

A Posteriori Error Analysis of the Method of Characteristics

Christine Bernardi, Frédéric Hecht, Rüdiger Verfürth

▶ **To cite this version:**

Christine Bernardi, Frédéric Hecht, Rüdiger Verfürth. A Posteriori Error Analysis of the Method of Characteristics. Mathematical Models and Methods in Applied Sciences, 2009, 21 (06), pp.1355-1377. 10.1142/S0218202511005337 . hal-00395280

HAL Id: hal-00395280
https://hal.science/hal-00395280

Submitted on 15 Jun 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

A Posteriori Error Analysis of the Method of Characteristics

Christine Bernardi, Frédéric Hecht
Laboratoire Jacques-Louis Lions
C.N.R.S. & Université Pierre et Marie Curie
B.C. 187
4 place Jussieu
F-75252 Paris Cedex 05
France
bernardi@ann.jussieu.fr, hecht@ann.jussieu.fr

and

Rüdiger Verfürth
Fakultät für Mathematik
Ruhr-Universität Bochum
D-44780 Bochum
Germany
ruediger.verfuerth@ruhr-uni-bochum.de

Abstract. We consider a two- or three-dimensional time-dependent diffusion-convection-reaction problem and its discretization by the method of characteristics and standard finite elements. We perform the a posteriori error analysis of this discretization and prove optimal error estimates, which lead to an efficient adaptivity strategy both for the time step and the spatial mesh. The estimates are robust with respect to the ratios of the diffusion to the reaction or convection. Some numerical experiments support the theoretical results.

Résumé. On considère un problème instationnaire de type diffusion-convection-réaction en dimension 2 ou 3 et sa discrétisation par la méthode des caractéristiques et des éléments finis usuels. On effectue l'analyse d'erreur a posteriori de cette discrétisation et on prouve des estimations optimales de l'erreur qui mènent à une stratégie efficace d'adaptation des pas de temps et des maillages. Les estimations sont uniformes par rapport aux taux de diffusion, de réaction ou de convection. Quelques expériences numériques confirment les résultats théoriques.

1. Introduction

In this note we consider general linear parabolic differential equations of second order, equation (2.1), and their discretization by finite elements based on the method of characteristics, equations (3.1) – (3.3). Our focus is on deriving reliable, efficient and robust a posteriori error estimates for these discretizations. Here, as usual, robust means that the quality of the error indicators, i.e. the ratio of the upper and lower bounds, does not depend on the relative size of the diffusion to the convection or reaction.

The method of characteristics was initiated in [16] as an alternative to standard space-time finite elements with or without upwinding or SUPG-stabilization. Its main advantages are that it implicitly includes an upwinding and thus stabilizing effect and that it requires only the solution of symmetric coercive problems. On the other hand it needs an efficient way for back-tracking the characteristics and re-interpolating finite element functions. Depending on the data structure of the implementation, this can be costly or not. The initial method was of order one but was soon extended to higher order in [9]. We refer, e.g., to [5, 6, 9, 17, 18, 19] for a more detailed description of the method of characteristics including a priori error estimates.

On the other hand, a large amount of work has been done concerning the a posteriori analysis of parabolic equations aimed at time and space adaptivity, see e.g. [4, 8, 11, 12, 21, 23, 24, 25] and the references therein. However, we are not aware of any a posteriori error analysis for the method of characteristics although, as we will see, it can be performed by adapting of the approach of [23, 24, 25].

The subsequent analysis follows the general framework of a posteriori error analysis and is performed in several steps. First, in Lemma 4.1, we prove that the energy norm of the error and the norm of its material derivative are equivalent to the dual norm of the residual (see section 2 for a detailed definition of these norms). This stability result is based on standard parabolic energy estimates and is independent of the discretization.

Next, in section 5, we decompose the residual into three contributions: a term representing temporal data oscillations, a temporal residual and a spatial residual. The first term can easily be estimated separately. For the estimation of two other terms it is crucial that the norm of their sum is equivalent to the sum of their norms, cf. Lemma 5.1. This result essentially is a strengthened Cauchy–Schwarz inequality. However, it is proved in a different way which is better adapted to dual norms.

The temporal residual is estimated in section 6. Here, the crucial step is Lemma 6.1 which allows to split the temporal residual into two

contributions reflecting the effects of the convection and of the combination of the diffusion and reaction. In case of a dominant diffusion or reaction, the contribution of the convection can be controlled by the diffusion-reaction term. In case of a dominant convection, however, this leads to a non robust estimate. To recover robustness, as in [25], the convection term must be controlled with the help of an auxiliary discrete diffusion-reaction problem.

The spatial residual is the term which reflects the difference between the method of characteristics and standard space-time finite element discretizations. It is a linear functional which admits an L^2-representation involving contributions by the elements and faces of the partition. Thus it falls into the abstract class of functionals which are amenable to a posteriori error estimation [20, 27]. However, it does not satisfy the Galerkin orthogonality, i.e. its kernel does not contain the lowest order conforming finite element space. Instead, its consistency error describes the difference between the method of characteristics and the backward Euler scheme. Contrary to the consistency errors of SUPG- or discontinuous Galerkin schemes, this term is crucial and cannot be controlled by the standard element and face residuals. Instead, in Lemma 7.1, we prove that this term can be controlled with the help of an auxiliary discrete diffusion-reaction problem. Thus its treatment is similar to the one of the convection in the convection-dominated regime.

In section 8 we collect all estimates and thus obtain our main result, Theorem 8.1. When compared with [25] the error indicator of Theorem 8.1 requires, in each time step, the solution of an additional discrete diffusion-reaction problem corresponding to the lowest order conforming finite element space. This extra cost is compensated by the less costly discretization. We stress that all estimates are robust with respect to the relative size of the diffusion compared to the reaction or convection. Moreover, they hold for every conforming spatial discretization based on Lagrangian elements and every ODE-solver for the characteristic equations as long as the endpoints of the approximate characteristics remain within the closure of the domain.

In section 9 we present some numerical experiments both for convection and diffusion-dominated problems in a square. The numerical results confirm the robustness of our estimates.

2. The differential equation

In what follows we consider general linear parabolic equations of second order:

$$\begin{aligned} \partial_t u - \operatorname{div}(D\nabla u) + \mathbf{a}\cdot\nabla u + bu &= f \quad \text{in } \Omega\times(0,T] \\ u &= 0 \quad \text{on } \Gamma\times(0,T] \\ u|_{t=0} &= u_0 \quad \text{in } \Omega. \end{aligned} \tag{2.1}$$

Here, $\Omega \subset \mathbb{R}^d$ is a bounded polygonal two or three dimensional domain with a Lipschitz-continuous boundary Γ. The final time T is arbitrary, but kept fixed in the sequel. We assume that the coefficients are smooth enough:

$$D \in C(0,T;L^\infty(\Omega)^{d\times d}),\ \mathbf{a} \in C(0,T;W^{1,\infty}(\Omega)^d),\ b \in C(0,T;L^\infty(\Omega)),$$

and satisfy the following conditions:

- The diffusion D is symmetric, uniformly positive definite and uniformly isotropic, i.e.

$$\varepsilon = \inf_{0<t\le T, x\in\Omega} \min_{z\in\mathbb{R}^d\setminus\{0\}} \frac{z^T D(x,t) z}{z^T z} > 0$$

  and the quantity

$$\kappa = \varepsilon^{-1} \sup_{0<t\le T, x\in\Omega} \max_{z\in\mathbb{R}^d\setminus\{0\}} \frac{z^T D(x,t) z}{z^T z}$$

  is of moderate size.
- The convection $\mathbf{a}$ is solenoidal and vanishes on the boundary:

$$\begin{aligned} \operatorname{div} \mathbf{a} &= 0 \quad \text{in } \Omega\times(0,T], \\ \mathbf{a} &= 0 \quad \text{on } \Gamma\times(0,T]. \end{aligned}$$

- The reaction b is non-negative and does not vary too much, i.e. there are two constants $\beta \ge 0$ and $c_b \ge 0$ such that

$$\beta \le b \le c_b\beta \quad \text{in } \Omega\times(0,T].$$

  Moreover, the constant c_b should be of moderate size.

These assumptions guarantee that, for any data f in $L^2(\Omega\times(0,T))$ and u_0 in $L^2(\Omega)$, problem (2.1) is a well-posed parabolic problem, see e.g. [3, Sections II.4 and V.2] and [10, Theorems XVIII.3.1 and XVIII.3.2]. The parameter κ is introduced in order to stress that the ratio of the constants in the error estimates depends on the condition number of the diffusion matrix. If this condition number is exceedingly large, length scales such as element diameters must be measured in a diffusion-dependent metric in order to recover robust estimates as described in [7] for elliptic equations. The assumptions cover a wide range of different regimes:

- *dominant diffusion*: $\|\mathbf{a}\|_{L^\infty(0,T;W^{1,\infty}(\Omega))} \le c_c\varepsilon$ and $\beta \le c_b'\varepsilon$ with constants of moderate size;
- *dominant reaction*: $\|\mathbf{a}\|_{L^\infty(0,T;W^{1,\infty}(\Omega))} \le c_c\varepsilon$ and $\beta \gg \varepsilon$ with a constant c_c of moderate size;
- *dominant convection*: $\|\mathbf{a}\|_{L^\infty(0,T;W^{1,\infty}(\Omega))} \gg \varepsilon$.

