Omega)^d$ into $L^2(\Omega)$ with norm $\sqrt{d}$, it is continuous from $H^s(\Omega)^d$ into $H^{s-1}(\Omega)$ with norm at most $d^{\frac{s}{2}}$. This yields

$$\begin{aligned} \|G(u)\|_{H^{s-1}(\Omega)} &= \|\operatorname{div} \left(\left(1 - \frac{\alpha}{\alpha_{\max}}\right) \, \mathbf{grad}\, u \right) \|_{H^{s-1}(\Omega)} \\ &\le d^{\frac{s}{2}} \| \left(1 - \frac{\alpha}{\alpha_{\max}}\right) \, \mathbf{grad}\, u \|_{H^s(\Omega)^d}. \end{aligned}$$

Since s is less than $\frac{1}{2}$, we have on the other hand

$$\left\|\left(1-\frac{\alpha}{\alpha_{\max}}\right) \mathbf{grad}\, u\right\|_{H^s(\Omega)^d} = \left\{\sum_{\ell=1}^{L}\left\|\left(1-\frac{\alpha_\ell}{\alpha_{\max}}\right)\ \mathbf{grad}\, u\right\|^2_{H^s(\Omega_\ell)^d}\right\}^{\frac{1}{2}}$$
$$\le \left(1-\frac{\alpha_{\min}}{\alpha_{\max}}\right)\ \|\,\mathbf{grad}\, u\|_{H^s(\Omega)^d}.$$

This implies that

$$\|G(u)\|_{H^{s-1}(\Omega)} \le d^{\frac{s}{2}}\left(1-\frac{\alpha_{\min}}{\alpha_{\max}}\right)\ \|u\|_{H^{s+1}(\Omega)\cap H^1_0(\Omega)}.$$

Combining this with the estimate of the norm of $\mathcal{L}$ yields the desired regularity property for all $s < \frac{1}{2}$ such that $d^{\frac{s}{2}}(1-\alpha_{\min}/\alpha_{\max})\,\chi^{s/s_1} < 1$. □

2.b The discrete problem

We consider a family $(\mathcal{T}_h)_h$ of partitions of Ω into a finite number of triangles if $d = 2$, or tetrahedra if $d = 3$, which satisfies the usual *admissibility condition*: any two elements share at most a vertex, or a whole edge, or, if $d = 3$, a complete face. In addition we assume that:

– The family $(\mathcal{T}_h)_h$ is *regular*, i.e. the ratio of the diameter of any element K in $\mathcal{T}_h$ to the diameter of its largest inscribed ball is bounded by a constant σ independent of K and of h.
– For all h, the boundaries of all subdomains Ω_ℓ are the union of edges resp. faces of elements in $\mathcal{T}_h$, i.e., any element K does not intersect two different subdomains Ω_ℓ.
  As usual, h stands for the maximal diameter of the elements K in $\mathcal{T}_h$.

Denote by $\mathcal{P}_1(K)$ the space of restrictions to K of affine functions in $\mathbb{R}^d$ and set

(2.2) $$X_h = \left\{v_h \in \mathcal{C}^0(\overline{\Omega});\ \forall K \in \mathcal{T}_h,\ v_{h\,|K} \in \mathcal{P}_1(K),\ v_h = 0 \text{ on } \partial\Omega\right\}.$$

The discrete problem then is: *find u_h in X_h such that*

(2.3) $$\forall v_h \in X_h,\quad \int_\Omega \alpha(\boldsymbol{x})\ \mathbf{grad}\, u_h \cdot\ \mathbf{grad}\, v_h\, d\boldsymbol{x} = \int_\Omega f(\boldsymbol{x}) v_h(\boldsymbol{x})\, d\boldsymbol{x}.$$

From the Lax-Milgram theorem we again obtain:

Proposition 2.3 *For any data f in $H^{-1}(\Omega)$, problem* (2.3) *has a unique solution u_h in X_h.*

In order to obtain error estimates which are independent of the ratio $\alpha_{\max}/\alpha_{\min}$, we will work with the natural energy norm for (2.1). It is defined for all functions v in $H_0^1(\Omega)$ by

$$(2.4)\quad \|v\|_\alpha = \left\|\alpha^{\frac{1}{2}}\ \mathbf{grad}\ v\right\|_{L^2(\Omega)^d} = \left\{\int_\Omega \alpha(\boldsymbol{x})\ \mathbf{grad}\ v \cdot\ \mathbf{grad}\ v\, d\boldsymbol{x}\right\}^{\frac{1}{2}}.$$

2.c A priori error analysis

Since $\|\cdot\|_\alpha$ is the energy norm of the bilinear form on the left-hand side of (2.1) and (2.3), Céa's lemma [1, Theorem 13.1] immediately implies that

$$(2.5)\qquad \|u-u_h\|_\alpha = \inf_{w_h\in X_h} \|u-w_h\|_\alpha.$$

Hence, we must evaluate the distance of u to X_h for the norm $\|\cdot\|_\alpha$. In order to do this we denote by h_ℓ the maximal diameter of the elements of $\mathcal{T}_h$ which are contained in $\overline{\Omega}_\ell$. For brevity we denote by Π_h^α the orthogonal projection from $H_0^1(\Omega)$ onto X_h for the norm $\|\cdot\|_\alpha$.

Proposition 2.4 *For any real number s with $1 \le s \le 2$, there exists a constant c, which neither depends on h nor on the ratio $\alpha_{\max}/\alpha_{\min}$, such that the following estimate holds for any v in $H^s(\Omega)\cap H_0^1(\Omega)$*

$$(2.6)\qquad \|v-\Pi_h^\alpha v\|_\alpha \le c\left\{\sum_{\ell=1}^L h_\ell^{2(s-1)}\,\alpha_\ell\,\|\ \mathbf{grad}\ v\|^2_{H^{s-1}(\Omega_\ell)^d}\right\}^{\frac{1}{2}}.$$

Proof. We first work with functions in $H^2(\Omega)$, next with general s.
If the function v belongs to $H^2(\Omega)$, it is continuous both in dimensions $d=2$ and $d=3$. Hence, we have

$$\|v-\Pi_h^\alpha v\|_\alpha \le \|v-\mathcal{I}_h v\|_\alpha,$$

where $\mathcal{I}_h$ denotes the Lagrange interpolation operator at the vertices of the elements in $\mathcal{T}_h$. The standard estimate [1, Theorem 16.1]

$$|v-\mathcal{I}_h v|_{H^1(\Omega_\ell)} \le c\,h_\ell\ \|\ \mathbf{grad}\ v\|_{H^1(\Omega_\ell)^d}$$

therefore establishes (2.6) for $s=2$.
Estimate (2.6) holds for $s=2$, according to the first step, and also for $s=1$, due to the definition of Π_h^α. Thus, the general result follows from an interpolation argument relying on the following remark [9, Chap. 1, Théorème 13.1]: each $H^{s-1}(\Omega_\ell)$ is the domain in $L^2(\Omega)$ of a positive self-adjoint operator S_ℓ, and the different S_ℓ commute. $\square$

Combining (2.5) with Proposition 2.4 leads to the a priori error estimate:

Theorem 2.5 *Assume that the solution u of problem (2.1) belongs to $H^s\ (\Omega)$, $1 \leq s \leq 2$. There exists a constant c, which neither depends on h nor on the ratio $\alpha_{\max}/\alpha_{\min}$, such that the following error estimate holds*

$$\|u - u_h\|_\alpha \leq c \left\{ \sum_{\ell=1}^{L} h_\ell^{2(s-1)}\, \alpha_\ell \, \|\, \textbf{grad}\ u\|^2_{H^{s-1}(\Omega_\ell)^d} \right\}^{\frac{1}{2}}. \tag{2.7}$$

Combining Theorem 2.5 with the regularity result of Proposition 2.2 yields

Corollary 2.6 *For any data f in $H^{-t}(\Omega)$, $0 \leq t < 1$, the following convergence holds*

$$\lim_{h \to 0} \|u - u_h\|_\alpha = 0.$$

This convergence result also holds in the standard $H^1(\Omega)$-norm but the convergence seems to be faster in the energy norm. Moreover, Theorem 2.5 shows that the convergence rate may be improved when working with triangulations $\mathcal{T}_h$ such that h_ℓ is small when α_ℓ is large.

2.d A posteriori error analysis

As usual for a posteriori error estimates we assume from now on that f belongs to $L^2(\Omega)$. Given any element K in $\mathcal{T}_h$, we denote by $\mathcal{E}_K$ the set of all its edges, if $d = 2$, resp. faces, if $d = 3$, that are *not contained* in the boundary $\partial\Omega$. The union of all $\mathcal{E}_K$, $K \in \mathcal{T}_h$, is denoted by $\mathcal{E}_h$. With each edge resp. face $e \in \mathcal{E}_h$ we associate a unit vector $\boldsymbol{n}_e$ orthogonal to e and denote by $[\varphi]_e$ the jump of any piecewise continuous function φ across e in direction $\boldsymbol{n}_e$.

From the general results in [11, Sect. 3.2] we know that — up to higher order perturbation terms — the H^1-norm of the error $u - u_h$ is bounded from below and from above by multiples of

$$\left\{ \sum_{K \in \mathcal{T}_h} h_K^2 \, \|f_h + \text{div}\ (\alpha\ \textbf{grad}\ u_h)\|^2_{L^2(K)} + \sum_{e \in \mathcal{E}_h} h_e \, \|\, [\alpha\, \partial_{\boldsymbol{n}_e} u_h]_e \,\|^2_{L^2(e)} \right\}^{\frac{1}{2}}.$$

Here, h_K and h_e denote the diameter of K and e, respectively, and f_h is *any* finite element approximation of f corresponding to $\mathcal{T}_h$. In the simplest case, f_h is the L^2-projection of f onto the space of piecewise constant functions. The higher order terms refer to the error $f - f_h$. The crucial point for the present analysis is that the multiplicative constants depend on the ratio $\alpha_{\max}/\alpha_{\min}$. The results of [12] indicate that this annoying drawback may perhaps be overcome by simultaneously passing to the energy norm and replacing the weights h_K and h_e by factors that appropriately take into account the function α. We therefore try to bound the energy norm $\|u-u_h\|_\alpha$ of the error from above and below by

$$\left\{ \sum_{K\in\mathcal{T}_h} \mu_K^2 \, \|f_h + \mathrm{div}\,(\alpha\ \mathbf{grad}\ u_h)\|_{L^2(K)}^2 + \sum_{e\in\mathcal{E}_h} \mu_e\ \|\,[\alpha\,\partial_{\boldsymbol{n}_e} u_h]_e\,\|_{L^2(e)}^2 \right\}^{\frac{1}{2}} \tag{2.8}$$

and to choose the weights μ_K and μ_e such that the corresponding multiplicative constants do not depend on the ratio $\alpha_{\max}/\alpha_{\min}$ (even if the term div $(\alpha\ \mathbf{grad}\ u_h)$ vanishes in this simple case, we keep it in view of the extension to higher order finite elements).

