SIAM J. NUMER. ANAL.
Vol. 39, No. 5, pp. 1648–1666

LEAST-SQUARES GALERKIN METHODS FOR PARABOLIC PROBLEMS II: THE FULLY DISCRETE CASE AND ADAPTIVE ALGORITHMS*

MOHAMMAD MAJIDI† AND GERHARD STARKE†

Abstract. In this second part of our series on least-squares Galerkin methods for parabolic initial-boundary value problems we study the full discretization in time and space. These methods are based on the minimization of a least-squares functional for an equivalent first-order system over space and time with respect to suitable discrete spaces. Based on the analysis of the semi-discretization in time carried out in the first part, we construct a posteriori error estimators for the approximation error components associated with time and space. For the space discretization error we use a hierarchical basis estimator with respect to the functional minimization problem. We prove a strengthened Cauchy–Schwarz inequality between coarse and hierarchical surplus space and a bound on the condition number of the auxiliary problem, both uniform in the mesh-size h and the time-step τ implying efficiency and reliability of the error estimator. This allows for an adaptive strategy for both the time-step size and the spatial triangulation, keeping a proper balance between these two components. Numerical experiments illustrate the performance of the resulting adaptive schemes in time and space.

Key words. least-squares Galerkin method, first-order system, a posteriori error estimate, adaptive time-step control, adaptive mesh refinement, parabolic problems

AMS subject classifications. 65M60, 65M15

PII. S0036142900379461

1. Introduction.

The efficient computation of accurate approximations to initial-boundary value problems for parabolic partial differential equations requires adaptive techniques. Both the step-size for the discretization in time as well as the choice of mesh for the spatial discretization need to be adapted to local features of the solution using appropriate error estimators. One of the strengths of the least-squares Galerkin approach presented in this series of papers is that it automatically provides such an a posteriori error estimator by evaluating the associated least-squares functional. This is achieved by rewriting the second-order parabolic problem as a first-order system for the primary scalar variable and the newly introduced flux. Both variables are approximated simultaneously in suitable spaces in each time-step. Based on the analysis of the semidiscrete case from the first part [12], we derive a posteriori error estimators for the approximation error components with respect to time and space. Since the least-squares functional comprises the error in space and time we need to employ, in addition, a different error estimator for the error associated with the space discretization. In our framework of functional minimization it is natural to use a hierarchical basis error estimator. We will prove that such an approach based on minimizing the functional with respect to the hierarchical surplus space leads to an error estimator, uniformly in the time-step size, assuming the usual saturation condition.

The specific method analyzed here and in the companion paper [12] is based on piecewise linear approximation spaces in time. Although the analysis and numerical

*Received by the editors October 12, 2000; accepted for publication (in revised form) August 3, 2001; published electronically January 11, 2002. This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG) under grant STA 402/7-1.
http://www.siam.org/journals/sinum/39-5/37946.html
†Institut für Angewandte Mathematik, Universität Hannover, Welfengarten 1, 30167 Hannover, Germany (mmajidi@ifam.uni-hannover.de, starke@ifam.uni-hannover.de).

tests are carried out for linear problems, it is our opinion that the real strength of this new methodology will show up in the treatment of nonlinear problems. The use of the least-squares approach for nonlinear parabolic initial-boundary value problems modeling variably saturated subsurface flow is the topic of the ongoing research work [11]. For the nonlinear elliptic systems arising after time discretization of such problems, the least-squares methodology has been studied in [13].

It is beyond the scope of this paper to give a full comparison of this methodology, in terms of computational work required to achieve a certain accuracy, with other approaches to time-space adaptivity for parabolic problems. However, some comments in this direction are in order. The so-called discontinuous Galerkin methods were intensively studied in the context of time-space adaptivity for parabolic problems (see, e.g., [14, Chap. 12] and [7]). The same is true for implicit Runge–Kutta methods, where adaptive time-stepping is based on some embedded formula (see, e.g., [9, Chap. IV]). Both of these approaches achieve comparable accuracy for the scalar unknown in the parabolic problem at less computational cost. Our least-squares approach computes, in addition to the primary variable, approximations for the associated flux, which is the quantity of main interest in many applications.

There is a conceptual difference between these methods regarding the a posteriori estimate of the time discretization error. Our least-squares approach estimates the error with respect to an integral norm over time (for scalar and flux variables) while the methods above are tailored to measure the maximal error over time of the primary variable. A discussion of what these methods achieve in terms of accuracy and error estimation per computational cost will be given at the end of section 5.

In section 2, we briefly review the least-squares formulation in time from [12]. Section 3 treats the fully discrete variational formulation, including quasi-optimality and a decomposition of the error components associated with time and space. The hierarchical estimator for the space discretization error is presented and analyzed in section 4. Section 5 contains the time-space adaptive algorithm and its relation to other adaptive time-stepping schemes for parabolic problems. Finally, section 6 presents the results of our adaptive method for some test examples.

