A posteriori error estimation techniques for non-linear elliptic and parabolic pdes

Rüdiger Verfürth

Fakultät für Mathematik, Rurh-Universität Bochum,
D-44780 Bochum, Germany
rv@num1.rurh-uni-bochum.de

ABSTRACT. *We give a brief overview of a posteriori error estimation techniques for nonlinear elliptic and parabolic pdes and point out some related questions which are not yet satisfactorily settled.*

RÉSUMÉ. *Un résumé des techniques d'estimation d'erreur a posteriori pour des edp elliptiques ou paraboliques est présenté. Simultanément, sont évoqués les problèmes relatifs à ces techniques n'ayant pas été résolus de façon satisfaisante.*

KEY WORDS: *a posteriori error estimates; non-linear problems; elliptic pdes; parabolic pdes.*

MOTS-CLÉS : *estimation d'erreur a posteriori ; problèmes non linéaires ; edp elliptiques ; edp paraboliques.*

1. Introduction

Adaptive finite element methods based on a posteriori error estimation techniques have become an indispensable tool in large scale scientific computation. Since the pioneering work of Babuška and Rheinboldt in the late seventies much work has been invested in putting a posteriori error estimation on a sound basis both theoretically and practically. Most of this work was done for linear elliptic pdes, much less for non-linear problems and for parabolic pdes, and very little for hyperbolic problems. Nowadays the theory seems rather mature for linear elliptic problems. Nevertheless there are still some important open questions. For non-linear and time-dependent problems there exists a general pathway which is rather promising but which must still be inspected more thoroughly. For hyperbolic poblems the field is still in its infancy.

It is the aim of this note to sketch briefly the general methodology which leads to a posteriori error estmiates for finite element discretisations of elliptic and parabolic pdes. At the same time we want to hint at some related problems which seem important to us and which are not yet completely solved. These are: treatment of non-linearities and sensivity estimates with regard to perturbations, estimation of constants, robustness with regard to parameters (in particular in the context of singularly perturbed problems), treatment of anisotropic equations or meshes. Of course this list is not complete and it reflects our personal point of view. Also we cannot present all existing error estimation techniques. Instead we will limit ourselves to two approaches: residual estimates and estimates based on the solution of auxiliary local problems. Although these have their particular benefits and drawbacks they are representative for other error estimation techniques. The above mentioned problems are relevant to all known error estimation techniques although they sometimes show up in varying disguises.

One should always keep in mind that any reasonable error estimator should satisfy at least three minimal requirements: *reliability, efficiency,* and *locality.* As usual, reliability means that the error estimator yields upper bounds on the error measured in some user-prescribed norm. Similarly, efficiency means that it also yields lower bounds on the error (of course measured in the same norm!). By locality we mean that the estimator should give information on the local (with regard to space and time) distribution of the error. Clearly, reliability is mandatory to guarantee a prescribed tolerance. Efficiency is needed to achieve this task with a (nearly) minimal amount of work. Locality is indispensable for the correct resolution of the relevant physical scales. The upper and lower bounds on the error always contain multiplicative constants. The product of these constants is a measure for the quality of the error estimator and is similar to a condition number. A good knowledge of this quantity is necessary for a correct calibration of the error estimator. If the differential equation contains critical parameters, e.g. if it is singularly perturbed, this quantity should stay decently bounded for a reasonably large range of these parameters.

2. Quasilinear elliptic pdes of second oder

In the next sections we will always consider quasilinear elliptic pdes of second order with homogeneous Dirichlet boundary conditions:

$$\begin{aligned} \partial_i a_i(x,u,\nabla u) &= b(x,u,\nabla u) \quad \text{in } \Omega \\ u &= 0 \qquad\qquad\quad \text{on } \Gamma. \end{aligned} \qquad [1]$$

Here and in what follows we use the summation convention, i.e. $\partial_i a_i := \sum_i \partial_i a_i$, $u_i v_i := \sum_i u_i v_i$ etc.. Ω is a bounded open polyhedron in $\mathbb{R}^n$, $n \geq 2$, with Lipschitz boundary Γ. The functions $a_1, \ldots, a_n$ and b are assumed to be sufficiently smooth and the matrix $(\partial_{p_j} a_i(x,y,p))_{1\leq i,j\leq n}$ must be uniformly positive definite on $\Omega \times \mathbb{R} \times \mathbb{R}^n$. The restriction to Dirichlet boundary conditions simplifies the exposition. With obvious modifications all results, however, also hold for Neumann or mixed Dirichlet-Neumann boundary conditions.

In order to obtain a well-posed weak formulation of [1] one generally has to consider Sobolev spaces $W^{1,p}(\Omega)$ with Lebesgue exponents $p > 2$ (cf. §3.3 in [VER 96]). In order to simplify the exposition and the notation we, however, restrict ourselves to the Hilbert-space setting. Correspondingly the weak formulation of [1] consists in finding $u \in H_0^1(\Omega)$ such that:

$$\int_\Omega a_i(x,u,\nabla u)\partial_i v = \int_\Omega b(x,u,\nabla u)v \quad \forall v \in H_0^1(\Omega). \qquad [2]$$

Here, $L^2(\Omega)$, $H^1(\Omega) := \{v \in L^2(\Omega) : \partial_i v \in L^2(\Omega), \forall 1 \leq i \leq n\}$, and $H_0^1(\Omega) := \{v \in H^1(\Omega) : v = 0 \,\text{on}\, \Gamma\}$ denote the usual Lebesgue and Sobolev spaces equipped with the standard norms, resp. semi-norm:

$$\|v\|_0 := \left\{ \int_\Omega v^2 \right\}^{1/2},$$

$$\|v\|_1 := \left\{ \|v\|_0^2 + \sum_{i=1}^n \|\partial_i v\|_0^2 \right\}^{1/2},$$

$$|v|_1 := \left\{ \sum_{i=1}^n \|\partial_i v\|_0^2 \right\}^{1/2}.$$

Recall, that $|.|_1$ is a norm on $H_0^1(\Omega)$ that is equivalent to $\|.\|_1$. $H^{-1}(\Omega)$ and $\|.\|_{-1}$ denote the dual space of $H_0^1(\Omega)$ and the corresponding norm.

If ω is an open subset of Ω with Lipschitz boundary γ, we denote by $\|.\|_{0;\omega}$, $\|.\|_{1;\omega}$, and $|.|_{1;\omega}$ the restrictions of the corresponding (semi-) norms to the set ω. Similarly, $\|.\|_\gamma$ denotes the norm of $L^2(\gamma)$.

3. Finite element discretization

We consider a family $\mathcal{T}_h$, $h > 0$, of partitions of Ω into n-simplices or n-cubes. Here, an n-cube is the image of the standard n-cube $[0,1]^n$ under an invertible *affine* mapping. The partitions must satisfy the following two conditions:

(1) *admissibility*: any two elements K, K' of $\mathcal{T}_h$ are either disjoint or share a complete k-face, $0 \le k \le n-1$.

(2) *shape-regularity*: $c_{\mathcal{T}} := \sup_{h>0} \sup_{K\in\mathcal{T}_h} h_K/\rho_K < \infty$.

Here, h_K denotes the diameter of K and ρ_K is the diameter of the largest ball which can be inscribed into K. In two dimensions, shape-regularity is equivalent to the minimal angle condition.

Consider a family X_h of finite element spaces associated with the family $\mathcal{T}_h$. We assume that $X_h \subset H_0^1(\Omega)$ and that X_h contains all continuous, piecewise linear or n-linear functions. Moreover, the functions in X_h should be piecewise (with regard to $\mathcal{T}_h$) twice continuously differentiable. Then the finite element discretization of [1] consists in finding $u_h \in X_h$ such that

$$\int_\Omega a_i(x, u_h, \nabla u_h)\partial_i v_h = \int_\Omega b(x, u_h, \nabla u_h) v_h \quad \forall v_h \in X_h. \qquad [3]$$

Note that we always consider affinely equivalent finite element spaces, i.e. each element is the image of a reference element under an affine mapping. Since the transformation is affine its Jacobi matrix and its functional determinant are constant. This is a crucial ingredient in many proofs and constructions, e.g. those of estimates [6] and [8] below. When using isoparametric elements, e.g. general quadrilaterals in two dimensions, this condition is no longer satisfied and one must resort to a perturbation argument, i.e. the transformation is close to an affine one. However, the treatment of finite element discretizations which are not affinely equivalent is not yet completely understood.

The shape regularity is also crucial for many estimates such as, e.g., [6] and [8] below. Shape regularity in particular implies that for each element all edges are of comparable length. In this sense shape regular meshes are isotropic. In many applications, however, one needs anisotropic meshes which have a much smaller length with regard to a certain direction than with regard to the other directions. In this case $c_{\mathcal{T}}$ becomes exceedingly large. Recently, Siebert [SIE 96] for cuboidal meshes and Kunert [KUN 98] for tetrahedral meshes have tried to extend the theory to anisotropic meshes. But much work has still to be done in this direction.