When compared with [25] we here consider a more general problem in that $D = \varepsilon I$ is assumed in [25]. On the other hand, the assumptions of [25] are slightly more general concerning the convection and the boundary conditions in that there the convection may not be solenoidal, may

not vanish on the boundary and that Neumann boundary conditions are allowed. The present assumptions are mainly made in order to avoid unnecessary technical difficulties with the method of characteristics.

We conclude this section with some notation and technical results that will be used in the sequel. For any bounded open subset ω of Ω with Lipschitz boundary γ, we denote by $H^k(\omega)$, $k \in \mathbb{N}$, $L^2(\omega) = H^0(\omega)$, and $L^2(\gamma)$ the usual Sobolev and Lebesgue spaces equipped with the standard norms $\|\cdot\|_{k;\omega} = \|\cdot\|_{H^k(\omega)}$ and $\|\cdot\|_\gamma = \|\cdot\|_{L^2(\gamma)}$ (cf. [1]). Similarly, $(\cdot,\cdot)_\omega$ and $(\cdot,\cdot)_\gamma$ denote the scalar products of $L^2(\omega)$ and $L^2(\gamma)$, respectively. If $\omega = \Omega$, we omit the subscript Ω.

As usual, $H_0^1(\Omega)$ is the Sobolev space of L^2-functions having their derivatives in $L^2(\Omega)$ and vanishing on the boundary Γ. It is equipped with the norm

$$|||v||| = \big\{\varepsilon\|\nabla v\|^2 + \beta\|v\|^2\big\}^{\frac{1}{2}}.$$

The dual space of $H_0^1(\Omega)$ is denoted by $H^{-1}(\Omega)$ and is equipped with the norm

$$|||\varphi|||_* = \sup_{v\in H_0^1(\Omega)\setminus\{0\}} \frac{\langle\varphi, v\rangle}{|||v|||}$$

where $\langle\cdot,\cdot\rangle$ stands for the standard duality pairing.

For any separable Banach space V with norm $\|\cdot\|_V$ and any two numbers $a < b$, we denote by $L^2(a,b;V)$ and $L^\infty(a,b;V)$ the spaces of measurable functions u defined on (a,b) with values in V such that the mapping $t \mapsto \|u(\cdot,t)\|_V$ is square integrable, respectively, essentially bounded. These are Banach spaces equipped with the norms

$$\|u\|_{L^2(a,b;V)} = \Big\{\int_a^b \|u(\cdot,t)\|_V^2 dt\Big\}^{\frac{1}{2}},$$
$$\|u\|_{L^\infty(a,b;V)} = \operatorname*{ess.\,sup}_{t\in(a,b)} \|u(\cdot,t)\|_V$$

(cf. [10, Chap. XVIII, §1]). For abbreviation we introduce the space

$$X(a,b) = \big\{u \in L^2(a,b;H_0^1(\Omega)) \cap L^\infty(a,b;L^2(\Omega)) : \partial_t u + \mathbf{a}\cdot\nabla u \in L^2(a,b;H^{-1}(\Omega))\big\}$$

and equip it with its graph norm

$$\|u\|_{X(a,b)} = \Big\{\operatorname*{ess.\,sup}_{t\in(a,b)} \|u(\cdot,t)\|^2 + \int_a^b |||u(\cdot,t)|||^2 dt + \int_a^b |||(\partial_t u + \mathbf{a}\cdot\nabla u)(\cdot,t)|||_*^2 dt\Big\}^{\frac{1}{2}}.$$

Here the derivative $\partial_t u$ has to be understood in the distributional sense [10, loc.cit.]. Note that $\partial_t u + \mathbf{a}\cdot\nabla u$ is the material derivative associated with the differential operator of (2.1). Furthermore, according to [10,

loc.cit.], every function $u \in X(a,b)$ has point-values $u(\cdot,a)$ and $u(\cdot,b)$ in $H^{-1}(\Omega)$.

With this notation, the variational formulation of problem (2.1) takes the form:

Find $u \in X(0,T)$ such that $u(\cdot,0) = u_0$ in $H^{-1}(\Omega)$ and for almost every $t \in (0,T)$ and all $v \in H^1_0(\Omega)$

$$(\partial_t u, v) + (D\nabla u, \nabla v) + (\mathbf{a}\cdot\nabla u, v) + (bu, v) = (f, v). \tag{2.2}$$

These assumptions imply that, for any data f in $L^2(0,T;H^{-1}(\Omega))$ and u_0 in $L^2(\Omega)$, problem (2.2) admits a unique solution, see e.g. [3, Sections II.4 and V.2] or [10, Theorems XVIII.3.1 and XVIII.3.2].

The key idea of the method of characteristics relies on another formulation of problem (2.1). Since the function $\mathbf{a}$ is Lipschitz continuous with respect to the spatial variable and vanishes on the boundary Γ, for every $(x^*,t^*) \in \Omega \times (0,T]$, standard global existence results for the flows of ordinary differential equations (cf., e.g., [2, Theorem II.7.6]) imply that the characteristic equation

$$\begin{aligned} \frac{d}{dt}x(t;x^*,t^*) &= \mathbf{a}(x(t;x^*,t^*),t), \quad t \in (0,t^*), \\ x(t^*;x^*,t^*) &= x^* \end{aligned} \tag{2.3}$$

has a unique solution $x(\cdot;x^*,t^*)$ which exists for all $t \in [0,t^*]$ and stays within $\Omega \cup \Gamma$. Hence, we may set $U(x^*,t) = u(x(t;x^*,t^*),t)$. The total derivative $d_t U$ satisfies

$$d_t U = \partial_t u + \mathbf{a}\cdot\nabla u.$$

Therefore, the first line of problem (2.1) can equivalently be written as

$$d_t U - \operatorname{div}(D\nabla u) + bu = f \quad \text{in } \Omega \times (0,T). \tag{2.4}$$

The discretization introduced in the next section relies on a separate treatment of equations (2.3) and (2.4).

For later use we finally note that integration by parts and the previous assumptions imply for all $v, w \in H^1_0(\Omega)$ and all $t \in (0,T]$:

$$(D\nabla v, \nabla v) + (\mathbf{a}\cdot\nabla v, v) + (bv, v) \geq |||v|||^2, \tag{2.5}$$

$$(D\nabla v, \nabla w) + (bv, w) \leq \max\{\kappa, c_b\}|||v|||\,|||w|||. \tag{2.6}$$

3. The finite element discretization

For the discretization of problem (2.1) we consider partitions $\mathcal{I} = \{(t_{n-1},t_n] : 1 \leq n \leq N_{\mathcal{I}}\}$ of the time-interval $[0,T]$ into subintervals satisfying $0 = t_0 < \ldots < t_{N_{\mathcal{I}}} = T$. For every n with $1 \leq n \leq N_{\mathcal{I}}$ we denote by $I_n = (t_{n-1},t_n]$ the n-th subinterval and by $\tau_n = t_n - t_{n-1}$ its length. With every intermediate time t_n, $0 \leq n \leq N_{\mathcal{I}}$, we associate a partition $\mathcal{T}_n$ of Ω and a corresponding finite element space $V_n \subset H^1_0(\Omega)$ which have to satisfy the following assumptions:

- The closure of Ω is the union of all elements in $\mathcal{T}_n$.

- Every element in $\mathcal{T}_n$ is either a *simplex* or a *parallelepiped*, i.e., it is the image of the d-dimensional *reference simplex* $\{x \in \mathbb{R}^d : x_1 \geq 0, \ldots, x_d \geq 0, x_1 + \ldots + x_d \leq 1\}$ or of the d-dimensional *reference cube* $[0,1]^d$ under an affine mapping (*affine equivalence*).
- Any two elements in $\mathcal{T}_n$ are either disjoint or share a complete lower dimensional face of their boundaries (*admissibility*).
- For every element K, the ratio of its diameter h_K to the diameter ρ_K of the largest ball inscribed into K is bounded uniformly with respect to all partitions $\mathcal{T}_n$ and $\mathcal{I}$ (*shape regularity*).
- For every n with $1 \leq n \leq N_{\mathcal{I}}$ there is an affine equivalent, admissible, and shape-regular partition $\widetilde{\mathcal{T}}_n$ such that it is a refinement of both $\mathcal{T}_n$ and $\mathcal{T}_{n-1}$ and such that

$$\sup_{1 \leq n \leq N_{\mathcal{I}}} \sup_{K \in \widetilde{\mathcal{T}}_n} \sup_{\substack{K' \in \mathcal{T}_n \\ K \subset K'}} \frac{h_{K'}}{h_K} < \infty$$

  uniformly with respect to all partitions $\mathcal{I}$ which are obtained by adaptive or uniform refinement of any initial partition of $[0,T]$ (*transition condition*).
- Each V_n consists of continuous functions which are polynomials on each element of $\mathcal{T}_n$, the degrees being at least one and being bounded uniformly with respect to all partitions $\mathcal{T}_n$ and $\mathcal{I}$ (*degree condition*).
- For every n there is a set $\mathcal{V}_n$ of *nodes* such that every function in V_n is uniquely defined by its values in the nodes (*Lagrange condition*).