2. Least-squares Galerkin formulation in time. As in the first part of this paper, we consider the first-order system formulation of the linear parabolic equation,

$$\begin{aligned} c\,\partial_t p + \operatorname{div} u &= f\,, \\ u + a\,\nabla p &= 0\,, \end{aligned} \tag{2.1}$$

where $f \in C([0,T], L^2(\Omega))$ in a bounded polygonal domain $\Omega \in \mathbb{R}^2$ for $t \in (0,T)$ with some $T > 0$. For $t = 0$ initial conditions, $p(0) \in H^1_{\Gamma_D}(\Omega)$ are prescribed. The boundary of Ω is divided into $\partial\Omega = \Gamma_D \cup \Gamma_N$, where homogeneous boundary conditions are prescribed for the normal component of the flux $\langle n, u\rangle$ on Γ_N and for p on Γ_D. We also assume that a and c are independent of t and that $a, c \in L^\infty(\Omega)$ and $a \geq \underline{a}, c \geq \underline{c}$ uniformly for $x \in \Omega$ with positive constants $\underline{a}, \underline{c}$. For the time discretization, we define a one-step method by minimizing the least-squares functional

$$\begin{aligned} \mathcal{F}(u,p;p(t),f) = \int_0^\tau \Big(&\tau \| c^{1/2}\,\partial_t p(t+\sigma) + c^{-1/2}\,(\operatorname{div} u(t+\sigma) - f)\|^2_{0,\Omega} \\ &+ \|a^{-1/2}\,u(t+\sigma) + a^{1/2}\,\nabla p(t+\sigma)\|^2_{0,\Omega} \Big)\, d\sigma \end{aligned} \tag{2.2}$$

subject to the initial condition $p(t)$ with respect to suitable spaces

$$V_\tau((0,T), H_{\Gamma_N}(\mathrm{div},\Omega)) \subset L^2((0,T), H_{\Gamma_N}(\mathrm{div},\Omega)),$$
$$Q_\tau((0,T), H^1_{\Gamma_D}(\Omega)) \subset H^1((0,T), H^1_{\Gamma_D}(\Omega))$$

for the approximation of u and p, respectively. For our analysis, we use the particular choice from [12], i.e., piecewise linear, not necessarily continuous, functions for V_τ and piecewise linear continuous functions for Q_τ on a partition $\{0 = t_0 < t_1 < \cdots < t_M = T\}$ of $[0,T]$. The analysis in [12] for the semidiscrete case was carried out under the assumption that f is independent of t. For this pair of spaces, the minimization is done with respect to

$$u_\tau^- := u_\tau(t+),\ u_\tau^+ := u_\tau(t+\tau-) \in H_{\Gamma_N}(\mathrm{div},\Omega),\ \text{and}\ p_\tau^+ := p_\tau(t+\tau) \in H^1_{\Gamma_D}(\Omega).$$

The least-squares functional becomes

$$\begin{aligned}
&\mathcal{F}(u_\tau, p_\tau; p_\tau(t), f)\\
&\quad = \int_0^\tau \tau \left\| c^{1/2}\frac{p_\tau^+ - p_\tau(t)}{\tau} + c^{-1/2}\left(\frac{\tau-\sigma}{\tau}\,\mathrm{div}\,u_\tau^- + \frac{\sigma}{\tau}\,\mathrm{div}\,u_\tau^+ - f\right)\right\|^2_{0,\Omega} d\sigma \\
&\quad + \int_0^\tau \left\| \frac{\tau-\sigma}{\tau}(a^{-1/2}u_\tau^- + a^{1/2}\nabla p_\tau(t)) + \frac{\sigma}{\tau}(a^{-1/2}u_\tau^+ + \nabla p_\tau^+)\right\|^2_{0,\Omega} d\sigma .
\end{aligned} \tag{2.3}$$

We abuse notation and write $\mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f)$ instead of $\mathcal{F}(u_\tau, p_\tau; p_\tau(t), f)$ in order to stress the degrees of freedom in the variational problem. Assuming that f is a constant function with respect to time in $(t, t+\tau)$, and using the fact that Simpson's rule is exact for polynomials of degree 2, we obtain

$$\begin{aligned}
\mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f) &= \frac{1}{6}\|c^{1/2}(p_\tau^+ - p_\tau(t)) + \tau\, c^{-1/2}(\mathrm{div}\,u_\tau^- - f)\|^2_{0,\Omega}\\
&+ \frac{2}{3}\|c^{1/2}(p_\tau^+ - p_\tau(t)) + \frac{\tau}{2}\, c^{-1/2}(\mathrm{div}\,u_\tau^- + \mathrm{div}\,u_\tau^+ - 2f)\|^2_{0,\Omega}\\
&+ \frac{1}{6}\|c^{1/2}(p_\tau^+ - p_\tau(t)) + \tau\, c^{-1/2}(\mathrm{div}\,u_\tau^+ - f)\|^2_{0,\Omega}\\
&+ \frac{\tau}{6}\|a^{-1/2}u_\tau^- + a^{1/2}\nabla p_\tau(t)\|^2_{0,\Omega} + \frac{\tau}{6}\|a^{-1/2}u_\tau^+ + a^{1/2}\nabla p_\tau^+\|^2_{0,\Omega}\\
&+ \frac{\tau}{6}\|a^{-1/2}(u_\tau^- + u_\tau^+) + a^{1/2}(\nabla p_\tau(t) + \nabla p_\tau^+)\|^2_{0,\Omega}.
\end{aligned}$$