We start with the lower bound of the error. From (2.1) we obtain for any function v in $H_0^1(\Omega)$

$$\begin{aligned} &\int_\Omega \alpha(\boldsymbol{x})\ \mathbf{grad}\,(u-u_h)\cdot\ \mathbf{grad}\,v\,d\boldsymbol{x} \\ &\quad = \int_\Omega f(\boldsymbol{x})v(\boldsymbol{x})\,d\boldsymbol{x} - \int_\Omega \alpha(\boldsymbol{x})\ \mathbf{grad}\,u_h\cdot\ \mathbf{grad}\,v\,d\boldsymbol{x}. \end{aligned} \tag{2.9}$$

Integration by parts elementwise yields

$$\begin{aligned} &\int_\Omega f(\boldsymbol{x})v(\boldsymbol{x})\,d\boldsymbol{x} - \int_\Omega \alpha(\boldsymbol{x})\ \mathbf{grad}\,u_h\cdot\ \mathbf{grad}\,v\,d\boldsymbol{x} \\ &\quad = \sum_{K\in\mathcal{T}_h} \int_K (f + \mathrm{div}\,(\alpha(\boldsymbol{x})\ \mathbf{grad}\,u_h))v\,d\boldsymbol{x} \\ &\qquad - \sum_{e\in\mathcal{E}_h} \int_e [\alpha(\boldsymbol{x})\partial_{\boldsymbol{n}_e} u_h]_e\,v\,d\tau. \end{aligned} \tag{2.10}$$

Denote by $\mathcal{N}_h$, $\mathcal{N}_K$, and $\mathcal{N}_e$ the sets of all vertices of all elements in $\mathcal{T}_h$, of a given element K, and of a given edge or face e, respectively. With each vertex $\boldsymbol{z}$ in $\mathcal{N}_h$ we associate the corresponding nodal basis function $\varphi_{\boldsymbol{z}}$. It is the unique continuous, piecewise affine function that takes the value 1 at

$\boldsymbol{z}$ and that vanishes at all other vertices. With every element K and every edge resp. face e we associate the bubble functions

$$\psi_K = (d+1)^{d+1} \prod_{z \in \mathcal{N}_K} \varphi_{\boldsymbol{z}} \quad \text{and} \quad \psi_e = d^d \prod_{z \in \mathcal{N}_e} \varphi_{\boldsymbol{z}}.$$

By transforming all quantities to the reference element and using the equivalence of norms on finite dimensional spaces there, one can prove the following estimates [11, Lemma 3.3]

$$\begin{aligned} \|v\|_{L^2(K)} &\le \gamma_1 \left\| \psi_K^{\frac{1}{2}} v \right\|_{L^2(K)}, \\ |\psi_K v|_{H^1(K)} &\le \gamma_2 h_K^{-1} \|v\|_{L^2(K)}, \\ \|\sigma\|_{L^2(e)} &\le \gamma_3 \left\| \psi_e^{\frac{1}{2}} \sigma \right\|_{L^2(e)}, \\ |\psi_e \sigma|_{H^1(K)} &\le \gamma_4 h_e^{-\frac{1}{2}} \|\sigma\|_{L^2(e)}, \\ \|\psi_e \sigma\|_{L^2(K)} &\le \gamma_5 h_e^{\frac{1}{2}} \|\sigma\|_{L^2(e)}. \end{aligned} \tag{2.11}$$

Here, K is an arbitrary element, e is an edge resp. face of K, and v and σ are arbitrary polynomials of degree at most k in d resp. $d-1$ variables. The constants $\gamma_1, \ldots, \gamma_5$ only depend on the polynomial degree k and on the shape parameter of K.

Fix an element K and insert the function $w_K = \psi_K(f_h + \text{div}\,(\alpha\ \textbf{grad}\ u_h))$ as a test-function v in (2.9) and (2.10). We then obtain

$$\begin{aligned} &\|f_h + \text{div}\,(\alpha\ \textbf{grad}\, u_h)\|^2_{L^2(K)} \\ &\quad \le \gamma_1^2 \int_K (f_h + \text{div}\,(\alpha(\boldsymbol{x})\ \textbf{grad}\, u_h))\, w_K d\boldsymbol{x} \\ &\quad = \gamma_1^2 \int_K \alpha(\boldsymbol{x})\ \textbf{grad}\,(u - u_h) \cdot\ \textbf{grad}\, w_K d\boldsymbol{x} \\ &\qquad + \gamma_1^2 \int_K (f_h - f)\, w_K d\boldsymbol{x} \\ &\quad \le \gamma_1^2 \|u - u_h\|_{\alpha;K} \|w_K\|_{\alpha;K} + \gamma_1^2 \|f - f_h\|_{L^2(K)} \|w_K\|_{L^2(K)} \\ &\quad \le \gamma_1^2 \left\{ \|u - u_h\|_{\alpha;K}\, \gamma_2 h_K^{-1} \alpha_K^{\frac{1}{2}} + \|f - f_h\|_{L^2(K)} \right\} \\ &\qquad\qquad \|f_h + \text{div}\,(\alpha\ \textbf{grad}\, u_h)\|_{L^2(K)}. \end{aligned}$$

Here, $\|\cdot\|_{\alpha;K}$ denotes the canonical restriction of the energy norm to K and α_K is the constant value of the function α on the element K. This estimate implies that

$$\begin{aligned}\mu_K\|f_h+ \operatorname{div}(\alpha\ \mathbf{grad}\ u_h)\|_{L^2(K)}\\ \leq \gamma_1^2\gamma_2\mu_K h_K^{-1}\alpha_K^{\frac{1}{2}}\|u-u_h\|_{\alpha;K}+\gamma_1^2\mu_K\|f-f_h\|_{L^2(K)}.\end{aligned}$$

Hence,

$$\mu_K = h_K\alpha_K^{-\frac{1}{2}} \tag{2.12}$$

seems to be a reasonable choice in (2.8).

Next consider an arbitrary edge resp. face e in $\mathcal{E}_h$. Denote by K_1 and K_2 the two elements which are adjacent to e (recall that e is not contained in $\partial\Omega$). Inserting the function $w_e = \psi_e\,[\alpha\,\partial_{\boldsymbol{n}_e}u_h]_e$ as a test function v in (2.9) and (2.10), we conclude that

$$\begin{aligned}\|\,[\alpha\,\partial_{\boldsymbol{n}_e}u_h]_e\,\|_{L^2(e)}^2 &\leq \gamma_3^2\int_e [\alpha(\boldsymbol{x})\,\partial_{\boldsymbol{n}_e}u_h]_e\,w_e d\tau\\
&= -\gamma_3^2\int_{K_1\cup K_2}\alpha(\boldsymbol{x})\ \mathbf{grad}\ (u-u_h)\cdot\ \mathbf{grad}\ w_e d\boldsymbol{x}\\
&\quad +\gamma_3^2\sum_{i=1}^{2}\int_{K_i}(f+\operatorname{div}(\alpha\ \mathbf{grad}\ u_h))\,w_e d\boldsymbol{x}\\
&\leq \gamma_3^2\sum_{i=1}^{2}\Big\{\|u-u_h\|_{\alpha;K_i}\|w_e\|_{\alpha;K_i}\\
&\quad +\|f_h+\operatorname{div}(\alpha\ \mathbf{grad}\ u_h)\|_{L^2(K_i)}\|w_e\|_{L^2(K_i)}\\
&\quad +\|f-f_h\|_{L^2(K_i)}\|w_e\|_{L^2(K_i)}\Big\}\\
&\leq \gamma_3^2\sum_{i=1}^{2}\Big\{\|u-u_h\|_{\alpha;K_i}\,\gamma_4 h_e^{-\frac{1}{2}}\alpha_{K_i}^{\frac{1}{2}}\\
&\quad +\|f_h+\operatorname{div}(\alpha\ \mathbf{grad}\ u_h)\|_{L^2(K_i)}\,\gamma_5 h_e^{\frac{1}{2}}\\
&\quad +\|f-f_h\|_{L^2(K_i)}\,\gamma_5 h_e^{\frac{1}{2}}\Big\}\,\|\,[\alpha\,\partial_{\boldsymbol{n}_e}u_h]_e\,\|_{L^2(e)}.\end{aligned}$$

Combined with the previous estimate this implies that

$$\begin{aligned}\mu_e^{\frac{1}{2}}\|\,[\alpha\,\partial_{\boldsymbol{n}_e}u_h]_e\,\|_{L^2(e)} \leq c\sum_{i=1}^{2}\Big\{&\mu_e^{\frac{1}{2}}h_e^{-\frac{1}{2}}\alpha_{K_i}^{\frac{1}{2}}\|u-u_h\|_{\alpha;K_i}\\ &+\mu_e^{\frac{1}{2}}h_e^{\frac{1}{2}}\|f-f_h\|_{L^2(K_i)}\Big\}.\end{aligned}$$

This suggests that

$$\mu_e = h_e \alpha_e^{-1} \qquad \text{with} \qquad \alpha_e = \max\{\alpha_{K_1}, \alpha_{K_2}\} \tag{2.13}$$

may be a reasonable choice in (2.8).

We now try to prove that (2.8) with the choices (2.12) and (2.13) yields the desired upper bounds on the error. For brevity set $w = u - u_h$. From the definition (2.4) of the energy norm, we immediately conclude that

$$\|u - u_h\|_\alpha^2 = \int_\Omega \alpha(\boldsymbol{x}) \ \mathbf{grad}\ (u - u_h) \cdot \ \mathbf{grad}\ w \, d\boldsymbol{x}. \tag{2.14}$$

Subtracting (2.1) and (2.3) we obtain Galerkin orthogonality

$$\forall w_h \in X_h, \quad \int_\Omega \alpha(\boldsymbol{x}) \ \mathbf{grad}\ (u - u_h) \cdot \ \mathbf{grad}\ w_h \, d\boldsymbol{x} = 0. \tag{2.15}$$