When compared with [25] the Lagrange condition is new. It simplifies the formulation of the method of characteristics. The transition condition is needed for the a posteriori error estimates of Lemmas 6.2 and 7.1.

Every set $\mathcal{V}_n$ splits into the subsets $\mathcal{V}_{n,\Omega}$ and $\mathcal{V}_{n,\Gamma}$ of all nodes inside Ω respective on Γ. For every intermediate time t_n, $1 \leq n \leq N_{\mathcal{I}}$, and every node $z \in \mathcal{V}_{n,\Omega}$ we compute an approximation x_z^{n-1} to $x(t_{n-1}; z, t_n)$ by applying an arbitrary but fixed ODE-solver such as e.g. the explicit Euler scheme to the characteristic equation (2.3) with $(x^*, t^*) = (z, t_n)$. *We assume that the time-step τ_n and the ODE-solver are chosen such that x_z^{n-1} lies within $\Omega \cup \Gamma$ for every $n \in \{1, \ldots, N_{\mathcal{I}}\}$ and every $z \in \mathcal{V}_{n,\Omega}$.* The assumptions on the convection $\mathbf{a}$ imply that this condition is satisfied for a single explicit Euler step if $\tau_n < 1/\|\mathbf{a}(\cdot, t_n)\|_{W^{1,\infty}(\Omega)}$ [18, Proposition 1]. Note that

$$x_z^{n-1} = z - \tau_n \mathbf{a}(z, t_n)$$

in this case. We stress that our results hold for *every* ODE-solver which satisfies the previous condition on the x_z^{n-1}.

As standard, all functions in V_n will be denoted with an index h. For every $n \in \{0, \dots, N_{\mathcal{I}}\}$ we denote by π_n the L^2-projection operator onto V_n and set $D^n = D(\cdot, t_n)$, $\mathbf{a}^n = \mathbf{a}(\cdot, t_n)$, $b^n = b(\cdot, t_n)$, $f^n = f(\cdot, t_n)$. With this notation the discretization of problem (2.1) is given by:

Set

$$u_h^0 = \pi_0 u_0. \tag{3.1}$$

For $n = 1, \dots, N_{\mathcal{I}}$ successively compute $\widetilde{u}_h^{n-1} \in V_n$ such that

$$\widetilde{u}_h^{n-1}(z) = \begin{cases} u_h^{n-1}(x_z^{n-1}) & \text{if } z \in \mathcal{V}_{n,\Omega}, \\ 0 & \text{if } z \in \mathcal{V}_{n,\Gamma}, \end{cases} \tag{3.2}$$

and find $u_h^n \in V_n$ such that

$$\big(\frac{u_h^n - \widetilde{u}_h^{n-1}}{\tau_n}, v_h\big) + (D^n \nabla u_h^n, \nabla v_h) + (b^n u_h^n, v_h) = (f^n, v_h) \tag{3.3}$$

holds for all $v_h \in V_n$.

The Lagrange condition and the property $x_z^{n-1} \in \Omega \cup \Gamma$ for all $n \in \{1, \dots, N_{\mathcal{I}}\}$ and all $z \in \mathcal{V}_{n,\Omega}$ imply that the $\widetilde{u}_h^{n-1}$ are well-defined. Similarly, inequality (2.5) implies that, for every n, problem (3.3) admits a unique solution u_h^n. Hence the above discretization yields a unique sequence $(u_h^n)_{0 \le n \le N_{\mathcal{I}}}$. With this we associate the function $u_{\mathcal{I}}$ which is continuous and linear with respect to time on each interval I_n and which equals u_h^n at time t_n, $0 \le n \le N_{\mathcal{I}}$. Note that

$$u_{\mathcal{I}} = u_h^{n-1} + \frac{t - t_{n-1}}{\tau_n}(u_h^n - u_h^{n-1}) \tag{3.4}$$

and

$$\partial_t u_{\mathcal{I}} = \frac{u_h^n - u_h^{n-1}}{\tau_n}$$

on every time-interval I_n.

We conclude this section with some notation which will be useful for the formulation of the a posteriori error estimates. For every $n \in \{0, \dots, N_{\mathcal{I}}\}$ we denote by $\mathcal{N}_n$ and $\mathcal{E}_n$ the collection of all vertices respective $(d-1)$-dimensional faces corresponding to the partition $\mathcal{T}_n$. A subscript Ω or Γ to any of these sets indicates that only those vertices or faces are considered which are contained in the respective set. Similarly, $\mathcal{N}_K$ and $\mathcal{E}_K$ denote the vertices and faces of a given element K and $\mathcal{N}_E$ is the collection of the vertices of a given face E.

With every face E we associate a unit vector $\mathbf{n}_E$ which is perpendicular to E and which coincides with the exterior normal of Ω if E is contained in Γ. For every piece-wise continuous function v and every face E which is not contained in Γ, we denote by $\mathbb{J}_E(v)$ the jump of v in direction $\mathbf{n}_E$. It depends on the orientation of $\mathbf{n}_E$, but quantities of the form $\mathbb{J}_E(\mathbf{n}_E D \nabla v)$ are independent thereof.

Finally, we denote by D_τ, $\mathbf{a}_\tau$, b_τ and f_τ functions which are *piece-wise constant* on the time-intervals and which, on every interval $(t_{n-1}, t_n]$,

equal D^n, $\mathbf{a}^n$, b^n and f^n, respectively. Similarly, $D_{\tau,h}$, $\mathbf{a}_{\tau,h}$, $b_{\tau,h}$ and $f_{\tau,h}$ are *piece-wise constant* functions on the time-intervals which, on every interval $(t_{n-1}, t_n]$, equal the L^2-projection of D^n, $\mathbf{a}^n$, b^n and f^n respectively onto the space of *piece-wise constant* functions corresponding to $\mathcal{T}_n$.

4. The equivalence of error and residual

With the function $u_{\mathcal{I}}$ defined from the solution of problems (3.1), (3.2) and (3.3) we associate the residual $R(u_{\mathcal{I}}) \in L^2(0,T;H^{-1}(\Omega))$ by setting for every $v \in H_0^1(\Omega)$

$$\begin{aligned}\langle R(u_{\mathcal{I}}), v\rangle = (f,v) - (\partial_t u_{\mathcal{I}}, v) - (D\nabla u_{\mathcal{I}}, \nabla v)\\ - (\mathbf{a}\cdot\nabla u_{\mathcal{I}}, v) - (b u_{\mathcal{I}}, v).\end{aligned} \tag{4.1}$$

The following lemma shows that $R(u_{\mathcal{I}})$ and the error $u - u_{\mathcal{I}}$ are equivalent. It is a slight modification of [25, Lemma 4.1], since a more general diffusion term is considered in the present work. Recall that $H_0^1(\Omega)$ and its dual space $H^{-1}(\Omega)$ are equipped with the energy-norm $|||\cdot|||$ and the dual norm $|||\cdot|||_*$ respectively.

Lemma 4.1. *For every $n \in \{1,\dots,N_{\mathcal{I}}\}$ the error can be bounded from below by*

$$\|\langle R(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} \le \sqrt{2}\max\{1,\kappa,c_b\}\|u-u_{\mathcal{I}}\|_{X(t_{n-1},t_n)}$$

and from above by

$$\begin{aligned}\|u-u_{\mathcal{I}}\|_{X(0,t_n)} \le \Big\{&2(1+\max\{\kappa,c_b\}^2)\|u_0-\pi_0 u_0\|^2\\ &+ 2(2+\max\{\kappa,c_b\}^2)\|R(u_{\mathcal{I}})\|^2_{L^2(0,t_n;H^{-1}(\Omega))}\Big\}^{\frac{1}{2}}\end{aligned}$$

where κ and c_b are introduced in section 2.

