Chapter 6

Construction of Adaptive Hierarchical Meshes

In the preceding chapters we discussed discretizations on uniform or at least shape regular meshes. In many practically important application problems, however, strongly localized phenomena arise, which can be treated more efficiently on problem adapted hierarchical meshes. Such meshes can be constructed by means of a posteriori estimators of the approximation errors in finite elements.

In many cases, even more important is the determination of the distribution of the actual local discretization errors to assure a sufficient accuracy of the numerical solution.[1] This gives a posteriori error estimators a crucial role in the termination criteria for mesh refinement processes.

In Sections 6.1 and 6.2 we work out the prevalent error estimators in a unified theoretical frame in some detail. On this basis, a subtle strategy of equilibration of discretization errors (Section 6.2.1) supplies local mesh markings, which, by heuristic refinement strategies, are translated into hierarchical meshes (see Section 6.2.2). The thus constructed adaptive meshes can then be combined both with (direct or iterative) solvers of linear algebra and with multigrid methods. For a model adaptive mesh refinement we give an elementary convergence proof (Section 6.3.1). In the subsequent Section 6.3.2 we illustrate how corner singularities are quasi-transformed away by adaptive hierarchical meshes. In the final Section 6.4 we document at a (quadratic) eigenvalue problem for a plasmon-polariton waveguide how far the presented concepts carry over into rough practice.

6.1 A Posteriori Error Estimators

An important algorithmic piece of *adaptive* PDE solvers are error estimators, by means of which *problem adapted hierarchical meshes* can be constructed – a topic that we will treat in the subsequent Section 6.2. In Section 4.4 we presented an approximation theory that described the global discretization error (in the energy norm $\|\cdot\|_a$ or the L^2-Norm $\|\cdot\|_{L^2}$) only on *uniform* or *quasi-uniform* meshes, respectively, via the order w.r.t. a characterizing mesh size h. For realistic problems from engineering and biosciences, however, *problem adapted*, which usually means *nonuniform* meshes are

[1] In 1991, the swimming fundament of the oil platform *Sleipner A* leaked and sank. The material broke at a strongly strained point with large mechanical stress, which had not been correctly computed by an FE simulation on a locally too coarse mesh. In the absence of an error estimator the wrong solution had been accepted and the design error had not been corrected. The total damage caused was about 700 million US Dollars [129].

of crucial importance. In this case, a characterization by a single global mesh size h is no longer reasonable; therefore, as in Section 4.3.1, we mean by h some local mesh size, but use an index h in $u_h \in S_h$ as a characterization of the discrete solution.

The error estimators treated in the following are valid for general elliptic boundary value problems. For ease of presentation, however, we pick the example of the Poisson equation with homogeneous Robin boundary conditions

$$-\operatorname{div}(\sigma \nabla u) = f, \quad n^T \sigma \nabla u + \alpha u = 0,$$

in weak formulation written as

$$a(u, v) = \langle f, v \rangle, \quad v \in H^1(\Omega), \tag{6.1}$$

and in discrete weak formulation as

$$a(u_h, v_h) = \langle f, v_h \rangle, \quad v_h \in S_h \subset H^1(\Omega). \tag{6.2}$$

Global Error Estimators. In actual computation we will only have access to an approximation $\tilde{u}_h$ in lieu of u_h. Consequently, we will have to distinguish the following errors:

- *algebraic error* $\delta_h = \|u_h - \tilde{u}_h\|_a$;
- *discretization error* $\epsilon_h = \|e_h\|_a$ with $e_h = u - u_h$;
- *approximation error* $\tilde{\epsilon}_h = \|\tilde{e}_h\|_a$ with $\tilde{e}_h = u - \tilde{u}_h$.

A usual accuracy check for the numerical solution of the Poisson problem will have the form

$$\tilde{\epsilon}_h \leq \mathrm{TOL} \tag{6.3}$$

for given error tolerance TOL. For this special purpose it is sufficient, to compute $\tilde{\epsilon}_h$ approximately, i.e., to *estimate* it approximately. From Theorem 4.3 above we have $a(u_h - \tilde{u}_h, e_h) = 0$, from which we directly obtain

$$\tilde{\epsilon}_h^2 = \epsilon_h^2 + \delta_h^2. \tag{6.4}$$

An efficient possibility to iteratively estimate the *algebraic* error δ_h within a PCG-method was already given in Section 5.3.3. Therefore, we restrict our subsequent attention to an estimation of the discretization error ϵ_h.

In a first seminal paper from 1972 I. Babuška and A. K. Aziz [9] introduced a what is now a generally accepted classification of estimators $[\epsilon_h]$ of discretization errors ϵ_h:

Definition 6.1. An *error estimator* $[\epsilon_h]$ for ϵ_h is called

- *reliable*, if, with a constant $\kappa_1 \geq 1$, one has

$$\epsilon_h \leq \kappa_1 \, [\epsilon_h]. \tag{6.5}$$

If κ_1 is known, then the nonevaluable test (6.3) (with $\delta_h = 0$) can be substituted

by the *evaluable check*

$$\kappa_1 \, [\epsilon_h] \leq \mathrm{TOL}, \tag{6.6}$$

from which then, with (6.5), the check (6.3) follows directly;

- *efficient*, if, with some $\kappa_2 \geq 1$, one additionally has

$$[\epsilon_h] \leq \kappa_2 \, \epsilon_h,$$

which, together with (6.5), can then be combined to

$$\frac{1}{\kappa_1} \, \epsilon_h \leq [\epsilon_h] \leq \kappa_2 \, \epsilon_h. \tag{6.7}$$

The product $\kappa_1 \kappa_2 \geq 1$ is also called *efficiency span*, since its deviation from the value 1 is a simple measure of the efficiency of an error estimator;

- *asymptotically exact*, if one additionally has

$$\frac{1}{\kappa_1} \, \epsilon_h \leq [\epsilon_h] \leq \kappa_2 \, \epsilon_h \quad \text{with } \kappa_{1,2} \to 1 \text{ for } h \to 0.$$

Localization of Global Error Estimators. Apart from the obvious purpose of determining the approximation quality of solutions u_h, error estimators can also be used for the construction of problem adapted meshes. This is possible in elliptic problems, as their Green's function (A.13) is strongly localized, i.e., local perturbations of the problem essentially remain local. Vice versa, on the same basis, global errors can be composed from local components, which we will elaborate in the following.

In finite element methods over a triangulation $\mathcal{T}$ there is a natural way to split the global error into its local components via the individual elements $T \in \mathcal{T}$ according to

$$\epsilon_h^2 = \sum_{T \in \mathcal{T}} \epsilon_h(T)^2 = \sum_{T \in \mathcal{T}} a(u - u_h, u - u_h)|_T. \tag{6.8}$$

The exact computation of the local error components $\epsilon_h(T)$ would, similar to the global case, require an amount comparable to the one for the solution of the whole problem. For this reason, one tries to construct local error estimators that are not only "efficient, reliable, and asymptotically exact", but also "cheap": Starting from a given approximation $\tilde{u}_h$ on a given finite element mesh $\mathcal{T}$ one acquires local *a posteriori* error estimates, i.e., error estimates that are locally evaluated *after* the computation of an approximation – in contrast to *a priori* error estimates that are available *before* the computation of an approximation, such as the results in Section 4.4.1.

An alternative option for a localization $\epsilon_h(T)$ to elements $T \in \mathcal{T}$ is via a localization on edges $E \in \mathcal{E}$ (or faces $F \in \mathcal{F}$, respectively) in the form of

$$\epsilon_h^2 = \sum_{E \in \mathcal{E}} \epsilon_h(E)^2$$

is equally useful. However, an interpretation of the kind $\epsilon_h(T)^2 = a(e_h, e_h)|_T$ is then no longer feasible.

Definition 6.2. Local error estimators $[\epsilon_h(T)]$ associated with the local errors $\epsilon_h(T)$ are called *error indicators*, if for two h-independent constants c_1, c_2 the relation

$$\frac{1}{c_1}\,\epsilon_h(T) \le [\epsilon_h(T)] \le c_2\,\epsilon_h(T), \quad T \in \mathcal{T}.$$

holds. Obviously, this is the localization of the concept of *efficiency* for global error estimators (see (6.7)).

Error Representation. On some occasions it is useful not only to have an estimate of the *norm* ϵ_h of the error, but also an explicit approximation $[e_h]$ of the discretization error e_h itself. The improved approximation $u_h + e_h$ can, in turn, be used to construct a goal oriented error estimator (see the subsequent Section 6.1.5).

Basic Formula. In order to estimate the discretization $\epsilon_h = \|e_h\|_a$ we resort to the representation via the dual norm. From the Cauchy–Schwarz inequality

$$a(e_h, v) \le \|e_h\|_a \|v\|_a \quad \text{for all } v \in H^1(\Omega)$$

we immediately derive the relation

$$\epsilon_h = \sup_{v \in H^1(\Omega), v \neq 0} \frac{a(e_h, v)}{\|v\|_a}. \tag{6.9}$$

The supremum is naturally attained at $v = e_h$, i.e., the exact determination of the supremum would again require the same amount of computation as the exact solution of the original problem (6.1). By some suitable choice of v, however, evaluable bounds of the error can be constructed. Note that the above error representation is localized for now.

6.1.1 Residual Based Error Estimator

This rather popular type of error estimator tries to gain a *global upper bound* for all v in (6.9). The derivation of this bound is a direct consequence of Corollary 4.1 above. There we had introduced the flow condition (4.13); it states that the normal components of the flow for the exact solution u are continuous at the internal faces (or edges) and thus their jumps $[\![n^T \sigma \nabla u]\!]_\Gamma$ vanish. As illustrated in Figure 4.10, this does not hold for the discrete solution u_h. Therefore the jumps of the discrete normal flows will constitute an essential part of the error.

Localization. This is achieved in two steps: as u_h is differentiable sufficiently often only on isolated elements of the triangulation, we first apply Green's Theorem (A.7)

to individual elements $T \in \mathcal{T}$ and sum up as follows:

$$a(e_h, v) = \sum_{T\in\mathcal{T}} a_T(e_h, v) \tag{6.10}$$
$$= \sum_{T\in\mathcal{T}} \left[-\int_T (\operatorname{div}(\sigma\nabla u_h) + f)v \,\mathrm{d}x + \int_{\partial T} v\, n^T \sigma \nabla e_h \,\mathrm{d}s \right] + \int_{\partial\Omega} \alpha v e_h \,\mathrm{d}s.$$

After that we sort the integrals w.r.t. ∂T to obtain a sum of integrals over faces $F \in \mathcal{F}$ (or edges $E \in \mathcal{E}$). Let T_1, T_2 denote two arbitrarily chosen neighboring elements with common face $F \in \mathcal{F}$. Then we arrive at local terms of the form

$$-\int_F v n^T (\sigma\nabla e_h|_{T_1} - \sigma\nabla e_h|_{T_2}) \,\mathrm{d}s = \int_F v [\![n^T \sigma \nabla u_h]\!]_F \,\mathrm{d}s$$

for internal faces. For boundary faces $F \subset \partial\Omega$ we obtain integrals of the form

$$\int_F v(n^T\sigma\nabla u_h + \alpha u_h) \,\mathrm{d}s,$$

so that we may, in order to unify the notation, define "boundary jumps" according to

$$[\![n^T \sigma \nabla u_h]\!]_F = n^T\sigma\nabla u_h + \alpha u_h.$$

Summarizing, we thus obtain the localized error representation

$$a(e_h, v) = \sum_{T\in\mathcal{T}} \int_T (\operatorname{div}(\sigma\nabla u_h) + f)v \,\mathrm{d}x + \sum_{F\in\mathcal{F}} \int_F v\, [\![n^T \sigma \nabla u_h]\!]_F \,\mathrm{d}s.$$

Applying the Cauchy–Schwarz inequality to all local integrals yields the estimate

$$a(e_h, v) \le \sum_{T\in\mathcal{T}} \| \operatorname{div}(\sigma\nabla u_h) + f \|_{L^2(T)} \|v\|_{L^2(T)} + \sum_{F\in\mathcal{F}} \| [\![n^T \sigma \nabla u_h]\!]_F \|_{L^2(F)} \|v\|_{L^2(F)}.$$

Due to (4.14) we have $a(e_h, v) = a(e_h, v - v_h)$ for all $v_h \in S_h$, so that the term $\|v\|$ above can be replaced by $\|v - v_h\|$. For the purpose of theory we now choose a v_h such that the contributions $\|v - v_h\|_{L^2(T)}$ and $\|v - v_h\|_{L^2(F)}$ are as small as possible. In fact, for each $v \in H^1(\Omega)$ there exists an (even computable) quasi-interpolant $v_h \in S_h$ with

$$\sum_{T\in\mathcal{T}} h_T^{-2} \|v - v_h\|^2_{L^2(T)} \le c^2 |v|^2_{H^1(\Omega)} \quad \text{and} \quad \sum_{F\in\mathcal{F}} h_F^{-1} \|v - v_h\|^2_{L^2(F)} \le c^2 |v|^2_{H^1(\Omega)}, \tag{6.11}$$

where h_T and h_F are again the diameters of T and F and c a constant depending only on the shape regularity of the triangulation (see, e.g., [186]). In view of this we get

$$a(e_h, v) = a(e_h, v - v_h)$$
$$\le \sum_{T\in\mathcal{T}} h_T \| \operatorname{div}(\sigma\nabla u_h) + f \|_{L^2(T)}\, h_T^{-1} \|v - v_h\|_{L^2(T)} + \sum_{F\in\mathcal{F}} h_F^{1/2} \| [\![n^T \sigma \nabla u_h]\!]_F \|_{L^2(F)}\, h_F^{-1/2} \|v - v_h\|_{L^2(F)}.$$

If we define local error estimators over T, F as

$$\eta_T := h_T \| \operatorname{div}(\sigma \nabla u_h) + f \|_{L^2(T)}, \quad \eta_F := h_F^{1/2} \| [\![n^T \sigma \nabla u_h]\!]_F \|_{L^2(F)},$$

then a second application of the Cauchy–Schwarz inequality, this time in finite dimensions, eventually leads to

$$\begin{aligned} a(e_h, v) &\le \Big(\sum_{T \in \mathcal{T}} \eta_T^2 \Big)^{\frac{1}{2}} \Big(\sum_{T \in \mathcal{T}} h_T^{-2} \|v - v_h\|_{L^2(T)}^2 \Big)^{\frac{1}{2}} \\ &\quad + \Big(\sum_{F \in \mathcal{F}} \eta_F^2 \Big)^{\frac{1}{2}} \Big(\sum_{F \in \mathcal{F}} h_F^{-1} \|v - v_h\|_{L^2(E)}^2 \Big)^{\frac{1}{2}} \\ &\le 2c |v|_{H^1(\Omega)} \Big(\sum_{T \in \mathcal{T}} \eta_T^2 + \sum_{F \in \mathcal{F}} \eta_F^2 \Big)^{\frac{1}{2}}. \end{aligned}$$

Due to the positivity of A there exists an $\alpha > 0$ such that $|v|_{H^1(\Omega)} \le \alpha \|v\|_a$. In the above inequality we thus replace $|v|_{H^1(\Omega)}$ by $\|v\|_a$ and swallow the constant α in the generic constant c. After this preparation, dividing by $\|v\|_a$ we can return to the basic formula (6.9) and eventually arrive at the classical error estimator due to I. Babuška and A. Miller [11]:

BM Error Estimator

$$[\epsilon_h] = 2c \Big(\sum_{T \in \mathcal{T}} \eta_T^2 + \sum_{F \in \mathcal{F}} \eta_F^2 \Big)^{\frac{1}{2}} \ge \epsilon_h. \tag{6.12}$$

By derivation, this is a global upper bound of the actual error. In order to algorithmically realize an elementwise localization of this error estimator in the sense of (6.8), the localized jump contributions at the faces $F \in \mathcal{F}$ must be redistributed to the elements $T \in \mathcal{T}$. One therefore defines

$$[\epsilon_h(T)]^2 = 4c^2 \alpha^2 \Big(\eta_T^2 + \sum_{F \in \mathcal{F}, F \subset \partial T} \alpha_{T,F}\, \eta_F^2 \Big), \tag{6.13}$$

where $\alpha_{T_1,F} + \alpha_{T_2,F} = 1$ for two elements T_1 and T_2 adjacent to some F. With the argument that the two adjacent elements must have the same share of the jump on F, mostly $\alpha_{T,F} = 1/2$ is set (see [39]). This argument is, at best, valid for uniform meshes, continuous diffusion coefficients and homogeneous approximations over the whole triangulation $\mathcal{T}$ (cf. the voluminous analysis by M. Ainsworth and J. T. Oden [3], which, however, offers only 1D-arguments for a locally variable choice of these coefficients).