Fix an arbitrary function w_h in X_h. Equations (2.9), (2.10), (2.14), and (2.15) together with the Cauchy-Schwarz inequality then imply that

$$\begin{aligned}
\|u - u_h\|_\alpha^2 &= \sum_{K \in \mathcal{T}_h} \int_K (f + \operatorname{div}\,(\alpha(\boldsymbol{x}) \ \mathbf{grad}\ u_h))\,(w - w_h)\, d\boldsymbol{x} \\
&\quad - \sum_{e \in \mathcal{E}_h} \int_e [\alpha(\boldsymbol{x})\, \partial_{\boldsymbol{n}_e} u_h]_e \ (w - w_h)\, d\tau \\
&\le \sum_{K \in \mathcal{T}_h} \mu_K \, \|f + \operatorname{div}\,(\alpha \ \mathbf{grad}\ u_h)\|_{L^2(K)} \, \mu_K^{-1} \, \|w - w_h\|_{L^2(K)} \\
&\quad + \sum_{e \in \mathcal{E}_h} \mu_e^{\frac{1}{2}} \, \|\, [\alpha\, \partial_{\boldsymbol{n}_e} u_h]_e \,\|_{L^2(e)} \, \mu_e^{-\frac{1}{2}} \, \|w - w_h\|_{L^2(e)} \\
&\le \Bigg\{ \sum_{K \in \mathcal{T}_h} \mu_K^2 \, \|f_h + \operatorname{div}\,(\alpha \ \mathbf{grad}\ u_h)\|_{L^2(K)}^2 \\
&\quad + \sum_{e \in \mathcal{E}_h} \mu_e \ \|\, [\alpha\, \partial_{\boldsymbol{n}_e} u_h]_e \,\|_{L^2(e)}^2 \Bigg\}^{\frac{1}{2}} \\
&\quad \Bigg\{ \sum_{K \in \mathcal{T}_h} \mu_K^{-2} \, \|w - w_h\|_{L^2(K)}^2 + \sum_{e \in \mathcal{E}_h} \mu_e^{-1} \, \|w - w_h\|_{L^2(K)}^2 \Bigg\}^{\frac{1}{2}} .
\end{aligned}$$

Consequently, we will have achieved our goal once we can choose w_h such that the approximation estimate

$$\left\{ \sum_{K \in \mathcal{T}_h} \mu_K^{-2} \|w - w_h\|_{L^2(K)}^2 + \sum_{e \in \mathcal{E}_h} \mu_e^{-1} \|w - w_h\|_{L^2(K)}^2 \right\}^{\frac{1}{2}} \le c \|w\|_\alpha \tag{2.16}$$

holds with a constant that does not depend on the ratio $\alpha_{\max}/\alpha_{\min}$.

To realize this we modify the quasi-interpolation operator of [13] (the operator of Clément [3] can be modified similarly). Given a vertex $\boldsymbol{z}$ in $\mathcal{N}_h$ we denote by $\omega_{\boldsymbol{z}}$ the support of the nodal basis function $\varphi_{\boldsymbol{z}}$. This is the union of all elements that have $\boldsymbol{z}$ as a vertex. With each vertex $\boldsymbol{z}$ we associate a number $\ell(\boldsymbol{z})$ in $\{1, \ldots, L\}$ such that

- $\boldsymbol{z}$ is contained in $\overline{\Omega}_{\ell(\boldsymbol{z})}$ and
- $\alpha_{\ell(\boldsymbol{z})}$ is maximal among all α_j such that $\overline{\Omega}_j$ contains $\boldsymbol{z}$.

Denote by

$$\fint_\omega v d\boldsymbol{x} = \frac{1}{\text{meas}_d(\omega)} \int_\omega v d\boldsymbol{x}$$

the mean-value of a given function v on a given measurable set ω in $\mathbb{R}^d$ with positive d-dimensional Lebesgue measure $\text{meas}_d(\omega)$. With this convention we set

$$\pi_{\boldsymbol{z}} v = \begin{cases} \fint_{\omega_{\boldsymbol{z}} \cap \Omega_{\ell(\boldsymbol{z})}} v d\boldsymbol{x} & \text{if } \boldsymbol{z} \in \Omega \\ 0 & \text{if } \boldsymbol{z} \in \partial\Omega \end{cases} \tag{2.17}$$

and define the quasi-interpolation operator $I_h : L^2(\Omega) \longrightarrow X_h$ by

$$I_h v = \sum_{\boldsymbol{z} \in \mathcal{N}_h} (\pi_{\boldsymbol{z}} v) \varphi_{\boldsymbol{z}}. \tag{2.18}$$

The operator I_h differs from the operator introduced in [13] by the treatment of vertices that are on the boundary of a subdomain. The following lemma shows that taking w_h equal to $I_h w$ realizes the desired estimate (2.16), provided the partition into subdomains satisfies:

Hypothesis 2.7 *For any two different subdomains $\overline{\Omega}_\ell$ and $\overline{\Omega}_k$, which share at least one point, there is a connected path passing from $\overline{\Omega}_\ell$ to $\overline{\Omega}_k$ through adjacent subdomains such that the function α is monotone along this path (adjacent means that the corresponding subdomains share an edge, if $d = 2$, or a face, if $d = 3$).*

Lemma 2.8 *Assume that Hypothesis 2.7 is satisfied. For every function v in $H_0^1(\Omega)$, every element K, and every edge resp. face e of K, the following estimates hold*

$$\|v - I_h v\|_{L^2(K)} \leq c_1 h_K \alpha_K^{-\frac{1}{2}} \|v\|_{\alpha;\Delta_K},$$
$$\|v - I_h v\|_{L^2(e)} \leq c_2 h_e^{\frac{1}{2}} \alpha_e^{-\frac{1}{2}} \|v\|_{\alpha;\Delta_e}.$$

Here, Δ_K and Δ_e denote the union of all elements that share at least one vertex with K or e, respectively. The constants c_1 and c_2 only depend on the shape parameter of $\mathcal{T}_h$.

Proof. We first consider an arbitrary element K. Since the nodal basis functions form a partition of unity we have

$$\|v - I_h v\|_{L^2(K)} = \left\| \sum_{z \in \mathcal{N}_K} \varphi_z (v - \pi_z v) \right\|_{L^2(K)} \le \sum_{z \in \mathcal{N}_K} \|\varphi_z (v - \pi_z v)\|_{L^2(K)}.$$

Consider a vertex z that is not contained in the boundary of any subdomain (including the boundary of Ω). From the Poincaré (also called Bramble–Hilbert) inequality and the regularity of $\mathcal{T}_h$ we conclude that

$$\begin{aligned} \|\varphi_z (v - \pi_z v)\|_{L^2(K)} &\le \|v - \pi_z v\|_{L^2(K)} \le \|v - \pi_z v\|_{L^2(\omega_z)} \\ &\le c \operatorname{diam}(\omega_z) \, |v|_{H^1(\omega_z)} \le c' h_K \alpha_K^{-\frac{1}{2}} \|v\|_{\alpha;\omega_z} \\ &\le c' h_K \alpha_K^{-\frac{1}{2}} \|v\|_{\alpha;\Delta_K}. \end{aligned}$$

The constants c and c' only depend on the shape parameter of $\mathcal{T}_h$. They are explicitly calculated in [13].
Next consider a vertex z on the boundary $\partial\Omega$. Since $\pi_z v$ is equal to zero and since v vanishes on $\partial\Omega$, the previous arguments remain valid using the Friedrichs (also called Poincaré–Friedrichs) inequality instead of the Poincaré inequality.
Finally consider a vertex which is not on the boundary $\partial\Omega$ but which is in $\partial\Omega_{\ell(K)}$ where $\ell(K)$ is such that K is contained in $\overline{\Omega}_{\ell(K)}$. If $\ell(K) = \ell(z)$ the previous arguments remain valid with ω_z replaced by $\omega_z \cap \Omega_{\ell(K)}$.
If $\ell(K) \neq \ell(z)$, we must argue differently. From the definition of π_z we now obtain

$$\begin{aligned} \|\varphi_z (v - \pi_z v)\|_{L^2(K)} &= \left\| \varphi_z \left(v - \fint_{\omega_z \cap \Omega_{\ell(z)}} v \, d\boldsymbol{x} \right) \right\|_{L^2(K)} \\ &\le \left\| \varphi_z \left(v - \fint_{\omega_z \cap \Omega_{\ell(K)}} v \, d\boldsymbol{x} \right) \right\|_{L^2(K)} \\ &\quad + \left\| \varphi_z \left(\fint_{\omega_z \cap \Omega_{\ell(K)}} v \, d\boldsymbol{x} - \fint_{\omega_z \cap \Omega_{\ell(z)}} v \, d\boldsymbol{x} \right) \right\|_{L^2(K)}. \end{aligned}$$

The first term can be estimated exactly as before. Using the regularity of $\mathcal{T}_h$, the second term may be estimated as follows

$$\begin{aligned}
&\left\| \varphi_z \left(\fint_{\omega_z \cap \Omega_{\ell(K)}} v d\boldsymbol{x} - \fint_{\omega_z \cap \Omega_{\ell(z)}} v d\boldsymbol{x} \right) \right\|_{L^2(K)} \\
&\quad = \|\varphi_z\|_{L^2(K)} \left| \fint_{\omega_z \cap \Omega_{\ell(K)}} v d\boldsymbol{x} - \fint_{\omega_z \cap \Omega_{\ell(z)}} v d\boldsymbol{x} \right| \\
&\quad \leq c h_K^{\frac{d}{2}} \left| \fint_{\omega_z \cap \Omega_{\ell(K)}} v d\boldsymbol{x} - \fint_{\omega_z \cap \Omega_{\ell(z)}} v d\boldsymbol{x} \right| .
\end{aligned}$$

Consider first the case where the subdomains $\Omega_{\ell(K)}$ and $\Omega_{\ell(z)}$ are adjacent, i.e. they share a common edge, if $d = 2$, or face, if $d = 3$, which is labeled e. Invoking the regularity of $\mathcal{T}_h$ once more we obtain

$$\begin{aligned}
&h_K^{\frac{d}{2}} \left| \fint_{\omega_z \cap \Omega_{\ell(K)}} v d\boldsymbol{x} - \fint_{\omega_z \cap \Omega_{\ell(z)}} v d\boldsymbol{x} \right| \\
&\quad \leq c h_e^{\frac{1}{2}} \left\| \fint_{\omega_z \cap \Omega_{\ell(K)}} v d\boldsymbol{x} - \fint_{\omega_z \cap \Omega_{\ell(z)}} v d\boldsymbol{x} \right\|_{L^2(e)} \\
&\quad \leq c h_e^{\frac{1}{2}} \left\| \fint_{\omega_z \cap \Omega_{\ell(K)}} v d\boldsymbol{x} - v \right\|_{L^2(e)} + c h_e^{\frac{1}{2}} \left\| v - \fint_{\omega_z \cap \Omega_{\ell(z)}} v d\boldsymbol{x} \right\|_{L^2(e)} .
\end{aligned}$$

Let k be any of the two indices $\ell(K)$ or $\ell(\boldsymbol{z})$ and denote by K' the element which is adjacent to e and contained in $\overline{\Omega}_k$. Invoking the trace theorem [13, Lemma 3.2]

$$\|\varphi\|_{L^2(e)} \leq c \left\{ h_e^{-\frac{1}{2}} \|\varphi\|_{L^2(K')} + h_e^{\frac{1}{2}} |\varphi|_{H^1(K')} \right\} \tag{2.19}$$

we arrive at

$$\begin{aligned}
h_e^{\frac{1}{2}} \left\| v - \fint_{\omega_z \cap \Omega_k} v \right\|_{L^2(e)} &\leq c \left\{ \left\| v - \fint_{\omega_z \cap \Omega_k} v d\boldsymbol{x} \right\|_{L^2(K')} + h_e |v|_{H^1(K')} \right\} \\
&\leq c' h_K \alpha_K^{-\frac{1}{2}} \|v\|_{\alpha;\Delta_K} .
\end{aligned}$$

When the domains $\Omega_{\ell(K)}$ and $\Omega_{\ell(z)}$ are not adjacent, by using Hypothesis 2.7, we introduce the domains Ω_ℓ which are on the path between them and apply the same arguments to the difference meanvalues on each pair of adjacent subdomains.