Proof. The proof of Lemma 4.1 is based on standard parabolic energy estimates and is a straightforward modification of the proof of Lemma 4.1 in [25]. We therefore only sketch the crucial steps. Equations (2.2) and (4.1) yield for every $v \in H_0^1(\Omega)$

$$\begin{aligned}\langle R(u_{\mathcal{I}}), v\rangle = (\partial_t(u-u_{\mathcal{I}}), v) + (D\nabla(u-u_{\mathcal{I}}), \nabla v)\\ + (\mathbf{a}\cdot\nabla(u-u_{\mathcal{I}}), v) + (b(u-u_{\mathcal{I}}), v).\end{aligned} \tag{4.2}$$

This identity, inequality (2.6) and the definition of the norms $|||\cdot|||$ and $|||\cdot|||_*$ yield for all $0 < t < T$ and all $v \in H_0^1(\Omega)$

$$\begin{aligned}\langle R(u_{\mathcal{I}}), v\rangle \le |||(\partial_t(u-u_{\mathcal{I}}) + \mathbf{a}\cdot\nabla(u-u_{\mathcal{I}}))(\cdot,t)|||_*|||v|||\\ + \max\{\kappa,c_b\}|||(u-u_{\mathcal{I}})(\cdot,t)|||\,|||v|||.\end{aligned}$$

Taking into account the definition of $X(t_{n-1},t_n)$ and of its norm, this estimate proves the first part of the lemma. To prove the second part,

we insert $v = (u - u_{\mathcal{I}})(\cdot, t)$ in (4.2). Taking into account inequality (2.5) this gives

$$\frac{1}{2}\frac{d}{dt}\|(u-u_{\mathcal{I}})(\cdot,t)\|^2 + |||(u-u_{\mathcal{I}})(\cdot,t)|||^2 \\ \leq |||R(u_{\mathcal{I}})(\cdot,t)|||_*|||(u-u_{\mathcal{I}})(\cdot,t)|||$$

and thus

$$\frac{d}{dt}\|(u-u_{\mathcal{I}})(\cdot,t)\|^2 + |||(u-u_{\mathcal{I}})(\cdot,t)|||^2 \leq |||R(u_{\mathcal{I}})(\cdot,t)|||_*^2.$$

Integrating this estimate with respect to time yields for every n

$$\|u-u_{\mathcal{I}}\|^2_{L^\infty(0,t_n;L^2(\Omega))} \leq \|u_0 - \pi_0 u_0\|^2 + \|R(u_{\mathcal{I}})\|^2_{L^2(0,t_n;H^{-1}(\Omega))}$$

and

$$\|u-u_{\mathcal{I}}\|^2_{L^2(0,t_n;H^1_0(\Omega))} \leq \|u_0 - \pi_0 u_0\|^2 + \|R(u_{\mathcal{I}})\|^2_{L^2(0,t_n;H^{-1}(\Omega))}.$$

Equation (4.2) and inequality (2.6), on the other hand, imply

$$|||\partial_t(u-u_{\mathcal{I}}) + \mathbf{a}\cdot\nabla(u-u_{\mathcal{I}})|||_* \leq |||R(u_{\mathcal{I}})|||_* + \max\{\kappa, c_b\}|||u-u_{\mathcal{I}}|||.$$

Taking the square of this inequality, integrating the result from 0 to t_n and combining this with the previous two estimates completes the proof of the lemma.

5. Decomposition of the residual

We may write the residual $R(u_{\mathcal{I}})$ as the sum of three contributions: a term $R_{\mathrm{osc}}(u_{\mathcal{I}})$ representing oscillations of the data and the coefficients with respect to time, a temporal residual $R_\tau(u_{\mathcal{I}})$ and a spatial residual $R_h(u_{\mathcal{I}})$. For every $v \in H^1_0(\Omega)$ and every interval $(t_{n-1}, t_n]$ these are defined by

$$\begin{aligned}
\langle R_{\mathrm{osc}}(u_{\mathcal{I}}), v\rangle &= (f - f^n, v) + ((D^n - D)\nabla u_{\mathcal{I}}, \nabla v) \\
&\quad + ((\mathbf{a}^n - \mathbf{a})\cdot\nabla u_{\mathcal{I}}, v) + ((b^n - b)u_{\mathcal{I}}, v), \\
\langle R_\tau(u_{\mathcal{I}}), v\rangle &= (D^n\nabla(u_h^n - u_{\mathcal{I}}), \nabla v) + (\mathbf{a}^n\cdot\nabla(u_h^n - u_{\mathcal{I}}), v) \\
&\quad + (b^n(u_h^n - u_{\mathcal{I}}), v), \\
\langle R_h(u_{\mathcal{I}}), v\rangle &= (f^n, v) - (\partial_t u_{\mathcal{I}}, v) - (D^n\nabla u_h^n, \nabla v) \\
&\quad - (\mathbf{a}^n\cdot\nabla u_h^n, v) - (b^n u_h^n, v).
\end{aligned} \tag{5.1}$$

Since $R_{\mathrm{osc}}(u_{\mathcal{I}})$ only involves the known discrete solution $u_{\mathcal{I}}$ and temporal oscillations of the data and coefficients of (2.1), the task of deriving upper and lower bounds for the norms $\|R(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}$ reduces to the estimation of the norms $\|R_\tau(u_{\mathcal{I}}) + R_h(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}$. The following lemma shows that this can be achieved by estimating the contributions of $R_\tau(u_{\mathcal{I}})$ and $R_h(u_{\mathcal{I}})$ separately. It is implicitly established in the proof of [25, Lemma 7.1].

Lemma 5.1. *For every $n \in \{1, \dots, N_{\mathcal{I}}\}$ we have*

$$\begin{aligned}
&\frac{1}{5}\big\{\|R_\tau(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} + \|R_h(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}\big\}^{\frac{1}{2}} \\
&\le \|R_\tau(u_{\mathcal{I}}) + R_h(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} \\
&\le \|R_\tau(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} + \|R_h(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}.
\end{aligned}$$

Proof. The second inequality is the triangle inequality. Thus, we only have to prove the first inequality. Since $R_h(u_{\mathcal{I}})$ is constant on $(t_{n-1}, t_n]$, we have

$$\|R_h(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} = \sqrt{\tau_n}|||R_h(u_{\mathcal{I}})|||_*. \tag{5.2}$$

Define $r^n \in H^{-1}(\Omega)$ by setting for every $v \in H^1_0(\Omega)$

$$\begin{aligned}
\langle r^n, v\rangle = (D^n\nabla(u_h^n - u_h^{n-1}), \nabla v) + (\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1}), v) \\
+ (b^n(u_h^n - u_h^{n-1}), v).
\end{aligned} \tag{5.3}$$

The definitions of $R_\tau(u_{\mathcal{I}})$ and $u_{\mathcal{I}}$ then yield

$$\langle R_\tau(u_{\mathcal{I}}), v\rangle = \frac{t_n - t}{\tau_n}\langle r^n, v\rangle.$$

Hence, we have

$$\|R_\tau(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} = \sqrt{\frac{\tau_n}{3}}|||r^n|||_*. \tag{5.4}$$

Next, owing to the Riesz isomorphism between the Hilbert spaces $H^1_0(\Omega)$ and $H^{-1}(\Omega)$, there exist functions v and w in $H^1_0(\Omega)$ such that

$$\begin{aligned}
|||v||| = |||R_h(u_{\mathcal{I}})|||_*, \quad \langle R_h(u_{\mathcal{I}}), v\rangle = |||R_h(u_{\mathcal{I}})|||^2_*, \\
|||w||| = |||r^n|||_*, \qquad\qquad \langle r^n, w\rangle = |||r^n|||^2_*.
\end{aligned}$$

For every $\alpha > 0$ we obtain

$$\begin{aligned}
&\|(\alpha+1)\big(\frac{t - t_{n-1}}{\tau_n}\big)^\alpha v + \frac{t_n - t}{\tau_n}w\|_{L^2(t_{n-1},t_n;H^1(\Omega))} \\
&\le \frac{\alpha+1}{\sqrt{2\alpha+1}}\sqrt{\tau_n}|||v||| + \sqrt{\frac{\tau_n}{3}}|||w||| \\
&= \frac{\alpha+1}{\sqrt{2\alpha+1}}\sqrt{\tau_n}|||R_h(u_{\mathcal{I}})|||_* + \sqrt{\frac{\tau_n}{3}}|||r^n|||_* \\
&\le \sqrt{\frac{\alpha^2+4\alpha+2}{2\alpha+1}}\big\{\|R_h(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} \\
&\qquad\qquad + \|R_\tau(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}\big\}^{\frac{1}{2}}.
\end{aligned}$$

We also have

$$\begin{aligned}
&\int_{t_{n-1}}^{t_n}\langle R_h(u_{\mathcal{I}}) + R_\tau(u_{\mathcal{I}}), (\alpha+1)\big(\frac{t - t_{n-1}}{\tau_n}\big)^\alpha v + \frac{t_n - t}{\tau_n}w\rangle dt \\
&= \tau_n\langle R_h(u_{\mathcal{I}}), v\rangle + \frac{\tau_n}{2}\langle R_h(u_{\mathcal{I}}), w\rangle
\end{aligned}$$

$$+ \langle r^n, v\rangle \int_{t_{n-1}}^{t_n} \frac{t_n - t}{\tau_n}(\alpha+1)\big(\frac{t-t_{n-1}}{\tau_n}\big)^\alpha dt + \frac{\tau_n}{3}\langle r^n, w\rangle.$$