4. Auxiliary results

Given $K \in \mathcal{T}_h$ we denote by $\mathcal{N}(K)$ and $\mathcal{E}(K)$ the sets of its vertices and of its $(n-1)$-faces, respectively. Set $\mathcal{N}_h := \bigcup_{K\in\mathcal{T}_h} \mathcal{N}(K)$ and $\mathcal{E}_h := \bigcup_{K\in\mathcal{T}_h} \mathcal{E}(K)$. Both sets can be decomposed as $\mathcal{N}_h = \mathcal{N}_{h,\Omega} \cup \mathcal{N}_{h,\Gamma}$ and $\mathcal{E}_h = \mathcal{E}_{h,\Omega} \cup \mathcal{E}_{h,\Gamma}$ into the set of interior vertices/faces and the set of boundary vertices/faces. With each face $E \in \mathcal{E}_{h,\Omega}$ we associate a unit vector n_E which is orthogonal to E. For any piecewise continuous function v we denote $[v]_E$ the jump of v across E in direction n_E. Of course $[v]_E$ depends on the orientation of n_E, but quantities like $[\partial_i v n_{E,i}]_E$ or $[a_i(x,v,\nabla v)n_{E_i}]_E$ are independent thereof.

For each element $K \in \mathcal{T}_h$, each $(n-1)$-face $E \in \mathcal{E}_h$, and each node $x \in \mathcal{N}_h$ we denote by:

ω_K the union of all elements that share an $(n-1)$-face with K,
$\tilde{\omega}_K$ the union of all elements that have at least one point in common with K,
ω_E the union of all elements that have E as an $(n-1)$-face,
$\tilde{\omega}_E$ the union of all elements that have at least one point in common with E,
ω_x the union of all elements that have x as a vertex.

With each element K and each $(n-1)$-face E we associate a cut-off function ψ_K and ψ_E wich satisfies the following properties:

$$\begin{aligned}
&0 \le \psi_K \le 1 && \text{on } K,\\
&\max_{x\in K} \psi_K(x) = 1, &&\\
&\psi_K = 0 && \text{on } \partial K,\\
&0 \le \psi_E \le 1 && \text{on } \omega_E,\\
&\max_{x\in E} \psi_E(x) = 1, &&\\
&\psi_E = 0 && \text{on } \partial\omega_E.
\end{aligned} \qquad [4]$$

One possibility to construct these functions is as follows. Given any node $x \in \mathcal{N}_h$ denote by λ_x the corresponding nodal bases function, i.e. the continuous, piecewise linear or n-linear function that takes the value 1 at x and that vanishes at all other nodes $y \in \mathcal{N}_h \setminus \{x\}$. Then there are real numbers α and β such that the functions

$$\begin{aligned}
\psi_K &= \alpha \prod_{x\in\mathcal{N}(K)} \lambda_x\\
\psi_E &= \beta \prod_{x\in\mathcal{N}(E)} \lambda_x
\end{aligned} \qquad [5]$$

satisfy the above requirements. Here, $\mathcal{N}(E)$ denotes the set of all vertices of E.

Given any integer k, one can prove (cf. §3.1 in [VER 96]) that there are constants $\gamma_1, \ldots, \gamma_5$ that only depend on k and on the parameter $c_{\mathcal{T}}$ such that the inequalities

$$\begin{aligned} \gamma_1 \|v\|_{0;K}^2 &\leq \int_K \psi_K v^2 \leq \|v\|_{0;K}^2, \\ \|\psi_K\|_{1;K} &\leq \gamma_2 h_K^{-1} \|v\|_{0;K}, \\ \gamma_3 \|\psi_E \varphi\|_E^2 &\leq \int_E \psi_E \varphi^2 \leq \|\varphi\|_E^2, \\ \|\psi_E \varphi\|_{1;\omega_E} &\leq \gamma_4 h_E^{-1/2} \|\varphi\|_E, \\ \|\psi_E \varphi\|_{0;\omega_E} &\leq \gamma_5 h_E^{1/2} \|\varphi\|_E, \end{aligned} \qquad [6]$$

hold for all elements K, all $(n-1)$-faces E and all polynomials v, φ of degree at most k defined on K and E, respectively.

Finally, we define a quasi-interpolation operator I_h by:

$$I_h v := \sum_{x \in \mathcal{N}_{h,\Omega}} \lambda_x \pi_x v \qquad [7]$$

where:

$$\pi_x v := \left\{ \int_{\omega_x} v dx \right\} / \left\{ \int_{\omega_x} dx \right\}$$

denotes the mean-value of v on ω_x. In particular, we have $I_h v \in X_h$ for any $v \in L^2(\Omega)$. One can prove (cf. [VER 99]) that there are two constants c_{I1} and c_{I2} which only depend on the parameter $c_{\mathcal{T}}$ such that the error estimates

$$\begin{aligned} \|v - I_h v\|_K &\leq c_{I1} h_K |v|_{1;\tilde{\omega}_K} \\ \|v - I_h v\|_E &\leq c_{I2} h_E^{1/2} |v|_{1;\tilde{\omega}_E} \end{aligned} \qquad [8]$$

hold for all $v \in H^1(\Omega)$, all $K \in \mathcal{T}_h$ and all $E \in \mathcal{E}_h$.

As we will see in subsequent sections, the constants $\gamma_1, \ldots, \gamma_5$, c_{I1}, c_{I2} are crucial for the quality of an error estimator and for its correct calibration. Correspondingly there is a strong need for sharp explicit estimates of these constants.

Estimates [6] are usually proven by passing to a reference element. Thus $\gamma_1, \ldots, \gamma_5$ can be decomposed into a contribution of $c_{\mathcal{T}}$ and of corresponding constants $\hat{\gamma}_1, \ldots, \hat{\gamma}_5$ referring to the reference element. For fixed polynomial degree k, the latter can explicitely be computed by solving an eigenvalue problem of moderate size which depends on k. On the other hand, a simple scaling argument shows that these quantities will be proportional to some power of k. Explicit bounds are derived in [VER 00] using a dimension-reduction argument.

When using anisotropic meshes the reference element must be chosen such that it correctly reflects the anisotropy. Correspondingly one has to work with a whole family of reference elements and to invoke an additional compactness argument. This is the approach of [KUN 98]. However, satisfactory quantitative results are still lacking.

The constants c_{I1}, c_{I2} are estimated in [VER 99] for shape regular meshes. These results are quite satisfactory but not yet optimal when compared with numerical estimates. Anisotropic meshes are tackled in [KUN 98]. However, quantitative results are again lacking in this case.

5. The equivalence of error and residual

Denote by u and u_h solutions of problems [2] and [3], respectively. These may not be unique, but are kept fixed in what follows. We want to estimate $\|u - u_h\|_1$. To this end we rewrite [2] and [3] as abstract non-linear equations. Define the mapping F of $H_0^1(\Omega)$ into $H^{-1}(\Omega)$ by:

$$\langle F(v), w\rangle := \int_\Omega \{a_i(x, v, \nabla v)\partial_i w - b(x, v, \nabla v)w\} \qquad [9]$$

Then, problem [2] is equivalent to $F(v) = 0$. Similarly, equation [3] is equivalent to $F_h(u_h) = 0$ where the mapping F_h of X_h into its dual space is given by:

$$\langle F_h(v_h), w_h\rangle := \int_\Omega \{a_i(x, v_h, \nabla v_h)\partial_i w_h - b(x, v_h, \nabla v_h)w_h\}. \qquad [10]$$

The Fréchet derivative of F at u is given by:

$$\langle DF(u)v, w\rangle := \int_\Omega \{\partial_{p_j} a_i(x, u, \nabla u)\partial_j v \partial_i w + \partial_u a_i(x, u, \nabla u) v \partial_i w - \partial_{p_i} b(x, u, \nabla u)\partial_i v w - \partial_u b(x, u, \nabla u) v w\}.$$

Under suitable differentiability and growth conditions on the functions $a_1, \ldots, a_n$ and b it is a bounded linear operator of $H_0^1(\Omega)$ in $H^{-1}(\Omega)$. Its norm is denoted by $A = A(u)$. Under similar conditions DF is locally Lipschitz continuous at u. I.e. there are numbers $R_0 > 0$ and $\beta > 0$ such that:

$$\|DF(u)w - DF(v)w\|_{-1} \le \beta\|u - v\|_1 \|w\|_1$$

holds for all $w \in H_0^1(\Omega)$ and all $v \in H_0^1(\Omega)$ with $\|u - v\|_1 \le R_0$.