This establishes the first estimate of the lemma.
The second one is proven in exactly the same way observing that

$$\|\varphi_{\boldsymbol{z}}\|_{L^2(e)} \leq c h_e^{\frac{d-1}{2}}$$

and invoking the trace theorem (2.19). For the latter, the element K adjacent to e must be chosen such that α is maximal. □

Summarizing all results we obtain the following a posteriori error estimates.

Theorem 2.9 *Denote by α_K the constant value of α on the element K in $\mathcal{T}_h$, and define α_e as the largest of the two α_K such that the element K is adjacent to the edge resp. face e in $\mathcal{E}_h$. For any element K in $\mathcal{T}_h$, set*

$$\eta_K = \Big\{ h_K^2 \alpha_K^{-1} \, \|f_h + \operatorname{div}\,(\alpha \ \mathbf{grad}\ u_h)\|^2_{L^2(K)} \tag{2.20}$$
$$+ \frac{1}{2} \sum_{e \in \mathcal{E}_K} h_e \alpha_e^{-1} \, \|[\alpha\, \partial_{\boldsymbol{n}_e} u_h]_e\|^2_{L^2(e)} \Big\}^{\frac{1}{2}}.$$

Assume that Hypothesis 2.7 is satisfied. There exist constants c_1 and c_2 which only depend on the shape parameter of $\mathcal{T}_h$ such that the estimates

$$\|u - u_h\|_\alpha \leq c_1 \left\{ \sum_{K \in \mathcal{T}_h} \left[\eta_K^2 + h_K^2 \alpha_K^{-1} \|f - f_h\|^2_{L^2(K)} \right] \right\}^{\frac{1}{2}} \tag{2.21}$$

and

$$\eta_K \leq c_2 \left\{ \|u - u_h\|^2_{\alpha;\omega_K} + \sum_{K' \subset \omega_K} h_{K'}^2 \alpha_{K'}^{-1} \|f - f_h\|^2_{L^2(K')} \right\}^{\frac{1}{2}} \tag{2.22}$$

hold for all finite element approximations f_h of f, all elements K, and all values of $\alpha_{\max}/\alpha_{\min}$. Here, ω_K denotes the union of all elements that share an edge, if $d = 2$, or a face, if $d = 3$, with K.

Remark 2.10 If Hypothesis 2.7 is violated, estimates (2.21) and (2.22) still hold. But the constant c_1 now depends on the ratio $\alpha_{\max}/\alpha_{\min}$, since the same now holds for the constants in Lemma 2.8. Note that a sufficient condition for Hypothesis 2.7 to hold in the case of dimension $d = 2$ is that at most 3 subdomains $\overline{\Omega}_\ell$ share a common point interior to Ω and that at most 2 subdomains $\overline{\Omega}_\ell$ share a common point on $\partial\Omega$.

2.e Treatment of non scalar piecewise constant coefficients

Now, we consider problem (1.1) with a piecewise constant function A with values in the space of square, symmetric, positive definite matrices of order d. Denote by A_ℓ the constant value of A on Ω_ℓ and by $\lambda_{\max}(A_\ell)$ and $\lambda_{\min}(A_\ell)$ the extremal eigenvalues of A_ℓ. Set

$$\alpha_{\min} = \min_{1\le\ell\le L} \lambda_{\min}(A_\ell), \quad \alpha_{\max} = \max_{1\le\ell\le L} \lambda_{\max}(A_\ell),$$
$$\kappa = \max_{1\le\ell\le L} \frac{\lambda_{\max}(A_\ell)}{\lambda_{\min}(A_\ell)}.$$

We are interested in the case where $\alpha_{\max}/\alpha_{\min}$ is large, but κ is of moderate size. The case of a large κ is treated in Sect. 3.

The variational formulation and the discrete problem are given by (2.1) and (2.3) with the function α replaced by the function A. Thanks to the Lax-Milgram lemma, they both admit a unique solution for any data f in $H^{-1}(\Omega)$. The regularity of the weak solution is as in Proposition 2.2. The energy norm is given by (2.4) with the function α replaced by the function A.

Obviously, estimate (2.5) still holds in the present situation. With the same definition of the operator Π_h^α and under the assumptions of Proposition 2.4, estimate (2.6), however, must be replaced by

$$\|v - \Pi_h^\alpha v\|_\alpha \le c \left\{ \sum_{\ell=1}^{L} h_\ell^{2(s-1)} \, \lambda_{\max}(A_\ell) \, \| \,\mathbf{grad}\, v\|^2_{H^{s-1}(\Omega_\ell)^d} \right\}^{\frac{1}{2}}.$$

Combining all this leads to the *a priori* error estimate: if the solution u of problem (1.1) belongs to $H^s(\Omega)$, $1 \le s \le 2$, there exists a constant c independent of h such that

$$\|u - u_h\|_\alpha \le c \left\{ \sum_{\ell=1}^{L} h_\ell^{2(s-1)} \, \lambda_{\max}(A_\ell) \, \| \,\mathbf{grad}\, u\|^2_{H^{s-1}(\Omega_\ell)^d} \right\}^{\frac{1}{2}}. \tag{2.23}$$

This estimate is optimal since, in contrast to Sect. 3, we are interested in the case that κ is of moderate size.

We now turn to *a posteriori* estimates. For a given element K of $\mathcal{T}_h$, denote by A_K the constant value of A on K and set now:

$$\alpha_K = \lambda_{\max}(A_K), \qquad \alpha_e = \max_{e\subset\partial K} \alpha_K.$$

Define the error estimator η_K as in (2.20) with these definitions of α_K and α_e, and with the function α replaced by the function A. Then the arguments, which led to Theorem 2.9, directly carry over and yield the same *a posteriori* error estimates. The constants c_1 and c_2 now, however, also depend on the parameter κ.

2.f Treatment of piecewise smooth coefficients

We now consider problem (1.1) with $A = \alpha I$ where α is a bounded and piecewise twice continuously differentiable function. The relevant parameters are now

$$\alpha_{\ell,\min} = \inf_{\boldsymbol{x}\in\Omega_\ell} \alpha(\boldsymbol{x}), \quad \alpha_{\ell,\max} = \sup_{\boldsymbol{x}\in\Omega_\ell} \alpha(\boldsymbol{x}), \quad \kappa = \max_{1\leq\ell\leq L} \frac{\alpha_{\ell,\max}}{\alpha_{\ell,\min}}$$

and

$$\alpha_{\min} = \min_{1\leq\ell\leq L} \alpha_{\ell,\min}, \quad \alpha_{\max} = \max_{1\leq\ell\leq L} \alpha_{\ell,\max}.$$

We are interested in the case that the ratio $\alpha_{\max}/\alpha_{\min}$ is large, but that the quantity κ is of moderate size.

The corresponding variational problem is as in Sect. 2.a. It again admits a unique solution which has the same regularity properties as in Proposition 2.2. The corresponding energy norm is given by (2.4).

For the discrete problem we denote by

$$\alpha_K = \fint_K \alpha d\boldsymbol{x}$$

the mean-value of α on K and by α_h the piecewise constant function that takes the value α_K on the element K. The discrete problem then is: *find* u_h *in* X_h *such that*

$$\text{(2.24)} \qquad \begin{aligned} &\forall v_h \in X_h, \\ &\int_\Omega \alpha_h(\boldsymbol{x}) \ \mathbf{grad}\, u_h \cdot \mathbf{grad}\, v_h \, d\boldsymbol{x} = \int_\Omega f(\boldsymbol{x}) v_h(\boldsymbol{x}) \, d\boldsymbol{x}. \end{aligned}$$

Thanks to the Lax-Milgram lemma it admits a unique solution for any data f in $H^{-1}(\Omega)$.

The following observation is crucial for the subsequent analysis. Let u_h be the solution of (2.24) and consider an arbitrary element v_h in X_h. Since **grad** u_h and **grad** v_h are piecewise constant, the definition of α_h implies that

$$\text{(2.25)} \qquad \begin{aligned} &\int_\Omega \alpha(\boldsymbol{x}) \ \mathbf{grad}\, u_h \cdot \mathbf{grad}\, v_h \, d\boldsymbol{x} \\ &\quad = \int_\Omega \alpha_h(\boldsymbol{x}) \ \mathbf{grad}\, u_h \cdot \mathbf{grad}\, v_h \, d\boldsymbol{x}. \end{aligned}$$

Thanks to (2.25) the a priori error analysis of Sect. 2.c directly extends to the present problem.

We now turn to the a posteriori error analysis. Consider first the upper bound on the error. Thanks to (2.25) equations (2.9), (2.10), (2.14), and (2.15) remain unchanged. The arguments of Sect. 2.d, in particular Lemma 2.8, therefore yield the following upper bound on the error

$$\|u-u_h\|_\alpha \le c\left\{\sum_{K\in\mathcal{T}_h} h_K^2\alpha_{K\max}^{-1}\|f_h+\operatorname{div}(\alpha\ \mathbf{grad}\,u_h)\|_{L^2(K)}^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_K} h_e\alpha_{e\max}^{-1}\,\|[\alpha\,\partial_{\boldsymbol{n}_e}u_h]_e\|_{L^2(e)}^2\right\}^{\frac{1}{2}},$$

where now

$$\alpha_{K\max} = \sup_{\boldsymbol{x}\in K}\alpha(\boldsymbol{x}), \qquad \alpha_{e\max} = \max_{K;e\in\mathcal{E}_K}\alpha_{K\max}.$$

The constant c only depends on the shape parameter of $\mathcal{T}_h$, but not on the ratio $\alpha_{\max}/\alpha_{\min}$.