The BM error estimator (6.12) is – in the presence of knowledge about the generic constant c – "reliable" in the sense of (6.5) with $\kappa_1 = 1$. However, the constant c is usually rather inaccessible, so that the check (6.6) performed with the above error estimator is indeed *not reliable* (see [19]). Moreover, the formal reliability of the error estimator (6.12) is based on a sufficiently accurate computation of the terms η_T and η_F. While η_F contains only well-known quantities of the discretization, the computation

of η_T requires an integration of f, which is usually realized via numerical quadrature (cf. Volume 1, Section 9). It can be shown, however (see R. Verfürth [209]), that for some class of problems the terms η_T can be dropped. For a characterization of this class we require the following definition, which dates back to W. Dörfler [79].

Definition 6.3 (Oscillation). Let ω_ξ denote some star-shaped domain around a node $\xi \in \mathcal{N}$ and

$$\bar{f_\xi} = \frac{1}{|\omega_\xi|} \int_{\omega_\xi} f \, \mathrm{d}x$$

the corresponding average value of f. Then the *oscillation* of f is given by

$$\operatorname{osc}(f;\mathcal{T})^2 = \sum_{\xi\in\mathcal{N}} h_\xi^2 \sum_{T\in\mathcal{T}:\xi\in T} \|f - \bar{f_\xi}\|^2_{L^2(T)},$$

where h_ξ again denotes the diameter of ω_ξ.

Theorem 6.4. *There exists a constant c independent of the local mesh size h such that for solutions u_h of* (6.2) *the following result holds:*

$$\sum_{T\in\mathcal{T}} \eta_T^2 \le c\Big(\sum_{F\in\mathcal{F}} \eta_F^2 + \operatorname{osc}\left(\operatorname{div}(\sigma\nabla u_h) + f;\mathcal{T}\right)^2\Big). \tag{6.14}$$

Proof. The piecewise residual is given by $r = \operatorname{div}(\sigma\nabla u_h) + f \in L^2(\Omega)$ in the interior of each element $T \in \mathcal{T}$. Let $\xi \in \mathcal{N}$ with associated nodal function $\phi_\xi \in S_h$. We define $\mathcal{F}_\xi = \{F \in \mathcal{F} : \xi \in F\}$. Because of (6.2) we have, after an elementary application of the Gaussian theorem,

$$\begin{aligned}\int_\Omega f\phi_\xi \, \mathrm{d}x &= \int_\Omega \nabla {u_h}^T \sigma \nabla\phi_\xi \, \mathrm{d}x \\ &= \sum_{F\in\mathcal{F}_\xi} \int_F \phi_\xi [\![n^T \sigma \nabla u_h]\!]_F \, \mathrm{d}s - \int_\Omega \phi_\xi \operatorname{div}(\sigma\nabla u_h)\, \mathrm{d}x\end{aligned}$$

and, by subsequent application of the Cauchy–Schwarz inequality in finite dimensions, eventually

$$\begin{aligned}\int_\Omega r\phi_\xi \, \mathrm{d}x = \sum_{F\in\mathcal{F}_\xi} \int_F \phi_\xi [\![n^T \sigma \nabla u_h]\!]_F \, \mathrm{d}s &\le \sum_{F\in\mathcal{F}_\xi} \|\phi_\xi\|_{L^2(F)} \|[\![n^T \sigma \nabla u_h]\!]_F\|_{L^2(F)} \\ &\le \Big(\sum_{F\in\mathcal{F}_\xi} h_F^{d-2} \Big)^{1/2} \Big(\sum_{F\in\mathcal{F}_\xi} h_F \|[\![n^T \sigma \nabla u_h]\!]_F\|^2_{L^2(F)} \Big)^{1/2} \\ &\le c h_\xi^{d/2-1} \Big(\sum_{F\in\mathcal{F}_\xi} \eta_F^2 \Big)^{1/2}. \end{aligned} \tag{6.15}$$

By the averaging

$$\bar{r}_\xi = \frac{1}{|\omega_\xi|}\int_{\omega_\xi} r\,\mathrm{d}x \le \frac{\|1\|_{L^2(\omega_\xi)}}{|\omega_\xi|}\|r\|_{L^2(\omega_\xi)} = \frac{\|r\|_{L^2(\omega_\xi)}}{|\omega_\xi|^{1/2}} \tag{6.16}$$

we have (this part of the proof has been sourced out to Exercise 6.1*c*)

$$\int_{\omega_\xi} 2r(1-\phi_\xi)\bar{r}_\xi\,\mathrm{d}x \le \frac{2\|1-\phi_\xi\|_{L^2(\omega_\xi)}}{|\omega_\xi|^{1/2}}\,\|r\|_{L^2(\omega_\xi)}\,|\omega_\xi|^{1/2}\bar{r}_\xi \le \|r\|^2_{L^2(\omega_\xi)}$$

and thus

$$\|\bar{r}_\xi\|^2_{L^2(\omega_\xi)} \le \int_{\omega_\xi}(r^2 + 2r(\phi_\xi-1)\bar{r}_\xi + \bar{r}_\xi^2)\,\mathrm{d}x = \int_{\omega_\xi}(2r\phi_\xi\bar{r}_\xi + (r-\bar{r}_\xi)^2)\,\mathrm{d}x.$$

Finally, we get, due to the L^2-orthogonality of the averaging

$$\|r\|^2_{L^2(\omega_\xi)} = \|r-\bar{r}_\xi\|^2_{L^2(\omega_\xi)} + \|\bar{r}_\xi\|^2_{L^2(\omega_\xi)} \le 2\|r-\bar{r}_\xi\|^2_{L^2(\omega_\xi)} + 2\int_{\omega_\xi} r\phi_\xi\bar{r}_\xi\,\mathrm{d}x.$$

From (6.16) we obtain

$$\bar{r}_\xi \le c h_\xi^{-d/2}\|r\|_{L^2(\omega_\xi)},$$

which with (6.15) gives rise to the quadratic inequality

$$\|r\|^2_{L^2(\omega_\xi)} \le \beta\|r\|_{L^2(\omega_\xi)} + \gamma,$$
$$\beta = c h_\xi^{-1}\Big(\sum_{F\in\mathcal{F}_\xi}\eta_F^2\Big)^{1/2},\ \gamma = 2\|r-\bar{r}_\xi\|^2_{L^2(\omega_\xi)}.$$

Its roots are given by

$$r_\pm = \tfrac{1}{2}\Big(\beta \pm \sqrt{\beta^2+4\gamma}\Big)$$

so that we have for the solutions

$$\begin{aligned}4\|r\|^2_{L^2(\omega_\xi)} &\le \big(\beta + \sqrt{\beta^2+4\gamma}\,\big)^2 = \beta^2 + 2\beta\sqrt{\beta^2+4\gamma} + \beta^2 + 4\gamma\\ &\le 2\beta^2 + 2\beta^2\left(1+\frac{2\gamma}{\beta^2}\right) + 4\gamma = 4(\beta^2+\gamma)\\ &\le c\Big(h_\xi^{-2}\sum_{F\in\mathcal{F}_\xi}\eta_F^2 + \|r-\bar{r}_\xi\|^2_{L^2(\omega_\xi)}\Big).\end{aligned}$$

Multiplication by h_ξ^2 and summation over $\xi\in\mathcal{N}$ then yields (6.14). □

Simplified BM Error Estimator. If the oscillation of the right-hand side f can be neglected, then, due to (6.14), the volume contribution in (6.12) can be replaced by the jump contribution. One thus obtains, again by neglecting the generic constants,

$$[\epsilon_h] = \Big(\sum_{F\in\mathcal{F}}\eta_F^2\Big)^{\frac{1}{2}}. \tag{6.17}$$

This error estimator is indeed reliable in the sense of (6.5). (The associated proof has been postponed until Exercise 7.2.)

Remark 6.5. Strictly speaking, the a priori restriction of the oscillation in Theorem 6.4 is not sufficient as a theoretical basis for the *practical* reliability of the error estimator (6.17): This theorem is based on the fact that u_h satisfies the weak formulation (6.2), which includes the presumption that the integrals $\langle f, v_h \rangle$ have been computed to sufficient accuracy by numerical quadrature. Even for functions f with small oscillation, this can be guaranteed only under additional smoothness assumptions.

Remark 6.6. In [11], I. Babuška and A. Miller presented their theoretical analysis not, as presented here, via the quasi-interpolant (6.11), but by a decomposition into local Dirichlet problems. A modification of such a decomposition will come up again in the immediately following Section 6.1.4, there, however, as an algorithmically realized option.

6.1.2 Triangle Oriented Error Estimators

The unsatisfactorily rough estimate (6.12) originates from the attempt to find a direct upper bound for all v in (6.9) by means of *a priori*-estimates. This insight motivates the alternative, to construct *a posteriori* a special $[e_h] \approx e_h$. For this purpose, the ansatz space S_h must be suitably expanded to a space S_h^+, which in the context of finite elements can be conveniently localized.

Localization. Following a suggestion of R. Bank and A. Weiser [19] from 1985, we return to the decomposition (6.10), but exploit the information in some way different from the one in the previous Section 6.1.1: in this approach, we may immediately read that the localized error $e_T = e_h|_T$ of the inhomogeneous Neumann problem

$$a_T(e_T, v) = b_T(v; n^T \sigma \nabla e_h) \quad \text{for all } v \in H^1(T) \tag{6.18}$$

with right-hand side

$$b_T(v; g) = \int_T (\operatorname{div}(\sigma \nabla u_h) + f) v \, \mathrm{d}x - \int_{\partial T} v g \, \mathrm{d}s$$

is all that is needed. This includes that, as wanted, $\epsilon_T^2 = a_T(e_T, e_T)$. For the purpose of error estimation again an approximate computation of e_T is sufficient. Due to the small size of the domain T it can be elementarily realized by some *locally expanded* ansatz space $S_h^+(T) = S_h(T) \oplus S_h^{\oplus}(T) \subset H^1(\Omega)|_T$. Typical examples under application of a polynomial ansatz space $S_h = S_h^p$ are (a) a uniform refinement $S_h^+ = S_{h/2}^p$ or (b) an increase of the polynomial order $S_h^+ = S_h^{p+1}$ (see Figure 6.1 for $p = 1$).

The Neumann problems over the individual triangles $T \in \mathcal{T}$ are, however, unique only up to an additive constant (see Theorem 1.3), and must, in addition, meet a compatibility condition similar to (1.3). Therefore we focus our attention to ansatz spaces

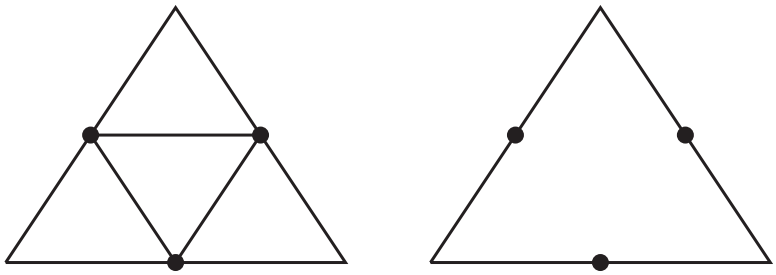

Figure 6.1. Choice of expansion $S_h^\oplus(T)$ of the ansatz space for linear finite elements in 2D. *Left:* piecewise linear expansion. *Right:* quadratic expansion.

$S_h^\oplus(T)$ that exclude constant functions; thus the bilinear form $a(u,v)$ is elliptic on $S_h^\oplus(T)$ (independent of the presence of a Helmholtz term).

Let $e_T^\oplus \in S_h^\oplus(T)$ be defined as solution of

$$a_T(e_T^\oplus, v) = b_T(v; n^T \sigma \nabla e_h) \quad \text{for all } v \in S_h^\oplus(T). \tag{6.19}$$

Because of the Galerkin orthogonality (4.14) we immediately obtain the *lower* bound

$$a(e_T^\oplus, e_T^\oplus) \le \epsilon_T^2.$$

A direct application of (6.19) to an error estimate fails, since, due to (4.13), the continuous normal flow $n^T g = n^T \sigma \nabla u$ as well as the noncontinuous normal flow $n^T \sigma \nabla e_h = n^T(g - \sigma \nabla u_h)$ are unknown, which requires the construction of an approximation g_h. Just as in the localization of residual based error estimators (see the choice $\alpha_{T,F} = 1/2$ in (6.13)), the simplest, though arbitrary choice of g_h is the *averaging*

$$g_h|_E(x) = \frac{1}{2} \lim_{\epsilon \to 0} \big(\sigma(x + \epsilon n) u_h(x + \epsilon n) + \sigma(x - \epsilon n) u_h(x - \epsilon n)\big). \tag{6.20}$$

This leads to

$$n^T(g_h - \sigma \nabla u_h) = \frac{1}{2} [\![n^T \sigma \nabla u_h]\!].$$

BW Error Estimator. The local error approximation $[e_T]$ is then defined by

$$a_T([e_T], v) = b_T\left(v; \frac{1}{2} [\![n^T \sigma \nabla u_h]\!]\right) \quad \text{for all } v \in S_h^\oplus(T),$$

with the associated local error estimator

$$[\epsilon_T] = \|[e_T]\|_a. \tag{6.21}$$

For linear finite elements this requires a local extension $S_h^\oplus(T)$ by quadratic shape functions apart from the evaluation of the jumps along edges and the strong residual $\operatorname{div}(\sigma \nabla u_h) + f$ as well as the solution of local equation systems of dimension $\frac{1}{2}d(d+1)$.

Error Representation. By composition of the single partial solutions $[e_T]$ one obtains a global approximation $[e_h] \approx e_h$ of the error. One has to observe, however, that in general $[e_h]$ exhibits jumps beyond edges and faces and therefore cannot be contained in $H^1(\Omega)$. A refinement of the solution by the ansatz $u_h + [e_h] \approx u$ would lead to nonconformal solutions.

Reliability. The restriction to error approximations in a finite dimensional subspace contains the risk to overlook part of the error: It is easy to construct examples, for which $b(v; \frac{1}{2}[\![n^T \sigma \nabla u_h]\!]) = 0$ holds for all $v \in S_h^{\oplus}(T)$, even though b itself does not vanish. For this reason, the verification of reliability makes the following saturation assumption necessary.

Definition 6.7. *Saturation assumption.* Let $\hat{S}_h^+ = \Pi_{T \in \mathcal{T}} S_h^+(T)$ and $S_h^+ = \hat{S}_h^+ \cap H^1(\Omega)$ denote an extended conformal ansatz space with corresponding Galerkin solution $u_h^+ \in S_h^+$ and error ϵ_h^+. Then there exists a $\beta < 1$ such that

$$\epsilon_h^+ \leq \beta \epsilon_h. \tag{6.22}$$

It should be explicitly mentioned that this saturation assumption is, in fact, a central *assumption*, since for each given space extension arbitrarily many right-hand sides f may be constructed such that $u_h^+ = u_h$ holds (see Exercise 7.2). Nevertheless it is useful in practice, since for each *fixed* and reasonably smooth right-hand side f it is satisfied on sufficiently fine meshes. In particular, the extension of linear to quadratic elements on simplicial triangulations satisfies the saturation assumption, if the oscillation of f is small, as shown by the following theorem of W. Dörfler and R. H. Nocchetto [80], which we present here without proof.