Computing the last integral by integration by parts yields

$$\begin{aligned}&\int_{t_{n-1}}^{t_n} \langle R_h(u_{\mathcal{I}}) + R_\tau(u_{\mathcal{I}}), (\alpha+1)\big(\frac{t-t_{n-1}}{\tau_n}\big)^\alpha v + \frac{t_n - t}{\tau_n} w\rangle dt\\ &= \tau_n |||R_h(u_{\mathcal{I}})|||_*^2 + \frac{\tau_n}{2}\langle R_h(u_{\mathcal{I}}), w\rangle + \frac{\tau_n}{\alpha+2}\langle r^n, v\rangle + \frac{\tau_n}{3}|||r^n|||_*^2.\end{aligned}$$

Next we use the inequality $ab \geq -\frac{\sqrt{3}}{2}a^2 - \frac{1}{2\sqrt{3}}b^2$ twice, which gives

$$\begin{aligned}&\int_{t_{n-1}}^{t_n} \langle R_h(u_{\mathcal{I}}) + R_\tau(u_{\mathcal{I}}), (\alpha+1)\big(\frac{t-t_{n-1}}{\tau_n}\big)^\alpha v + \frac{t_n - t}{\tau_n} w\rangle dt\\ &\geq \big(1 - \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{2\alpha+4}\big)\big\{\|R_h(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}\\ &\qquad\qquad + \|R_\tau(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}\big\}.\end{aligned}$$

Combining these estimates we arrive at

$$\begin{aligned}&\|R_\tau(u_{\mathcal{I}}) + R_h(u_{\mathcal{I}})\|_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}\\ &\geq \sqrt{\frac{2\alpha+1}{\alpha^2+4\alpha+2}}\big(1 - \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{2\alpha+4}\big)\big\{\|R_h(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}\\ &\qquad\qquad + \|R_\tau(u_{\mathcal{I}})\|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))}\big\}^{\frac{1}{2}}.\end{aligned}$$

Since $\alpha > 0$ is arbitrary, we may in particular insert $\alpha = 2$ in this inequality. Since $\sqrt{\frac{5}{14}}\big(1 - \frac{3\sqrt{3}}{8}\big) \geq \frac{1}{5}$, this proves the first inequality of the lemma.

6. Estimation of the temporal residual

Equation (5.4) shows that the estimation of the temporal residual requires a reliable, efficient and robust error indicator for the functionals r^n of (5.3). The following lemma is a first step in this direction. It is a slight modification of [26, Lemma 3.1] in that we consider a more general diffusion here.

Lemma 6.1. *For every $n \in \{1, \dots, N_{\mathcal{I}}\}$ the dual norm of the functional r^n defined in equation (5.3) can be bounded from above and from below by*

$$\begin{aligned}&\frac{1}{2+\max\{\kappa, c_b\}}\big\{|||u_h^n - u_h^{n-1}||| + |||\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1})|||_*\big\}\\ &\leq |||r^n|||_*\\ &\leq \max\{\kappa, c_b\}\big\{|||u_h^n - u_h^{n-1}||| + |||\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1})|||_*\big\},\end{aligned}$$

where κ and c_b are introduced in section 2.

Proof. Since the proof of Lemma 6.1 is very similar to the one of [26, Lemma 3.1], we only sketch the crucial steps. The second inequality immediately follows from equation (5.3) and inequality (2.6). To prove the first inequality, we observe that equation (5.3) and inequality (2.5) yield

$$|||r^n|||_* \geq \frac{\langle r^n, u_h^n - u_h^{n-1}\rangle}{|||u_h^n - u_h^{n-1}|||} \geq |||u_h^n - u_h^{n-1}|||.$$

On the other hand, owing to the definition of the dual norm, using the Riesz isomorphism implies that there exists a function $v \in H_0^1(\Omega)$ with

$$|||v||| = 1 \quad \text{and} \quad (\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1}), v) = |||\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1})|||_*.$$

We insert v in (5.3) and obtain with the help of (2.6)

$$|||r^n|||_* \geq \langle r^n, v\rangle \geq |||\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1})|||_* - \max\{\kappa, c_b\}|||u_h^n - u_h^{n-1}|||.$$

The first inequality of the lemma now follows by considering all convex combinations of both estimates for $|||r^n|||_*$ and taking the supremum with respect to the convex weight.

In contrast to $|||u_h^n - u_h^{n-1}|||$, the quantity $|||\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1})|||_*$ cannot be used as an error indicator since it involves a dual norm. In order to bound this quantity by a computable one denote by

$$c_\Omega = \sup_{v \in H_0^1(\Omega)\setminus\{0\}} \frac{\|v\|}{\|\nabla v\|}$$

the optimal constant in the Poincaré–Friedrichs inequality for Ω. Then, since $\mathbf{a}^n$ is divergence-free, we have

$$|||\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1})|||_* \leq \varepsilon^{-\frac{1}{2}} \min\{\beta^{-\frac{1}{2}}, \varepsilon^{-\frac{1}{2}} c_\Omega\} \|\mathbf{a}^n\|_\infty |||u_h^n - u_h^{n-1}|||.$$

In the case of a dominant diffusion or dominant reaction, using once more (2.5) and (2.6), we therefore obtain

$$\begin{aligned}\frac{1}{2 + \max\{\kappa, c_b\}} |||u_h^n - u_h^{n-1}||| &\leq |||r^n|||_* \\ &\leq \max\{\kappa, c_b\}\{1 + c_c c_\Omega\}|||u_h^n - u_h^{n-1}|||,\end{aligned}$$

where the constants κ, c_b and c_c are introduced in section 2. If the convection is dominant, however, none of these estimates is robust. In this case we have to proceed as in [25, Lemma 8.1] (note that the present situation corresponds to the case $\theta = 1$, $\widetilde{\mathcal{T}}_{h,n} = \mathcal{T}_n$ in this reference).

For every $n \in \{1, \ldots, N_{\mathcal{I}}\}$ denote by Y_n the space of continuous piecewise linear finite element functions vanishing on Γ associated with $\mathcal{T}_n$ and by $v_h^n \in Y_n$ the unique solution of the discrete diffusion-reaction problem

$$\varepsilon(\nabla v_h^n, \nabla \varphi_h) + \beta(v_h^n, \varphi_h) = (\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1}), \varphi_h) \qquad (6.1)$$

for all $\varphi_h \in Y_n$. For every $S \in \mathcal{T}_n \cup \mathcal{E}_n$ set

$$\alpha_S = \min\{\varepsilon^{-\frac{1}{2}} \operatorname{diam}(S), \beta^{-\frac{1}{2}}\}$$

and define the error indicator η_τ^n by

$$\begin{aligned}\eta_\tau^n = \Big\{ \sum_{K\in\mathcal{T}_n} \alpha_K^2 \|\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1}) + \varepsilon\Delta v_h^n - \beta v_h^n\|_K^2 \\ + \sum_{E\in\mathcal{E}_n} \varepsilon^{-\frac{1}{2}}\alpha_E \|\mathbb{J}_E(\mathbf{n}_E \cdot \nabla v_h^n)\|_E^2 \Big\}^{\frac{1}{2}}.\end{aligned} \tag{6.2}$$

The transition condition ensures that u_h^n and u_h^{n-1}, which are defined on different meshes, are piece-wise polynomials on the common refinement which has a local mesh-size comparable to the one of $\mathcal{T}_n$. With the same arguments as in the proof of [25, Lemma 8.1] we therefore obtain:

Lemma 6.2. *There are two constants c_* and c^* which depend only on the ratios h_K/ρ_K such that the following estimates are valid*

$$c_*\big\{|||v_h^n||| + \eta_\tau^n\big\} \le |||\mathbf{a}^n \cdot \nabla(u_h^n - u_h^{n-1})|||_* \le c^*\big\{|||v_h^n||| + \eta_\tau^n\big\}.$$

7. Estimation of the spatial residual

Equation (5.2) shows that the estimation of the spatial residual requires, on each time-interval I_n separately, a reliable, efficient and robust error indicator for the restriction of $R_h(u_\mathcal{I})$ to I_n. Using integration by parts element-wise we easily conclude from the definition (5.1) of $R_h(u_\mathcal{I})$ that it admits an L^2-representation consisting of element residuals and inter-element jumps. Thus it falls into the class of functionals amenable to a posteriori error estimates (cf., e.g., [20, 27]). A comparison of equations (3.3) and (5.1), however, shows that $R_h(u_\mathcal{I})$ does not satisfy the Galerkin orthogonality, i.e. its kernel does not contain the space Y_n introduced in Section 6. Instead, for every $n \in \{1, \dots, N_\mathcal{I}\}$ and every $v_h \in V_n$, we have

$$\langle R_h(u_\mathcal{I}), v_h\rangle = \big(\frac{u_h^{n-1} - \widetilde{u}_h^{n-1}}{\tau_n} - \mathbf{a}^n \cdot \nabla u_h^n, v_h\big). \tag{7.1}$$

This term describes the difference between the method of characteristics and the backward Euler scheme and must be taken into account by the error indicator. The following lemma shows a possible way to achieve this.