Our essential assumption is that $DF(u)$ is invertible and has a bounded inverse. This means that the linearization of the pde [1] at u admits for each right-hand side $f \in H^{-1}(\Omega)$ a unique weak solution $w_f \in H_0^1(\Omega)$ which depends continuously on f. Denote by $\alpha := \alpha(u)$ the *inverse* of the norm of $DF(u)^{-1}$. Note that:

$$A = \sup_{v \in H_0^1(\Omega)\setminus\{0\}} \sup_{w \in H_0^1(\Omega)\setminus\{0\}} \frac{\langle DF(u)v, w\rangle}{\|v\|_1 \|w\|_1}$$

$$\alpha = \inf_{v \in H_0^1(\Omega)\setminus\{0\}} \sup_{w \in H_0^1(\Omega)\setminus\{0\}} \frac{\langle DF(u)v, w\rangle}{\|v\|_1 \|w\|_1}.$$

Assume that u_h is sufficiently close to u in the sense that:

$$\|u - u_h\|_1 \leq R := \min\{R_0, 2\alpha\beta^{-1}, \frac{1}{2}A\beta^{-1}\}.$$

Since $F(u) = 0$ and since $DF(u)$ is invertible we have

$$\begin{aligned} &u_h - u \\ &= DF(u)^{-1}\left\{F(u_h) + \int_0^1 [DF(u) - DF(u + t(u_h - u))](u_h - u)dt\right\}. \end{aligned} \qquad [11]$$

From equation [11] and the previous assumptions we easily conclude (cf. Proposition 2.1 in [VER 96]) that:

$$\frac{1}{2}A^{-1}\|F(u_h)\|_{-1} \leq \|u - u_h\|_1 \leq 2\alpha^{-1}\|F(u_h)\|_{-1}. \qquad [12]$$

This means that *the error* $\|u - u_h\|_1$ *is equivalent to the residual* $\|F(u_h)\|_{-1}$. The condition number of this equivalence is $4\alpha^{-1}A$. The residual is measured with regard to the dual norm $\|.\|_{-1}$. Hence its exact calculation would require the solution of an infinite dimensional variational problem. All error estimators try to approximate $\|F(u_h)\|_{-1}$ by a quantity which is as close as possible and which is much easier to compute.

The main assumption of this section is the invertibility of $DF(u)$. If $DF(u)$ is not invertible, but if its index is known a priori, one can still deduce the equivalence of error and residual by augmenting the space X and the function F (cf. §2.2 in [VER 96]). An example is the computation of simple eigenvalues and of corresponding eigenfunctions. However, up to now, there is no fully satisfactory strategy to determine the index of $DF(u)$ from the computed numerical solution u_h.

The quantities α and A are crucial for the equivalence of error and residual. There are various strategies which try to estimate these quantities from the numerical solution u_h. One approach consists in computing approximately the extremal eigenvalues of $DF_h(u_h)$. Another way consists in solving a related discrete adjoint problem (cf. [BEC 95]).

The sizes of α and A of course also depend on the norm of $H_0^1(\Omega)$. In order to see how a suitable choice of this norm may influence favorably these constants, assume that $DF(u)$ corresponds to the singularly perturbed, constant coefficient, reaction-diffusion operator $L_\varepsilon v := -\varepsilon\partial_i\partial_i v + v$ with $0 < \varepsilon \ll 1$. If we equip $H_0^1(\Omega)$ with its standard norm $\|.\|_1$, we easily conclude that:

$$\alpha \sim \varepsilon\,,\; A \sim \varepsilon + 1\,,\; \alpha^{-1}A \sim \varepsilon^{-1}.$$

Correspondingly, the relation between error and residual is very poor. On the other hand, the norm $|||v||| := \{\varepsilon|v|_1^2 + \|v\|_0^2\}^{1/2}$ is the natural energy norm for the operator L_ε. If we equip $H_0^1(\Omega)$ with this norm, we conclude that:

$$\alpha \sim 1\,,\; A \sim 1\,,\; \alpha^{-1}A \sim 1.$$

When doing this we must of course replace $\|F(u_h)\|_{-1}$ by the corresponding quantity:

$$|||F(u_h)|||_{-1} := \sup_{v \in H_0^1(\Omega)\setminus\{0\}} \frac{\langle F(u_h), v\rangle}{|||v|||}.$$

As we will see in the next section, this severely influences the computation of the residual.

A similar situation arises when $DF(u)$ corresponds to an anisotropic differential operator such as, e.g. $Lu := -\partial_i(A_{ij}\partial_j u) + u$ with $0 < \lambda_{\min}(A_{ij}) \ll \lambda_{\max}(A_{ij})$. The corresponding energy norm then is the anisotropic H^1-norm $|||u||| := \{\sum_i \|A_{ij}\partial_j u\|_0^2 + \|u\|_0^2\}^{1/2}$. When replacing $\|.\|_1$ and $\|.\|_{-1}$ by this norm and the corresponding dual norm resp. one again obtains

$$\alpha \sim 1\,,\; A \sim 1\,,\; \alpha^{-1}A \sim 1.$$

As we will see in the next section, this will require anisotropic analoga of estimates [6] and [8].

6. A residual error estimator

In this section we try to bound the H^{-1}-norm $\|F(u_h)\|_{-1}$ of the residual from above and from below by a mesh-dependent L^2-norm of element and face residuals. To this end consider a function $v \in H_0^1(\Omega)$ with $\|v\|_1 = 1$. Integration by parts elementwise yields an L^2-representation of the residual:

$$\begin{aligned}\langle F(u_h), v\rangle = &- \sum_{K\in\mathcal{T}_h} \int_K \{\partial_i a_i(x,u_h,\nabla u_h) + b(x,u_h,\nabla u_h)\}v \\ &+ \sum_{E\in\mathcal{E}_{h,\Omega}} \int_E [n_{E,i}a_i(x,u_h,\nabla u_h)]_E v \\ =: &\sum_{K\in\mathcal{T}_h} \int_K R_K(u_h)v + \sum_{E\in\mathcal{E}_{h,\Omega}} \int_E R_E(u_h)v.\end{aligned} \quad [13]$$

From equations [9] and [10], we obtain Galerkin orthogonality:

$$\langle F(u_h), v_h\rangle = 0 \quad \forall v_h \in X_h. \quad [14]$$

Since X_h contains the space S_h of all continuous, piecewise linear functions, we can replace v on the right-hand side of [13] by $v - v_h$ where $v_h \in S_h$ is arbitrary. This together with the Cauchy-Schwarz inequality for integrals yields:

$$\langle F(u_h), v\rangle \le \sum_{K\in\mathcal{T}_h} \|R_K(u_h)\|_{0;K}\|v - v_h\|_{0;K} + \sum_{E\in\mathcal{E}_{h,\Omega}} \|R_E(u_h)\|_E\|v - v_h\|_E.$$

Invoking the Cauchy-Schwarz inequality for sums, this implies:

$$\begin{aligned}\langle F(u_h), v\rangle \le &\left\{\sum_{K\in\mathcal{T}_h} h_K^2\|R_K(u_h)\|_{0;K}^2 + \sum_{E\in\mathcal{E}_{h,\Omega}} h_E\|R_E(u_h)\|_E^2\right\}^{1/2} \times \\ &\left\{\sum_{K\in\mathcal{T}_h} h_K^{-2}\|v - v_h\|_{0;K}^2 + \sum_{E\in\mathcal{E}_{h,\Omega}} h_E^{-1}\|v - v_h\|_E^2\right\}^{1/2}.\end{aligned}$$

From estimate [8] we conclude that:

$$\begin{aligned}&\inf_{v_h\in S_h} \left\{\sum_{K\in\mathcal{T}_h} h_K^{-2}\|v - v_h\|_{0;K}^2 + \sum_{E\in\mathcal{E}_{h,\Omega}} h_E^{-1}\|v - v_h\|_E^2\right\}^{1/2} \\ \le &\left\{\sum_{K\in\mathcal{T}_h} h_K^{-2}\|v - I_h v\|_{0;K}^2 + \sum_{E\in\mathcal{E}_{h,\Omega}} h_E^{-1}\|v - v_h\|_E^2\right\}^{1/2} \\ \le &c_0 \max\{c_{I1}, c_{I2}\}\|v\|_1.\end{aligned}$$

Here, c_0 is the maximal number of elements that share an arbitrary vertex. This number depends on $c_{\mathcal{T}}$.

With the abbreviation:

$$\eta_{R,K} := \left\{ h_K^2 \|R_K(u_h)\|_{0;K}^2 + \frac{1}{2} \sum_{E \in \mathcal{E}_{h,\Omega} \cap \mathcal{E}(K)} \|R_E(u_h)\|_E^2 \right\}^{1/2} \qquad [15]$$

we have thus proven that:

$$\|F(u_h)\|_{-1} \leq c_0 \max\{c_{I1}, c_{I2}\} \left\{ \sum_{K \in \mathcal{T}_h} \eta_{R,K}^2 \right\}^{1/2}. \qquad [16]$$

Together with the results of the previous section this implies that:

$$\eta_R := \left\{ \sum_{K \in \mathcal{T}_h} \eta_{R,K}^2 \right\}^{1/2}$$

is a reliable a posteriori error estimator for $\|u - u_h\|_1$. The corresponding constant is $2\alpha^{-1} c_0 \max\{c_{I1}, c_{I2}\}$.