Theorem 6.8. *Let $S_h = S_h^1$ and $S_h^+ = S_h^2 \subset H_0^1(\Omega)$. Then there exists a constant $\mu < 1$, which only depends on the shape regularity of the triangulation $\mathcal{T}$, so that a small oscillation*

$$\operatorname{osc}(f; \mathcal{T}) \leq \mu \epsilon_h$$

implies the saturation assumption (6.22) *with* $\beta = \sqrt{1 - \mu^2}$.

After employing this assumption, we are now ready to verify that $e_h^{\oplus} = u_h^+ - u_h \in S_h^+$ supplies a reliable error approximation $\epsilon_h^{\oplus} = \|e_h^{\oplus}\|_a$.

Theorem 6.9. *Under the saturation assumption* (6.22) *the following result holds:*

$$\epsilon_h^{\oplus} \leq \epsilon_h \leq \frac{\epsilon_h^{\oplus}}{\sqrt{1 - \beta^2}}. \tag{6.23}$$

Proof. Based on the orthogonality of the error e_h^+ in the ansatz space S_h^+ we have

$$\|e_h\|_a^2 = \|e_h^+\|_a^2 + \|e_h^{\oplus}\|_a^2 \geq \|e_h^{\oplus}\|_a^2,$$

which confirms the lower bound in (6.23). For a verification of the upper bound we need the saturation assumption:

$$\|e_h\|_a^2 = \|e_h^+\|_a^2 + \|e_h^\oplus\|_a^2 \le \beta^2 \|e_h\|_a^2 + \|e_h^\oplus\|_a^2.$$

By repositioning of terms we immediately get

$$(1-\beta^2)\|e_h\|_a^2 \le \|e_h^\oplus\|_a^2,$$

which is the missing inequality. □

Now, by comparison with $\epsilon_h^\oplus$, the reliability of the estimator (6.21) can be shown.

Theorem 6.10. *Let, in addition to the assumption* (6.22)*, some* $\gamma < 1$ *exist such that the strengthened Cauchy–Schwarz inequality*

$$a_T(v_h, v_h^\oplus) \le \gamma \|v_h\|_{a_T} \|v_h^\oplus\|_{a_T}$$

is satisfied for all $T \in \mathcal{T}$ *and* $v_h \in S_h(T)$, $v_h^\oplus \in S_h^\oplus(T)$. *Then*

$$(1-\gamma^2)\sqrt{1-\beta^2}\,\epsilon_h \le [\epsilon_h]. \tag{6.24}$$

Proof. The proof is done in two steps via auxiliary quantities $\hat{e}_T \in S_h^+(T)$ defined by

$$a_T(\hat{e}_T, v) = b_T(I_h^\oplus v) \quad \text{for all } v \in S_h^+(T)\,, \quad \int_T \hat{e}_T \, dT = 0,$$

where we dropped the flow approximation as parameters in b_T. Here $I_h^\oplus : S_h^+(T) \to S_h^\oplus(T)$ is the projector uniquely defined by the decomposition $S_h^+(T) = S_h(T) \oplus S_h^\oplus(T)$; in addition, we define $I_h = I - I_h^\oplus$. We first have

$$a_T([e_T], v) = b_T(v) = b_T(I_h^\oplus v) = a_T(\hat{e}_T, v) \quad \text{for all } v \in S_h^\oplus(T).$$

Due to $I_h^\oplus \hat{e}_T \in S_h^\oplus$ we now obtain

$$\|[e_T]\|_{a_T} \|I_h^\oplus \hat{e}_T\|_{a_T} \ge a_T([e_T], I_h^\oplus \hat{e}_T) = a_T(\hat{e}_T, I_h^\oplus \hat{e}_T).$$

We have

$$a_T(\hat{e}_T, I_h \hat{e}_T) = b_T(I_h^\oplus I_h \hat{e}_T) = 0 \tag{6.25}$$

and with the strengthened Cauchy–Schwarz inequality

$$\begin{aligned}
\|[e_T]\|_{a_T} \|I_h^\oplus \hat{e}_T\|_{a_T} &\ge a_T(\hat{e}_T, I_h \hat{e}_T) + a_T(\hat{e}_T, I_h^\oplus \hat{e}_T) \\
&= \|I_h \hat{e}_T\|_{a_T}^2 + 2a_T(I_h \hat{e}_T, I_h^\oplus \hat{e}_T) + \|I_h^\oplus \hat{e}_T\|_{a_T}^2 \\
&\ge \|I_h \hat{e}_T\|_{a_T}^2 - 2\gamma \|I_h \hat{e}_T\|_{a_T} \|I_h^\oplus \hat{e}_T\|_{a_T} + \|I_h^\oplus \hat{e}_T\|_{a_T}^2 \\
&= (\|I_h \hat{e}_T\|_{a_T} - \gamma \|I_h^\oplus \hat{e}_T\|_{a_T})^2 + (1-\gamma^2)\|I_h^\oplus \hat{e}_T\|_{a_T}^2 \\
&\ge (1-\gamma^2)\|I_h^\oplus \hat{e}_T\|_{a_T}^2.
\end{aligned}$$

Division by $\|I_h^{\oplus}\hat{e}_T\|_{a_T}$ leads to the estimate $(1-\gamma^2)\|I_h^{\oplus}\hat{e}_T\|_{a_T} \le \|[e_T]\|_{a_T}$. From (6.25) we may conclude that

$$\|\hat{e}_T\|_{a_T}^2 = a_T(\hat{e}_T, I_h^{\oplus}\hat{e}_T) \le \|\hat{e}_T\|_{a_T}\|I_h^{\oplus}\hat{e}_T\|_{a_T},$$

so that in total we get

$$(1-\gamma^2)\|\hat{e}_T\|_{a_T} \le \|[e_T]\|_{a_T} . \tag{6.26}$$

In the second step we turn to global error estimation. Simple summation over the elements $T \in \mathcal{T}$ leads to the global error approximation $\hat{e}_h \in \hat{S}_h^+$ with $\hat{e}_h|_T = \hat{e}_T$ and

$$a(\hat{e}_h, v) = b(I_h^{\oplus}v) \quad \text{for all } v \in \hat{S}_h^+ \supset S_h^+.$$

Note that the summation of b_T leads directly to the localization (6.10), which is why $a(e_h, v) = b(v)$ holds for all $v \in H^1(\Omega)$, and, with the Galerkin orthogonality $a(e_h^+, v) = 0$ for all $v \in S_h^+$, also

$$a(e_h^{\oplus}, v) = a(e_h - e_h^+, v) = b(v) \quad \text{for all } v \in S_h^+.$$

Because of $I_h^{\oplus}e_h^{\oplus} \in S_h^+$ we thus obtain

$$\begin{aligned}\|\hat{e}_h\|_a\|e_h^{\oplus}\|_a \ge a(\hat{e}_h, e_h^{\oplus}) &= b(I_h^{\oplus}e_h^{\oplus}) = a(e_h^{\oplus}, I_h^{\oplus}e_h^{\oplus}) \\ &= a(e_h^{\oplus}, e_h^{\oplus} - I_h e_h^{\oplus}) = a(e_h^{\oplus}, e_h^{\oplus}) = \|e_h^{\oplus}\|_a^2\end{aligned}$$

and, after division by $\|e_h^{\oplus}\|_a$ with Theorem 6.9

$$\|\hat{e}_h\|_a \ge \|e_h^{\oplus}\|_a \ge \sqrt{1-\beta^2}\,\epsilon_h. \tag{6.27}$$

The combination of (6.26) and (6.27) then leads to the assertion (6.24). □

Remark 6.11. The *triangle oriented* error estimator due to R. Bank and A. Weiser [19] from 1985 was the first error estimator in which a localization of the quadratic ansatz space was realized. For a long time it remained the basis of adaptivity in the finite element code PLTMG [16], but has been replaced in newer versions by the gradient recovery treated in the next Section 6.1.3. The auxiliary quantity $\hat{e}_h$ used in the proof of Theorem 6.10 was also considered in the paper [19] as a possible error estimator. It gives rise to a better constant κ_1 in the estimate for the reliability, but has shown to be less efficient in practical tests.

6.1.3 Gradient Recovery

We return to the averaging (6.20) and try to replace this rather simple approximation of the discontinuous discrete flows $g_h = \sigma\nabla u_h \approx \sigma\nabla u = g$ by an improved version called gradient recovery. For that purpose, we project, as shown in Figure 6.2, the gradient $g_h \in S_{h,\text{grad}} = \{\nabla v : v \in S_h\}$ by some suitable projector Q onto an

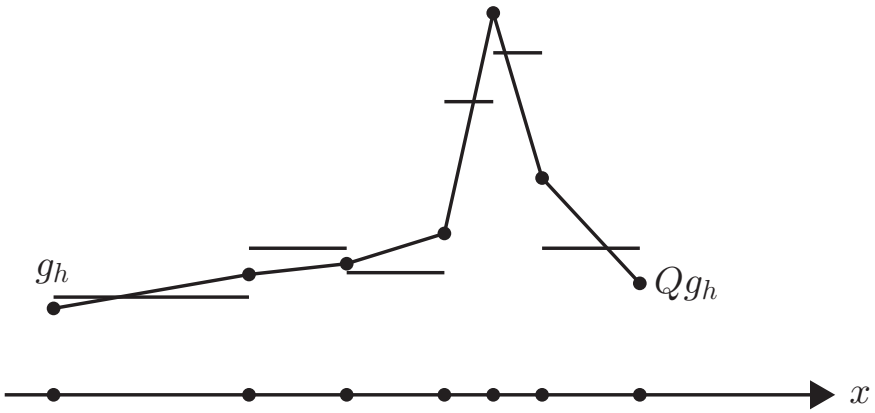

Figure 6.2. Gradient recovery for linear finite elements in $\mathbb{R}^1$ via L^2-projection.

ansatz space $\bar{S}_h$ of *continuous* functions. For the time being, we focus our attention on continuous diffusion coefficients, where both the exact gradients ∇u and the flows g are continuous. So there is no conceptual difference between a projection of the flows and one of the gradients.

We consider the homogeneous Neumann problem with

$$\|e_h\|_a = \|\sigma^{1/2}(\nabla u - \nabla u_h)\|_{L^2(\Omega)}$$

and $\alpha = 0$. Then we replace ∇u by

$$Q\nabla u_h = \operatorname*{arg\,min}_{q_h \in \bar{S}_h} \|\sigma^{1/2}(\nabla u_h - q_h)\|_{L^2(\Omega)},$$

so that we have for all $q_h \in \bar{S}_h$:

$$\begin{aligned}\|\sigma^{1/2}(\nabla u_h - Q\nabla u_h)\|_{L^2(\Omega)} &\le \|\sigma^{1/2}(\nabla u_h - q_h)\|_{L^2(\Omega)} \\ &\le \|\sigma^{1/2}(\nabla u - \nabla u_h)\|_{L^2(\Omega)} + \|\sigma^{1/2}(\nabla u - q_h)\|_{L^2(\Omega)} \\ &= \epsilon_h + \|\sigma^{1/2}(\nabla u - q_h)\|_{L^2(\Omega)}. \qquad (6.28)\end{aligned}$$

First, we neglect the second term $\|\sigma^{1/2}(\nabla u - q_h)\|_{L^2(\Omega)}$ in the sum and obtain

$$[\epsilon_h] = \|\sigma^{1/2}(\nabla u_h - Q\nabla u_h)\|_{L^2(\Omega)}. \qquad (6.29)$$

Localization. From (6.29) we may recognize a retrospective localization immediately. We obtain the error indicators

$$[\epsilon_h(T)] = \|\sigma^{1/2}(\nabla u_h - Q\nabla u_h)\|_{L^2(T)}.$$

Reliability. The reliability of the error estimator (6.29) is a direct consequence of Theorem 6.4 and of the following theorem.

Theorem 6.12. *There exists a constant c independent of h and f such that for* (6.29) *the result*

$$\sum_{F\in\mathcal{F}} \eta_F^2 \le c[\epsilon_h]^2 \tag{6.30}$$

holds.

Proof. For some $\xi \in \mathcal{N}$ we consider the quotient spaces

$$S_\xi = (S_{h,\mathrm{grad}}(\omega_\xi) + \bar{S}_h(\omega_\xi))/\bar{S}_h(\omega_\xi)$$

and on them the seminorm

$$|\nabla u_h|_\xi^2 = \sum_{F\in\mathcal{F}, F\subset\omega_\xi} h_F \|[\![n^T \nabla u_h]\!]_F\|_{L^2(F)}^2$$

as well as the norm

$$\|\nabla u_h\|_\xi = \min_{q_h\in\bar{S}_h} \|\sigma^{1/2}(\nabla u_h - q_h)\|_{L^2(\omega_\xi)}.$$

Because of the bounded dimension of S_ξ there exists a constant c depending on the shape regularity of the elements $T \subset \omega_\xi$, but not on h, with $|\nabla u_h|_\xi \le c\|\nabla u_h\|_\xi$. Summation over all $\xi \in \mathcal{N}$ leads, due to the bounded overlap of Ω with ω_ξ, to the inequality (6.30). □

As in the simplified BM error estimator (6.17) the reliability here is only given under the a priori assumption of a small oscillation of f and for exact computation of the scalar products.

Efficiency. If the neglected term $\|\sigma^{1/2}(\nabla u - q_h)\|_{L^2(\Omega)}$ from (6.28) can be suitably bounded, then the error estimator (6.29) is a lower bound of the error and thus efficient. For this we need a variant of the saturation assumption (6.22).

Theorem 6.13. *Suppose there exists an h-independent constant $\hat{\beta}$ such that*

$$\min_{q_h\in\bar{S}_h} \|\sigma^{1/2}(\nabla u - q_h)\|_{L^2(\Omega)} \le \hat{\beta}\epsilon_h \tag{6.31}$$

holds. Then $[\epsilon_h] \le (1+\hat{\beta})\epsilon_h$.

Proof. Let $w_h = \arg\min_{q_h\in\bar{S}_h} \|\sigma^{1/2}(\nabla u - q_h)\|_{L^2(\Omega)} \in \bar{S}_h$. Then $[\epsilon_h] \le \epsilon_h + \|\sigma^{1/2}(\nabla u - w_h)\|_{L^2(\Omega)} \le \epsilon_h + \hat{\beta}\epsilon_h$ holds due to (6.28) and the saturation assumption. □

For piecewise polynomial finite elements $u_h \in S_h^p$ the gradient ∇u_h lies in

$$S_{h,\mathrm{grad}}^p = \{v \in L^2(\Omega)^d : v|_T \in \mathbb{P}_{p-1} \text{ for all } T \in \mathcal{T}\}.$$

In order to assure the saturation assumption asymptotically for sufficiently smooth u, i.e., for $h \to 0$ with $\kappa_2 \to 1$ in Definition 6.1, the order for $\bar{S}_h = (S_h^p)^d$ is increased – for linear elements this order increase is anyway necessary to achieve continuity.