The underlying idea is the same as for Lemma 6.2. Making use of the space Y_n and of the piece-wise constant functions $f_{\tau,h}$, $D_{\tau,h}$, $\mathbf{a}_{\tau,h}$ and $b_{\tau,h}$ defined at the end of section 3, we introduce the unique solution of the discrete diffusion-reaction problem

$$\varepsilon(\nabla w_h^n, \nabla\varphi_h) + \beta(w_h^n, \varphi_h) = \big(\frac{u_h^{n-1} - \widetilde{u}_h^{n-1}}{\tau_n} - \mathbf{a}^n \cdot \nabla u_h^n, \varphi_h\big) \tag{7.2}$$

for all $\varphi_h \in Y_n$ and define the error indicator η_h^n by

$$\eta_h^n = \Big\{ \sum_{K\in\mathcal{T}_n} \alpha_K^2 \|f_{\tau,h} - \frac{u_h^n - u_h^{n-1}}{\tau_n} + \operatorname{div}(D_{\tau,h}\nabla u_h^n) \\ - \mathbf{a}_{\tau,h}\cdot\nabla u_h^n - b_{\tau,h}u_h^n + \varepsilon\Delta w_h^n - \beta w_h^n\|_K^2 \\ + \sum_{E\in\mathcal{E}_n} \varepsilon^{-\frac{1}{2}}\alpha_E \|\mathbb{J}_E(\varepsilon^{-1}\mathbf{n}_E\cdot(D_{\tau,h}\nabla u_h^n) - \mathbf{n}_E\cdot\nabla w_h^n)\|_E^2 \Big\}^{\frac{1}{2}} \tag{7.3}$$

Lemma 7.1. *There are two constants $c_\dagger$ and $c^\dagger$ which depend only on the ratios h_K/ρ_K such that the following estimates are valid for every $n \in \{1,\dots,N_{\mathcal{I}}\}$*

$$|||w_h^n||| + c_\dagger\{\eta_h^n - \Theta_h^n\} \le |||R_h(u_{\mathcal{I}})|||_* \le |||w_h^n||| + c^\dagger\{\eta_h^n + \Theta_h^n\},$$

where the term Θ_h^n, defined by

$$\Theta_h^n = \Big\{ \sum_{K\in\mathcal{T}_n} \alpha_K^2 \|f^n - f_{\tau,h} + \operatorname{div}((D^n - D_{\tau,h})\nabla u_h^n) \\ - (\mathbf{a}^n - \mathbf{a}_{\tau,h})\cdot\nabla u_h^n - (b^n - b_{\tau,h})u_h^n\|_K^2 \\ + \sum_{E\in\mathcal{E}_n} \varepsilon^{-\frac{1}{2}}\alpha_E \|\mathbb{J}_E(\varepsilon^{-1}\mathbf{n}_E\cdot((D^n - D_{\tau,h})\nabla u_h^n))\|_E^2 \Big\}^{\frac{1}{2}}, \tag{7.4}$$

only depends on the oscillations of the data and coefficients.

Proof. Since Y_n is contained in V_n we conclude from (7.1) that

$$\varepsilon(\nabla w_h^n, \nabla\varphi_h) + \beta(w_h^n, \varphi_h) = \langle R_h(u_{\mathcal{I}}), \varphi_h\rangle$$

holds for all $\varphi_h \in Y_n$. This in particular implies

$$|||w_h^n||| \le |||R_h(u_{\mathcal{I}})|||_*.$$

Denote by $w^n \in H_0^1(\Omega)$ the unique solution of the weak diffusion-reaction equation

$$\varepsilon(\nabla w^n, \nabla\varphi) + \beta(w^n, \varphi) = \langle R_h(u_{\mathcal{I}}), \varphi\rangle$$

for all $\varphi \in H_0^1(\Omega)$. Then we have

$$|||w^n||| = |||R_h(u_{\mathcal{I}})|||_*$$

and therefore

$$\frac{1}{3}\{|||w_h^n||| + |||w^n - w_h^n|||\} \le |||R_h(u_{\mathcal{I}})|||_* \\ \le \{|||w_h^n||| + |||w^n - w_h^n|||\}.$$

Integration by parts element-wise yields the following L^2-representation for the residual of w_h^n

$$\begin{aligned}\langle R_h(u_{\mathcal{I}}),\varphi\rangle - \varepsilon(\nabla w_h^n,\nabla\varphi) - \beta(w_h^n,\varphi)\\ = \sum_{K\in\mathcal{T}_n}\int_K \big(f^n - \frac{u_h^n-u_h^{n-1}}{\tau_n} + \operatorname{div}(D^n\nabla u_h^n)\\ - \mathbf{a}^n\cdot\nabla u_h^n - b^n u_h^n + \varepsilon\Delta w_h^n - \beta w_h^n\big)\varphi\\ + \sum_{E\in\mathcal{E}_n}\int_E \mathbb{J}_E(\mathbf{n}_E\cdot(D^n\nabla u_h^n) - \varepsilon\mathbf{n}_E\cdot\nabla w_h^n)\varphi.\end{aligned}$$

The transition condition once more ensures that u_h^n and u_h^{n-1}, which are defined on different meshes, are piece-wise polynomials on the common refinement which has a local mesh-size comparable to the one of $\mathcal{T}_n$. Standard techniques [22] therefore yield

$$c_\dagger\big\{\eta_h^n - \Theta_h^n\big\} \le |||w^n - w_h^n||| \le c^\dagger\big\{\eta_h^n + \Theta_h^n\big\}$$

and thus prove the lemma.

8. The a posteriori error estimates

We may now collect the results of the preceding sections and formulate our main result.

Theorem 8.1. *Denote by u the solution of the variational formulation* (2.2) *of the differential equation* (2.1) *and by $u_{\mathcal{I}}$ its discrete approximation defined by problems* (3.1), (3.2), (3.3) *and equation* (3.4). *Then the error $u - u_{\mathcal{I}}$ can be bounded from above by*

$$\begin{aligned}\|u-u_{\mathcal{I}}\|_{X(0,T)} \le c_\sharp\Big\{&\|u_0-u_h^0\|^2\\ &+ \|f - f_\tau + \operatorname{div}((D-D_\tau)\nabla u_{\mathcal{I}}\\ &\qquad - (\mathbf{a}-\mathbf{a}_\tau)\cdot\nabla u_{\mathcal{I}} - (b-b_\tau)u_{\mathcal{I}}\|^2_{L^2(0,T;H^{-1}(\Omega))}\\ &+ \sum_{n=1}^{N_{\mathcal{I}}}\tau_n\big[|||u_h^n - u_h^{n-1}|||^2 + |||v_h^n|||^2 + (\eta_\tau^n)^2\big]\\ &+ \sum_{n=1}^{N_{\mathcal{I}}}\tau_n\big[|||w_h^n|||^2 + (\eta_h^n)^2 + (\Theta_h^n)^2\big]\Big\}^{\frac12}\end{aligned}$$

and, for every $n \in \{1, \dots, N_{\mathcal{I}}\}$*, from below by*

$$\begin{aligned}
&\tau_n^{\frac{1}{2}} \big[|||u_h^n - u_h^{n-1}||| + |||v_h^n||| + \eta_\tau^n + |||w_h^n||| + \eta_h^n \big] \\
&\qquad \leq c^\sharp \Big\{ \|u - u_{\mathcal{I}}\|^2_{X(t_{n-1},t_n)} \\
&\qquad\qquad + \|f - f_\tau + \operatorname{div}((D - D_\tau)\nabla u_{\mathcal{I}} \\
&\qquad\qquad\qquad - (\mathbf{a} - \mathbf{a}_\tau) \cdot \nabla u_{\mathcal{I}} - (b - b_\tau) u_{\mathcal{I}} \|^2_{L^2(t_{n-1},t_n;H^{-1}(\Omega))} \\
&\qquad\qquad + \tau_n (\Theta_h^n)^2 \Big\}^{\frac{1}{2}}.
\end{aligned}$$

The functions v_h^n *and* w_h^n *and the quantities* η_τ^n*,* η_h^n *and* Θ_h^n *are defined in equations* (6.1)*,* (7.2)*,* (6.2)*,* (7.3) *and* (7.4) *respectively. The constants* $c_\sharp$ *and* $c^\sharp$ *depend on the ratios* h_K/ρ_K *but are independent of any relative size of the diffusion with respect to the convection or reaction. If the diffusion or reaction are dominant, the quantities* $|||v_h^n|||$ *and* η_τ^n *can be omitted.*