In order to obtain an error estimator which is easy to compute and to be able to derive lower bounds on the error, we introduce the discontinuous, piecewise affine function $\tilde{\alpha}_h$ which, on a given element K, is equal to the L^2-projection of α onto the affine functions on K. Set

$$\tilde{\eta}_K = \left\{h_K^2\alpha_{K\max}^{-1}\,\|f_h+\operatorname{div}(\tilde{\alpha}_h\ \mathbf{grad}\,u_h)\|_{L^2(K)}^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_K} h_e\alpha_{e\max}^{-1}\,\|[\tilde{\alpha}_h\,\partial_{\boldsymbol{n}_e}u_h]_e\|_{L^2(e)}^2\right\}^{\frac{1}{2}}, \tag{2.26}$$

where f_h is *any* finite element approximation of f. We then obtain the upper bound

$$\|u-u_h\|_\alpha \le c\left\{\sum_{K\in\mathcal{T}_h}\tilde{\eta}_K^2 + \sum_{K\in\mathcal{T}_h} h_K^2\alpha_{K\max}^{-1}\|f-f_h\|_{L^2(K)}^2 + \sum_{K\in\mathcal{T}_h} h_K^2\alpha_{K\max}^{-1}\|\operatorname{div}((\alpha-\tilde{\alpha}_h)\ \mathbf{grad}\,u_h)\|_{L^2(K)}^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_K} h_e\alpha_{e\max}^{-1}\,\|[(\alpha-\tilde{\alpha}_h)\,\partial_{\boldsymbol{n}_e}u_h]_e\|_{L^2(e)}^2\right\}^{\frac{1}{2}}. \tag{2.27}$$

When replacing the quantities α_K, α_e, and α by $\alpha_{K\max}$, $\alpha_{e\max}$, and $\tilde{\alpha}_h$, respectively, the arguments of Sect. 2.d immediately yield the following lower bound on the error

$$\tilde{\eta}_K \le c\|u-u_h\|_{\alpha;\omega_K} + c\Bigg\{\sum_{K'\subset\omega_K} h_{K'}^2\alpha_{K'\max}^{-1}\|f-f_h\|_{L^2(K')}^2 + h_{K'}^2\alpha_{K'\max}^{-1}\|\mathrm{div}\,((\alpha-\tilde{\alpha}_h)\ \mathbf{grad}\ u_h)\|_{L^2(K')}^2\Bigg\}^{\frac{1}{2}}$$

$$+c\Bigg\{\sum_{e\in\mathcal{E}_K} h_e\alpha_{e\max}^{-1}\,\|[(\alpha-\tilde{\alpha}_h)\,\partial_{\boldsymbol{n}_e}u_h]_e\|_{L^2(e)}^2\Bigg\}^{\frac{1}{2}}. \tag{2.28}$$

It remains to bound the terms involving $\alpha-\tilde{\alpha}_h$ in (2.27) and (2.28).

Consider first an arbitrary element K. Since u_h is affine on K, we conclude from standard error estimates that

$$\begin{aligned}\|\mathrm{div}\,((\alpha-\tilde{\alpha}_h)\ \mathbf{grad}\ u_h)\|_{L^2(K)} &\le |\alpha-\tilde{\alpha}_h|_{W^{1,\infty}(K)}|u_h|_{H^1(K)}\\ &\le ch_K|\alpha|_{W^{2,\infty}(K)}\,\alpha_{K\min}^{-\frac{1}{2}}\|u_h\|_{\alpha;K},\end{aligned} \tag{2.29}$$

where

$$\alpha_{K\min} = \inf_{\boldsymbol{x}\in K}\alpha(\boldsymbol{x}).$$

Next consider an arbitrary edge resp. face e and denote by K_1 and K_2 the elements adjacent to e. Using a standard inverse estimate we then obtain

$$\begin{aligned}\|[(\alpha-\tilde{\alpha}_h)\,\partial_{\boldsymbol{n}_e}u_h]_e\|_{L^2(e)} &\le |\alpha-\tilde{\alpha}|_{L^\infty(e)}\,\|[\partial_{\boldsymbol{n}_e}u_h]_e\|_{L^2(e)}\\ &\le |\alpha-\tilde{\alpha}|_{L^\infty(e)}\sum_{i=1}^{2} ch_e^{-\frac{1}{2}}|u_h|_{H^1(K_i)}\\ &\le \sum_{i=1}^{2} c'h_e^{-\frac{1}{2}}\alpha_{K_i\min}^{-\frac{1}{2}}h_{K_i}^2|\alpha|_{W^{2,\infty}(K_i)}\|u_h\|_{\alpha;K_i}.\end{aligned}$$

Observing that

$$\alpha_{e\max}^{-\frac{1}{2}} \le \alpha_{K_i\max}^{-\frac{1}{2}}$$

for $i=1,2$ and taking into account the stability estimate

$$\|u_h\|_\alpha \le \alpha_{\min}^{-\frac{1}{2}}\|f\|_{H^{-1}(\Omega)},$$

we thus arrive at the following result:

Theorem 2.11 *Define the estimator $\tilde{\eta}_K$ by (2.26). Assume that Hypothesis 2.7 is satisfied. There exist constants $c_1, \dots, c_4$, which only depend on the shape parameter of $\mathcal{T}_h$, such that the estimates*

$$\|u - u_h\|_\alpha \le c_1 \left\{ \sum_{K \in \mathcal{T}_h} \tilde{\eta}_K^2 + \sum_{K \in \mathcal{T}_h} h_K^2 \alpha_{K\max}^{-1} \|f - f_h\|_{L^2(K)}^2 \right\}^{\frac{1}{2}}$$

$$+ c_2 \alpha_{\min}^{-1} \kappa^{\frac{1}{2}} \max_{K \in \mathcal{T}_h} \left\{ h_K^2 \alpha_{K\max}^{-\frac{1}{2}} |\alpha|_{W^{2,\infty}(K)} \right\} \|f\|_{H^{-1}(\Omega)} \tag{2.30}$$

and

$$\tilde{\eta}_K \le c_3 \left\{ \|u - u_h\|_{\alpha;\omega_K}^2 + \sum_{K' \subset \omega_K} h_{K'}^2 \alpha_{K'\max}^{-1} \|f - f_h\|_{L^2(K')}^2 \right\}^{\frac{1}{2}}$$

$$+ c_4 \alpha_{\min}^{-1} \max_{K' \subset \omega_K} \left\{ h_{K'}^2 \alpha_{K'\max}^{-\frac{1}{2}} |\alpha|_{W^{2,\infty}(K')} \right\} \|f\|_{H^{-1}(\Omega)} \tag{2.31}$$

hold for all finite element approximations f_h of f, all elements K, and all values of $\alpha_{\max}/\alpha_{\min}$.

Remark 2.12 The previous analysis extends to higher order finite elements of order $k \ge 2$. In this case the function α_h must be chosen as the L^2-projection of α onto the space of discontinuous, piecewise polynomials of degree $2k - 2$. Moreover, when establishing (2.29), one has in addition to invoke the inverse estimate

$$\|\Delta u_h\|_{L^2(K)} \le c h_K^{-1} |u_h|_{H^1(K)}.$$

Of course, the case of non scalar, piecewise smooth coefficients can be treated by combining the arguments of Sects. 2.e and 2.f.

3. Anisotropic coefficients

In this section we consider problem (1.1) with a constant, symmetric, positive definite matrix A such that the ratio of its largest eigenvalue $\lambda_{\max}$ to its smallest one $\lambda_{\min}$ is large. We want to derive a priori and a posteriori error estimates which are independent of this ratio.

3.a Variational formulation and regularity

The variational formulation is standard: *find u in $H_0^1(\Omega)$ such that*

$$\forall v \in H_0^1(\Omega), \quad \int_\Omega A \ \mathbf{grad}\, u \cdot \ \mathbf{grad}\, v \, d\boldsymbol{x} = \int_\Omega f(\boldsymbol{x}) v(\boldsymbol{x}) \, d\boldsymbol{x}. \tag{3.1}$$

Thanks to the Lax-Milgram lemma it admits a unique solution for any right-hand side f in $H^{-1}(\Omega)$. The solution enjoys the following regularity:

Proposition 3.1 *Assume that f belongs to $L^2(\Omega)$. There is a real number $s \geq \frac{1}{2}$ such that the unique solution of problem (3.1) belongs to $H^{s+1}(\Omega) \cap H_0^1(\Omega)$. If Ω is convex, s is equal to 1 and*

$$|u|_{H^2(\Omega)} \leq \lambda_{\min}^{-1} \|f\|_{L^2(\Omega)}. \tag{3.2}$$

Proof. Consider the transformation $\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ with $\tilde{\boldsymbol{x}} = \Phi(\boldsymbol{x}) = A^{-\frac{1}{2}}\boldsymbol{x}$ and denote by $\tilde{\Omega}$ the image of Ω under Φ. An elementary calculation shows that u is a solution of (3.1) if and only if $\tilde{u} = u \circ \Phi^{-1}$ is a weak solution of the Laplace equation on $\tilde{\Omega}$ with right-hand side $\tilde{f} = f \circ \Phi^{-1}$ and homogeneous Dirichlet boundary conditions. Noting that $\tilde{\Omega}$ is convex if and only if the same holds for Ω, invoking standard regularity results for the Laplace equation on $\tilde{\Omega}$ [5], and transforming back to Ω, establishes the desired regularity result for u.

In order to prove (3.2), assume first that $\partial\tilde{\Omega}$ is smooth. Since $\tilde{\Omega}$ is assumed to be convex, the curvature of $\partial\tilde{\Omega}$ is positive. An elementary calculation using integration by parts therefore implies that

$$|\tilde{u}|_{H^2(\tilde{\Omega})} \leq \|\tilde{f}\|_{L^2(\tilde{\Omega})}. \tag{3.3}$$

Exhausting a convex polygon or polyhedron by smoothly bounded convex domains, shows that (3.3) also holds for a convex polyhedral domain $\tilde{\Omega}$. Transforming back to Ω and observing that

$$|u|_{H^2(\Omega)} \leq \lambda_{\min}^{-1} \det(A)^{\frac{1}{4}}\, |\tilde{u}|_{H^2(\tilde{\Omega})}, \quad \det(A)^{\frac{1}{4}}\, \|\tilde{f}\|_{L^2(\tilde{\Omega})} = \|f\|_{L^2(\Omega)},$$

we derive (3.2) from (3.3). □

3.b The discrete problem

For the discrete problem we consider a family $(\mathcal{T}_h)_h$ of admissible partitions of Ω into triangles or tetrahedra. *In contrast to Sect. 2 we no longer require that it is regular.* Thus the aspect ratio of the elements is allowed to be large.

With the notation of Sect. 2, the discrete problem then is: *find u_h in X_h such that*

$$\forall v_h \in X_h, \quad \int_\Omega A \;\textbf{grad}\, u_h \cdot \textbf{grad}\, v_h \, d\boldsymbol{x} = \int_\Omega f(\boldsymbol{x}) v_h(\boldsymbol{x})\, d\boldsymbol{x}. \tag{3.4}$$

Thanks to the Lax-Milgram lemma it has a unique solution.