Error Representation. As in the residual based error estimator (6.13), we cannot directly deduce an error approximation $[e_h]$ from (6.29). But we have the defect equation

$$-\operatorname{div}(\sigma \nabla e_h) = -\operatorname{div}(\sigma \nabla u - \sigma \nabla u_h) \approx -\operatorname{div}(\sigma (Q \nabla u_h - \nabla u_h),$$

whose localization to individual elements T similar to (6.18) leads to

$$a_T(e_T, v) = \int_T \big(\operatorname{div}(\sigma \nabla u_h) - \operatorname{div}(\sigma Q \nabla u_h)\big) v \, \mathrm{d}x - \int_{\partial T} v n^T \sigma (Q \nabla u_h - \nabla u_h) \, \mathrm{d}s. \tag{6.32}$$

The replacement of $f = -\operatorname{div}(\sigma \nabla u)$ by $-\operatorname{div}(\sigma Q \nabla u_h)$ ensures that the compatibility condition (1.3) is satisfied by the flow reconstruction $\sigma Q \nabla u_h$, and thus saves us the arbitrariness of the averaging (6.20). By virtue of the solution of local Neumann problems in the extended local ansatz space $S_h^{p+1}(T)$ one then obtains – up to constants – an approximate error representation $[e_h]$. Alternatively, a hierarchical ansatz space may be applied for the solution of (6.32) just as in the BW error estimator.

Robin Boundary Conditions. For $\alpha > 0$ an additional term enters into the energy norm

$$\|e_h\|_a^2 = \|\sigma^{1/2} \nabla e_h\|_{L^2(\Omega)}^2 + \|\alpha^{1/2} e_h\|_{L^2(\partial\Omega)}^2 ,$$

which is not covered by the error estimator (6.29). Nevertheless $[\epsilon_h]$ remains formally efficient, as the boundary term is dominated by the domain term.

Theorem 6.14. *There exists a constant $c > 0$ with*

$$c\epsilon_h \le \|\sigma^{1/2} \nabla e_h\|_{L^2(\Omega)} \le \epsilon_h.$$

Proof. We define the average

$$\bar{e}_h = \frac{1}{|\Omega|} \int_\Omega e_h \, \mathrm{d}x$$

and regard it as a constant function in S_h. Then, due to the Galerkin orthogonality (4.14) and the continuity A.10 of the bilinear form a, the following result holds:

$$\epsilon_h^2 \le \|e_h\|_a^2 + \|\bar{e}_h\|_a^2 = \|e_h - \bar{e}_h\|_a^2 \le C_a \|e_h - \bar{e}_h\|_{H^1(\Omega)}^2.$$

Form the Poincaré inequality (A.25) we then get

$$\epsilon_h \le C_P \sqrt{C_a} |e_h|_{H^1} \le C_P \frac{\sqrt{C_a}}{\inf \sqrt{\sigma}} \|\sigma^{1/2} \nabla e_h\|_{L^2(\Omega)}. \qquad \square$$

Note, however, that even for asymptotically exact gradient recovery, the error estimator will, in general, not be asymptotically exact, due to the neglection of the boundary terms.

Discontinuous Coefficients. In the case of discontinuous diffusion coefficients σ the *tangential component* of the exact flow – just as the normal component of the gradient – is also discontinuous. The gradient "recovery" by projection to the continuous functions would necessarily lead to a deviation $\nabla u_h - Q\nabla u_h$ of the order of the coefficient jump $[\![\sigma]\!]$ and, correspondingly, to an unnecessarily strong mesh refinement along the discontinuity faces. In these cases a piecewise projection with discontinuities of the recovered gradient $Q\nabla u_h$ is possible. For domains with many interior faces, at which diffusion coefficients are discontinuous, such a concept is rather unhandy, let alone that the recovery effect should, in principle, be achieved by the projection to a continuous space.

ZZ Error Estimator. Due to the h-independent bounded condition number of the mass matrix the computation of the L^2-projection $Q\nabla u_h$ is easily possible by the CG-method, which, however, requires a global computational amount. One of the simplest local recovery formulas is also the oldest one: an averaging of piecewise constant discrete gradients, which dates back to O. C. Zienkiewicz and J. Z. Zhu [231, 232]:

$$(Q\nabla u_h)(\xi) = \frac{1}{|\omega_\xi|} \sum_{T\in\mathcal{T}:\,\xi\in T} |T|\ \nabla u_h|_T(\xi) \quad \text{for all } \xi \in \mathcal{N}.$$

The asymptotic exactness is based on special *superconvergence results on structured meshes*, which can be extended to unstructured meshes by means of certain smoothing operations (see R. E. Bank and J. Xu [20, 21]). In the survey article [49] C. Carstensen showed that equivalent error estimators can be constructed by means of different but nevertheless simple local averaging operators.

6.1.4 Hierarchical Error Estimators

A rather straightforward way of error estimation does not originate from the localization (6.10), but from the global defect equation

$$a(e_h, v) = a(u - u_h, v) = \langle f, v\rangle - a(u_h, v) \quad \text{for all } v \in H^1(\Omega), \tag{6.33}$$

from which an error approximation $[e_h] \approx u - u_h$ is constructed. This approach dates back to P. Deuflhard, P. Leinen, and H. Yserentant [73] from 1989, who suggested an *edge oriented* error estimator.

Embedded Error Estimator. As before in (6.19), the defect equation (6.33) is solved approximately in some (here) globally extended ansatz space $S_h^+ \supset S_h$. We thus define the error approximation $[e_h] = u_h^+ - u_h \in S_h^+$ as the solution of

$$a([e_h], v) = a(u_h^+ - u_h, v) = \langle f, v\rangle - a(u_h, v) \quad \text{for all } v \in S_h^+ \tag{6.34}$$

and, correspondingly, the embedded error estimator

$$[\epsilon_h] = \|[e_h]\|_a. \tag{6.35}$$

Upon application of the piecewise polynomial ansatz space $S_h = S_h^p$ we are essentially led to the extensions $S_h^+ = S_{h/2}^p$ by uniform refinement and $S_h^+ = S_h^{p+1}$ by order increase (cf. Figure 6.1 for the special case $p = 1$).

Localization and Error Representation. With the computation of $[e_h] \in H^1(\Omega)$ as Galerkin solution of (6.34) a consistent approximation of the error is readily available. The localization to individual elements $T \in \mathcal{T}$ is obvious:

$$[\epsilon_h(T)] = a_T([e_h], [e_h]).$$

At the same time, with $[e_h]$ a more accurate Galerkin solution $u_h^+ = u_h + [e_h] \in S_h^+$ is available, which might mean that an estimation of the error $\epsilon_h^+ = \|u - u_h^+\|_a$ would be more interesting – which, however, is unavailable. We thus find ourselves in the same kind of dilemma as in the adaptive timestep control for ordinary differential equations (cf. Volume 2, Section 5).

Reliability and Efficiency. In Theorem 6.9 we already showed that the value $[\epsilon_h]$ from (6.35) supplies an efficient error estimator under the saturation assumption $\beta < 1$; as for the reliability, we again refer to Theorem 6.8. The comparison of the result (6.23) with Definition 6.1 directly leads to the constants $\kappa_1 = 1/\sqrt{1-\beta^2}$ and $\kappa_2 = 1$.

For asymptotic exactness we require additionally that $\beta \to 0$ for $h \to 0$, which – given sufficient regularity of the solution u – is guaranteed by the higher order extension $S_h^+ = S_h^{p+1}$.

DLY Error Estimator. We are now ready to introduce the edge oriented error estimator due to [73] for linear finite elements in some detail. For realistic applications, the complete solution of (6.34) is still too costly (cf. Exercise 7.3). We therefore replace the Galerkin solution of (6.34) by the solution of a 'sparsed' equation system: We restrict ourselves to the complement $S_h^\oplus$ and merely use the diagonal of the stiffness matrix so that only local (1, 1)-"systems" are to be solved. Let $S_h^\oplus$ be spanned by the basis $\Psi^\oplus$; then we obtain the approximate error representation

$$[e_h] := \sum_{\psi \in \Psi^\oplus} \frac{\langle f - Au_h, \psi \rangle}{\|\psi\|_A^2} \psi \in S_h^\oplus \subset H^1(\Omega)$$

and the associated error estimator

$$[\epsilon_h]^2 := \sum_{\psi \in \Psi^\oplus} \frac{\langle f - Au_h, \psi \rangle^2}{\|\psi\|_A^2}. \tag{6.36}$$

In view of an efficient computation, the basis functions $\psi \in \Psi^\oplus$ should again have a support over few elements $T \in \mathcal{T}$ only. The simplest choice is an extension of linear elements by quadratic "bubbles" on the edges (see Figure 6.3). The localization necessary for an adaptive mesh refinement can then be directly identified from (6.36).

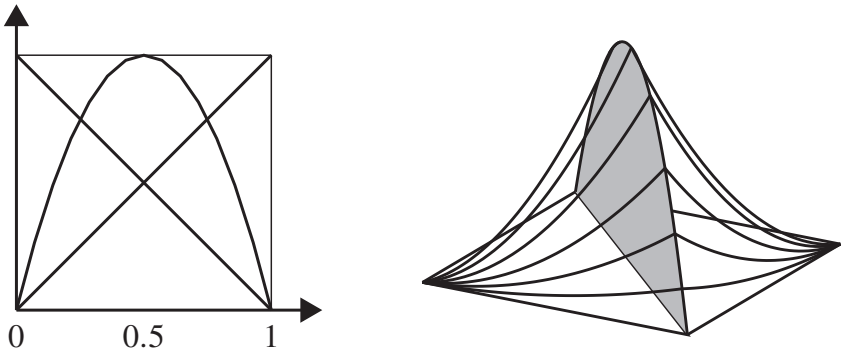

Figure 6.3. Quadratic "bubble" in edge oriented error estimator. *Left:* shape functions on edge (cf. Figure 4.8). *Right:* ansatz function in $\mathbb{R}^2$.

However, in contrast to (6.19), we now get an error estimator localized to individual edges $E \in \mathcal{E}$:

$$[\epsilon_h(E)] = \langle r, \psi_E \rangle / \|\psi_E\|_A. \tag{6.37}$$

This edge-oriented hierarchical error estimator constitutes the conceptual fundament of the adaptive finite element codes KASKADE (see the software page at the end of the book).

Reliability and Efficiency. By solving the reduced system (6.34) only approximately, we have not forfeited efficiency, but the efficiency span gets larger and, at the same time, we also lose asymptotic exactness, as shown by the following theorem.

Theorem 6.15. *Under the saturation assumption* (6.22) *the error estimator* (6.36) *is efficient according to* (6.7)*, i.e., there exist h-independent constants* $K_1, K_2 < \infty$ *such that*

$$\sqrt{\frac{1-\beta^2}{K_1}}\,\epsilon_h \le [\epsilon_h] \le \sqrt{K_2}\,\epsilon_h.$$

The proof is postponed until Section 7.1.5, where we can work it out much more simply in the context of subspace correction methods.

Connection with Residual-based Error Estimator. For linear finite elements and piecewise constant data σ and f an interesting connection between the BM error estimator (6.12) and the hierarchical error estimator (7.28) has been found in [37]. For this type of problems the error estimator

$$\sum_{T\in\mathcal{T}} \eta_T^2 + \sum_{F\in\mathcal{F}} \eta_F^2$$

is equivalent to the hierarchical error estimator

$$\sum_{\psi\in\Psi_{TF}^{\oplus}} \frac{\langle f - Au_h, \psi\rangle^2}{\|\psi\|_A^2},$$

if one extends the ansatz space S_h appropriately:

- for $d = 2$ by locally quadratic "bubbles" on the midpoints of the edges (as in DLY) and by additional cubic "bubbles" in the barycenter of the triangles, whereby the saturation property is satisfied, i.e., the error estimator is efficient; in actual computation this space extension beyond DLY has shown not to pay off;
- for $d = 3$ beyond DLY locally cubic "bubbles" in the barycenters of the triangular faces and quartic "bubbles" in the barycenters of the tetrahedra, whereby, however, the saturation assumption is still not satisfied.

6.1.5 Goal-oriented Error Estimation

Up to now, we have tacitly defined the error ϵ_h as the energy norm of the deviation e_h, which, due to the best approximation property of the Galerkin solution u_h in elliptic PDEs, seemed to be self-evident. In practical applications, however, this is often not the quantity of interest. In structural mechanics, e.g., one might be interested in the *maximal* occurring stresses and strains (see Section 2.3) to be able to assess the risk of a crack or a rupture in material, whereas integral norms conceal the occurrence of stress cusps. In fluid dynamics simulations one is interested in the lift and the drag of wings, whereas an accurate computation of the airflow might be less important. Against this background, R. Rannacher et al. [15, 25] suggested estimating an error, not, as usual, w.r.t. an arbitrary norm, but w.r.t. a *quantity of interest*; this directs us to the concept of *goal-oriented* error estimation.

Let J denote such a quantity of interest. In the simplest case, J is a *linear functional* of the solution u so that we may write the deviation as

$$\epsilon_h = J(u) - J(u_h) = J(u - u_h) = \langle j, e_h \rangle$$

with a unique j due to the Theorem of Fischer–Riesz. The thus defined selectivity may be interpreted geometrically: error components orthogonal to j do not enter into the error estimation and thus need not be resolved by a finer and therefore more expensive discretization. With respect to the quantity of interest, a significantly smaller effort will be sufficient to achieve a prescribed accuracy; of course, the error in other directions may then well be larger.

In case an explicit error approximation $[e_h]$ is available, such as from (7.28), the goal-oriented error can be directly estimated by $[\epsilon_h] = \langle j, [e_h] \rangle$. From this, the localization $[\epsilon_h](T) = |\langle j, [e_h]|_T \rangle|$ can be immediately found. As already mentioned in the introductory section, this form of localization is reasonable only for elliptic problems with spatially smooth j, where, due to the strongly localized Green's function, the error distribution can be simultaneously used for mesh refinement. In other problems, however, e.g., when a discontinuous j occurs or when an explicit error approximation is missing, the goal-oriented error estimator can be more properly expressed by the

residual $r_h = f - Au_h$:

$$\epsilon_h = \langle j, e_h \rangle = \langle j, A^{-1} r_h \rangle = \langle A^{-*} j, r_h \rangle$$

In view of this relation we define the *weight function* $z \in H^1(\Omega)$ as the solution of the *dual* problem and thus obtain for the error

$$A^* z = j, \quad \epsilon_h = \langle z, r_h \rangle.$$

Because of this equation the whole approach is called the *dual weighted residual (DWR)* method. Normally z itself needs to be computed again. The elliptic model problems considered in this section are self-adjoint, i.e., we have $A^* = A$, which simplifies the computation of z. Due to the Galerkin orthogonality (4.14), however, an approximation $z_h \in S_h$ would just lead to the useless error estimate $[\epsilon_h] = 0$, so that useful approximations can only be determined from an extended FE space. In fact, approximations $z_h \in S_h^+$ can be obtained by the very methods treated above, such as hierarchical error estimators as described above (see Section 6.1.4), or gradient recovery (see Section 6.1.3).

Remark 6.16. In the course of time, goal-oriented error estimators have been suggested for a series of problem classes, among them for parabolic equations (see Section 9.2.3), for fluid dynamics [23] or for more general optimization problems [24, 212]. Comprehensive surveys are [15, 25].

6.2 Adaptive Mesh Refinement

In this section we present the actual construction of adaptive meshes, i.e., meshes adapted to the problem to be solved. We restrict our attention to the special case of hierarchical simplicial meshes. Starting from a coarse mesh with a moderate number of nodes, a sequence of nested, successively finer meshes is constructed. For that purpose we need two algorithmic modules:

- *marking strategies* on the basis of *local a posteriori error estimators* (see Section 6.1), by means of which we select edges, triangles, or tetrahedra for refinement; as an efficient example we present the method of local error extrapolation in Section 6.2.1;
- *refinement strategies*, i.e., a system of rules, by which we determine how meshes with marked elements should be conformally refined, without generating too obtuse interior angles (cf. Section 4.4.3); in Section 6.2.2 we introduce a selection of efficient techniques for *simplicial elements*, i.e., for triangles ($d = 2$) or tetrahedra ($d = 3$).