In order to prove the efficiency of η_R we approximate the functions $a_1, \ldots, a_n$, and b by functions $a_{1,h}, \ldots, a_{n,h}, b_h$ which are piecewise (with regard to $\mathcal{T}_h$) polynomials. $\tilde{R}_K(u_h)$ and $\tilde{R}_E(u_h)$ denote the element and face residuals computed with $a_{1,h}, \ldots, a_{n,h}, b_h$ instead of $a_1, \ldots, a_n, b$. Consider an arbitrary element K. From inequality [6] we know that:

$$\gamma_1 \|\tilde{R}_K(u_h)\|_{0;K}^2 \leq \int_K \tilde{R}_K(u_h) \psi_K \tilde{R}_K(u_h) =: \int_K \tilde{R}_K(u_h) w_K.$$

Since $w_K := \tilde{R}_K(u_h)\psi_K$ vanishes on ∂K we obtain from equation [13] that:

$$\begin{aligned} \int_K \tilde{R}_K(u_h) w_K &= \int_K [\tilde{R}_K(u_h) - R_K(u_h)] w_K + \langle F(u_h), w_K \rangle \\ &\leq \|\tilde{R}_K(u_h) - R_K(u_h)\|_{0;K} \|w_K\|_{0;K} + \|F(u_h)\|_{-1} \|w_K\|_{1;K}. \end{aligned}$$

Combining these estimates with inequality [6] and recalling that $0 \leq \psi_K \leq 1$, we arrive at:

$$h_K \|\tilde{R}_K(u_h)\|_{0;K} \leq \frac{1}{\gamma_1} h_K \|\tilde{R}_K(u_h) - R_K(u_h)\|_{0;K} + \frac{\gamma_2}{\gamma_1} \|F(u_h)\|_{-1}$$

and therefore:

$$h_K\|R_K(u_h)\|_{0;K} \le (1+\frac{1}{\gamma_1})h_K\|\tilde{R}_K(u_h)-R_K(u_h)\|_{0;K}+\frac{\gamma_2}{\gamma_1}\|F(u_h)\|_{-1}. \quad [17]$$

Next consider an arbitrary face E. Using the same arguments as above, inequaltiy [6], equation [13], and estimate [17] imply that:

$$\begin{aligned} h_E^{1/2}\|R_E(u_h)\|_E \le & \frac{1}{\gamma_3}[\gamma_4+\frac{\gamma_2\gamma_5}{\gamma_1}]\|F(u_h)\|_{-1} \\ & +(1+\frac{1}{\gamma_3})h_E^{1/2}\|\tilde{R}_E(u_h)-R_E(u_h)\|_E \\ & +\frac{\gamma_5}{\gamma_3}(1+\frac{1}{\gamma_1})\sum_{K\subset\omega_E} h_E\|\tilde{R}_K(u_h)-R_K(u_h)\|_{0;K}. \end{aligned} \quad [18]$$

Since the quantities $h_K\|\tilde{R}_K(u_h)-R_K(u_h)\|_{0;K}$ and $h_E^{1/2}\|\tilde{R}_E(u_h)-R_E(u_h)\|_E$ are higher order perturbations which only depend on the smoothness of the functions $a_1,\dots,a_n$ and b, estimates [17] and [18] together with the results of the previous section imply the efficiency and the locality of η_R. The corresponding constant is $2A\max\{\frac{\gamma_2}{\gamma_1},\frac{\gamma_4}{\gamma_3},\frac{\gamma_2\gamma_5}{\gamma_1\gamma_3}\}$.

We have seen at the end of the previous section that it may be advisable to replace $\|.\|_1$ by a problem dependent norm of the form $|||.||| = \{\varepsilon|.|_1^2+\|.\|_0^2\}^{1/2}$ with $0<\varepsilon\ll 1$. Then $\|.\|_{-1}$ has to be replaced by the corresponding dual norm. Similarly, one has to replace $\|.\|_1$ by $|||.|||$ and its corresponding local version throughout this section. When doing this in a naive and straightforward way by retaining the scalings of the element and face residuals, one arrives at upper and lower bounds on the error $|||u-u_h|||$ such that their ratio is proportional to $\varepsilon^{-1/2}$. This means that the corresponding error estimator is not robust with regard to the parameter ε. This unpleasant situation can be remedied by a more refined analysis (cf. [VER 98a, VER 98b]). When replacing the scaling factors h_K and $h_E^{1/2}$ of the element and face residuals by $\alpha_K := \min\{1,h_K\varepsilon^{-1/2}\}^{1/2}$ and $\beta_E := \varepsilon^{-1/4}\min\{1,h_E\varepsilon^{-1/2}\}^{1/2}$, resp. one arrives at an a posteriori error estimate of $|||u-u_h|||$ which is reliable, efficient, local and robust with regard to ε.

When using a problem adapted anisotropic H^1-norm as described at the end of the previous section, one needs anisotropic analoga of estimates [6] and [8] in order to arrive at an a posteriori error estimate which is robust with regard to the anisotropy. By rescaling the coordinates one sees that this problem is strongly related to the treatment of anisotropic meshes. First results in this direction are obtained in [BER 00, KUN 98, KUN 00, SIE 96].

7. Error estimators based on the solution of auxiliary local problems

The idea is to lift $F(u_h) \in H^{-1}(\Omega)$ to a function $v \in H_0^1(\Omega)$ by solving a suitable elliptic pde of second order and to use $\|v\|_1$ as an error estimator. In order to render this idea operative, the compution of v must be done on a local and discrete level. To make things more precise we will consider a variant which has its roots in [BAB 78].

Choose a vertex $x_0 \in \mathcal{N}_{h,\Omega}$. Set

$$A_{ij} := \partial_{p_j} a_i(x_0, u_h(x_0), (\pi_h \nabla u_h)(x_0)).$$

Here, $\pi_h \nabla u_h$ is some average or projection of the possibly discontinuous gradient of u_h. Choose a finite element space $V_{x_0} \subset H_0^1(\omega_{x_0})$ corresponding to $\mathcal{T}_h$ which consists of piecewise polynomials of a sufficiently high degree. This means that the polynomial degree of the functions in V_{x_0} should be larger than the one of u_h. One possible choice consists in taking all functions $\psi_K \tilde{R}_K(u_h)$ and $\psi_E \tilde{R}_E(u_h)$ where K and E are elements and faces having x_0 as a vertex and where $\tilde{R}_K(u_h)$ and $\tilde{R}_E(u_h)$ are as in the previous section. Denote by $v_{x_0} \in V_{x_0}$ the unique solution of:

$$\int_{\omega_{x_0}} A_{ij} \partial_j v_{x_0} \partial_i w = \langle F(u_h), w \rangle \quad \forall w \in V_{x_0} \qquad [19]$$

and set:

$$\eta_{D,x_0} := |v_{x_0}|_{1;\omega_{x_0}}. \qquad [20]$$

Problem [19] admits a unique solution since the matrix $(\partial_{p_j} a_i(x, v, p))_{1 \le i,j \le n}$ is assumed to be uniformly positive definite. Denote by $\lambda > 0$ the minimal eigenvalue of the matrix (A_{ij}). Inserting v_{x_0} as a test function in [19] we immediately obtain that:

$$\eta_{D,x_0} \le \frac{1}{\lambda} \|F(u_h)\|_{-1}. \qquad [21]$$

Together with the results of Section 6 this implies that η_{D,x_0} is an efficient and local error estimator for $\|u - u_h\|_1$. The corresponding constant is $2\lambda^{-1}A$.

It is much more tedious to prove the reliability of this error estimator. If V_{x_0} contains the functions $\psi_K \tilde{R}_K(u_h)$ and $\psi_E \tilde{R}_E(u_h)$, one may compare η_{D,x_0} with the estimator of the previous section. The definition of η_{D,x_0} and the second part of the previous section then imply that:

$$\left\{ \sum_{K \in \mathcal{T}_h} \eta_{R,K} \right\}^{1/2} \le c \left\{ \sum_{x \in \mathcal{N}_{h,\Omega}} \eta_{D,x}^2 \right\}^{1/2}.$$

The constant c is proportional to γ_1^{-1}, γ_3^{-1} and to γ_2, γ_4, γ_5. This estimate and the results of Sections 5 and 6 establish the reliability of the error estimator:

$$\eta_D := \left\{ \sum_{x \in \mathcal{N}_{h,\omega}} \eta_{D,x}^2 \right\}^{1/2} .$$

The strategy just described also applies to singularly perturbed problems, provided the matrix (A_{ij}), the functions in V_{x_0}, and the norm in the definiton of η_{D,x_0} take into account the singular perturbation (cf. [VER 98a, VER 98b]). First results for anisotropic meshes can be found in [KUN 98].

Problem [19] is a discrete analogue of the Dirichlet problem:

$$\begin{aligned} -\partial_i(A_{ij}\partial_j u) &= F(u_h) && \text{in } \omega_{x_0} \\ u &= 0 && \text{on } \partial\omega_{x_0}. \end{aligned}$$

Similarly, one can also consider error estimators which are based on the solution of discrete analoga of the Neumann problem (cf. § 3.3 in [VER 96]):

$$\begin{aligned} -\partial_i(A_{ij}\partial_j u) &= R_K(u_h) && \text{in } K \\ n_{K,i}A_{ij}\partial_j u &= R_E(u_h) && \text{on } \partial K. \end{aligned}$$

This idea was first introduced in [BAN 85].

8. Quasilinear parabolic pdes of second order

The parabolic counterpart of problem [1] is

$$\begin{aligned} \partial_t u - \partial_i a_i(x, u, \nabla u) &= b(x, u, \nabla u) && \text{in } \Omega \times (0, T) \\ u &= 0 && \text{on } \Gamma \times (0, T) \\ u(.,0) &= u_0 && \text{in } \Omega. \end{aligned} \qquad [22]$$

Here we assume for simplicity that the functions $a_1, \ldots, a_n$ and b do not depend on the time t. T is a given, fixed, finite final time.