Remark 8.2. According to Section 6, the terms $\tau_n^{\frac{1}{2}} \big[|||u_h^n - u_h^{n-1}||| + |||v_h^n||| + \eta_\tau^n \big]$ control the restriction to $(t_{n-1}, t_n]$ of the temporal residual $R_\tau(u_{\mathcal{I}})$. Hence, they can be used for adapting the time-step τ_n. Evidently, they split into element-contributions:

$$\begin{aligned}
&\tau_n^{\frac{1}{2}} \big[|||u_h^n - u_h^{n-1}||| + |||v_h^n||| + \eta_\tau^n \big] \\
&\qquad = \Big\{ \tau_n \sum_{K \in \mathcal{T}_n} \big[|||u_h^n - u_h^{n-1}|||_K^2 + |||v_h^n|||_K^2 + \big(\eta_{\tau,K}^n\big)^2 \big] \Big\}^{\frac{1}{2}}
\end{aligned}$$

with

$$\begin{aligned}
\eta_{\tau,K}^n = \Big\{ &\alpha_K^2 \|\mathbf{a}^n \cdot \nabla (u_h^n - u_h^{n-1}) + \varepsilon \Delta v_h^n - \beta v_h^n\|_K^2 \\
&+ \frac{1}{2} \sum_{E \in \mathcal{E}_{K,\Omega}} \varepsilon^{-\frac{1}{2}} \alpha_E \|\mathbb{J}_E(\mathbf{n}_E \cdot \nabla v_h^n)\|_E^2 \Big\}^{\frac{1}{2}}.
\end{aligned}$$

Thus, the terms $\tau_n^{\frac{1}{2}} \big[|||u_h^n - u_h^{n-1}|||_K + |||v_h^n|||_K + \eta_{\tau,K}^n \big]$ are natural candidates for an eventual local (in space) time-step control. We once more stress that the terms $|||v_h^n|||$ and η_τ^n and their local analogues can be omitted in the case of dominant reaction or diffusion. If, however, the convection is dominant these terms are mandatory for conserving the full robustness of the error estimates.

Remark 8.3. According to Section 7, the terms $\tau_n^{\frac{1}{2}} \big[|||w_h^n||| + \eta_h^n \big]$ control the restriction to $(t_{n-1}, t_n]$ of the spatial residual $R_h(u_{\mathcal{I}})$ up to the data-oscillation $\tau_n^{\frac{1}{2}} \Theta_h^n$. Evidently, these terms split into element-contributions:

$$\tau_n^{\frac{1}{2}} \big[|||w_h^n||| + \eta_h^n \big] = \Big\{ \tau_n \sum_{K \in \mathcal{T}_n} \big[|||w_h^n|||_K^2 + \big(\eta_{h,K}^n\big)^2 \big] \Big\}^{\frac{1}{2}},$$

with

$$\eta_{h,K}^n = \Big\{ \alpha_K^2 \| f_{\tau,h} - \frac{u_h^n - u_h^{n-1}}{\tau_n} + \operatorname{div}(D_{\tau,h} \nabla u_h^n) \\ - \mathbf{a}_{\tau,h} \cdot \nabla u_h^n - b_{\tau,h} u_h^n + \varepsilon \Delta w_h^n - \beta w_h^n \|_K^2 \\ + \frac{1}{2} \sum_{E \in \mathcal{E}_{K,\Omega}} \varepsilon^{-\frac{1}{2}} \alpha_E \| \mathbb{J}_E (\varepsilon^{-1} \mathbf{n}_E \cdot (D_{\tau,h} \nabla u_h^n) - \mathbf{n}_E \cdot \nabla w_h^n) \|_E^2 \Big\}^{\frac{1}{2}}.$$

Taking into account that the spatial residual is piece-wise constant on the time intervals, combining the arguments used in the proof of Lemma 7.1 with the techniques of [22] for establishing lower error bounds and defining $\Theta_{h,K}^n$ in an obvious way as the local analogue of Θ_h^n, we can easily prove that

$$|||w_h^n|||_K + \eta_{h,K}^n \leq c \Big\{ \sup_{v \in H_0^1(\widetilde{\omega}_K) \setminus \{0\}} \frac{\langle R_h(u_{\mathcal{I}}), v \rangle}{|||v|||_K} + \Theta_{h,K}^n \Big\}$$

holds for every interval $(t_{n-1}, t_n]$ and every element $K \in \mathcal{T}_n$. Here, $\widetilde{\omega}_K$ denotes the union of all elements in $\mathcal{T}_n$ that share at least a vertex with K. The constant c only depends on the shape parameter of $\mathcal{T}_n$ but is independent of any relative size of the diffusion with respect to the convection or reaction. This shows that the terms $|||w_h^n|||_K + \eta_{h,K}^n$ can efficiently be used for adapting the spatial meshes.

Remark 8.4. We finally want to mention that the approach of [4, 8] could also be adapted to the present problem. Instead of splitting the residual as in Lemma 5.1, this would mean to split the error in the form $\|u - u_\tau\|_{X(0,T)} + \|u_\tau - u_{\mathcal{I}}\|_{X(0,T)}$ by introducing a semi-discrete approximation u_τ as follows:

> Set $u^0 = \pi_0 u_0$. For $n = 1, \ldots, N_{\mathcal{I}}$ define $\widetilde{u}^{n-1}$ by $\widetilde{u}^{n-1}(y) = u^{n-1}(x(t_{n-1}; y, t_n))$ for all $y \in \Omega$ and find u^n in $H_0^1(\Omega)$ such that
>
> $$\big(\frac{u^n - \widetilde{u}^{n-1}}{\tau_n}, v \big) + (D^n \nabla u^n, \nabla v) + (b^n u^n, v) = (f^n, v)$$
>
> holds for all v in $H_0^1(\Omega)$. Then, u_τ is the continuous, piece-wise linear (with respect to time) function which coincides with u^n at time t_n, $0 \leq n \leq N_{\mathcal{I}}$.

Standard techniques can then be used to obtain upper and lower bounds for the terms $\|u - u_\tau\|_{X(0,T)}$ and $\|u_\tau - u_{\mathcal{I}}\|_{X(0,T)}$ separately. At first sight this approach is appealing since it yields an obvious uncoupling of spatial and temporal error contributions. But it has the severe drawback that it leads to error estimates which are not fully robust with respect to the relative sizes of diffusion, convection and reaction. Essentially, this is due to the fact that $\|u - u_{\mathcal{I}}\|_{X(0,T)}$ can be bounded from above by $\|u - u_\tau\|_{X(0,T)} + \|u_\tau - u_{\mathcal{I}}\|_{X(0,T)}$ but not from below. The present

approach does not suffer from this drawback since Lemma 5.1 provides two-sided bounds.

As a conclusion, when replacing D^n, $\mathbf{a}^n$, b^n and f^n by their finite element approximations $D_{\tau,h}$, $\mathbf{a}_{\tau,h}$, $b_{\tau,h}$ and $f_{\tau,h}$, the quantities v_h^n, w_h^n, η_τ^n and η_h^n (or their local analogues $\eta_{\tau,K}^n$ and $\eta_{h,K}^n$) are very easy to compute. Hence, they are an efficient tool for mesh and time adaptivity, for all types of diffusion-convection-reaction equations.

9. Some numerical experiments

The numerical simulation that we now present has been performed with the code FreeFem++ due to F. Hecht and O. Pironneau, see [14]. We only consider a model domain, the square $\Omega = (-1,1)^2$. The space V_n is the space of continuous piece-wise linear finite element functions vanishing on Γ associated with $\mathcal{T}_n$. Thus, it coincides with the space Y_n introduced in section 6.

The diffusion coefficient D is equal to ε times the identity, and we intend to work with different values of ε. The coefficient b is taken equal to zero and the convection coefficient $\mathbf{a}$ is constant in time, given by

$$\mathbf{a}(x) = \mathrm{curl}(1-x_1^2)^2(1-x_2^2)^2 = 4(1-x_1^2)(1-x_2^2)\begin{pmatrix} -(1-x_1^2)x_2 \\ x_1(1-x_2^2)\end{pmatrix}.$$

The data f and u_0 are chosen such that the exact solution is

$$u(x,t) = v(\frac{1}{\varepsilon}|x-z(t)|^2),$$

with

$$z(t) = \begin{pmatrix} \frac{1}{2}\cos t \\ \frac{1}{2}\sin t\end{pmatrix}, \qquad v(s) = \begin{cases} (1+2s)(1-s)^2 & \text{if } 0 \le s \le 1 \\ 0 & \text{if } s > 1. \end{cases}$$

The final time T is equal to 4. For simplicity, we do not perform time adaptivity and take all the τ_n equal to 10^{-1}. Mesh adaptivity is performed according to the following simple strategy where $\ell \in \{0,1,2,3\}$:

- Given the triangulation $\mathcal{T}_{10\ell}$, for $n \in \{10\ell, \dots, 10\ell+9\}$, we compute the solution u_h^n, the indicators $\eta_{\tau,K}^n$ and $\eta_{h,K}^n$, the solutions v_h^n and w_h^n together with their local norms $|||v_h^n|||_K$ and $|||w_h^n|||_K$ and set

  $$H_K^n = |||u_h^n - u_h^{n-1}|||_K + \eta_{\tau,K}^n + |||v_h^n|||_K + \eta_{h,K}^n + |||w_h^n|||_K.$$

- Next, we define $\overline{H}_K^\ell$ as the maximum of the H_K^n, for $n \in \{10\ell, \dots, 10\ell+9\}$.
- Then we construct a new triangulation $\mathcal{T}_{10(\ell+1)}$ such that the diameter of any element in $\mathcal{T}_{10(\ell+1)}$ which contains or is contained in an element K of $\mathcal{T}_{10\ell}$ is proportional to the diameter of K times the ratio $H^*/\overline{H}_K^\ell$, where H^* is a fixed threshold.