The special role of simplicial elements comes from two properties:

- simplices in higher dimensions can be built from those of lower dimensions, e.g., tetrahedra from four triangles; this property is an advantage in the discretization of surfaces of three-dimensional bodies;
- simplices are comparably easy to subdivide into smaller simplices of the same dimension, e.g., triangles into four smaller similar triangles (i.e., with the same interior angles); this property is important for the construction of hierarchical meshes; in Section 6.2.2 we will give suitable illustrative examples.

6.2.1 Equilibration of Local Discretization Errors

Before we turn to methods of mesh refinement, we want to investigate theoretically what an adaptive mesh should look like. The basis for this is a generalization of the error estimates from Section 4.4.1.

Adaptive Mesh Model. Assertions on nonuniform meshes are hard to make because of their necessarily discrete structure. That is why we turn to a simple model and consider continuous *mesh size functions* $h : \Omega \to \mathbb{R}$. For a number N of mesh nodes we obtain

$$N \approx c \int_\Omega h^{-d} \, \mathrm{d}x \quad \text{with} \quad c > 0.$$

This continuous model includes a *pointwise* variant of the interpolation error estimate for linear finite elements from Lemma 4.18:

$$|\nabla(u - I_h u)| \leq c h |u''| \quad \text{almost everywhere in } \Omega \, .$$

Integration and application of the Poincaré inequality supplies the global error estimate

$$\|u - I_h u\|^2_{H^1(\Omega)} \leq c^2 \int_\Omega h^2 |u''|^2 \, \mathrm{d}x.$$

Let us now fix the number of nodes and search for a mesh size function h that minimizes this error bound. For this purpose we need to solve the constrained optimization problem

$$\min_h \int_\Omega h^2 |u''|^2 \, \mathrm{d}x \quad \text{subject to} \quad c \int_\Omega h^{-d} \, \mathrm{d}x = N.$$

According to Appendix A.6 we couple the constraints by some positive Lagrange multiplier $\lambda \in \mathbb{R}$ to the objective functional and thus obtain the saddle point problem

$$\max_\lambda \min_h \left[\int_\Omega h^2 |u''| \, \mathrm{d}x - \lambda \left(c \int_\Omega h^{-d} \, \mathrm{d}x - N \right) \right].$$

For fixed λ differentiation of the integrand supplies the pointwise optimality condition

$$2h|u''|^2 - c\lambda d \; h^{-(d+1)} = 0.$$

The core of this relation is the proportionality

$$h^{d+2} = s^2 |u''|^{-2} \quad \text{with} \quad s^2 = c\lambda d/2 \in \mathbb{R}. \tag{6.38}$$

After this preparation we again turn to discrete meshes and replace the continuous mesh size function h by a piecewise constant function. Accordingly, we replace (6.38) by the approximate local relation

$$h^{d+2} \approx s^2 |u''|^{-2}.$$

For the *local error contributions* on an element $T \in \mathcal{T}$ we then get

$$\|u - I_h u\|^2_{H^1(T)} \le c \int_T h_T^2 |u''|^2 \, \mathrm{d}x = c h_T^{-d} \int_T h_T^{d+2} |u''|^2 \, \mathrm{d}x \approx c s^2, \tag{6.39}$$

where we have used $h_T^{-d} \int_T \mathrm{d}x = c$ with the usual generic constant c.

Conclusion. *In an optimal adaptive mesh all local error contributions have (roughly) the same size.*

This means that any strategy for adaptive mesh refinement should aim at an *equilibration of the local discretization errors*: Starting from some given mesh, exactly those elements or edges should be marked whose contribution to the error estimate exceeds some *threshold value*. Various marking strategies differ only in the choice of such a threshold.

In the absence of a uniform mesh size an error estimate as in Lemma 4.18 is inappropriate for adaptive meshes. However, if we choose the number of nodes as a measure, then, on the basis of the above relation (6.39), we obtain the following generalization of Lemma 4.18.

Theorem 6.17. *Let $\Omega \subset \mathbb{R}^d$ be a polytope, $u \in H^2(\Omega)$ and $\mathcal{T}$ a mesh over Ω, which satisfies the equilibration condition*

$$c^{-1} s \le \|u - I_h u\|_{H^1(T)} \le c s \quad \textit{for all } T \in \mathcal{T} \tag{6.40}$$

for some $s > 0$ and a constant c depending only on the interior angles of the elements. Then the following error estimate holds:

$$\|u - I_h u\|_{H^1(\Omega)} \le c N^{-1/d} |u|_{H^2(\Omega)}. \tag{6.41}$$

Proof. First we define a function z, piecewise constant on $\mathcal{T}$, by

$$z^2|_T = |T|^{-1} \|u\|^2_{H^2(T)} \quad \text{where} \quad |T| = c h_T^d.$$

Then Lemma 4.18, which we here apply locally on each element $T \in \mathcal{T}$, yields

$$s \le c \|u - I_h u\|_{H^1(T)} \le c h_T \|u\|_{H^2(T)} \le c h_T^{1+d/2} z.$$

From this, we obtain for the number $M = |\mathcal{T}|$ of elements

$$M = \sum_{T\in\mathcal{T}} 1 = c \sum_{T\in\mathcal{T}} \int_T h_T^{-d}\,\mathrm{d}x = c \int_\Omega h^{-d}\,\mathrm{d}x \le c \int_\Omega \left(\frac{s}{z}\right)^{-\frac{2d}{2+d}}\,\mathrm{d}x$$

and thus

$$s \le c \left(M^{-1} \int_\Omega z^{\frac{2d}{2+d}}\,\mathrm{d}x \right)^{\frac{2+d}{2d}}.$$

Because of the equilibration condition (6.40) the total error can be estimated as

$$\|u - I_h u\|^2_{H^1(\Omega)} = s^2 M \le c M^{1-\frac{2+d}{d}} \left(\int_\Omega z^{\frac{2d}{2+d}}\,\mathrm{d}x \right)^{\frac{2+d}{d}} \le c M^{-2/d} \|z\|^2_{L^q(\Omega)}$$

where $q = 2d/(2+d) < 2$. With $M \ge cN$ and $\|z\|_{L^q(\Omega)} \le c\|z\|_{L^2(\Omega)} = |u|_{H^2(\Omega)}$ we finally arrive at the assertion (6.41). □

Attentive readers may have observed that the estimate

$$\|z\|_{L^q(\Omega)} \le c\|z\|_{L^2(\Omega)}$$

at the end of the above proof is rather generous. In fact, the result (6.41) can be shown for less regular functions $u \in W^{2,q}(\Omega)$, so that domains with reentrant corners are also covered (see our corresponding Example 6.23 in the subsequent Section 6.3.2). However, we do not want to go deeper into this theory, but instead refer to [162].

Based on the Galerkin optimality (4.15), Theorem 6.17 immediately supplies the corresponding generalization of Theorem 4.19.

Corollary 6.1. *Under the assumptions of Theorem* 6.17 *the energy error of an FE solution* u_h *of*

$$a(u_h, v_h) = \langle f, v_h\rangle, \quad v_h \in S_h$$

satisfies the error estimate

$$\|u - u_h\|_a \le c N^{-1/d} |u|_{H^2(\Omega)} = c N^{-1/d} \|f\|_{L^2(\Omega)}. \tag{6.42}$$

Local Error Extrapolation. In what follows we introduce an iterative algorithmic strategy for realizing the desired equilibration approximately; this was suggested in 1978 by I. Babuška and W. C. Rheinboldt [13] for elliptic problems. In Volume 1, Section 9.7, we have already presented this strategy using the simple example of numerical quadrature.

As a basis we require one of the local error estimators worked out in Section 6.1 above. For ease of presentation, we first treat finite elements in 1D and later consider the difference in higher space dimensions for different estimators $[\epsilon(T)]$, defined on triangles T, and $[\epsilon(E)]$, defined on edges. We want to construct a hierarchical refinement strategy aiming at an equilibration of the discretization error over the elements

$T \in \mathcal{T}$ of a given mesh. Let $T_0 = (t_l, t_m, t_r)$ denote a selected element with left and right boundary nodes t_l, t_r and a midpoint node t_m. By subdivision of this element subelements T_l and T_r are generated, where

$$T_l := \left(t_l, \frac{t_l + t_m}{2}, t_m\right) \quad \text{and} \quad T_r := \left(t_m, \frac{t_r + t_m}{2}, t_r\right).$$

Upon refining twice, we obtain a *binary tree*, depicted in Figure 6.4.

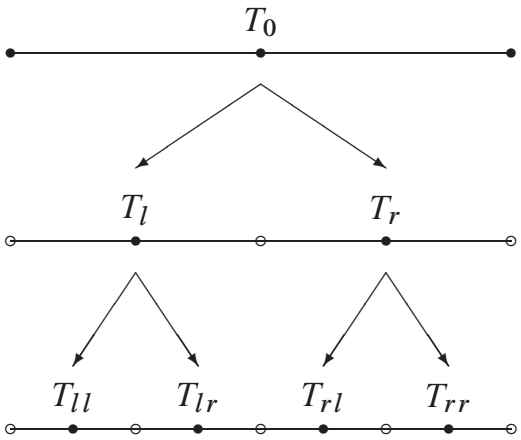

Figure 6.4. Double refinement of edge $T_0 = (t_l, t_m, t_r)$.

The figure selects three levels of a hierarchical refinement, which we will denote by $\mathcal{T}^-, \mathcal{T}, \mathcal{T}^+$ so that

$$T_0 \in \mathcal{T}^-, \quad \{T_l, T_r\} \in \mathcal{T}, \quad \{T_{ll}, T_{lr}, T_{rl}, T_{rr}\} \in \mathcal{T}^+.$$

With this notation, we now turn to the local error estimation. In Section 4.4.1 the error behavior of linear finite elements over a triangulation with mesh size h was given: Theorem 4.19 supplied $\mathcal{O}(h)$ for the error in the energy norm, Theorem 4.21 $\mathcal{O}(h^2)$ in the L^2-norm; in both theorems, however, the usually unrealistic assumption $u \in H^2(\Omega)$ was set, which would require convex domains Ω or those with smooth boundaries. For general domains, however, the local error behavior is hard to predict: in the interior of a domain one will expect $\mathcal{O}(h)$ for the local energy error, close to reentrant corners $\mathcal{O}(h^{1/\alpha-\epsilon})$ with a factor determined by the interior angle $\alpha\pi > \pi$, and finally $\mathcal{O}(h^\beta)$ for *rough* right-hand sides f with an a priori unknown exponent β. In this rather unclear situation I. Babuška and W. C. Rheinboldt [13] suggested making the ansatz

$$\epsilon(T) \doteq C h^\gamma, \quad T \in \mathcal{T} \tag{6.43}$$

for the discretization error, where h denotes the local mesh size, γ a local order and C a local problem dependent constant. The two unknowns C, γ can be determined from the numerical results of $\mathcal{T}^-$ and $\mathcal{T}$. Similar to (6.43) one then obtains for the edge level $\mathcal{T}^-$

$$\epsilon(T^-) \doteq C(2h)^\gamma, \quad T^- \in \mathcal{T}^-$$

as well as for level $\mathcal{T}^+$

$$\epsilon(T^+) \doteq C(h/2)^\gamma, \quad T^+ \in \mathcal{T}^+.$$

Without being forced to estimate the errors on the finest level $\mathcal{T}^+$ explicitly, these three relations permit, by *local extrapolation*, to gain the look-ahead estimate

$$[\epsilon(T^+)] \doteq \frac{[\epsilon(T)]^2}{[\epsilon(T^-)]}.$$

This gives us an idea in advance what effect a refinement of the elements in $\mathcal{T}$ would have. As the threshold value, above which we refine an element, we take the maximum local error that we would get by *global uniform refinement*, i.e., by refinement of *all* elements. This leads to the definition

$$\kappa^+ := \max_{T^+ \in \mathcal{T}^+} [\epsilon(T^+)]. \tag{6.44}$$

The situation is illustrated in Figure 6.5, which shows that the errors towards the

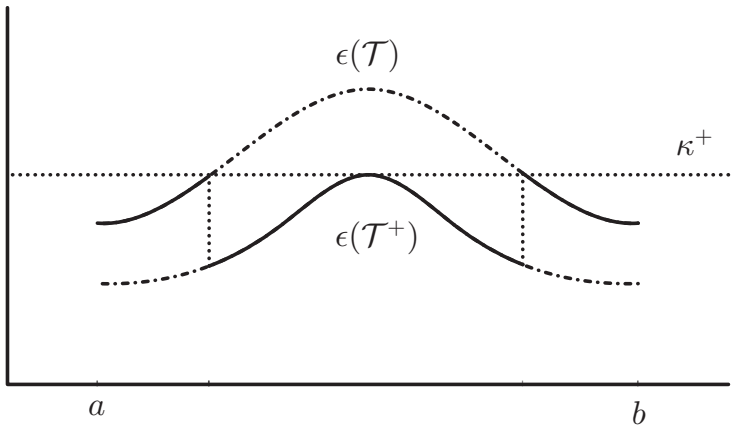

Figure 6.5. Iterative step towards discretization error equilibration: error distribution $\{\epsilon(\mathcal{T})\}$, extrapolated distribution $\{\epsilon(\mathcal{T}^+)\}$ for (not realized) uniform refinement, expected distribution (fat line) for realized nonuniform refinement.

right and left boundary already lie below the level that would be expected by further refinement; therefore we do not need to refine there. Only in the center part might a refinement pay off. So we arrive at the following *refinement rule*: Refine only such elements $T \in \mathcal{T}$, for which

$$[\epsilon(T)] \geq \kappa^+, \quad T \in \mathcal{T}.$$

Application of this rule supplies an expected error distribution, depicted in Figure 6.5 by thick lines. Compared with the original distribution, we recognize that this strategy has brought us one step closer to the desired equilibration of the local errors – assuming that the actual local errors arising in the next refinement step roughly behave as

expected. Upon repetition of the procedure in subsequent refinement steps the distribution will get successively "smaller", i.e., we will iteratively approach the equilibration up to a width of 2^{γ} (see the analysis in Volume 1, Section 9.7.2).

The result of local error extrapolation is a *marking* of elements which, in view of the nested structure of hierarchical meshes, directly carries over to higher space dimensions. For $d > 1$, however, the *edges* of a triangulation may be not completely nested, since by subdivision of elements new edges may be generated not contained in the edge set on coarser levels (cf. the subsequent section). When using the edge-oriented DLY error estimator (6.36) for local error extrapolation, the question of how to deal with such edges arises. The simplest way of dealing with parentless edges has turned out to be to skip them in the maximum search (6.44), which will merely lead to a possibly slightly lower threshold value κ^{+}.

6.2.2 Refinement Strategies

In this section we start, as an example, from a *triangulation with marked edges*. The aim now is to generate a refined mesh which is both *conformal* and *shape regular*. In two or three space dimensions this task is all but trivial, since the local subdivision of an element will, in general, lead to a nonconformal mesh. To construct a conformal mesh, the neighboring elements have to be suitably subdivided such that this process does not spread over the whole mesh. At the same time, one must take care that the approximation properties of the FE ansatz spaces on the meshes do not deteriorate, i.e., that the maximal interior angles remain bounded (cf. (4.65) in Section 4.4.3). In some cases, a constraint on the minimal interior angles is of interest (see Section 7.1.5). In view of an extension of multigrid methods from Section 5.5 to adaptive meshes one is also interested in the construction of nested mesh hierarchies.