For the weak formulation of problem [22] we must introduce some function spaces. Let V and W be two Banach spaces with corresponding norms $\|.\|_V$ and $\|.\|_W$ such that $V \hookrightarrow W$ is a continuous and dense injection. $L^2(0, T; V)$

denotes the space of all functions v on $(0,T)$ with values in V such that the function $t \longrightarrow \|v(.,t)\|_V$ is square integrable. The corresponding norm is:

$$\|v\|_{L^2(0,T;V)} := \left\{ \int_0^T \|v(.,t)\|_V^2 dt \right\}^{1/2} .$$

Set $W^2(0,T;V,W) := \{v \in L^2(0,T;V) : \partial_t v \in L^2(0,T;W)\}$ and equip it with the norm:

$$\|v\|_{W^2(0,T;V,W)} := \left\{ \int_0^T \|v(.,t)\|_V^2 dt + \int_0^T \|\partial_t v(.,t)\|_W^2 dt \right\}^{1/2} .$$

Here, the time derivative $\partial_t v$ must be understood in the distributional sense. $L^2(0,T;V)$ and $W^2(0,T;V,W)$ are Banach spaces. One can prove that for any $v \in W^2(0,T;V,W)$ the quantity $v(.,T)$ exists and is an element of W. Therefore:

$$W_0^2(0,T;V,W) := \{v \in W^2(0,T;V,W) : v(.,T) = 0\}$$

is well-defined. For abbreviation we set:

$$\begin{aligned} X &:= L^2(0,T;H_0^1(\Omega)) \\ Y &:= W_0^2(0,T;H_0^1(\Omega), H^{-1}(\Omega)). \end{aligned} \qquad [23]$$

Then we may consider the following weak formulation of problem [22]: Find $u \in X$ such that:

$$\begin{aligned} &- \int_0^T \int_\Omega u \partial_t v - \int_\Omega u_0 v(.,0) \\ &+ \int_0^T \int_\Omega \{a_i(x,u,\nabla u)\partial_i v - b(x,u,\nabla u)v\} = 0 \quad \forall v \in Y. \end{aligned} \qquad [24]$$

Define a mapping F of X into the dual space of Y by:

$$\begin{aligned} \langle F(u), v \rangle := &- \int_0^T \int_\Omega u \partial_t v - \int_\Omega u_0 v(.,0) \\ &+ \int_0^T \int_\Omega \{a_i(x,u,\nabla u)\partial_i v - b(x,u,\nabla u)v\} . \end{aligned} \qquad [25]$$

Then problem [24] is equivalent to $F(u) = 0$. Thus it fits into the abstract framework of Section 5. $DF(u)$ is locally Lipschitz continuous at u if the functions $a_1, \ldots, a_n$ and b satisfy appropriate differentiability and growth conditions (cf. [VER 98c, VER 98d]).

9. Discretization with space-time finite elements

There are three main approaches to discretize parabolic pdes: the *method of lines*, *Rothe's method*, and *space-time finite elements*. Often, all approaches lead to the same discrete problem. But the analysis is completely different, in particular concerning the necessary regularity requirements. This is also reflected by the a posteriori error analysis. In [ADJ 88, BIE 82a, BIE 82b] an a posteriori error analysis for the method of lines is given; Rothe's method is investigated in [BOR 90, BOR 91, BOR 92]. Here we will concentrate on space-time finite element methods. This approach has several advantages: it requires minimal regularity assumptions and it has a variational structure. The latter allows us to put this discretization into the abstract framework of Section 5.

We first subdivide the interval $[0,T]$ into N_τ subintervals $J_1 = [t_1, t_2)$, ..., $J_{N_\tau} = [t_{N_\tau}, t_{N_\tau+1})$ with respective length $\tau_1, \ldots, \tau_{N_\tau}$. The subintervals are arranged in a natural way, $0 = t_1 < t_2 < \ldots < t_{N_\tau} < t_{N_\tau+1} = T$. We assume that this partition is *shape regular*, i.e. the ratios τ_i/τ_{i+1} and τ_{i+1}/τ_i are bounded from above uniformly with regard to i and τ. With each $j \in \{1, \ldots, N_\tau\}$ we associate a partition $\mathcal{T}_j$ of Ω, which satisfies the assumptions of Section 3, and a corresponding finite element space $V_j \subset H_0^1(\Omega)$. Denote by λ_j the continuous, piecewise linear function that takes the value 1 at t_j and vanishes at all other points t_i, $i \neq j$. Set:

$$\lambda_j^{(\theta)} := \lambda_j + (6\theta - 3)\lambda_j[\lambda_{j+1} - \lambda_{j-1}]$$

with the obvious modification for $j = 1$. Here, $\theta \in [0,1]$ is a parameter which will be chosen later. Denote by χ_j the characteristic function of the j-th subinterval and set:

$$\begin{aligned} X_h &:= \operatorname{span}\{\chi_j(t)v_j(x) \quad : 1 \le j \le N_\tau,\, v_j \in V_j\}, \\ Y_h &:= \operatorname{span}\{\lambda_j^{(\theta)}(t)v_j(x) : 1 \le j \le N_\tau,\, v_j \in V_j\}. \end{aligned}$$

Note that the functions in X_h are piecewise constant with regard to time and that the functions in Y_h are continuous, piecewise quadratic functions with regard to time which vanish at the final time T. These properties ensure that $X_h \subset X$ and $Y_h \subset Y$.

The space-time finite element discretization of problem [22] then consists in finding $u_h \in X_h$ such that:

$$\begin{aligned} &-\int_0^T \int_\Omega u_h \partial_t v_h - \int_\Omega u_0 v_h(.,0) \\ &+ \int_0^T \int_\Omega \{a_i(x, u_h, \nabla u_h)\partial_i v_h - b(x, u_h, \nabla u_h)v_h\} = 0 \qquad \forall v_h \in Y_h. \end{aligned} \quad [26]$$

Problem [26] fits into the abstact framework of Section 5. The function F_h of X_h into the dual space of Y_h is given by:

$$\begin{aligned}\langle F_h(u_h), v_h\rangle = &- \int_0^T \int_\Omega u_h \partial_t v_h - \int_\Omega u_0 v_h(.,0) \\ &+ \int_0^T \int_\Omega \{a_i(x, u_h, \nabla u_h)\partial_i v_h - b(x, u_h, \nabla u_h) v_h\} \\ =&\langle F(u_h), v_h\rangle.\end{aligned} \qquad [27]$$

At first sight problem [26] may look rather strange. But it corresponds to the popular θ-scheme. The parameters $\theta = 0$, $\theta = 1$, and $\theta = \frac{1}{2}$ in particular yield the explicit Euler, implicit Euler, and Crank-Nicolson scheme, respectively. To see this, denote by u_h^j the constant value of u_h on the j-th subinterval and insert $\lambda_j^{(\theta)}(t)v_j(x)$, $v_j \in V_j$, as a test-function v_h in [26]. Since:

$$\int_{t_{j-1}}^{t_j} \lambda_j^\theta(t)dt = (1-\theta)\tau_{j-1}, \quad \int_{t_j}^{t_{j+1}} \lambda_j^{(\theta)}(t)dt = \theta\tau_j,$$

intergration by parts with regard to time on the subintervals yields:

$$\begin{aligned}&\int_\Omega u_h^j v_j + \theta\tau_j \int_\Omega \{a_i(x, u_h^j, \nabla u_h^j)\partial_i v_j - b(x, u_h^j, \nabla u_h^j)v_j\}, \\ = &\int_\Omega u_h^{j-1} v_j + (1-\theta)\tau_{j-1} \int_\Omega \{a_i(x, u_h^{j-1}, \nabla u_h^{j-1})\partial_i v_j - b(x, u_h^{j-1}, \nabla u_h^{j-1})v_j\},\end{aligned}$$

if $j \geq 2$, and:

$$\int_\Omega (u_h^1 - u_0)v_1 = 0,$$

if $j = 1$.

The previous approach can be extended to polynomials of degree $k \geq 1$ with regard to time (cf. [VER 98c, VER 98d]). It then corresponds to an implicit $k+1$-stage Runge-Kutta method which, for $\theta = \frac{1}{2}$, has the corresponding diagonal Padé approximation as its stability function. Hence, the time-discretization is A-stable and of order $2k+2$. The previous approach strongly resembles the popular discontinuous Galerkin method [ERI 85]. The latter, however, uses the same space of discontinuous (with regard to time) functions as test and trial spaces. In particular, the lowest order method corresponds to the implicit Euler scheme. The higher order methods correspond to implicit $k+1$-stage Runge-Kutta schemes which have the corresponding sub-diagonal Padé approximation as their stability function. Hence, this time-discretization is L-stable and of order $2k+1$. Since both test and trial functions are discontinuous with regard to time, the discontinuous Galerkin method is non-conforming with regard to any variational formulation of [22]. This makes its a posteriori error analysis more difficult.