The energy norm is now defined by

$$\|v\|_A = \|A^{\frac{1}{2}} \;\textbf{grad}\, v\|_{L^2(\Omega)^d} = \left\{ \int_\Omega A \;\textbf{grad}\, v \cdot \textbf{grad}\, v \, d\boldsymbol{x} \right\}^{\frac{1}{2}}.$$

3.c A priori error analysis

We define an A-dependent norm $|\cdot|_A$ on $\mathbb{R}^d$ by

$$|\boldsymbol{x}|_A = |A^{-\frac{1}{2}}\boldsymbol{x}|, \tag{3.5}$$

where $|\cdot|$ denotes the standard Euclidean norm on $\mathbb{R}^d$. Given any element K in $\mathcal{T}_h$, we set

$$h_{\mathrm{A},K} = \sup_{\boldsymbol{x},\boldsymbol{y}\in K} |\boldsymbol{x}-\boldsymbol{y}|_A\ , \qquad \rho_{\mathrm{A},K} = 2\sup_{\boldsymbol{x}\in K}\inf_{\boldsymbol{y}\in\partial K} |\boldsymbol{x}-\boldsymbol{y}|_A. \tag{3.6}$$

When A is the identity matrix, these quantities reduce to the standard diameter of K resp. the diameter of the largest ball inscribed into K. Using these definitions we obtain the following a priori error estimate:

Theorem 3.2 *Assume that the solution u of problem (3.1) belongs to $H^s(\Omega)$, $1 \le s \le 2$. There exists a constant c, which neither depends on h nor on the ratio $\lambda_{\max}/\lambda_{\min}$, such that the solution u_h of (3.4) satisfies the following error estimate*

$$\|u-u_h\|_A \le c\left\{\max_{K\in\mathcal{T}_h}\frac{h^2_{\mathrm{A},K}}{\rho_{\mathrm{A},K}}\right\}^{s-1}\lambda_{\max}^{\frac{s}{2}}|u|_{H^s(\Omega)}. \tag{3.7}$$

Proof. From Céa's lemma we conclude that

$$\|u-u_h\|_A \le \inf_{v_h\in X_h}\|u-v_h\|_A.$$

Since

$$\|u\|_A \le \lambda_{\max}^{\frac{1}{2}}|u|_{H^1(\Omega)}$$

this proves (3.7) for $s=1$. Thus it remains to establish (3.7) for the case $s=2$ since the general case then follows by interpolation. To this aim we invoke the transformation Φ which was introduced in the proof of Proposition 3.1. It maps the admissible partition $\mathcal{T}_h$ of Ω into the admissible partition $\tilde{\mathcal{T}}_h = \{\Phi(K)\,;\, K\in\mathcal{T}_h\}$ of $\tilde{\Omega}$. One easily checks that $\tilde{u}_h = u_h\circ\Phi^{-1}$ is the unique solution of the corresponding finite element discretization of the Laplace equation on $\tilde{\Omega}$ with right-hand side $\tilde{f} = f\circ\Phi^{-1}$ and homogeneous Dirichlet boundary conditions. Set $\tilde{u} = u\circ\Phi^{-1}$ and denote by $\tilde{\mathcal{I}}_h$ the nodal interpolation operator corresponding to $\tilde{\mathcal{T}}_h$. Standard interpolation error estimates [1, Theorem 16.1] then imply that

$$\begin{aligned}\inf_{v_h\in X_h}\|u-v_h\|_A &\le \|u-(\tilde{\mathcal{I}}_h\tilde{u})\circ\Phi\|_A \\ &= \det(A)^{\frac{1}{4}}|\tilde{u}-\tilde{\mathcal{I}}_h\tilde{u}|_{H^1(\tilde{\Omega})}\end{aligned}$$

where

$$\inf_{v_h \in X_h} \|u - v_h\|_A \le c \max_{\tilde{K} \in \tilde{\mathcal{T}}_h} \frac{h_{\tilde{K}}^2}{\rho_{\tilde{K}}} \det(A)^{\frac{1}{4}} |\tilde{u}|_{H^2(\tilde{\Omega})}$$
$$\le c \max_{\tilde{K} \in \tilde{\mathcal{T}}_h} \frac{h_{\tilde{K}}^2}{\rho_{\tilde{K}}} \lambda_{\max} |u|_{H^2(\Omega)}.$$

Here $h_{\tilde{K}}$ and $\rho_{\tilde{K}}$ denote the diameter of $\tilde{K}$ and the diameter of the largest ball inscribed into $\tilde{K}$, respectively, both measured with the standard Euclidean norm. Consider an arbitrary element $\tilde{K}$ of $\tilde{\mathcal{T}}_h$. By definition there is an element K of $\mathcal{T}_h$ with $\tilde{K} = \Phi(K)$. Recalling the definitions of Φ and of $|\cdot|_A$, we conclude that

$$h_{\tilde{K}} = \sup_{\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}} \in \tilde{K}} |\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}}| = \sup_{\boldsymbol{x}, \boldsymbol{y} \in K} |A^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{y})| = h_{\mathrm{A},K}$$
$$\rho_{\tilde{K}} = 2 \sup_{\tilde{\boldsymbol{x}} \in \tilde{K}} \inf_{\tilde{\boldsymbol{y}} \in \partial \tilde{K}} |\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}}| = 2 \sup_{\boldsymbol{x} \in K} \inf_{\boldsymbol{y} \in \partial K} |A^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{y})| = \rho_{\mathrm{A},K}.$$

This establishes (3.7) for the case $s = 2$. □

Estimate (3.7) shows that $\max_{K \in \mathcal{T}_h} h_{\mathrm{A},K}/\rho_{\mathrm{A},K}$ should be of order 1 or, equivalently, that the partition $\tilde{\mathcal{T}}_h$ of $\tilde{\Omega}$ should be uniform. Since $\lambda_{\max}/\lambda_{\min}$ is assumed to be large, this means that the partition $\mathcal{T}_h$ must be anisotropic with an anisotropy correctly aligned with A. This is illustrated by the following example.

Example 3.3 We consider problem (1.1) in the unit square with

$$A = \begin{pmatrix} \varepsilon & 0 \\ 0 & 1 \end{pmatrix}$$

and a partition $\mathcal{T}_h$ which consists of right-angled triangles with short sides parallel to the coordinate axes having respective lengths h_x and h_y and longest sides parallel to the line $x = y$. Figure 2 shows the domains Ω and $\tilde{\Omega}$ with the corresponding partitions. An elementary calculation yields

$$h_{\mathrm{A},K} = \sqrt{\varepsilon^{-1} h_x^2 + h_y^2}\,, \quad \rho_{\mathrm{A},K} = \varepsilon^{-\frac{1}{2}} h_x + h_y - \sqrt{\varepsilon^{-1} h_x^2 + h_y^2}.$$

Since the function

$$z \mapsto \frac{\sqrt{1 + z^2}}{1 + z - \sqrt{1 + z^2}}$$

attains its minimum at $z = 1$, this shows that the partition $\mathcal{T}_h$ is optimal if $h_x = \varepsilon^{\frac{1}{2}} h_y$.

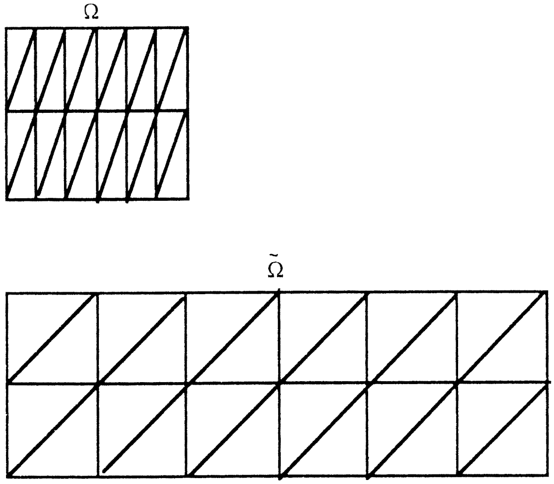

Fig. 2. Domains Ω and $\tilde{\Omega} = \Phi(\Omega)$

3.d A posteriori error analysis

The transformation technique used in the proof of Theorem 3.2 suggests that we may obtain a good a posteriori error estimator for problem (3.4) by correctly transforming an error estimator for the discretization of the Laplace equation on $\tilde{\Omega}$. Since the partitions may be anisotropic, we should look for an estimator which is suitable for isotropic and anisotropic partitions as well. This is satisfied by the estimator introduced in [6, 7].

In order to describe this estimator we need some additional notation. Recall that a $\tilde{}$ always refers to transformed quantities on $\tilde{\Omega}$.

In two dimensions we enumerate the vertices $\tilde{P}_0, \tilde{P}_1, \tilde{P}_2$ of a given triangle $\tilde{K}$ such that:

- $\tilde{P}_0\tilde{P}_1$ is the longest edge,
- $\tilde{P}_0\tilde{P}_2$ is the shortest one.

Denote by:

- $\tilde{\boldsymbol{p}}_1$ the vector $\overrightarrow{\tilde{P}_0\tilde{P}_1}$,
- $\tilde{\boldsymbol{p}}_2$ the vector perpendicular to $\tilde{P}_0\tilde{P}_1$ pointing to $\tilde{P}_2$.

Set $h_{i,\tilde{K}} = |\tilde{\boldsymbol{p}}_i|$ and $h_{\min,\tilde{K}} = \min\{h_{1,\tilde{K}}, h_{2,\tilde{K}}\} = h_{2,\tilde{K}}$.
In three dimensions we enumerate the vertices $\tilde{P}_0, \ldots, \tilde{P}_3$ of a given tetrahedron such that:

- $\tilde{P}_0\tilde{P}_1$ is the longest edge,
- the triangle $\Delta\tilde{P}_0\tilde{P}_1\tilde{P}_2$ has the largest area of the two triangles adjacent to $\tilde{P}_0\tilde{P}_1$,
- $\tilde{P}_0\tilde{P}_2$ is the shortest edge of $\Delta\tilde{P}_0\tilde{P}_1\tilde{P}_2$.

We define three vectors $\tilde{\boldsymbol{p}}_1, \dots, \tilde{\boldsymbol{p}}_3$ as follows:

- $\tilde{\boldsymbol{p}}_1 = \overrightarrow{\tilde{P}_0\tilde{P}_1}$;
- $\tilde{\boldsymbol{p}}_2$ is the vector in the plane $\tilde{P}_0\tilde{P}_1\tilde{P}_2$ which is perpendicular to $\tilde{P}_0\tilde{P}_1$ and which points to $\tilde{P}_2$;
- $\tilde{\boldsymbol{p}}_3$ is the vector which is perpendicular to the plane $\tilde{P}_0\tilde{P}_1\tilde{P}_2$ and which points to $\tilde{P}_3$.

Set $h_{i,\tilde{K}} = |\tilde{\boldsymbol{p}}_i|$ and $h_{\min;\tilde{K}} = \min\{h_{1,\tilde{K}}, h_{2,\tilde{K}}, h_{3,\tilde{K}}\} = h_{3,\tilde{K}}$.
Given any edge resp. face $\tilde{e}$ we denote by $\tilde{K}_{\tilde{e}}$ the element of $\tilde{\mathcal{T}}_h$ which is adjacent to $\tilde{e}$ and which has minimal $h_{\min,\tilde{K}}$. Set

$$h_{\min,\tilde{e}} = h_{\min,\tilde{K}_{\tilde{e}}}\,, \qquad h_{\tilde{e}}^{\perp} = d\,\mathrm{meas}_d(\tilde{K}_{\tilde{e}})/\mathrm{meas}_{d-1}(\tilde{e}).$$

Note that the quantity $h_{\tilde{e}}^{\perp}$ is called $h_{\tilde{e}}$ in [6,7].