Definition 6.18. *Mesh hierarchy.* A family $(\mathcal{T}_k)_{k=0,\dots,j}$ of meshes over a common domain $\Omega \subset \mathbb{R}^d$ is called *mesh hierarchy* of depth j, if for each $T_k \in \mathcal{T}_k$ there exists a set $\hat{T} \subset \mathcal{T}_{k+1}$ with $T_k = \bigcup_{T \in \hat{T}} T$. The *depth* of an element $T \in \bigcup_{k=0,\dots,j} \mathcal{T}_k$ is given by $\min\{k \in \{0, \dots, j\} : T \in \mathcal{T}_k\}$; the depth for faces, edges, or nodes is defined in a similar way.

An interesting property of mesh hierarchies is the embedding of the nodal set: $\mathcal{N}_k \subset \mathcal{N}_{k+1}$ (see Exercise 7.7). In view of the objective criteria conformity, shape regularity, and mesh hierarchy essentially two strategies have evolved, the bisection and so-called red-green refinement.

Bisection

The simplest method, already worked out in 1984 by M.-C. Rivara [175], is the bisection of marked edges coupled with a decomposition into two neighboring simplices (see Figure 6.6).

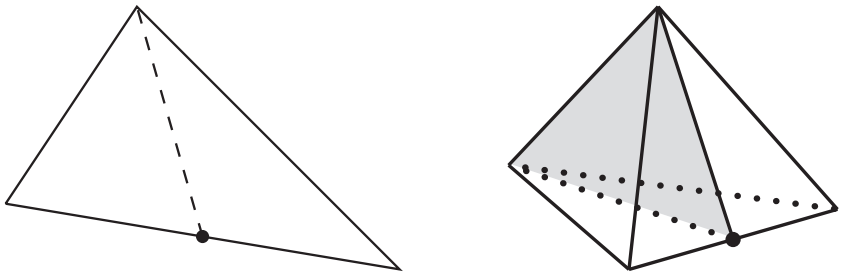

Figure 6.6. Bisection of triangles (*left*) and tetrahedra (*right*).

In order to maintain shape regularity, the longest edge is usually selected for bisection. The refinement is continued as long as hanging nodes still exist (see Figure 6.7).

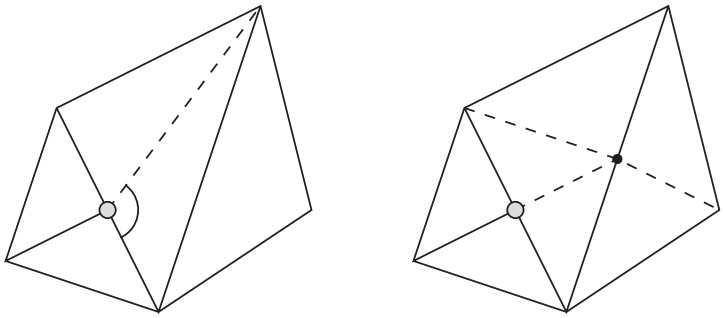

Figure 6.7. Conformal mesh refinement by continued bisection in 2D.

In general, the refinement process terminates before the mesh has been uniformly refined (cf. Exercise 6.2): If a hanging node is situated on a "longest" edge, then further bisection does not generate another new node. Otherwise a new hanging node is generated on a longer edge. Hence, the continuation of the refinement can only affect elements with longer edges.

Red-green refinement

A more tricky method dates back to R. E. Bank et al. [18] from 1983, who had introduced some "red-green" refinement strategy in his adaptive FE package PLTMG (in 2D).[2]

[2] R. E. Bank started, however, from the algorithmic basis of triangle oriented error estimators in 2D.

Red Refinement. In this kind of refinement (named as such by R. E. Bank) the elements are subdivided into 2^d smaller simplices, where exactly the midpoints of all edges are introduced as new nodes. In 2D, this yields a unique decomposition of a triangle into four geometrically similar triangles; as the interior angles do not change, the shape regularity is preserved (see Figure 6.8, left).

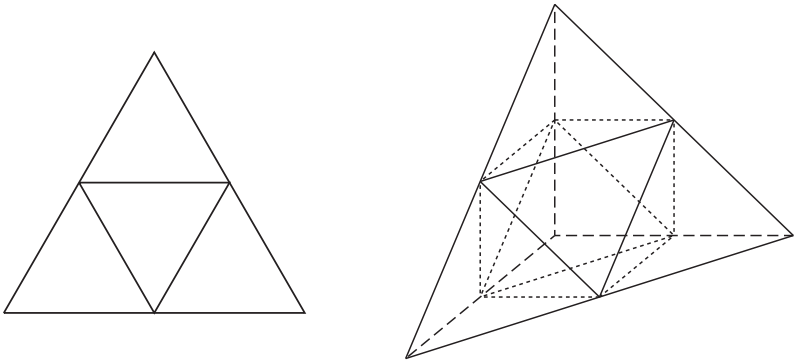

Figure 6.8. Red refinement of triangles (*left*) and tetrahedra (*right*).

In 3D, things are more complex geometrically: here one first gets a unique decomposition into four tetrahedra at the vertices of the original tetrahedron as well as an octahedron in the center (see Figure 6.8, right). By selecting one of the diagonals as the common new edge, the octahedron may be decomposed into four further tetrahedra, which, however, are no longer similar to the original tetrahedron, so that the shape regularity will now depend on the refinement depth. In fact, the selection of the diagonal to be subdivided must be done with great care in order to be able to prove that shape regularity is preserved. As a rule, the shortest diagonal is selected.

In order to preserve conformity of the triangulation, elements neighboring already subdivided elements must also be subdivided. In such a strategy, however, the red refinement would inevitably spread over the whole mesh and end up with a global uniform refinement. That is why a remedy has been thought of early: Red refinements are performed locally only, if (in 2D) *at least two* edges of a triangle are marked. By this rule, the refinement process will spread, but will remain sufficiently local.

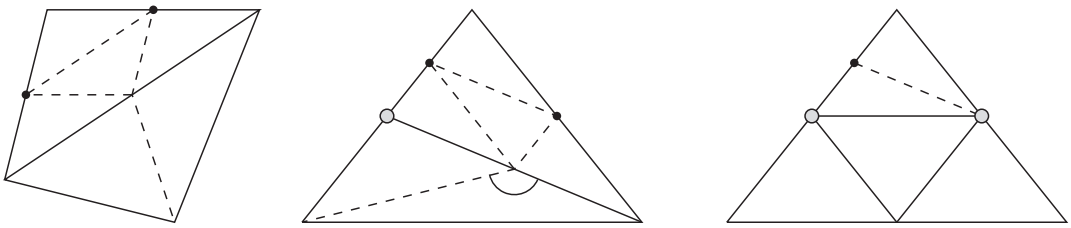

Figure 6.9. Red-green refinement strategy in 2D. *Left:* combination of a red and a green refinement at a marked edge with only one hanging node. *Center:* increase of the interior angle at continued green refinement. *Right:* shape regularity by removal of green edges before refinement.

Green Refinement. We now assume (for illustration again in 2D) that in a local triangle only one edge has been marked for refinement. In this case we introduce a new so-called *green* edge (again named by R. Bank) and subdivide the triangle into two triangles (see Figure 6.9, left). In this way, the refinement is not continued any further, which is why the green edges are called *green completion*. The interior angles of triangles from green refinements may be both larger and smaller than the ones of the original triangle. Hence, by continuation of green refinements, the shape regularity might suffer (see Figure 6.9, center). In order to assure shape regularity nevertheless, all green completions are removed from the mesh before the next refinement, possibly adding green edges once red refinement has been performed (see Figure 6.9, right). Along this line, shape regularity can be preserved.

In 3D, the process runs in a similar way, but requires three different types of green completions (see [36]), depending on whether one, two, or three edges are marked (see Figure 6.10). With four marked edges red refinement can again take place.

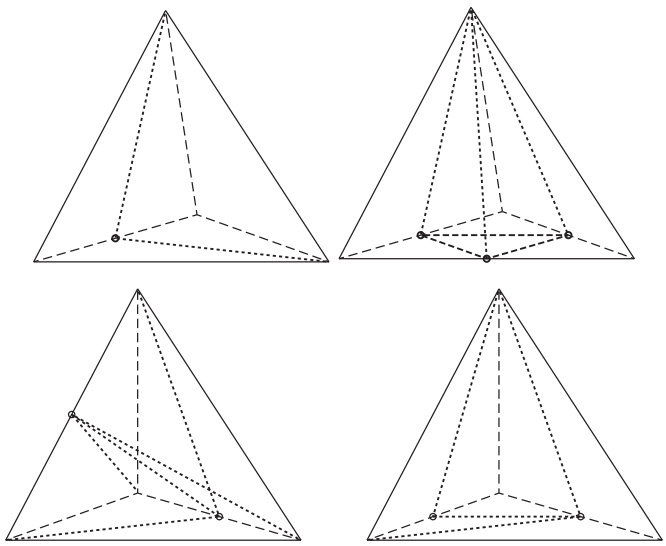

Figure 6.10. Four different types of green completions with tetrahedra [36]. *Top left:* one edge marked. *Top right:* three edges of a face marked. *Bottom:* two edges marked. In all other cases red refinement is performed.

Remark 6.19. The preservation of shape regularity by red refinement of simplices of arbitrary space dimension in finitely many similarity classes was already proven in 1942 by H. Freudenthal [92], for the adaptive mesh refinement in 3D, however, rediscovered only 50 years later [27, 102]. The stability of selecting the shortest diagonal has been shown in [136, 230].

Remark 6.20. If in the course of the refinement of meshes simultaneously both the mesh size h and the local order p of the finite elements are adapted, one speaks of hp-refinement. We have skipped this efficient method, since throughout most of this beek we treat elements of lower order. We refer interested readers to the book of C. Schwab [183].

6.2.3 Choice of Solvers on Adaptive Hierarchical Meshes

Up to this point we have elaborated how to construct adaptive hierarchical meshes. Let us now briefly outline the kind of solvers that can be combined with this kind of meshes. As in Section 5.5 above, we build upon a mesh hierarchy and nested FE spaces:

$$S_0 \subset S_1 \subset \cdots \subset S_j \subset H_0^1(\Omega).$$

For boundary value problems this corresponds to a hierarchical system of linear equations

$$A_k u_k = f_k, \quad k = 0, 1, \ldots, j \tag{6.45}$$

with symmetric positive definite matrices A_k of dimension N_k. In contrast to the situation in Section 5.5, where we started from a fixed fine mesh $k = j$, for *adaptive* hierarchical meshes the order from coarse mesh $k = 0$ to fine mesh $k = j$ is now compulsory. In particular, the linear equation systems (6.45) have to be solved on *each* level k in turn, in order to enable the adaptive construction of the mesh hierarchy.

This successive solution can be performed in two principal ways:

1. by methods from numerical linear algebra, such as the *direct elimination methods* presented in Section 5.1, or iterative eigenvalue solvers (see, e.g., Volume 1, Section 8.5), which we here will only demonstrate in a nontrivial example (see the subsequent Section 6.4);
2. by adaptive multigrid methods, which we will discuss in the subsequent Chapter 7; there it is sufficient to reduce the algebraic error $\|u_j - \hat{u}_j\|$ down to the size of the discretization error $\|u - u_j\|$ – an obviously reasonable complexity reduction. Note that a nested iteration approach for multigrid as discussed in Exercise 5.8 comes in quite natural in the context of adaptive mesh refinement. As usual, we suppose that at least the system $A_0 u_0 = f_0$ can be solved by direct elimination methods. Of course, all kinds of intermediate forms are realized where direct solvers are used up to refinement level $0 < k_{\max} < j$.

As a general rule, direct elimination methods, due to their superlinear growth of the computational amount are the methods of choice for "smaller" problems, whereas iterative methods with linearly growing amount are inevitable for "large" problems (see Section 5.5.3 above). As already mentioned there, adaptive meshes are particularly crucial for direct methods. This is illustrated in the technological example given in the subsequent Section 6.4.

6.3 Convergence on Adaptive Meshes

Up to now, we have tacitly assumed that the meshes refined by means of localized error estimators actually generate a sequence $\{\mathcal{T}_k\}$ of meshes with corresponding FE

solutions $u_k \in S_k = S_h^1(\mathcal{T}_k)$ that essentially realize the optimal convergence order

$$\|u - u_k\|_A \leq c N_k^{-1/d} \|f\|_{L^2(\Omega)}$$

from Corollary 6.1. Despite convincing results, this had not been theoretically confirmed for a long time – not even the pure convergence $u_k \to u$. After 1996 W. Dörfler [79] managed to prove convergence of adaptive solvers by introducing his concept of *oscillation* (see Definition 6.3). The *optimal* convergence rate was shown in 2004 by P. Binev, W. Dahmen, and R. De Vore [29], although by application of not very practicable coarsening steps. In 2007 R. Stevenson [193] was able to prove convergence for the typical algorithms in common use.

In the following Section 6.3.1 we give such a convergence proof. In the subsequent Section 6.3.2 we illustrate the computational situation by an elementary 2D example with reentrant corners.

6.3.1 A Convergence Proof

In order to work out the main aspects of a convergence proof for adaptive methods, we first define a model algorithm for hierarchical refinement.

Adaptive Model Algorithm. We focus our attention to an idealized situation with linear finite elements S_h^1 on a triangular mesh ($d = 2$). For the marking of edges we use a hierarchical error estimator. As extension $S_h^{\oplus}$ instead of quadratic "bubbles" as in the DLY estimator (see Figure 6.1, right), we here use piecewise linear "bubbles" (see Figure 6.1, left). As refinement rule we combine red refinement with hanging nodes: Whenever a hanging node $\xi \in \mathcal{N}$ is midpoint of an edge (ξ_l, ξ_r) that on one side has an unrefined, on the other side a refined triangle, then we use interpolation by linear finite elements and require that $u_h(\xi) = (u_h(\xi_l) + u_h(\xi_r))/2$. Ee allow at most one hanging node per edge. The thus generated meshes are – up to the missing green completions – identical with red-green refined meshes.

By this model of an adaptive algorithm, we ensure, on the one hand, that the generated sequence $\{S_k\}$ of ansatz spaces is nested, i.e., that $S_k \subset S_{k+1}$ holds, and, on the other hand, that the ansatz functions applied for error estimation are contained in the ansatz space of the next level of mesh refinement. In order to construct a convergent method in this frame, it seems reasonable to impose a refinement of those triangles with the largest error indicator. However, not too many triangles should be refined, in particular those with especially small error indicator, since then the fast growing number N_k of nodes would not coincide with an associated error reduction and thus the convergence rate would be suboptimal only. We therefore choose as limiting case in each refinement step an edge $E \in \mathcal{E}$ with maximum error indicator, i.e., where $[\epsilon_h(E)] \geq [\epsilon_h(e)]$ holds for all $e \in \mathcal{E}$, and mark exactly the triangles incident with E for refinement. Hence, due to the locality of the refinement, the number of additional nodes is bounded by some constant m.

For this idealized case we are now able to show convergence of the FE solution $u_k \in S_k$. For the proof, we employ the *saturation* instead of the *oscillation* assumption, both of which were shown to be equivalent by W. Dörfler and R. H. Nochetto [80] in 2002 (see Theorem 6.8).