10. Auxiliary results

We adapt the notations of Section 4. In particular, an index j indicates that the given quantity corresponds to the partition $\mathcal{T}_j$ of Ω. The intervals J_j and the partitions $\mathcal{T}_j$ induce a partition $\mathcal{P}_\tau$ of the space-time cylinder $Q_T := \Omega \times (0,T)$ into prisms of the form $Q = K \times J_j$ with $K \in \mathcal{T}_j$. Given any of these prisms, we denote by $\partial Q_L := \partial K \times J_j$ its lateral boundary and by $\partial Q_B := K \times \{t_j\}$ its bottom. The corresponding jumps are labeled by an index ∂Q_L or ∂Q_B, respectively. The jumps across lateral boundaries are again in the direction n_E, those across the bottoms are in the direction of increasing time.

With the help of the basis functions λ_j of the previous section and of the cut-off functions ψ_K, ψ_E of Section 4 we define cut-off functions with regard to space and time by:

$$\begin{aligned} \psi_Q &:= 4\lambda_j(t)\lambda_{j+1}(t)\psi_K(x) && ,Q = K \times J_j, \\ \psi_{E,j} &:= 4\lambda_j(t)\lambda_{j+1}(t)\psi_E(x) && ,E \in \mathcal{E}_j, \\ \psi_{K,j} &:= \lambda_j(t)\psi_K(x) && ,K \in \mathcal{T}_j. \end{aligned} \qquad [28]$$

With these functions one can prove the following analogue of estimate [6] (cf. [VER 98c, VER 98d]):

$$\begin{aligned} \delta_1 \|v\|^2_{L^2(Q)} &\le \int_Q \psi_Q v^2 \le \|v\|^2_{L^2(Q)}, \\ \|\psi_Q v\|_{L^2(J_j;H^1(K))} &\le \delta_2 h_K^{-1} \|v\|_{L^2(Q)}, \\ \|\partial_t(\psi_Q v)\|_{L^2(Q)} &\le \delta_3 \tau_j^{-1} \|v\|_{L^2(Q)}, \\ \|\partial_t(\psi_Q v)\|_{L^2(J_j;H^{-1}(K))} &\le \delta_4 \sigma_n(h_K) \|\partial_t(\psi_Q v)\|_{L^2(Q)}, \\ \delta_5 \|\varphi\|^2_{L^2(E\times J_j)} &\le \int_{J_j\times E} \psi_{E,j}\varphi^2 \le \|\varphi\|^2_{L^2(E\times J_j)}, \\ \|\psi_{E,j}\varphi\|_{L^2(J;H^1(\omega_E))} &\le \delta_6 h_E^{-1/2} \|\varphi\|_{L^2(E\times J_j)}, \\ \|\partial_t(\psi_{E,j}\varphi)\|_{L^2(\omega_E\times J_j)} &\le \delta_7 \tau_j^{-1} h_E^{1/2} \|\varphi\|_{L^2(E\times J_j)}, \\ \|\partial_t(\psi_{E,j}\varphi)\|_{L^2(J_j;H^{-1}(\omega_E))} &\le \delta_8 \sigma_n(h_E) \|\partial_t(\psi_{E,j}\varphi)\|_{L^2(\omega_E\times J_j)}, \\ \delta_9 \|w\|^2_{L^2(K)} &\le \int_K \psi_{K,j} w^2 \le \|w\|^2_{L^2(K)}, \\ \|\psi_{K,j} w\|_{L^2(J_{j-1}\cup J_j;H^1(K))} &\le \delta_{10} h_K^{-1} \tau_j^{1/2} \|w\|_{L^2(K)}, \\ \|\partial_t(\psi_{K,j} w)\|_{L^2(K\times[J_{j-1}\cup J_j])} &\le \delta_{11} \tau_j^{-1/2} \|w\|_{L^2(K)}, \\ \|\partial_t(\psi_{K,j} w)\|_{L^2(J_{j-1}\cup J_j;H^{-1}(K))} &\le \delta_{12} \sigma_n(h_K) \|\partial_t(\psi_{K,j} w)\|_{L^2(K\times[J_{j-1}\cup J_j])}. \end{aligned} \qquad [29]$$

Here, $Q = K \times J_j$ is an arbitrary prism, $E \subset \partial K$ is a face of K, v, φ, w are polynomials of an arbitrary but fixed degree and:

$$\sigma_n(h) := \begin{cases} h|\ln h| & \text{if } n = 2 \\ h & \text{if } n \geq 3. \end{cases} \qquad [30]$$

The constants $\delta_1, \ldots, \delta_{12}$ depend on the polynomial degree of v, φ, w and on the shape-regularity of the $\mathcal{T}_j$ via $c_\mathcal{T} = \sup_j \sup_{K \in \mathcal{T}_j} h_K / \rho_K$. The factor σ_n is due to the non-local nature of the H^{-1}-spaces (cf. Lemma 3.5 and Remarks 3.1 and 3.2 in [VER 98c]).

Finally, we must define an interpolation operator with regard to space and time. Denote by I_j the interpolation operator of Section 4 corresponding to $\mathcal{T}_j$. Given $j \in \{1, \ldots, N_\tau\}$ we define a projection operator π_j of Y into $H_0^1(\Omega)$ by:

$$\pi_j v := \frac{1}{\tau_{j-1} + \tau_j} \int_{t_{j-1}}^{t_{j+1}} v(.,t)dt$$

with the obvious modification if $j = 1$. The operators π_j and I_j commute. We define the interpolation operator in space and time by:

$$I_\tau v := \sum_{j=1}^{N_\tau} \lambda_j^{(\theta)}(t) \pi_j I_j v. \qquad [31]$$

One can prove (cf. [VER 98c,VER 98d]) that this interpolation operator satisfies the following analogue of estimate [8]:

$$\begin{aligned} \|v - I_\tau v\|_{L^2(Q)} &\leq \tilde{c}_{I1} \{ h_K \|v\|_{L^2(J_j, H^1(\tilde{\omega}_K))} \\ &\qquad + \tau_j h_K^{-1} \|\partial_t v\|_{L^2(t_{j-1}, t_{j+2}; H^{-1}(K))} \}, \\ \|v - I_\tau v\|_{L^2(E \times J_j)} &\leq \tilde{c}_{I2} \{ h_E^{1/2} \|v\|_{L^2(J_j, H^1(\tilde{\omega}_E))} \\ &\qquad + \tau_j h_E^{-3/2} \|\partial_t v\|_{L^2(t_{j-1}, t_{j+2}; H^{-1}(\tilde{\omega}_E))} \}, \\ \int_K |w||(v - I_\tau v)(., t_j)| &\leq \tilde{c}_{I3} \{ \tau_j^{1/2} \|w\|_{H^1(K)} \|\partial_t v\|_{L^2(t_{j-1}, t_{j+1}; H^{-1}(K))} \\ &\qquad + \tau_j^{-1/2} h_K \|w\|_{L^2(K)} \\ &\qquad\qquad \|v\|_{L^2(t_{j-1}, t_{j+1}; H^1(\tilde{\omega}_K))} \}. \end{aligned} \qquad [32]$$

Here, $Q = K \times J_j$ is an arbitrary prism. E is a face of K, v is an element of Y, and $w \in H^1(K)$ is arbitrary. The constants $\tilde{c}_{I1}, \ldots, \tilde{c}_{I3}$ only depend on the shape-regularity via c_τ defined above.

11. Equivalence of error and residual

Since Problems [24] and [26] fit into the general framework of Section 5, we immediately obtain the equivalence of the error $\|u - u_h\|_X = \|u - u_h\|_{L^2(0,T;H^1(\Omega))}$ and of the residual:

$$\|F(u_h)\|_{Y'} = \sup_{v \in Y \setminus \{0\}} \frac{\langle F(u_h), v \rangle}{\|v\|_Y}.$$

(Recall that $\|.\|_Y = \|.\|_{W^2(0,T;H^1(\Omega),H^{-1}(\Omega))}$.)

The Lipschitz continuity of $DF(u)$ is satisfied if the functions $a_1, \ldots, a_n$ and b fulfill suitable smoothness and growth conditions. The invertibility of $DF(u)$ is equivalent to the unique solvability of the linearized parabolic pde [22]. The corresponding constants α and A now also depend on the final time T. Usually $\alpha^{-1}A$ will be a monotonically increasing function of T. This introduces new difficulties. In particular it in generally excludes estimates which are global in time. A satisfactory general theory for long-time a posteriori error estimates is still lacking; first results are given in [ERI 91, ERI 95a, ERI 95b].

12. A residual error estimator

In order to obtain computable upper and lower bounds for the residual $\|F(u_h)\|_{Y'}$ we proceed as in Section 6. Consider a function $v \in Y$ with $\|v\|_Y = \|v\|_{W^2(0,T;H^1(\Omega),H^{-1}(\Omega))} = 1$. Performing integration by parts with regard to space and time on each prism Q we conclude that:

$$\begin{aligned} &\langle F(u_h), v \rangle \\ &= \sum_{Q \in \mathcal{P}_\tau} \Big\{ \int_Q \{\partial_t u_h - \partial_i a_i(x, u_h, \nabla u_h) - b(x, u_h, \nabla u_h)\} v \\ &\qquad + \frac{1}{2} \int_{\partial Q_L} [n_{E,i} a_i(x, u_h, \nabla u_h)]_{\partial Q_L} v + \int_{\partial Q_B} [u_h]_{\partial Q_B} v \Big\} \\ &=: \sum_{Q \in \mathcal{P}_\tau} \Big\{ \int_Q R_Q(u_h) v + \frac{1}{2} \int_{\partial Q_L} R_{\partial Q_L}(u_h) v + \int_{\partial Q_B} R_{\partial Q_B}(u_h) v \Big\}. \end{aligned} \quad [33]$$

The integrals along the lateral boundaries are weighted with a factor one half since each face is counted twice. Note that for $Q = K \times J$:

$$[\psi]_{\partial Q_B} = \psi(., \tau_j + 0) - \psi(., \tau_j - 0).$$

In order to obtain the compact form of [33] we therefore used the convention that:

$$u_h(.,0-0) := u_0 \qquad [34]$$

where u_0 is the given initial value.