If we define the bilinear form

$$\begin{aligned}
&\mathcal{B}(u_\tau^-, u_\tau^+, p_\tau^+; v_\tau^-, v_\tau^+, q_\tau)\\
&\quad = \frac{1}{3}(c^{1/2}p_\tau^+ + \tau\, c^{-1/2}\mathrm{div}\,u_\tau^-, c^{1/2}q_\tau + \tau\, c^{-1/2}\mathrm{div}\,v_\tau^-)_{0,\Omega}\\
&\quad + \frac{1}{3}(c^{1/2}p_\tau^+ + \tau\, c^{-1/2}\mathrm{div}\,u_\tau^+, c^{1/2}q_\tau + \tau\, c^{-1/2}\mathrm{div}\,v_\tau^+)_{0,\Omega}\\
&\quad + \frac{1}{6}(c^{1/2}p_\tau^+ + \tau\, c^{-1/2}\mathrm{div}\,u_\tau^-, c^{1/2}q_\tau + \tau\, c^{-1/2}\mathrm{div}\,v_\tau^+)_{0,\Omega}\\
&\quad + \frac{1}{6}(c^{1/2}p_\tau^+ + \tau\, c^{-1/2}\mathrm{div}\,u_\tau^+, c^{1/2}q_\tau + \tau\, c^{-1/2}\mathrm{div}\,v_\tau^-)_{0,\Omega}\\
&\quad + \frac{\tau}{6}(a^{-1/2}u_\tau^-, a^{-1/2}v_\tau^-)_{0,\Omega} + \frac{\tau}{6}(a^{-1/2}u_\tau^+ + a^{1/2}\nabla p_\tau^+, a^{-1/2}v_\tau^+ + a^{1/2}\nabla q_\tau)_{0,\Omega}\\
&\quad + \frac{\tau}{6}(a^{-1/2}(u_\tau^- + u_\tau^+) + a^{1/2}\nabla p_\tau^+, a^{-1/2}(v_\tau^- + v_\tau^+) + a^{1/2}\nabla q_\tau)_{0,\Omega}
\end{aligned} \tag{2.4}$$

corresponding to $\mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f)$, then $(u_\tau^-, u_\tau^+, p_\tau^+)$ satisfies the variational problem

$$
\begin{aligned}
&\mathcal{B}(u_\tau^-, u_\tau^+, p_\tau^+; v^-, v^+, q) \\
&\quad = (c^{1/2}\, p_\tau(t) + \tau\, c^{-1/2}\, f, c^{1/2}\, q)_{0,\Omega} \\
&\quad + \frac{\tau}{2}(c^{1/2}\, p_\tau(t) + \tau\, c^{-1/2}\, f, c^{-1/2}\, \operatorname{div} v^-)_{0,\Omega} \\
&\quad + \frac{\tau}{2}(c^{1/2}\, p_\tau(t) + \tau\, c^{-1/2}\, f, c^{-1/2}\, \operatorname{div} v^+)_{0,\Omega} \\
&\quad - \frac{\tau}{6}\,(a^{1/2}\nabla p_\tau(t), a^{-1/2}(2v^- + v^+))_{0,\Omega} - \frac{\tau}{6}\,(a^{1/2}\nabla p_\tau(t), a^{1/2}\nabla q)_{0,\Omega}
\end{aligned}
\tag{2.5}
$$

for all $v^-, v^+ \in H_{\Gamma_N}(\operatorname{div}, \Omega)$ and $q \in H^1_{\Gamma_D}(\Omega)$.

Remark. This justifies our assumption $p(0) \in H^1_{\Gamma_D}(\Omega)$ in the statement of the initial-boundary value problem at the beginning of this section. Note that the discontinuous Galerkin methods require only $p(0) \in L^2(\Omega)$ since the variational problem does not depend on $\nabla p(t)$ at the previous time-step. Other methods—for example, the Crank–Nicolson method—evaluate a second-order bilinear form at the old time-step similar to our method and therefore require $p(0) \in H^1_{\Gamma_D}(\Omega)$ (see [14, Chap. 12]).