Theorem 6.21. *Let the saturation assumption* (6.22) *be satisfied. Then the FE solution u_j converge to u and there exist constants $c, s > 0$ such that*

$$\|u - u_j\|_A \leq c N_j^{-s} \|u - u_0\|_A, \quad j = j_0, \ldots.$$

Proof. We use the notation of Section 6.1.4. From the error localization (6.37) and the Galerkin orthogonality we obtain due to $\psi_E \in S_{k+1}$ the following result for the edge oriented error estimator with linear finite elements ($k = 0, \ldots, j$)

$$\begin{aligned}[\epsilon_k(E)]\,\|\psi_E\|_A = \langle f - Au_k, \psi_E\rangle = \langle A(u - u_k), \psi_E\rangle = \langle A(u_{k+1} - u_k), \psi_E\rangle \\ \leq \|u_{k+1} - u_k\|_A \|\psi_E\|_A,\end{aligned}$$

which also holds for the modulus, of course. Thus we can estimate the error reduction. Again due to the Galerkin orthogonality we have

$$\|u - u_k\|_A^2 = \|u - u_{k+1}\|_A^2 + \|u_{k+1} - u_k\|_A^2 \geq \|u - u_{k+1}\|_A^2 + [\epsilon_k(E)]^2,$$

which immediately supplies

$$\epsilon_{k+1}^2 \leq \epsilon_k^2 - [\epsilon_k(E)]^2. \tag{6.46}$$

For the following let $c > 0$ again denote a generic constant. As $[\epsilon_k(E)]$ is maximal, Theorem 6.15 yields

$$[\epsilon_k(E)]^2 \geq \frac{1}{|\mathcal{E}_k|} \sum_{e \in \mathcal{E}_k} [\epsilon_k(e)]^2 \geq c \frac{[\epsilon_k]^2}{N_k} \geq c \frac{\epsilon_k^2}{N_k},$$

where we have bounded the number of edges by the number of nodes. We further get $N_{k+1} \leq N_k + m$, i.e., $N_{k+1} \leq (k+1)m + N_0$, as well as $N_0 \geq 1$, which leads to

$$[\epsilon_k(E)]^2 \geq c \frac{\epsilon_k^2}{k+1}.$$

Upon inserting this into (6.46) we obtain after some brief calculation

$$\epsilon_j^2 \leq \left(1 - \frac{c}{j}\right) \epsilon_j^2 \leq \prod_{k=0}^{j-1} \left(1 - \frac{c}{k+1}\right) \epsilon_0^2 \leq c j^{-c} \epsilon_0^2 \leq c N_j^{-s} \epsilon_0^2$$

and thus convergence as stated above. □

For the model algorithm studied here, Theorem 6.21 supplies convergence, but the convergence rate remains unknown. With considerably more effort in the proof, one

may also obtain convergence rates that are optimal in the sense of principal approximability of the solution u (see, e.g., [193]). However, even with optimal convergence rate, the method described does not achieve optimal computational amount: As in each refinement step, only a bounded number of new degrees of freedom is added, the number of refinement steps is very large so that the total amount W_j is of the order $\mathcal{O}(N_j^2)$. In order to keep an optimal amount of the order $\mathcal{O}(N_j)$, in practice sufficiently many triangles are refined, which, however, requires slightly more complex proof techniques.

Remark 6.22. A theoretical concept as described here underlies the adaptive finite element code ALBERTA (for 2D and 3D), written by ALfred Schmidt and KuniBERT Siebert[3] and published in [180]; this program uses the residual based error estimator (see Section 6.1.1), coupled with local error extrapolation due to [13] and *bisection* for refinement (see Section 6.2.2).

6.3.2 An Example with a Reentrant Corner

We have repeatedly indicated that the asymptotic behavior of the discretization error with uniform meshes of mesh size h changes at reentrant corners due to the corner singularities occurring there. For $d = 2$ we characterize the interior angle by $\alpha\pi$ with $\alpha > 1$. From Remark 4.20 we take that the error behavior for *Dirichlet problems* in the *energy norm* as

$$\|u - u_h\|_a \leq c h^{\frac{1}{\alpha}-\epsilon}, \quad \epsilon > 0. \tag{6.47}$$

For one-sided *homogeneous Neumann problems* the value α in the above exponent has to be substituted by 2α, if the problem can be extended symmetrically to a Dirichlet problem. Analogously, from Remark 4.22 we have for the error behavior in the L^2-*norm*

$$\|u - u_h\|_{L^2(\Omega)} \leq c h^{\frac{2}{\alpha}-\epsilon}, \quad \epsilon > 0. \tag{6.48}$$

Here, too, one-sided Neumann problems would require the factor α in the above exponent to be replaced by 2α. For an illustration of adaptive meshes with reentrant corners we modify an example that originally was originally constructed by R. E. Bank [16].

Example 6.23. *Poisson model problem on slit domain.* Consider the simple Poisson equation

$$-\Delta u = 1$$

on a circular domain with slit along the positive axis, as depicted in Figure 6.11. On the left, homogeneous Dirichlet boundary conditions are prescribed on both sides of the slit. On the right, however, homogeneous Neumann boundary conditions are prescribed at the lower side of the slit (observe the orthogonality of the level lines), while

[3] Originally, the program was called ALBERT, which, however, caused some legal objection from economy.

on the upper side homogeneous Dirichlet boundary conditions are prescribed. For both cases, the numerical solution is represented by level lines (see bottom row of Figure 6.11). In the rows above the meshes $\mathcal{T}_j$ generated adaptively by linear finite elements are shown for $j = 0, 5, 11$. The circular boundary is approximated by successively finer polygonal boundaries. For the generation of the adaptive meshes the DLY error estimator from Section 6.1.4, local extrapolation from Section 6.2.1 and red-green refinement from Section 6.2.2 have been selected. The numerical solution was done by one of the multigrid solvers which will be described in Section 7.3, but which does not play a role in the present context.

For this example, approximation theory (see (6.47) and (6.48)) supplies the following error behavior ($\epsilon \approx 0$):

- *Dirichlet* problem: $\alpha = 2$, i.e.,

$$\text{ERRD-EN} := \|u - u_h\|_a^{(D)} \sim \mathcal{O}(h^{\frac{1}{2}}), \quad \text{ERRD-L2} := \|u - u_h\|_{L^2}^{(D)} \sim \mathcal{O}(h);$$

- *Neumann* problem: $2\alpha = 4$, i.e.,

$$\text{ERRN-EN} := \|u - u_h\|_a^{(N)} \sim \mathcal{O}(h^{\frac{1}{4}}), \quad \text{ERRN-L2} := \|u - u_h\|_{L^2}^{(N)} \sim \mathcal{O}(h^{\frac{1}{2}}).$$

The behavior of these four error measures can be nicely observed in double-logarithmic scale in Figure 6.12, top.

In Theorem 6.17 we presented a theory for *adaptive* meshes, the core of which was the characterization of the FE approximation error vs. the number N of unknowns instead of a characterization versus a mesh size h, which really works only in the uniform case. This theory also holds for reentrant corners. To apply it, we consider the above Poisson problem *without* reentrant corners, no matter whether with Dirichlet or with Neumann boundary conditions. For this case, *uniform* meshes as in Section 4.4.1 with $N = ch^d, d = 2$ would have led to an energy error

$$\|u - u_h\|_a \sim \mathcal{O}(h) \sim \mathcal{O}(N^{-\frac{1}{2}})$$

and an L^2-error

$$\|u - u_h\|_{L^2} \sim \mathcal{O}(h^2) \sim \mathcal{O}(N^{-1}) \ .$$

In fact, this same behavior is observed if one draws the *adaptive* meshes vs. $N = N_j$, as in Figure 6.12, bottom. The errors ERRD-EN and ERRN-EN only differ by some constant, which exactly reflects assertion (6.42). We have not given an analogous proof for the L^2-norm; nevertheless, the same behavior can be also seen for the error measures ERRD-L2 and ERRN-L2.

In conclusion, from these figures we obtain:

> *Adaptive meshes "remove" corner singularities by some implicit transformation.*

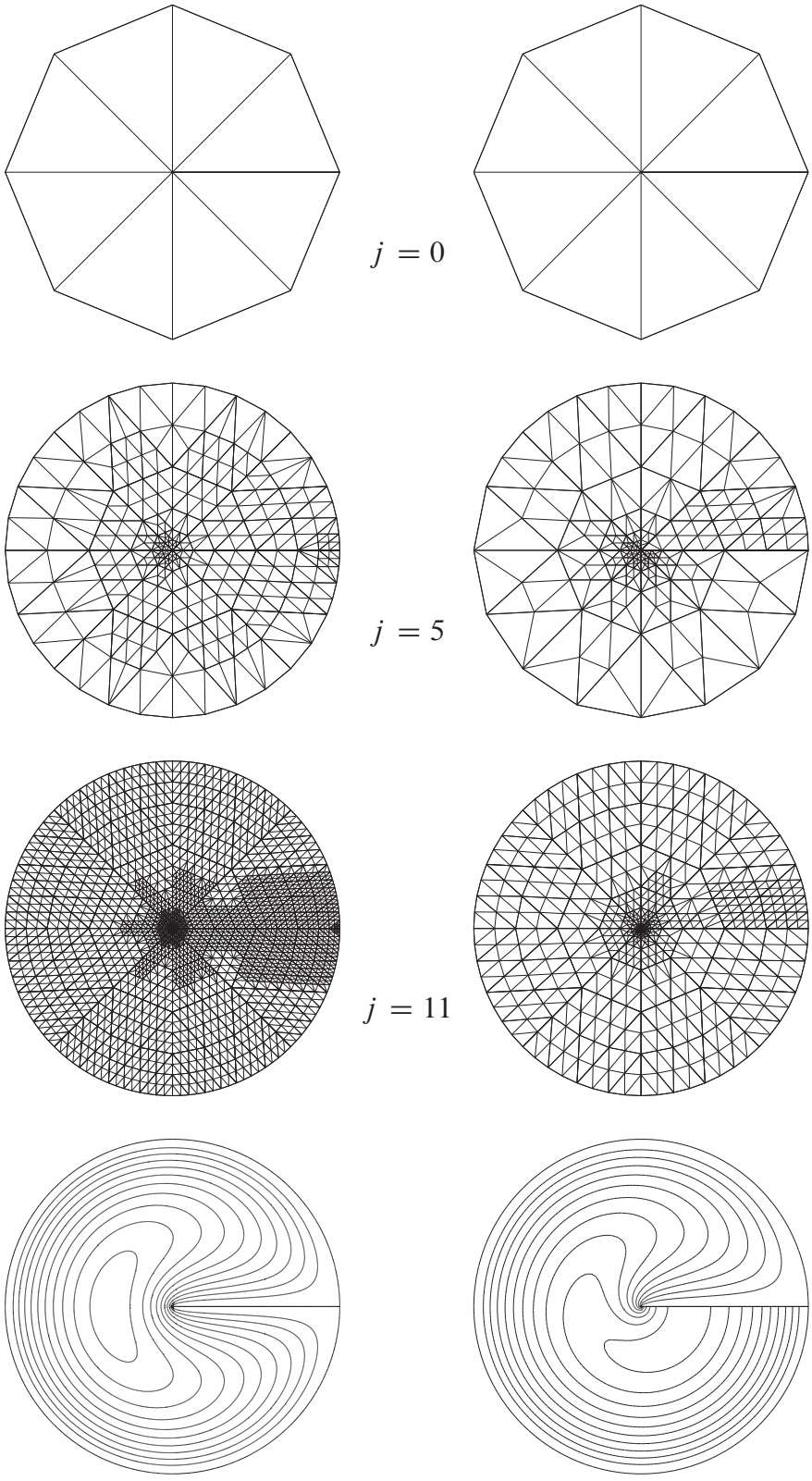

Figure 6.11. Example 6.23: adaptive mesh refinement history. *Left:* Dirichlet boundary conditions; coarse mesh with $N_0 = 10$ nodes, selected refined meshes with $N_5^{(D)} = 253$, $N_{11}^{(D)} = 2152$, level lines for the solution. *Right:* Neumann boundary conditions; same coarse mesh, refined meshes with $N_5^{(N)} = 163$, $N_{11}^{(N)} = 594$, level lines for the solution.

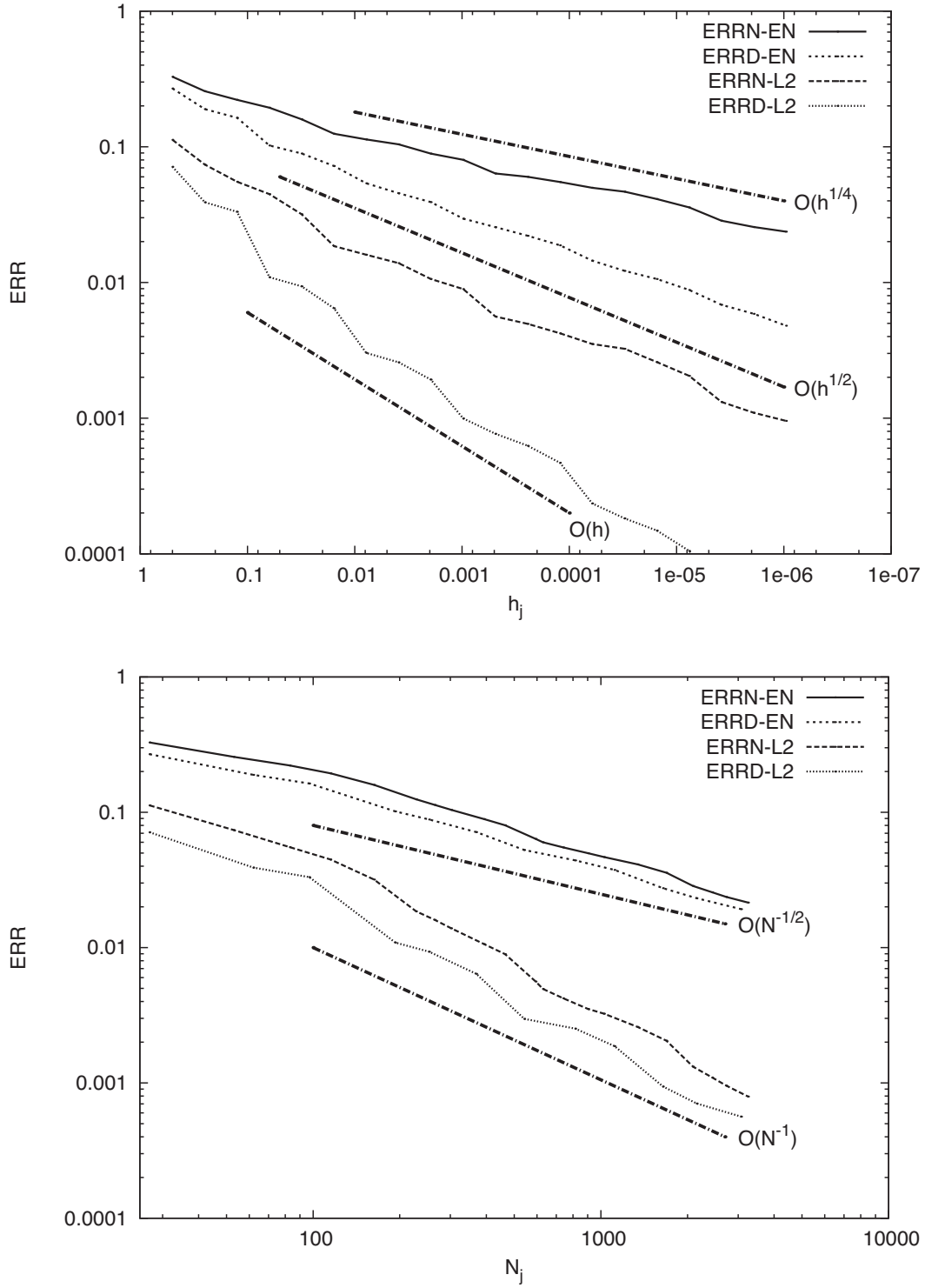

Figure 6.12. Example 6.23: adaptive mesh refinement history. *Top:* error measures vs. smallest mesh size $h = h_j$. *Bottom:* error measures versus number of nodes $N = N_j$. Cf. Theorem 6.17.

6.4 Design of a Plasmon-Polariton Waveguide

This type of semiconductor device has only rather recently become a focus of interest. It serves as biosensor as well as interconnect on integrated semiconductor chips. The nontrivial example presented here due to [46] will give a glimpse into the problem world of nanotechnology. As will turn out, we are led to a quadratic eigenvalue problem.