Thanks to equation [27] we still have the Galerkin orthogonality [14]. Hence, we may replace v on the right-hand side of equation [33] by $v - I_\tau v$. Applying the Cauchy-Schwarz inequality for integrals we thus arrive at:

$$\begin{aligned}\langle F(u_h), v\rangle \leq \sum_{Q\in P_\tau} \Big\{ & \|R_Q(u_h)\|_{L^2(Q)}\|v - I_\tau v\|_{L^2(Q)} \\ & + \frac{1}{2}\|R_{\partial Q_L}(u_h)\|_{L^2(\partial Q_L)}\|v - I_\tau v\|_{L^2(\partial Q_L)} \\ & + \int_{\partial Q_B} |R_{\partial Q_B}(u_h)||v - I_\tau v| \Big\}.\end{aligned}$$

Inserting the estimates [32] and using the inverse inequality:

$$\|R_{\partial Q_B}(u_h)\|_{H^1(K)} \leq c h_K^{-1}\|R_{\partial Q_B}(u_h)\|_{L^2(K)}$$

we conclude that:

$$\begin{aligned}\langle F(u_h), v\rangle \leq \sum_{Q\in P_\tau} \Big\{ & \tilde{c}_{I1}[h_K + \tau_j h_K^{-1}]\|R_Q(u_h)\|_{L^2(Q)}\|v\|_{Y|\tilde{\omega}_Q} \\ & + \frac{1}{2}\tilde{c}_{I2}[h_E^{1/2} + \tau_j h_E^{-3/2}]\|R_{\partial Q_L}(u_h)\|_{L^2(\partial Q_L)}\|v\|_{Y|\tilde{\omega}_Q} \\ & + \tilde{c}_{I3}[\tau_j^{-1/2} h_K + \tau_j^{1/2} h_K^{-1}]\|R_{\partial Q_B}(u_h)\|_{L^2(\partial Q_B)}\|v\|_{Y|\tilde{\omega}_Q} \Big\}.\end{aligned}$$

Here, $\|.\|_{Y|\tilde{\omega}_Q}$ denotes the natural restriction of $\|.\|_Y$ to the set $\tilde{\omega}_Q$ which is the union of all prims that have at most one point in common with Q. Using the Cauchy-Schwarz inequality for finite sums we finally arrive at the upper bound:

$$\|F(u_h)\|_{Y'}, \leq c \max\{\tilde{c}_{I1}, \tilde{c}_{I2}, \tilde{c}_{I3}\} \left\{ \sum_{Q\in P_\tau} \eta_{R,Q}^2 \right\}^{1/2}, \qquad [35]$$

where c is the maximal number of prisms contained in $\tilde{\omega}_Q$ and where:

$$\begin{aligned}\eta_{R,Q} := \Big\{ & [h_K + \tau_j h_K^{-1}]^2 \|R_Q(u_h)\|_{L^2(Q)}^2 \\ & + \frac{1}{2} h_E^{-1}[h_E + \tau_j h_E^{-1}]^2 \|R_{\partial Q_L}(u_h)\|_{L^2(\partial Q_L)}^2 \\ & + \tau_j^{-1}[h_K + \tau_j h_K^{-1}]^2 \|R_{\partial Q_B}(u_h)\|_{L^2(\partial Q_B}^2 \Big\}^{1/2}.\end{aligned} \qquad [36]$$

In order to prove the efficiency of the error estimator we proceed in exactly the same way as in Section 6. We first define modified residuals $\tilde{R}_Q(u_h)$ and $\tilde{R}_{\partial Q_L}(u_h)$ by approximating the functions $a_1, \ldots, a_n$, and b by functions $a_{1,h}, \ldots, a_{n,h}$, and b_h which are piecewise polynomials. The residual $R_{\partial Q_B}(u_h)$ may not be modified since it only involves jumps of u_h which is a piecewise polynomial. Then we use estimate [29] to bound the contributions to $\eta_{R,Q}$. For the element residual, e.g., we thus proceed as follows. From estimate [29] we get:

$$\delta_1 \|\tilde{R}_Q(u_h)\|^2_{L^2(Q)} \leq \int_Q \tilde{R}_Q(u_h)\psi_Q \tilde{R}_Q(u_h) =: \int_Q \tilde{R}_Q(u_h) w_Q.$$

Inserting $w_Q := \psi_Q \tilde{R}_Q(u_h)$ as a test-function v in eqation [33] we obtain:

$$\begin{aligned}\int_Q \tilde{R}_Q(u_h) w_Q =& \langle F(u_h), w_Q\rangle + \int_Q [\tilde{R}_Q(u_h) - R_Q(u_h)] w_Q \\ \leq& \|F(u_h)\|_{Y'} \|w_Q\|_Y + \|\tilde{R}_Q(u_h) - R_Q(u_h)\|_{L^2(Q)} \|w_Q\|_{L^2(Q)}.\end{aligned}$$

Estimate [29] yields:

$$\begin{aligned}\|w_Q\|_Y =& \left\{ \|w_Q\|^2_{L^2(J_j, H^1(K))} + \|\partial_t w_Q\|^2_{L^2(J_j, H^{-1}(K))} \right\}^{1/2} \\ \leq& \{\delta_2^2 h_K^{-2} + \delta_4^2 \sigma_n(h_K)^2 \delta_3^2 \tau_j^{-2}\}^{1/2} \|\tilde{R}_Q(u_h)\|_{L^2(Q)} \\ \leq& \max\{\delta_2, \delta_3, \delta_4\}\{h_K^{-1} + \sigma_n(h_K)\tau_j^{-1}\} \|\tilde{R}_Q(u_h)\|_{L^2(Q)}.\end{aligned}$$

Since $\|w_Q\|_{L^2(Q)} \leq \|\tilde{R}_Q(u_h)\|_{L^2(Q)}$ we obtain:

$$\begin{aligned}&[h_K + \tau_j h_K^{-1}] \|R_Q(u_h)\|_{L^2(Q)} \\ \leq& \frac{1}{\delta_1} \max\{\delta_2, \delta_3, \delta_4\} [h_K + \tau_j h_K^{-1}][h_K^{-1} + \sigma_n(h_K)\tau_j^{-1}] \|F(u_h)\|_{Y'} \\ &+ (1 + \frac{1}{\delta_1})[h_K + \tau_j h_K^{-1}] \|R_Q(u_h) - \tilde{R}_Q(u_h)\|_{L^2(Q)}.\end{aligned}$$

The remaining terms in $\eta_{R,Q}$ are treated in exactly the same way. Summarizing all estimates we finally obtain the estimate:

$$\begin{aligned}\eta_{R,Q} \leq& c_1 [h_K + \tau_j h_K^{-1}][h_K^{-1} + \sigma_n(h_K)\tau_j^{-1}] \|F(u_h)\|_{Y'} \\ &+ c_2 [h_K + \tau_j h_K^{-1}] \|R_Q(u_h) - \tilde{R}_Q(u_h)\|_{L^2(Q)} \\ &+ c_3 [h_E + \tau_j h_E^{-1}] h_E^{-1/2} \|R_{\partial Q_L}(u_h) - \tilde{R}_{\partial Q_L}(u_h)\|_{L^2(\partial Q_L)}.\end{aligned} \qquad [37]$$

The second and third terms on the right-hand side of estimate [37] are again higher order perturbations which only depend on the smoothness of the functions $a_1, \ldots, a_n$ and b. The constants $c_1, \ldots, c_3$ depend on $\delta_1, \ldots, \delta_{12}$. We thus obtain the efficiency and locality of the error estimator.

But in contrast to corresponding results for elliptic problems, we now have an additional factor $[h_K + \tau_j h_K^{-1}][h_K^{-1} + \sigma_n(h_K)\tau_j^{-1}]$. This factor is of order one if and only if the *CFL* condition $\tau_j \sim h_K^2$ is satisfied. This condition is annoying when recalling that the choice $\theta = \frac{1}{2}$ yields the Crank-Nicolson scheme which is of second order. On the other hand this condition is very natural when recalling that the pde [22] is of second order with regard to space but of first order with regard to time. In this sense [22] is the limit case of a second order equation which is singularly perturbed with regard to time. Thus one may perhaps avoid this CFL condition if one succeeds in adopting the methods for singularly perturbed elliptic pdes. Finally, we stress that the above CFL condition does not show up in the existing literature, e.g. [ERI 91, ERI 95a, ERI 95b], since only upper bounds on the error are established there.