3. Least-squares Galerkin formulation in time and space. For the actual computation, we perform the minimization with respect to finite-dimensional subspaces $V_h \subseteq H_{\Gamma_N}(\operatorname{div}, \Omega)$ and $Q_h \subseteq H^1_{\Gamma_D}(\Omega)$, which are based on a triangulation $\mathcal{T}_h$. Since our ultimate interest is in adaptively refined triangulations, we do not assume quasi-uniformity but only shape regularity of our sequence of triangulations. The computation on the previous time-step will, in general, also be done approximately, which means that we need to replace $p_\tau(t)$ by $p_{\tau,h}(t)$ in the least-squares formulation. Denote by $(\bar{u}_\tau^-, \bar{u}_\tau^+, \bar{p}_\tau^+)$ the minimum of

$$\mathcal{F}(v_\tau^-, v_\tau^+, q_\tau^+; p_{\tau,h}(t), f)$$

with respect to $(v_\tau^-, v_\tau^+, q_\tau^+) \in H_{\Gamma_N}(\operatorname{div}, \Omega)^2 \times H^1_{\Gamma_D}(\Omega)$, where $p_{\tau,h}(t)$ is the fully discrete approximation at the previous time-step. Therefore, $(\bar{u}_\tau^-, \bar{u}_\tau^+, \bar{p}_\tau^+)$ is the solution of the variational problem (2.5) with $p_\tau(t)$ replaced by $p_{\tau,h}(t)$. The fully discrete version of the least-squares method is then given by the minimum of

$$\mathcal{F}(v_{\tau,h}^-, v_{\tau,h}^+, q_{\tau,h}^+; p_{\tau,h}(t), f)$$

with respect to $(v_{\tau,h}^-, v_{\tau,h}^+, q_{\tau,h}^+) \in V_h^2 \times Q_h$. The minimization is equivalent to the variational problem of finding $(u_{\tau,h}^-, u_{\tau,h}^+, p_{\tau,h}^+) \in V_h^2 \times Q_h$ such that

$$
\begin{aligned}
&\mathcal{B}(u_{\tau,h}^-, u_{\tau,h}^+, p_{\tau,h}^+; v_h^-, v_h^+, q_h) = (c^{1/2}\, p_{\tau,h}(t) + \tau\, c^{-1/2}\, f, c^{1/2}\, q_h)_{0,\Omega} \\
&\quad + \frac{\tau}{2}(c^{1/2}\, p_{\tau,h}(t) + \tau\, c^{-1/2}\, f, c^{-1/2}\, \operatorname{div} v_h^-)_{0,\Omega} \\
&\quad + \frac{\tau}{2}(c^{1/2}\, p_{\tau,h}(t) + \tau\, c^{-1/2}\, f, c^{-1/2}\, \operatorname{div} v_h^+)_{0,\Omega} \\
&\quad - \frac{\tau}{6}\,(a^{1/2}\nabla p_{\tau,h}(t), a^{-1/2}(2v_h^- + v_h^+))_{0,\Omega} - \frac{\tau}{6}\,(a^{1/2}\nabla p_{\tau,h}(t), a^{1/2}\nabla q_h)_{0,\Omega}
\end{aligned}
\tag{3.1}
$$

holds for all $v_h^-, v_h^+ \in V_h$, and $q_h \in Q_h$. In order to establish the well-posedness of this variational problem, we proved coercivity and continuity of the above bilinear

form with respect to τ-weighted norms in [12, Thm. 6.1]. We repeat this result in a slightly stronger form.

THEOREM 3.1. *With respect to the weighted norm*

$$|||(v^-,v^+,q)||| := \Big(\frac{\tau}{3}\|a^{-1/2}v^-\|_{0,\Omega}^2+\frac{\tau^2}{3}\|c^{-1/2}\mathrm{div}\,v^-\|_{0,\Omega}^2+\frac{\tau}{3}\|a^{-1/2}v^+\|_{0,\Omega}^2 + \frac{\tau^2}{3}\|c^{-1/2}\mathrm{div}\,v^+\|_{0,\Omega}^2+\|c^{1/2}q\|_{0,\Omega}^2+\frac{\tau}{3}\|a^{1/2}\nabla q\|_{0,\Omega}^2\Big)^{1/2}, \tag{3.2}$$

the bilinear form (2.4) *satisfies*

$$\begin{aligned}\mathcal{B}(v^-,v^+,q;v^-,v^+,q) &\geq \frac{1}{4}|||(v^-,v^+,q)|||^2,\\ \mathcal{B}(u^-,u^+,p;v^-,v^+,q) &\leq 2|||(u^-,u^+,p)|||\cdot|||(v^-,v^+,q)|||\end{aligned} \tag{3.3}$$

for all $(u^-,u^+,p),(v^-,v^+,q)\in H_{\Gamma_N}(\mathrm{div},\Omega)^2\times H^1_{\Gamma_D}(\Omega)$.