Given any element $\tilde{K}$ in $\tilde{\mathcal{T}}_h$ set

$$\tilde{\eta}_{\tilde{K}} = \left\{ h_{\min,\tilde{K}}^2 \|\tilde{f} + \Delta\tilde{u}_h\|_{L^2(\tilde{K})}^2 \right.$$

$$\left. + \frac{1}{2} \sum_{\tilde{e}\in\mathcal{E}_{\tilde{K}}} h_{\min,\tilde{K}}^2 (h_{\tilde{e}}^{\perp})^{-1} \| \left[\partial_{\boldsymbol{n}_{\tilde{e}}} \tilde{u}_h\right]_{\tilde{e}} \|_{L^2(\tilde{e})}^2 \right\}^{\frac{1}{2}}. \tag{3.8}$$

Here, $\tilde{f} = f \circ \Phi^{-1}$ and $\tilde{u}_h = u_h \circ \Phi^{-1}$ are as in the proof of Theorem 3.2. Given any function φ in $H^1(\tilde{\Omega})$ set

$$\tilde{m}_1(\varphi, \tilde{\mathcal{T}}_h) = \left\{ \sum_{\tilde{K}\in\tilde{\mathcal{T}}_h} \sum_{i=1}^{d} h_{\min,\tilde{K}}^{-2} \|\tilde{\boldsymbol{p}}_i \cdot \ \mathbf{grad}\, \varphi\|_{L^2(\tilde{K})}^2 \right\}^{1/2} |\varphi|_{H^1(\tilde{\Omega})}^{-1}.$$

This function is called *matching function* in [6,7]. If the partition $\tilde{\mathcal{T}}_h$ is regular in the sense of Sect. 2, this function is bounded from above by the shape parameter of the partition $\tilde{\mathcal{T}}_h$. Recall that the latter quantity is equal to $\max_{K\in\mathcal{T}_h} h_{\mathrm{A},K}/\rho_{\mathrm{A},K}$ and that this one is of order 1 if and only if $\tilde{\mathcal{T}}_h$ is regular.

With this notation we obtain from [6, Theorem 3.4] the following a posteriori error estimates for the Laplace equation on $\tilde{\Omega}$

$$|\tilde{u} - \tilde{u}_h|_{H^1(\tilde{\Omega})} \le c_1\, \tilde{m}_1(\tilde{u} - \tilde{u}_h, \tilde{\mathcal{T}}_h)$$

$$\left\{ \sum_{\tilde{K}\in\tilde{\mathcal{T}}_h} \tilde{\eta}_{\tilde{K}}^2 + h_{\min,\tilde{K}}^2 \|\tilde{f} - \tilde{f}_h\|_{L^2(\tilde{K})}^2 \right\}^{\frac{1}{2}},$$

$$\tilde{\eta}_{\tilde{K}} \le c_2 \left\{ |\tilde{u} - \tilde{u}_h|^2_{H^1(\omega_{\tilde{K}})} + \sum_{\tilde{K}' \subset \omega_{\tilde{K}}} h^2_{\min,\tilde{K}'} \|\tilde{f} - \tilde{f}_h\|^2_{L^2(\tilde{K}')} \right\}^{\frac{1}{2}}. \tag{3.9}$$

Here, $\tilde{f}_h$ is *any* finite element approximation of $\tilde{f}$ corresponding to $\tilde{\mathcal{T}}_h$. The constants c_1 and c_2 neither depend on h nor on any shape parameter of $\tilde{\mathcal{T}}_h$.

Estimate (3.9) suggests that the quantities $\tilde{\eta}_{\tilde{K}}$ may be well suited for our purposes. Recalling that $\|u - u_h\|_A = \det(A)^{\frac{1}{4}} |\tilde{u} - \tilde{u}_h|_{H^1(\tilde{\Omega})}$ we therefore define for any element K in $\mathcal{T}_h$

$$\eta_K = \det(A)^{\frac{1}{4}}\, \tilde{\eta}_{\Phi(K)}.$$

Next, we want to express η_K by quantities which only refer to the element K and which do not resort to the transformation Φ.

We start with the weights. Denote by $P_0, \dots P_d$ the vertices of a given element K such that they are the pre-images of the vertices $\tilde{P}_0, \dots, \tilde{P}_d$ which correspond to $\tilde{K} = \Phi(K)$ and which are defined as above. In two dimensions, we immediately conclude that

$$\begin{aligned} |P_0 - P_1|_A &= \max_{0 \le i < j \le 2} |P_i - P_j|_A, \\ h_{\min,\Phi(K)} = h_{\mathrm{A},\min,K} &= \inf_{\boldsymbol{y} \in P_0P_1} |P_2 - \boldsymbol{y}|_A. \end{aligned} \tag{3.10}$$

In three dimensions, we observe that, among two faces of a tetrahedron sharing an edge, that face has maximal area which has the maximal height above the common edge. Hence, we conclude that

$$\begin{aligned} |P_0 - P_1|_A &= \max_{0 \le i < j \le 3} |P_i - P_j|_A, \\ \inf_{\boldsymbol{y} \in P_0P_1} |P_2 - \boldsymbol{y}|_A &= \max_{2 \le i \le 3} \inf_{\boldsymbol{y} \in P_0P_1} |P_i - \boldsymbol{y}|_A, \\ h_{\min,\Phi(K)} = h_{\mathrm{A},\min,K} &= \inf_{\boldsymbol{y} \in \Delta P_0P_1P_2} |P_3 - \boldsymbol{y}|_A. \end{aligned} \tag{3.11}$$

Given an edge resp. face e in $\mathcal{E}_h$ denote by K_e the element adjacent to e which has minimal $h_{\mathrm{A},\min,K}$ and set

$$h_e^{\perp} = d\, \mathrm{meas}_d(K_e)/\mathrm{meas}_{d-1}(e). \tag{3.12}$$

Note that $h_e^{\perp}$ is the height of K_e above e measured in the *Euclidean* norm and that it is *not* $h_{\tilde{e}}^{\perp}$ expressed in quantities referring to e.

Next, we consider the element residuals. From the transformation rule we immediately obtain

$$\begin{aligned} &\det(A)^{\frac{1}{4}} h_{\min,\Phi(K)} \|\tilde{f}_h + \Delta \tilde{u}_h\|_{L^2(\Phi(K))} \\ &\quad = h_{\mathrm{A},\min,K} \|f_h + \mathrm{div}\,(A\ \mathbf{grad}\ u_h)\|_{L^2(K)}. \end{aligned}$$

Now we turn to the edge resp. face residuals. We need a technical lemma.

Lemma 3.4 *The following identity holds for any edge resp. face e in $\mathcal{E}_h$ and $\tilde{e} = \Phi(e)$:*

$$\partial_{\boldsymbol{n}_{\tilde{e}}} \tilde{u}_h = \frac{\mathrm{meas}_{d-1}(e)}{\mathrm{meas}_{d-1}(\tilde{e})} \det(A)^{-\frac{1}{2}} \, \boldsymbol{n}_e \cdot A \ \mathbf{grad}\, u_h. \tag{3.13}$$

Proof. We treat separately the cases $d = 2$ and $d = 3$.
In the case $d = 2$, denote by $\vec{e}$ a vector that has the same length as e, is parallel to e, and satisfies $\det(\boldsymbol{n}_e\,,\, \vec{e}) > 0$. The vector $\vec{\tilde{e}}$ is defined correspondingly with $\tilde{e}$ instead of e. Set

$$P = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Since

$$PBP^T = \det(B)\, B^{-1}$$

for any regular, symmetric matrix B of order 2, we obtain

$$\begin{aligned}
\partial_{\boldsymbol{n}_{\tilde{e}}} \tilde{u}_h &= \boldsymbol{n}_{\tilde{e}} \cdot \ \mathbf{grad}\, \tilde{u}_h \\
&= \mathrm{meas}_1(\tilde{e})^{-1} P\vec{\tilde{e}} \cdot A^{\frac{1}{2}} \ \mathbf{grad}\, u_h \\
&= \mathrm{meas}_1(\tilde{e})^{-1} P A^{-\frac{1}{2}} \vec{e} \cdot A^{\frac{1}{2}} \ \mathbf{grad}\, u_h \\
&= \mathrm{meas}_1(\tilde{e})^{-1} P A^{-\frac{1}{2}} P^T P \vec{e} \cdot A^{\frac{1}{2}} \ \mathbf{grad}\, u_h \\
&= \mathrm{meas}_1(\tilde{e})^{-1} \mathrm{meas}_1(e) \det(A)^{-\frac{1}{2}} \, A^{\frac{1}{2}} \boldsymbol{n}_e \cdot A^{\frac{1}{2}} \ \mathbf{grad}\, u_h \\
&= \frac{\mathrm{meas}_1(e)}{\mathrm{meas}_1(\tilde{e})} \det(A)^{-\frac{1}{2}} \, \boldsymbol{n}_e \cdot A \ \mathbf{grad}\, u_h.
\end{aligned}$$

In the case $d = 3$, we choose two different edges α and β of the face e. Denote by $\vec{\alpha}$ and $\vec{\beta}$ two vectors that are parallel to α and β, have the same length as α and β, and satisfy $\det(\vec{\alpha}\,,\, \vec{\beta}\,,\, \boldsymbol{n}_e) > 0$. Set $\tilde{\alpha} = \Phi(\alpha)$, $\tilde{\beta} = \Phi(\beta)$, and denote by $\vec{\tilde{\alpha}}$ and $\vec{\tilde{\beta}}$ the corresponding vectors. Denoting by $\times$ the vector product in $\mathbb{R}^3$, we then get

$$\begin{aligned}
\partial_{\boldsymbol{n}_{\tilde{e}}} \tilde{u}_h &= \boldsymbol{n}_{\tilde{e}} \cdot \ \mathbf{grad}\, \tilde{u}_h \\
&= \mathrm{meas}_2 \left(\vec{\tilde{\alpha}} \times \vec{\tilde{\beta}} \right)^{-1} \left(\vec{\tilde{\alpha}} \times \vec{\tilde{\beta}} \right) \cdot A^{\frac{1}{2}} \ \mathbf{grad}\, u_h \\
&= \mathrm{meas}_2 \left(\vec{\tilde{\alpha}} \times \vec{\tilde{\beta}} \right)^{-1} \left(A^{-\frac{1}{2}} \vec{\alpha} \times A^{-\frac{1}{2}} \vec{\beta} \right) \cdot A^{-\frac{1}{2}} \, A \ \mathbf{grad}\, u_h.
\end{aligned}$$