- Finally, we perform a new computation of the solution u_h^n, $n \in \{10\ell, \ldots, 10\ell + 9\}$, on the new mesh.

We refer to [15] for more details on this algorithm and to [13, Chap. 21] for the way of constructing such a mesh. Note also that this adaptation step can be iterated several times when necessary.

Figures 9.1 – 9.3 show the adapted mesh and the corresponding solution for several values of ε. In all cases the singularity of the solution is well detected and resolved.

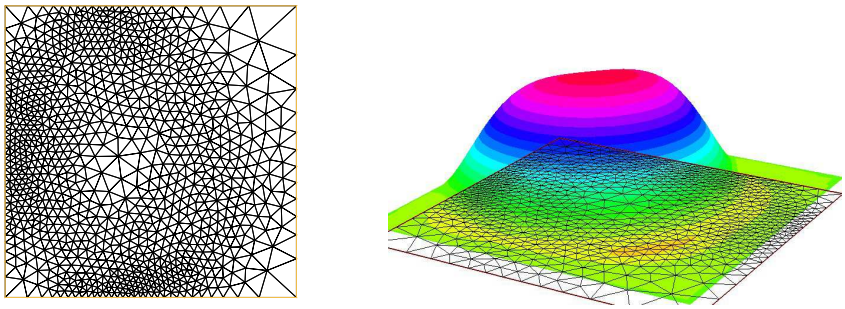


FIGURE 9.1. The adapted mesh and the solution at time T for $\varepsilon = 1$

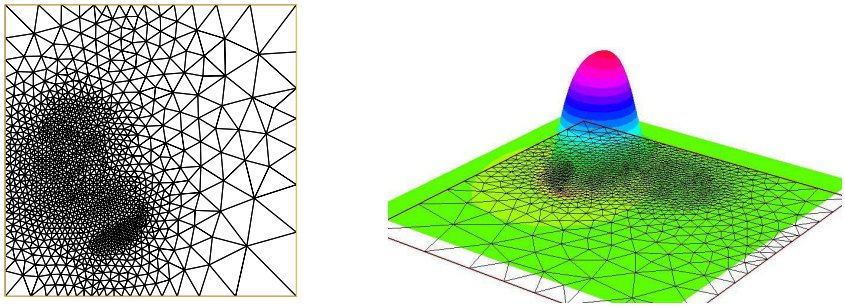


FIGURE 9.2. The adapted mesh and the solution at time T for $\varepsilon = 10^{-1}$

Acknowledgements: We are very grateful to Brigitte Métivet for a first discussion on the subject of this paper and to Martin Vohralík for his kind invitation of the third author to visit the Laboratoire Jacques-Louis Lions.

REFERENCES

[1] R. A. Adams, *Sobolev Spaces*, Academic Press, 1975.

[2] H. Amann, *Ordinary differential equations*, de Gruyter Studies in Mathematics, vol. 13, Walter de Gruyter & Co., Berlin, 1990, An introduction to nonlinear analysis, Translated from the German by Gerhard Metzen.

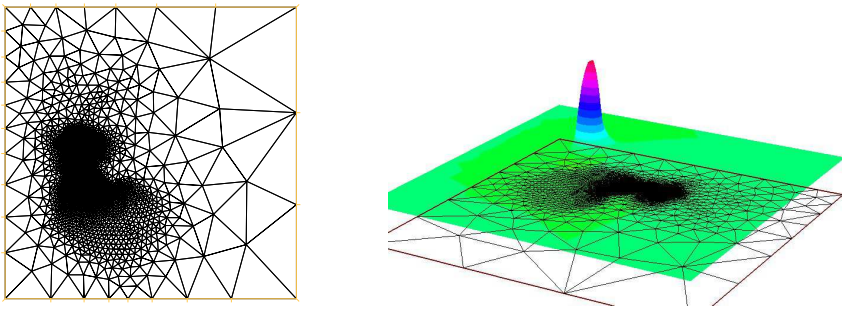


FIGURE 9.3. The adapted mesh and the solution at time T for $\varepsilon = 10^{-2}$

[3] ———, *Linear and Quasilinear Parabolic Problems. Volume I: Abstract Linear Theory*, Birkhäuser, Basel, 1995.

[4] A. Bergam, C. Bernardi, and Z. Mghazli, *A posteriori analysis of the finite element discretization of some parabolic equations*, Math. Comp. **74** (2005), no. 251, 1117–1138.

[5] A. Bermùdez, M.R. Nogueiras, and C. Vàzquez, *Numerical analysis of convection-diffusion-reaction problems with higher order characteristics/finite elements. Part I. Time discretization*, SIAM J. Numer. Anal. **44** (2006), 1829–1853.

[6] ———, *Numerical analysis of convection-diffusion-reaction problems with higher order characteristics/finite elements. Part II. Fully discretized schemes and quadrature formulas*, SIAM J. Numer. Anal. **44** (2006), 1854–1876.

[7] C. Bernardi and R. Verfürth, *Adaptive finite element methods for elliptic equations with non-smooth coefficients*, Numer. Math. **85** (2000), no. 4, 579–608.

[8] ———, *A posteriori error analysis of the fully discretized time-dependent Stokes equations*, M2AN Math. Model. Numer. Anal. **38** (2004), no. 3, 437 – 455.

[9] K. Boukir, Y. Maday, and B. Métivet, *A high order characteristics method for the incompressible Navier–Stokes equations*, Comput. Methods Appl. Mech. Engrg. **116** (1994), no. 1-4, 211–218.

[10] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology V: Evolution Problems I*, Springer, Berlin - Heidelberg - New York, 1992.

[11] K. Eriksson and C. Johnson, *Adaptive finite element methods for parabolic problems. I. A linear model problem*, SIAM J. Numer. Anal. **28** (1991), 43–77.

[12] ———, *Adaptive finite element methods for parabolic problems. IV. Nonlinear problems*, SIAM J. Numer. Anal. **32** (1995), 1729–1749.

[13] P.J. Frey and P.-L. George, *Maillages, applications aux éléments finis*, Hermès, 1999.

[14] F. Hecht, A. Le Hyaric, K. Ohtsuka, and O. Pironneau, *Freefem++, second edition, v 3.0-1*, Université Pierre et Marie Curie, Paris, 2007, on the web at `http://www.freefem.org/ff++/ftp/freefem++doc.pdf`.

[15] F. Hecht and B. Mohammadi, *Mesh adaptation for time dependent simulation and optimization*, Revue Européenne des Éléments Finis **10** (2001), 575–593.

[16] O. Pironneau, *On the transport-diffusion algorithm and its applications to the Navier–Stokes equations*, Numer. Math. **38** (1981/82), no. 3, 309–332.

[17] ———, *Finite element characteristics methods requiring no quadrature*, J. Scientific Computing (2009).

[18] H. Rui and M. Tabata, *A second order characteristic finite element scheme for convection-diffusion problems*, Numer. Math. **92** (2002), 161–177.

[19] Y. Tourigny and E. Süli, *The finite element method with nodes moving along the characteristics for convection-diffusion equations*, Numer. Math. **59** (1991), no. 4, 399–412.

[20] A. Veeser and R. Verfürth, *Explicit upper bounds for dual norms of residuals*, report, Ruhr-Universität, Bochum, October 2008.

[21] R. Verfürth, *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. Methods Partial Differential Equations **14** (1998), no. 4, 487 – 518.

[22] ———, *Robust a posteriori error estimators for a singularly perturbed reaction-diffusion equation*, Numer. Math. **78** (1998), no. 3, 479 – 493.

[23] ———, *A posteriori error estimates for finite element discretizations of the heat equation*, Calcolo **40** (2003), no. 3, 195 – 212.

[24] ———, *A posteriori error estimates for linear parabolic equations*, report, Ruhr-Universität, Bochum, July 2004.

[25] ———, *Robust a posteriori error estimates for nonstationary convection-diffusion equations*, SIAM J. Numer. Anal. **43** (2005), no. 4, 1783 – 1802.

[26] ———, *Robust a posteriori error estimates for stationary convection-diffusion equations*, SIAM J. Numer. Anal. **43** (2005), no. 4, 1766 – 1782.

[27] ———, *A note on constant-free a posteriori error estimates*, report, Ruhr-Universität, Bochum, June 2008.