Proof. From integration by parts we obtain

$$(v^-,\nabla q)_{0,\Omega}+(\mathrm{div}\,v^-,q)_{0,\Omega}=(v^+,\nabla q)_{0,\Omega}+(\mathrm{div}\,v^+,q)_{0,\Omega}=0$$

for all $v^-,v^+\in H_{\Gamma_N}(\mathrm{div},\Omega)$, and $q\in H^1_{\Gamma_D}(\Omega)$. Combined with the definition of the bilinear form in (2.4), this leads to

$$\begin{aligned}\mathcal{B}(v^-,v^+,q;v^-,v^+,q) &= \mathcal{B}(v^-,v^+,q;v^-,v^+,q)\\ &\quad -\frac{\tau}{3}\left((v^-,\nabla q)_{0,\Omega}+(\mathrm{div}\,v^-,q)_{0,\Omega}\right)-\frac{\tau}{3}\left((v^+,\nabla q)_{0,\Omega}+(\mathrm{div}\,v^+,q)_{0,\Omega}\right)\\ &= \left(\mathbf{z},\begin{pmatrix}\frac{\tau}{3}&0&\frac{\tau}{6}&0&0&-\frac{\tau}{6}\\ 0&\frac{\tau^2}{3}&0&\frac{\tau^2}{6}&\frac{\tau}{6}&0\\ \frac{\tau}{6}&0&\frac{\tau}{3}&0&0&0\\ 0&\frac{\tau^2}{6}&0&\frac{\tau^2}{3}&\frac{\tau}{6}&0\\ 0&\frac{\tau}{6}&0&\frac{\tau}{6}&1&0\\ -\frac{\tau}{6}&0&0&0&0&\frac{\tau}{3}\end{pmatrix}\mathbf{z}\right)_{0,\Omega}\end{aligned}$$

with $\mathbf{z}=(a^{-1/2}v^-,\,c^{-1/2}\mathrm{div}\,v^-,\,a^{-1/2}v^+,\,c^{-1/2}\mathrm{div}\,v^+,\,c^{1/2}q,\,a^{1/2}\nabla q)^T$. The lower and upper bounds in (3.3) are obtained from the smallest and largest eigenvalue, respectively, of

$$\begin{pmatrix}\frac{\tau}{3}&0&\frac{\tau}{6}&0&0&-\frac{\tau}{6}\\ 0&\frac{\tau^2}{3}&0&\frac{\tau^2}{6}&\frac{\tau}{6}&0\\ \frac{\tau}{6}&0&\frac{\tau}{3}&0&0&0\\ 0&\frac{\tau^2}{6}&0&\frac{\tau^2}{3}&\frac{\tau}{6}&0\\ 0&\frac{\tau}{6}&0&\frac{\tau}{6}&1&0\\ -\frac{\tau}{6}&0&0&0&0&\frac{\tau}{3}\end{pmatrix}\begin{pmatrix}z_1\\z_2\\z_3\\z_4\\z_5\\z_6\end{pmatrix}=\lambda\begin{pmatrix}\frac{\tau}{3}&0&0&0&0&0\\ 0&\frac{\tau^2}{3}&0&0&0&0\\ 0&0&\frac{\tau}{3}&0&0&0\\ 0&0&0&\frac{\tau^2}{3}&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&\frac{\tau}{3}\end{pmatrix}\begin{pmatrix}z_1\\z_2\\z_3\\z_4\\z_5\\z_6\end{pmatrix}.$$

These generalized eigenvalues are given by the eigenvalues of the matrix

$$\begin{pmatrix}1&0&\frac{1}{2}&0&0&-\frac{1}{2}\\ 0&1&0&\frac{1}{2}&\frac{1}{2\sqrt{3}}&0\\ \frac{1}{2}&0&1&0&0&0\\ 0&\frac{1}{2}&0&1&\frac{1}{2\sqrt{3}}&0\\ 0&\frac{1}{2\sqrt{3}}&0&\frac{1}{2\sqrt{3}}&1&0\\ -\frac{1}{2}&0&0&0&0&1\end{pmatrix}$$

which can be shown to be located in the interval $[\frac{1}{4},2]$. ∎

Theorem 3.1 implies quasi-optimality of the solution of the variational problem (3.1). Cea's lemma (cf. [4, 5]) implies

$$|||(\bar{u}_\tau^- - u_{\tau,h}^-, \bar{u}_\tau^+ - u_{\tau,h}^+, \bar{p}_\tau^+ - p_{\tau,h}^+)||| \leq \sqrt{8} \inf_{v_h^- \in V_h, v_h^+ \in V_h, q_h \in Q_h} |||(\bar{u}_\tau^- - v_h^-, \bar{u}_\tau^+ - v_h^+, \bar{p}_\tau^+ - q_h)|||. \tag{3.4}$$

Another interpretation of Theorem 3.1 is that the bilinear form in (2.4) defines a norm,

$$\mathcal{B}(v^-, v^+, q; v^-, v^+, q)^{1/2} =: |||(v^-, v^+, q)|||_{\mathcal{B}}, \tag{3.5}$$

which is equivalent to $|||(\cdot, \cdot, \cdot)|||$ on $H_{\Gamma_N}(\mathrm{div}, \Omega)^2 \times H^1_{\Gamma_D}(\Omega)$.