Denote by $a^1, \ldots, a^3$ the columns of $A^{-\frac{1}{2}}$ and by $e^1, \ldots, e^3$ the standard unit vectors of $\mathbb{R}^3$. Since $A^{-\frac{1}{2}}$ is symmetric and since the mapping

$$\boldsymbol{x}\,,\, \boldsymbol{y}\,,\, \boldsymbol{z} \mapsto \boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z}) = \det(\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z})$$

is an alternating trilinear form on $\mathbb{R}^3$, we conclude that

$$A^{-\frac{1}{2}}\left(A^{-\frac{1}{2}}\vec{\alpha}\times A^{-\frac{1}{2}}\vec{\beta}\right)=\sum_{1\le i,j,k\le 3}e^i\alpha_j\beta_k a^i\cdot(a^j\times a^k)$$
$$=\det(A)^{-\frac{1}{2}}\vec{\alpha}\times\vec{\beta}.$$

We therefore obtain

$$\partial_{\boldsymbol{n}_{\tilde{e}}}\tilde{u}_h=\mathrm{meas}_2(\vec{\tilde{\alpha}}\times\vec{\tilde{\beta}})^{-1}\det(A)^{-\frac{1}{2}}(\vec{\alpha}\times\vec{\beta})\cdot A\ \mathbf{grad}\ u_h$$
$$=\frac{\mathrm{meas}_2(\vec{\alpha}\times\vec{\beta})}{\mathrm{meas}_2(\vec{\tilde{\alpha}}\times\vec{\tilde{\beta}})}\det(A)^{-\frac{1}{2}}\boldsymbol{n}_e\cdot A\ \mathbf{grad}\ u_h$$
$$=\frac{\mathrm{meas}_2(e)}{\mathrm{meas}_2(\tilde{e})}\det(A)^{-\frac{1}{2}}\boldsymbol{n}_e\cdot A\ \mathbf{grad}\ u_h,$$

which concludes the proof. □

With the help of (3.13) we may rewrite the edge resp. face residuals as follows

$$\det(A)^{\frac{1}{4}}h_{\min,\Phi(K)}\left(h^{\perp}_{\Phi(e)}\right)^{-\frac{1}{2}}\left\|\left[\partial_{\boldsymbol{n}_{\Phi(e)}}\tilde{u}_h\right]_{\Phi(e)}\right\|_{L^2(\Phi(e))}$$
$$=h_{\mathrm{A,min},K}\left\{\frac{\mathrm{meas}_{d-1}(e)}{h^{\perp}_{\Phi(e)}\,\mathrm{meas}_{d-1}(\Phi(e))\det(A)^{\frac{1}{2}}}\right\}^{\frac{1}{2}}$$
$$\|[\boldsymbol{n}_e\cdot A\ \mathbf{grad}\ u_h]_e\|_{L^2(e)}\,.$$

Since

$$d\,\mathrm{meas}_d(\Phi(K))=h^{\perp}_{\Phi(e)}\,\mathrm{meas}_{d-1}(\Phi(e)),$$
$$d\,\mathrm{meas}_d(K)=h^{\perp}_e\,\mathrm{meas}_{d-1}(e),$$
$$\mathrm{meas}_d(K)=\det(A)^{\frac{1}{2}}\mathrm{meas}_d(\Phi(K)),$$

this yields the identity

$$\det(A)^{\frac{1}{4}}h_{\min,\Phi(K)}\left(h^{\perp}_{\Phi(e)}\right)^{-\frac{1}{2}}\left\|\left[\partial_{\boldsymbol{n}_{\Phi(e)}}\tilde{u}_h\right]_{\Phi(e)}\right\|_{L^2(\Phi(e))}$$
$$=h_{\mathrm{A,min},K}(h^{\perp}_e)^{-\frac{1}{2}}\ \|[\boldsymbol{n}_e\cdot A\ \mathbf{grad}\ u_h]_e\|_{L^2(e)}\,.$$

Finally, we consider the matching function. Denote by $\boldsymbol{p}_{K,1},\dots,\boldsymbol{p}_{K,d}$ the pre-images of the vectors $\tilde{\boldsymbol{p}}_1,\dots,\tilde{\boldsymbol{p}}_d$. Without resorting to the transformation Φ, these can be computed as follows:

– $\boldsymbol{p}_{K,1}$ is parallel to P_0P_1 and points to P_1;
– $\boldsymbol{p}_{K,2}$ lies in the plane $P_0P_1P_2$, is A^{-1}-orthogonal to $\boldsymbol{p}_{K,1}$ and points to P_2;
– if $d = 3$: $\boldsymbol{p}_{K,3}$ is A^{-1}-orthogonal to the plane $P_0P_1P_2$ and points to P_3.

Here, A^{-1}-orthogonality of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ means that $\boldsymbol{x} \cdot A^{-1}\,\boldsymbol{y} = 0$. From these properties we conclude that

$$\tilde{\boldsymbol{p}}_i \cdot \ \mathbf{grad}\ \tilde{u}_h = \boldsymbol{p}_{K,i} \cdot \ \mathbf{grad}\ u_h.$$

This yields

$$\begin{aligned} &\tilde{m}_1(\tilde{u} - \tilde{u}_h, \tilde{\mathcal{T}}_h) \\ &= \left\{ \sum_{K\in\mathcal{T}_h} \sum_{i=1}^{d} h_{\mathrm{A,min},K}^{-2} \|\boldsymbol{p}_{K,i} \cdot \ \mathbf{grad}\ (u - u_h)\|_{L^2(K)}^2 \right\}^{\frac{1}{2}} \|u - u_h\|_A^{-1}. \end{aligned} \tag{3.14}$$

Since the vectors $\boldsymbol{p}_{K,i}$ are mutually A^{-1}-orthogonal and satisfy $|\boldsymbol{p}_{K,i}|_A \le h_{\mathrm{A,max},K}$ with

$$h_{\mathrm{A,max},K} = \max_{\boldsymbol{x},\boldsymbol{y}\in K} |\boldsymbol{x} - \boldsymbol{y}|_A, \tag{3.15}$$

the right-hand side of (3.14) can be bounded by

$$\max_{K\in\mathcal{T}_h} h_{\mathrm{A,max},K}/h_{\mathrm{A,min},K}.$$

Summarizing all these results, we arrive at the following a posteriori error estimate:

Theorem 3.5 *Define the quantities* $h_{\mathrm{A,min},K}$ *and* $h_e^{\perp}$ *as in* (3.10) – (3.12) *and set*

$$\begin{aligned} \eta_K = \Bigg\{ & h_{\mathrm{A,min},K}^2 \|f_h + \mathrm{div}\,(A\ \mathbf{grad}\ u_h)\|_{L^2(K)}^2 \\ & + \frac{1}{2} \sum_{e\in\mathcal{E}_K} h_{\mathrm{A,min},K}^2 (h_e^{\perp})^{-1} \|\,[\boldsymbol{n}_e \cdot A\ \mathbf{grad}\ u_h]_e\,\|_{L^2(e)}^2 \Bigg\}^{\frac{1}{2}}. \end{aligned} \tag{3.16}$$

Then the following a posteriori error estimates hold

$$\begin{aligned} \|u - u_h\|_A &\le c_1\, m_1(u - u_h, \mathcal{T}_h) \\ & \left\{ \sum_{K\in\mathcal{T}_h} \left[\eta_K^2 + h_{\mathrm{A,min},K}^2 \|f - f_h\|_{L^2(K)}^2 \right] \right\}^{\frac{1}{2}} \end{aligned} \tag{3.17}$$

and

$$\eta_K \leq c_2 \left\{ \|u - u_h\|^2_{A;\omega_K} + \sum_{K' \subset \omega_K} h^2_{\mathrm{A,min},K'} \|f - f_h\|^2_{L^2(K')} \right\}^{\frac{1}{2}} . \tag{3.18}$$

Here, f_h is any finite element approximation of f corresponding to $\mathcal{T}_h$. The constants c_1 and c_2 neither depend on h, nor on any shape parameter of $\mathcal{T}_h$, nor on the ratio $\lambda_{\max}/\lambda_{\min}$. The term $m_1(u - u_h, \mathcal{T}_h)$, given by

$$m_1(u - u_h, \mathcal{T}_h) = \left\{ \sum_{K \in \mathcal{T}_h} \sum_{i=1}^{d} h^{-2}_{\mathrm{A,min},K} \|\boldsymbol{p}_{K,i} \cdot \textbf{grad}\ (u - u_h)\|^2_{L^2(K)} \right\}^{\frac{1}{2}} \|u - u_h\|^{-1}_A, \tag{3.19}$$

is bounded from above by $\max_{K \in \mathcal{T}_h} h_{\mathrm{A,max},K}/h_{\mathrm{A,min},K}$ with $h_{\mathrm{A,max},K}$ defined in (3.15).

Estimates (3.17) and (3.18) are fully optimal, in the sense that the constants are independent of $\lambda_{\max}/\lambda_{\min}$ for an appropriate but not standard choice of the family of triangulations.

References

1. Ciarlet, P.G. (1991): Basic Error Estimates for Elliptic Problems. In: Ciarlet, P.G., Lions, J.-L. (eds.) Handbook of Numerical Analysis, vol. II, North Holland
2. Clain, S., Touzani, R. (1997): Solution of a two-dimensional stationary induction heating problem without boundedness of the coefficients. Modél. Math. et Anal. Numér. **31**, 845–870
3. Clément, P. (1975): Approximation by finite element functions using local regularization. R.A.I.R.O. Anal. Numér. **9**, 77–84
4. Dauge, M. (1992): Neumann and mixed problems on curvilinear polyhedra. Integr. Equat. and Oper. Th. **15**, 227–261
5. Grisvard, P. (1985): Elliptic Problems in Nonsmooth Domains, Pitman
6. Kunert, G. (1998): A posteriori error estimation for anisotropic tetrahedral and triangular finite element meshes. Thesis, TU-Chemnitz
7. Kunert, G. A residual error estimator for the finite element method on anisotropic tetrahedral or triangular meshes. In preparation
8. Lewandowski, R. (1997): Analyse mathématique et océanographie. Collection Recherches en Mathématiques Appliquées, Masson
9. Lions, J.-L., Magenes, E. (1968): Problèmes aux limites non homogènes et applications, vol. 1, Dunod
10. Meyers, N.G. (1963): An L^p-estimate for the gradient of solutions of second order elliptic divergence equations. Ann. Sc. Norm. Sup. Pisa **17**, 189–206
11. Verfürth, R. (1996): A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques. Wiley & Teubner (1996)
12. Verfürth, R. (1998): Robust a posteriori error estimators for a singularly perturbed reaction-diffusion equation. Numer. Math. **78**, 479–493
13. Verfürth, R. (1999): Error estimates for some quasi-interpolation operators. Modél. Math. et Anal. Numér. **33**, 695–713