The error estimator contains contributions from the lateral faces and form the bottoms of the space-time prisms. Thus their relative sizes could be used for an anisotropic refinement with regard to space and time. This would correspond to a local time-stepping. But up to now the correct treatment of local time stepping within a variational framework is a completely open problem.

13. Error estimators based on the solution of auxiliary local problems

The techniques of Section 7 may be extended to parabolic pdes too (cf. [LON 98]). This gives rise to error estimators which are based on the approximate solution of auxiliary local parabolic problems. To give an example, consider an arbitrary space-time prism $Q = K \times J_j$, $K \in \mathcal{T}_j$. Denote by ω_Q the union of all prisms that share at least one point with the lateral boundary and the bottom of Q. Choose a finite element space V_Q consisting of piecewise polynomials of a sufficiently high degree which vanish on $\partial\omega_Q$. For example V_Q may be choosen such that it contains the functions $\psi_Q \tilde{R}_{Q'}(u_h)$, $\psi_{E,j}\tilde{R}_{\partial Q_L}(u_h)$, $\psi_{K,j}\tilde{R}_{\partial Q_B}(u_h)$ with $Q' \subset \omega_Q$, $E \subset \partial K$. Set $A_{ij} = \partial_{p_j}\pi_Q a_i(x, u_h(x), \nabla u_h(x))$, where π_Q denotes a suitable average on ω_Q, i.e. the L^2-projection onto $\mathbb{R}$. Since the functions in V_Q vanish on $\partial\omega_Q$, the problem:

$$-\int_{\omega_Q} v\partial_t w + \int_{\omega_Q} A_{ij}\partial_i v \partial_j w = \langle F(u_h), w\rangle \quad \forall w \in Y_Q \qquad [38]$$

admits a unique solution $v_Q \in V_Q$. Set:

$$\eta_{D,Q} := \|v_Q\|_{L^2(J_{j-1}\cup J_j, H^1(\tilde{\omega}_K))}. \qquad [39]$$

Inserting v_Q as a test-function w in [38] and using a scaling argument, one concludes that:

$$\eta_{D,Q} \le c[1 + \tau_j^{-1} h_K^2]\|F(u_h)\|_{Y'}.$$

Inserting the functions $\psi_Q \tilde{R}_Q(u_h)$, $\psi_{E,j} \tilde{R}_{\partial Q_L}(u_h)$, and $\psi_{K,j} \tilde{R}_{\partial Q_B}(u_h)$ as test-functions in [38] and using the estimates of the previous section, one obtains that:

$$\begin{aligned}\eta_{R,Q} \leq & c_1[1 + \tau_j h_K^{-2} + \sigma_n(h_K) h_K \tau_j^{-1}]\eta_{D,Q} \\ & + c_2[h_K + \tau_j h_K^{-1}] \|R_Q(u_h) - \tilde{R}_Q(u_h)\|_{L^2(Q)} \\ & + c_3[h_E + \tau_j h_E^{-1}] h_E^{-1/2} \|R_{\partial Q_L}(u_h) - \tilde{R}_{\partial Q_L}(u_h)\|_{L^2(\partial Q_L)}.\end{aligned}$$

Hence, $\eta_D := \left\{ \sum_{Q \in \mathcal{P}_\tau} \eta_{D,Q}^2 \right\}^{1/2}$ is a reliable, efficient and local error estimator for $\|u - u_h\|_X$ provided the CFL condition $\tau_j \sim h_K^2$ is satisfied.

Note that problem [38] is a discrete analogue of the parabolic pde:

$$\begin{aligned}\partial_t u - \partial_i(A_{ij}\partial_j u) &= F(u_h) && \text{in } \tilde{\omega}_K \times (t_{j-1}, t_{j+1}) \\ u &= 0 && \text{on } \partial\tilde{\omega}_K \times (t_{j-1}, t_{j+1}) \\ u(., t_{j-1}) &= 0 && \text{in } \tilde{\omega}_K.\end{aligned}$$

References

[ADJ 88] ADJERID S., FLAHERTY J. E., ≪A local refinement finite element method for two dimensional parabolic problems≫, *SIAM J. Sci. Stat. Comput.*, vol. 9, 1988, p. 792 – 811.

[BAB 78] BABUŠKA I., RHEINBOLDT W. C., ≪Error estimates for adaptive finite element computations≫, *SIAM J. Numer. Anal.*, vol. 15, 1978, p. 736 – 754.

[BAN 85] BANK R. E., WEISER A., ≪Some a posteriori error estimators for elliptic partial differential equations≫, *Math. Comput.*, vol. 44, 1985, p. 283 – 301.

[BEC 95] BECKER R., RANNACHER R., ≪Weighted a posteriori error control in FE methods≫, *Proc. ENUMATH-95*, Paris, September 1995

[BER 00] BERNARDI CH., VERFÜRTH R., ≪Adaptive finite element methods for elliptic equations with non-smooth coefficients≫, *Numer. Math.*, 2000, (to appear)

[BIE 82a] BIETERMAN M., BABUŠKA I., ≪The finite element method for parabolic equations I. A posteriori error estimation≫, *Numer. Math.*, vol. 40, 1982, p. 339 – 371.

[BIE 82b] BIETERMAN M., BABUŠKA I., ≪The finite element method for parabolic equations II. A posteriori error estimation and adaptive approach≫, *Numer. Math.*, vol. 40, 1982, p. 373 - 406.

[BOR 90] BORNEMANN F., ≪An adaptive multilevel approach to parabolic equations I. General theory and 1D implementation≫, *IMPACT of Comp. in Sci. and Engrg.*, vol. 2, 1990, 279 - 317.

[BOR 91] BORNEMANN F., ≪An adaptive multilevel approach to parabolic equations II. Variable order time discretization based on a multiplicative error correction≫, *IMPACT of Comp. in Sci. and Engrg.*, vol. 3, 1991, p. 93 - 122.

[BOR 92] BORNEMANN F., ≪An adaptive multilevel approach to parabolic equations III. 2D error estimation and multilevel preconditioning≫, *IMPACT of Comp. in Sci. and Engrg.*, vol. 4, 1992, p. 1 - 45.

[ERI 91] ERIKSSON K., JOHNSON C., ≪Adaptive finite element methods for parabolic problems I. A linear model problem≫, *SIAM J. Numer. Anal.*, vol. 28, 1991, p. 43 - 77.

[ERI 95a] ERIKSSON K., JOHNSON C., ≪Adaptive finite element methods for parabolic problems IV. Nonlinear problems≫, *SIAM J. Numer. Anal.*, vol. 32, 1995, p. 1724 - 1749.

[ERI 95b] ERIKSSON K., JOHNSON C., ≪Adaptive finite element methods for parabolic problems V. Long-time integration≫, *SIAM J. Numer. Anal.*, vol. 32, 1995, p. 1750 - 1763.

[ERI 85] ERIKSSON K., JOHNSON C., THOMÉE V., ≪Time discretization of parabolic problems by the discontinuous Galerkin method≫, *Modélisation Mathématique et Analyse Numérique*, vol. 19, 1985, p. 611 - 643.

[KUN 98] KUNERT G., A posteriori error estimation for anisotropic tetrahedral and triangular finite element meshes. PhD thesis, TU Chemnitz, 1998

[KUN 00] KUNERT G., VERFÜRTH R., ≪Edge residuals dominate a posteriori error estimates for linear finite element methods on anisotropic triangular and tetrahedral meshes≫, *Numer. Math.*, 2000, (to appear)

[LON 98] LONSING M., Zwei a posteriori Fehlerschätzer für die Wärmeleitungsgleichung. Diplomarbeit, Ruhr-Universität Bochum, 1998

[SIE 96] SIEBERT K. G., ≪An a posteriori error estimator for anisotropic refinement≫, *Numer. Math.*, vol. 73, 1996, p. 373 -398.

[VER 96] VERFÜRTH R., *A Review of A Posteriori Error Estimation and Adaptive Mesh - Refinement Techniques*, Wiley-Teubner, Stuttgart, 1996.

[VER 98a] VERFÜRTH R., ≪Robust a posteriori error estimators for a singularly perturbed reaction-diffusion equation≫, *Numer. Math.*, vol. 78, 1998, 479 - 493.

[VER 98b] VERFÜRTH R., ≪A posteriori error estimators for convection-diffusion equations≫, *Numer. Math.*, vol. 80, 1998, p. 641 - 663.

[VER 98c] VERFÜRTH R., ≪A posteriori error estimates for nonlinear problems. $L^r(0, T; W^{1,\rho}(\Omega))$-error estimates for finite element discretizations of parabolic equations≫, *Numer. Meth. for PDE*, vol. 14, 1998, p. 487 - 518.

[VER 98d] VERFÜRTH R., ≪A posteriori error estimates for nonlinear problems. $L^r(0, T; L^\rho(\Omega))$-error estimates for finite element discretizations of parabolic equations≫, *Math. Comput.*, vol. 67, 1998, p. 1335 - 1360.

[VER 99] VERFÜRTH R., ≪Error estimates for some quasi-interpolation operators≫, *Modélisation Mathématique et Analyse Numérique*, vol. 33, 1999, p. 695 - 713.

[VER 00] VERFÜRTH R., ≪On the constants in some inverse inequalities for finite element functions≫, *Modélisation Mathématique et Analyse Numérique* (submitted)