THEOREM 3.2. *We have the following decomposition of the functional minimum into the discretization error components associated with time and with space:*

$$\begin{aligned}\mathcal{F}(u_{\tau,h}^-, u_{\tau,h}^+, p_{\tau,h}^+; p_\tau(t), f) &= \mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f) \\ &\quad + |||(\bar{u}_\tau^- - u_{\tau,h}^-, \bar{u}_\tau^+ - u_{\tau,h}^+, \bar{p}_\tau^+ - p_{\tau,h}^+)|||_{\mathcal{B}}^2 .\end{aligned} \tag{3.6}$$

Proof. Let $\bar{p} := E(t+\tau; t) p_{\tau,h}(t) \in H^1_{\Gamma_D}(\Omega)$ denote the true evolution of the parabolic problem and let $\bar{u} := -a \nabla \bar{p}$ denote the corresponding fluxes. We define the bilinear form

$$\begin{aligned}\mathcal{B}_0(u, p; v, q) = \int_t^{t+\tau} \Big(&\tau (c^{1/2} \partial_t p(s) + c^{-1/2} \mathrm{div}\, u(s), c^{1/2} \partial_t q(s) + c^{-1/2} \mathrm{div}\, v(s))_{0,\Omega} \\ &+ (a^{-1/2} u(s) + a^{1/2} \nabla p(s), a^{-1/2} v(s) + a^{1/2} \nabla q(s))_{0,\Omega} \Big)\, ds\end{aligned}$$

for all $u, v \in L^2((0,T), H_{\Gamma_N}(\mathrm{div}, \Omega))$, and $p, q \in H^1((0,T), H^1_{\Gamma_D}(\Omega))$. By comparison with (2.4) we see that

$$\begin{aligned}\mathcal{B}_0(\bar{u}_\tau, \bar{p}_\tau; v_\tau, q_\tau) &= \mathcal{B}(\bar{u}_\tau^-, \bar{u}_\tau^+, \bar{p}_\tau^+; v_\tau^-, v_\tau^+, q_\tau^+) \\ &\quad - \Big(c^{1/2} p_{\tau,h}(t), c^{1/2} q_\tau^+ + \frac{\tau}{2} c^{-1/2} (\mathrm{div}\, v_\tau^- + \mathrm{div}\, v_\tau^+) \Big)_{0,\Omega} \\ &\quad + \frac{\tau}{6} \Big(a^{1/2} \nabla p_{\tau,h}(t), a^{1/2} \nabla q_\tau^+ + a^{-1/2} (2 v_\tau^- + v_\tau^+) \Big)_{0,\Omega}\end{aligned} \tag{3.7}$$

with

$$\bar{u}_\tau(t+\sigma) = \Big(1 - \frac{\sigma}{\tau}\Big) \bar{u}_\tau^- + \frac{\sigma}{\tau} \bar{u}_\tau^+ , \quad v_\tau(t+\sigma) = \Big(1 - \frac{\sigma}{\tau}\Big) v_\tau^- + \frac{\sigma}{\tau} v_\tau^+,$$

for $\bar{u}_\tau^-, \bar{u}_\tau^+, v_\tau^-, v_\tau^+ \in H_{\Gamma_N}(\mathrm{div}, \Omega)$,

$$\bar{p}_\tau(t+\sigma) = \Big(1 - \frac{\sigma}{\tau}\Big) p_{\tau,h}(t) + \frac{\sigma}{\tau} \bar{p}_\tau^+ , \quad q_\tau(t+\sigma) = \frac{\sigma}{\tau} q_\tau^+,$$

for $\bar{p}_\tau^+, q_\tau^+ \in H^1_{\Gamma_D}(\Omega))$. Moreover,

$$\mathcal{F}(v_\tau^-, v_\tau^+, q_\tau^+; p_{\tau,h}(t), f) = \mathcal{B}_0(\bar{u} - v_\tau, \bar{p} - q_\tau; \bar{u} - v_\tau, \bar{p} - q_\tau)$$

holds for all $(v_\tau^-, v_\tau^+, q_\tau^+) \in H_{\Gamma_N}(\mathrm{div}, \Omega)^2 \times H^1_{\Gamma_D}(\Omega)$. If we define the subspace

$$Q_\tau^\circ \left((t, t+\tau), H^1_{\Gamma_D}(\Omega)\right) = \left\{ q_\tau(t+\sigma) = \frac{\sigma}{\tau} q_\tau^+ : q_\tau^+ \in H^1_{\Gamma_D}(\Omega), \sigma \in (0, \tau) \right\},$$

then comparing (2.5) with (3.7) implies that the solution $(\bar{u}_\tau, \bar{p}_\tau)$ of the minimization problem satisfies

$$\mathcal{B}_0(\bar{u} - \bar{u}_\tau, \bar{p} - \bar{p}_\tau; v_\tau, q_\tau) = 0$$

for all $(v_\tau, q_\tau) \in V_\tau((t, t+\tau), H_{\Gamma_N}(\mathrm{div}, \Omega)) \times Q_\tau^\circ((t, t+\tau), H^1_{\Gamma_D}(\Omega))$. The identity in this theorem then follows from