Figure 6.13 presents a scheme of such a special optical wave guide by showing some 2D cross section (see [26]): on a dielectric material 1 (with relative dielectric constant $\varepsilon_{r,1} = 4.0$) there lies a very thin silver strip with thickness $h = 24.5\,\text{nm}$ and width $b = 1\,\mu\text{m}$. A dielectric material 2 (with $\varepsilon_{r,2} = 3.61$) covers the device from above. The geometry is invariant in z-direction, i.e., the device is modelled as infinitely long.

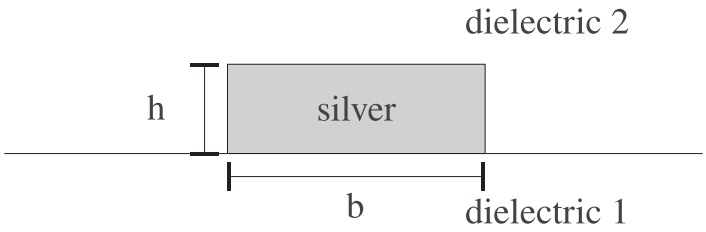

Figure 6.13. Schematic representation of a plasmon-polariton waveguide.

It is of technological interest that light fields spread along the z-direction in this structure but remain localized at the surface of the thin silver strip. As already introduced in Section 2.1.2, such fields are called *waveguide modes*. If light spreads on a metal surface, one speaks of *plasmon-polariton waveguides*.

For our present case we have $\omega = 2\pi/\lambda_0\, c$ with $\lambda_0 = 633\,\text{nm}$ and the light velocity c. For this choice of ω one has in silver (chemical notation: Ag) the relative dielectric constant $\varepsilon_{r,\text{Ag}} = -19 + 0.53\text{i}$. The large real part and nonvanishing imaginary part characterize the damping of the optical wave.

Mathematical Modelling. A hierarchy of optical models was given in Section 2.1.2, among them also a special model for waveguides. In the present case we require a variant of the model introduced there. For this purpose, we recall the *time-harmonic* Maxwell equations. We start from equation (2.5), merely exchanging $(E, \epsilon) \leftrightarrow (H, \mu)$. If we write the differential operator componentwise as vector product of the gradient (cf. Exercise 2.3), then we come up with the following equation for the electric field E:

$$\begin{bmatrix} \partial_x \\ \partial_y \\ \partial_z \end{bmatrix} \times \frac{1}{\mu} \begin{bmatrix} \partial_x \\ \partial_y \\ \partial_z \end{bmatrix} \times E(x,y,z) - \omega^2 \varepsilon(x,y,z) E(x,y,z) \;=\; 0.$$

The dielectric constant depends solely on the cross section coordinates, which means that $\varepsilon = \varepsilon(x, y)$. In our specific example we had $\mu_r = 1$, but in the general case the relative magnetic permeability would be $\mu_r = \mu_r(x, y)$.

For the waveguide modes we make an ansatz just as in (2.9), i.e.,

$$E(x, y, z) = \hat{E}(x, y)\mathrm{e}^{\mathrm{i}k_z z}$$

with a propagation coefficient k_z. As determining equation for the eigenvalue k_z we then get

$$\begin{bmatrix} \partial_x \\ \partial_y \\ \mathrm{i}k_z \end{bmatrix} \times \frac{1}{\mu} \begin{bmatrix} \partial_x \\ \partial_y \\ \mathrm{i}k_z \end{bmatrix} \times \hat{E}(x, y) - \omega^2 \varepsilon(x, y)\hat{E}(x, y) = 0. \tag{6.49}$$

Upon expanding the above terms (which we leave for Exercise 6.4) we find that the eigenvalue occurs both linearly and quadratically. Thus we encounter a *quadratic eigenvalue problem*. By some elementary transformation (also left for Exercise 6.4) this problem can be reformulated as a linear eigenvalue problem of double dimension (see, e.g., the survey article by F. Tisseur and K. Meerbergen [199]).

Discrete Formulation. In principle, the eigenvalue problem is stated in the whole $\mathbb{R}^2$. For its numerical discretization the problem must be restricted to some finite domain of interest. The aim is to compute the waveguide mode with the strongest localization around the silver strip, the so-called *fundamental mode*. For this purpose, it is enough to allocate homogeneous Dirichlet boundary conditions on a sufficiently large section around the silver strip such that the "truncation error" can be neglected. For discretization, *edge elements* of second order were applied, which we have skipped in this book (see J. C. Nédélec [158]). Along this line a very large algebraic linear eigenvalue problem with some nonsymmetric matrix arises. For its numerical solution the "inverse shifted Arnoldi method" is applied, a well-proven method of numerical linear algebra (see [143]). During the course of this algorithm sparse linear equations of up to several million unknowns have to be solved, which is done by the program package PARDISO [178].

Numerical Results. The design of a waveguide requires obviously the selection of an eigenvalue via some property of the corresponding eigenfunction: it should be "strongly localized" around the silver strip. In order to approach this goal, first 100 or so discrete eigenvalue problems have been simultaneously approximated on relatively coarse meshes and the localization of their eigenfunctions quantified. From this set, successively fewer eigenvalues have been selected on finer and finer meshes. The eigenvalue finally left over by this iterative procedure is, for technological reasons, required to high accuracy, i.e., it needs to be computed on rather fine meshes. In Figure 6.14 we show isolines of the computed intensity $\|E\|^2$, in logarithmic scale, the representation preferred by chip designers. One recognizes the desired localization

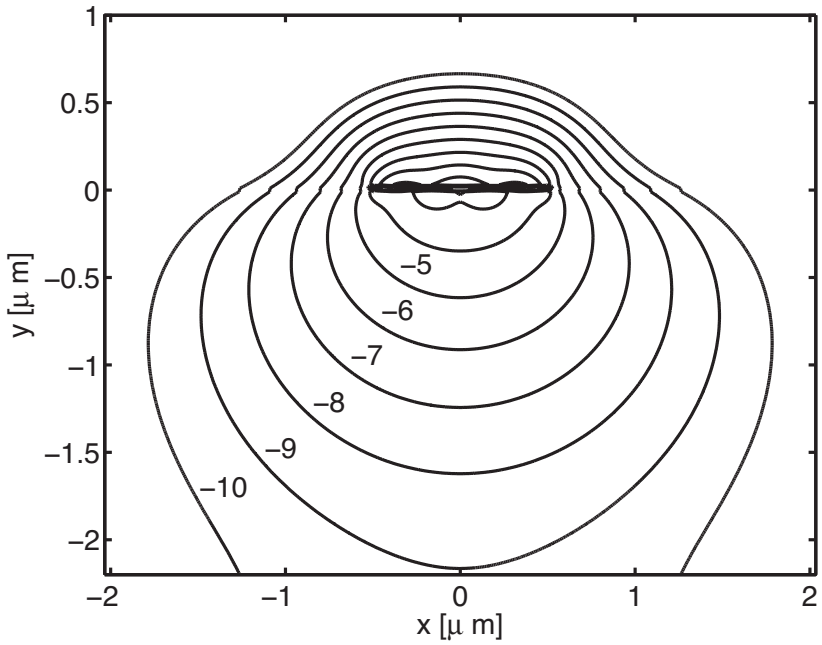

Figure 6.14. Plasmon-polariton waveguide: isolines of the field intensity $\|E\|^2$ (in logarithmic scale).

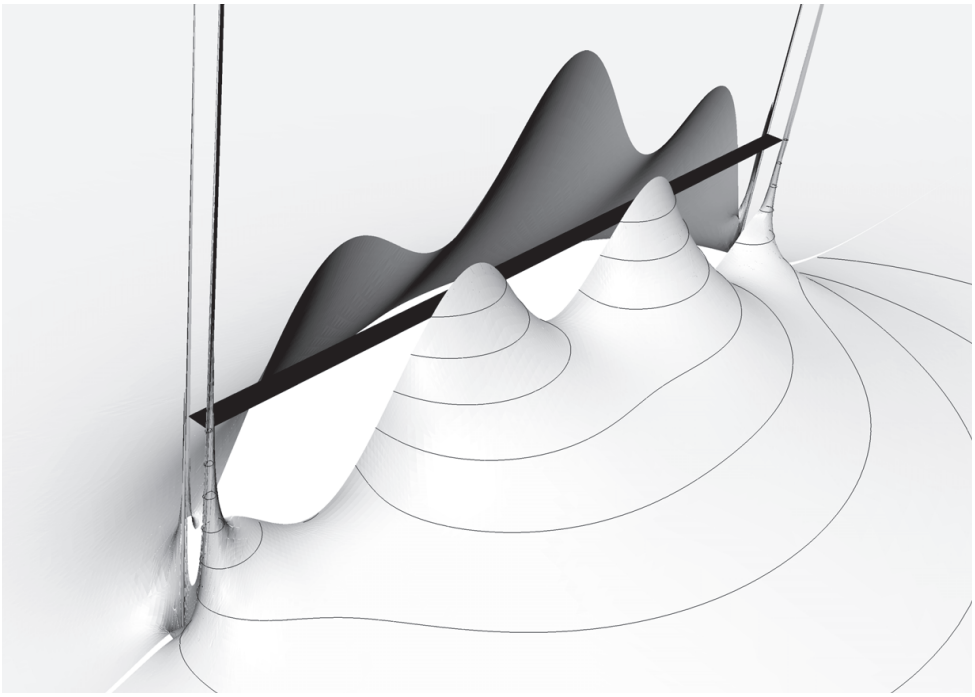

Figure 6.15. Plasmon-polariton waveguide: dominant y-component of the field E. The black line marks the silver strip. The singularities at the ends of the silver strip have been cut off.

around the silver strip. Moreover, the field exhibits an unpleasant number of *singularities* at the ends of the silver strip, which has been made especially visible in Figure 6.15.

Adaptive Meshes. Because of the singularities involved, an adaptive refinement of the FE mesh is of utmost importance. In [46], the residual based error estimator from Section 6.1.1 has been chosen; even though it is only determined up to some constant, it is very well suited as an error indicator in the presence of singularities.

In Table 6.1 the computed eigenvalues k_z vs. the number N of unknowns are listed. In practical applications, one needs the real part of k_z up to the fourth decimal place after the dot. The imaginary part should be computed up to two significant decimal digits.

Table 6.1. Plasmon-polariton waveguide: eigenvalue approximations k_z vs. adaptive mesh refinement ($k_0 = 2\pi/\lambda_0$).

N	$\Re(k_z)/k_0$	$\Im(k_z)/k_0$
1 830	2.003961e+00	1.031e−03
4 150	2.004239e+00	1.125e−03
8 757	2.003659e+00	1.071e−03
18 955	2.003339e+00	1.034e−03
38 329	2.003155e+00	1.011e−03
74 574	2.003061e+00	9.992e−04
148 872	2.003012e+00	9.929e−04
271 281	2.002986e+00	9.894e−04
528 178	2.002973e+00	9.875e−04
1 010 541	2.002965e+00	9.864e−04
1 903 730	2.002962e+00	9.858e−04
3 518 031	2.002960e+00	9.855e−04

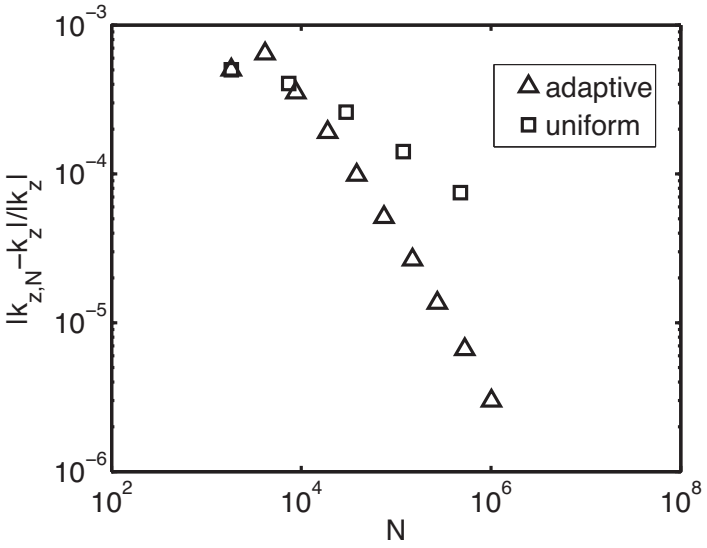

Figure 6.16. Plasmon-polariton waveguide: comparison of convergence of eigenvalue approximations k_z for adaptive versus uniform meshes.

From this, a bound for the relative discretization error of 10^{-6} is obtained. Figure 6.16 demonstrates a comparison of the discretization errors from the generated adaptive FE meshes compared with those from uniform FE meshes. Even with roughly one million unknowns, the uniform mesh refinement does not achieve the desired accuracy. In Table 6.1 the number of unknowns increases to even more than 3.5 million. The memory (32 GB) of the computer used for this part of the simulation did not permit any further uniform refinement beyond 4 million nodes. In view of the comparison of direct vs. iterative solvers discussed in Section 5.5.3 we find that for direct solvers in particular, due to their rapidly increasing array storage, the application of adaptive meshes is reasonable.

6.5 Exercises

Exercise 6.1. Show that for nodal functions φ_ξ over simplicial triangulations the estimate $2\|1-\varphi_\xi\|_{L^2(\omega_\xi)} \le |\omega_\xi|^{1/2}$ holds.

Exercise 6.2. Construct a nontrivial mesh which by subdivision of one element by means of bisection refines globally in the attempt to restore conformity by continued refinement.

Exercise 6.3. *Local extrapolation.* Let u denote a function on $[0,1]$ and u^j an associated approximation over the mesh

$$\mathcal{T}^j = \{t_i^j : 0 \le i < n_j\} \quad \text{with} \quad t_i^j :=]x_i^j, x_{i+1}^j[$$

with $0 = x_0^j < x_1^j < \cdots < x_{n_j}^j = 1$. As an initial mesh let $\mathcal{T}^0 = \{]0,1/2[,]1/2,1[\}$ be given, and let the mesh $\mathcal{T}^1$ be generated by uniform refinement. Let the other meshes $\mathcal{T}^j$, $j > 1$ be suggested by bisection in combination with a refinement strategy on the basis of local extrapolation (see Section 6.2.1). Suppose η_i^j are the local error indicators, defined by

$$\eta_i^j = \|u - u^j\|^2_{t_i^j}.$$

Two model assumptions will be compared:

1. *Same exponent, different prefactor.* Let the local error on a subinterval $t_h^j :=]z, z+h[$ satisfy the condition

$$\|u-u^j\|^2_{t_h^j} = \begin{cases} c_1 h^\gamma & \text{for } t_h^j \subset]0,1[, \\ c_2 h^\gamma & \text{for } t_h^j \subset]1/2,1[\end{cases}$$

with $0 < c_1 < c_2$. Show that there exists a $j_0 > 0$ at which the error is "nearly uniform" in the following sense: for $j > j_0$ two further refinement steps lead to exactly one refinement in each subinterval.

2. *Same prefactor, different exponent.* Let the local error on a subinterval t_h^j satisfy the condition

$$\|u - u^j\|^2_{t_h^j} = \begin{cases} ch^{\gamma_1} & \text{for } t_h^j \subset]0, 1/2[, \\ ch^{\gamma_2} & \text{for } t_h^j \subset]1/2, 1[\end{cases}$$

with $0 < \gamma_1 < \gamma_2$. What changes in this case?

Exercise 6.4. We return to equation (6.49) for the plasmon-polariton waveguide. Expand the differential operators and write the quadratic eigenvalue problem in weak formulation in terms of matrices. Perform a transformation of the kind $v = k_z u$ to obtain a linear eigenvalue problem of double dimension. Two principal reformulations exist. What properties do the arising matrices of double dimension have?