$$\begin{aligned}
&\mathcal{F}(u^-_{\tau,h}, u^+_{\tau,h}, p^+_{\tau,h}; p_\tau(t), f) \\
&\quad = \mathcal{B}_0(\bar{u} - u_{\tau,h}, \bar{p} - p_{\tau,h}; \bar{u} - u_{\tau,h}, \bar{p} - p_{\tau,h}) \\
&\quad = \mathcal{B}_0(\bar{u} - \bar{u}_\tau, \bar{p} - \bar{p}_\tau; \bar{u} - \bar{u}_\tau, \bar{p} - \bar{p}_\tau) \\
&\quad + 2\mathcal{B}_0(\bar{u} - \bar{u}_\tau, \bar{p} - \bar{p}_\tau; \bar{u}_\tau - u_{\tau,h}, \bar{p}_\tau - p_{\tau,h}) \\
&\quad + \mathcal{B}(\bar{u}^-_\tau - u^-_{\tau,h}, \bar{u}^+_\tau - u^+_{\tau,h}, \bar{p}^+_\tau - p^+_{\tau,h}; \bar{u}^-_\tau - u^-_{\tau,h}, \bar{u}^+_\tau - u^+_{\tau,h}, \bar{p}^+_\tau - p^+_{\tau,h}) \\
&\quad = \mathcal{F}(\bar{u}^-_\tau, \bar{u}^+_\tau, \bar{p}^+_\tau; p_{\tau,h}(t), f) + |||(\bar{u}^-_\tau - u^-_{\tau,h}, \bar{u}^+_\tau - u^+_{\tau,h}, \bar{p}^+_\tau - p^+_{\tau,h})|||^2_{\mathcal{B}}\,,
\end{aligned}$$

where, similarly as above, $u^-_{\tau,h} = u_{\tau,h}(t), u^+_{\tau,h} = u_{\tau,h}(t+\tau)$, and $p^+_{\tau,h} = p_{\tau,h}(t+\tau)$. Note that $(\bar{u}_\tau - u_{\tau,h}, \bar{p}_\tau - p_{\tau,h}) \in V_\tau((t, t+\tau), H_{\Gamma_N}(\mathrm{div}, \Omega)) \times Q_\tau^\circ((t, t+\tau), H^1_{\Gamma_D}(\Omega))$, and therefore the middle term vanishes. □

In other words, Theorem 3.2 states that the space discretization error would be orthogonal to the error with respect to time. As a consequence of Theorem 3.2, the functional at the fully discrete approximation $\mathcal{F}(u^-_{\tau,h}, u^+_{\tau,h}, p^+_{\tau,h}; p_{\tau,h}(t), f)$ is decreasing if the spaces V_h and Q_h are expanded. In the course of our computations, we evaluate this functional value and augment it with the result of an a posteriori estimator for $|||(u^-_\tau - u^-_{\tau,h}, u^+_\tau - u^+_{\tau,h}, p^+_\tau - p^+_{\tau,h})|||^2_{\mathcal{B}}$. Only the consistency error resulting from the (time and space) discretization of the new time-step is controlled with this approach. The error from previous time-steps ($p(t) - p_\tau(t)$ in the semidiscrete case, $p(t) - p_{\tau,h}(t)$ in the fully discrete case) is not taken into account in our a posteriori estimator. However, the stability result established in [12] ensures that this initial error is not amplified, and convergence of the fully discrete method can be shown similarly as in [12, Thm. 5.2].

4. Hierarchical basis a posteriori error estimator in space. Our aim is to construct an a posteriori error estimator for

$$|||(\bar{u}^-_\tau - u^-_{\tau,h}, \bar{u}^+_\tau - u^+_{\tau,h}, \bar{p}^+_\tau - p^+_{\tau,h})|||_{\mathcal{B}}\,,$$

the error associated with the space discretization. To this end, we use a hierarchical basis approach based on minimizing the functional with respect to the hierarchical surplus space associated with one step of uniform refinement. The principle of hierarchical a posteriori error estimators is based on the hierarchical decomposition of the finite element spaces on a uniformly refined triangulation $\mathcal{T}_{h/2}$,

$$V_{h/2} = V_h \oplus Z_h\,,\ Q_{h/2} = Q_h \oplus Y_h$$

or, alternatively, based on using higher order polynomials (cf. [2]). The functional is then minimized with respect to the hierarchical surplus space,

$$\min_{(z^-_h, z^+_h, y_h) \in Z_h^2 \times Y_h} \mathcal{F}(u^-_{\tau,h} + z^-_h, u^+_{\tau,h} + z^+_h, p^+_{\tau,h} + y_h; p_\tau(t), f)\,. \tag{4.1}$$

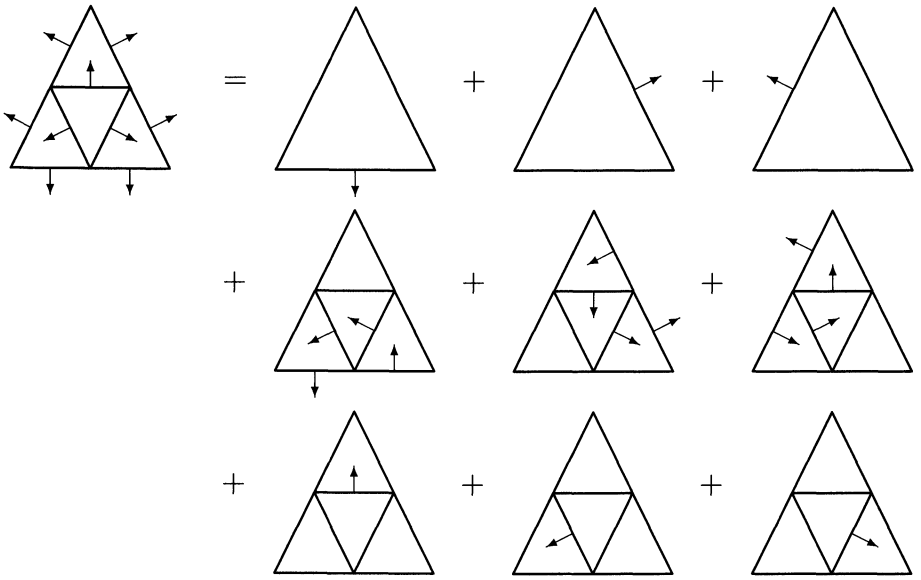

FIG. 4.1. *Hierarchical basis for lowest-order Raviart–Thomas elements.*

The variational formulation of (4.1) is to find $(d_h^-, d_h^+, e_h) \in Z_h^2 \times Y_h$ such that

$$
\begin{aligned}
&\mathcal{B}(d_h^-, d_h^+, e_h; z_h^-, z_h^+, y_h) \\
&\quad = \Big(c^{1/2}\, p_\tau(t) + \tau\, c^{-1/2}\, f, c^{1/2}\, y_h + \frac{\tau}{2}\, c^{-1/2}\, (\operatorname{div} z_h^- + \operatorname{div} z_h^+)\Big)_{0,\Omega} \\
&\quad - \frac{\tau}{6}\, (a^{1/2}\nabla p_\tau(t), a^{-1/2}(2z_h^- + z_h^+))_{0,\Omega} - \frac{\tau}{6}\, (a^{1/2}\nabla p_\tau(t), a^{1/2}\nabla y_h)_{0,\Omega} \\
&\quad - \mathcal{B}(u_{\tau,h}^-, u_{\tau,h}^+, p_{\tau,h}; z_h^-, z_h^+, y_h)
\end{aligned}
\tag{4.2}
$$

holds for all $z_h^-, z_h^+ \in Z_h$, and $y_h \in Y_h$. We restrict our attention to the lowest-order Raviart–Thomas spaces for V_h combined with standard piecewise linear continuous finite elements for Q_h. As we shall see below, choosing the basis functions for Z_h and Y_h suitably, the variational formulation (4.2) results in a system of linear equations which is uniformly well conditioned with respect to h and τ. A similar approach was proposed in [16] for an error estimator based on the hierarchical extension to polynomials of higher degree.

The hierarchical basis that we use for the Raviart–Thomas spaces is depicted in Figure 4.1. The three basis functions in the top row are those associated with the coarse triangulation. The middle row of Figure 4.1 spans the divergence-free part of Z_h, and the bottom row complements the basis for Z_h.

We follow the usual steps for analyzing hierarchical basis error estimators (see [15, section 1.4]). First, the following strengthened Cauchy–Schwarz inequality, uniformly in the time-step τ, is shown.

LEMMA 4.1. *There exists a* $\gamma \in [0, 1)$ *such that*

$$
\mathcal{B}(v_h^-, v_h^+, q_h; z_h^-, z_h^+, y_h) \le \gamma |||(v_h^-, v_h^+, q_h)|||_{\mathcal{B}} |||(z_h^-, z_h^+, y_h)|||_{\mathcal{B}} \tag{4.3}
$$

for all $(v_h^-, v_h^+, q_h) \in V_h^2 \times Q_h$ *and* $(z_h^-, z_h^+, y_h) \in Z_h^2 \times Y_h$, *uniformly for all* $\tau > 0$.

Proof. (i) We define the auxiliary bilinear form

$$\mathcal{A}(u^-,u^+,p;v^-,v^+,q) = \tau(u^-,v^-)_{0,\Omega} + \tau^2(\operatorname{div} u^-,\operatorname{div} v^-)_{0,\Omega} + \tau(u^+,v^+)_{0,\Omega} + \tau^2(\operatorname{div} u^+,\operatorname{div} v^+)_{0,\Omega} + (p,q)_{0,\Omega} + \tau(\nabla p,\nabla q)_{0,\Omega}$$

and observe that Theorem 3.1 combined with the assumptions on a and c implies

$$\underline{\alpha}\mathcal{A}(v^-,v^+,q;v^-,v^+,q) \le \mathcal{B}(v^-,v^+,q;v^-,v^+,q) \le \bar{\alpha}\mathcal{A}(v^-,v^+,q;v^-,v^+,q)$$

for all $(v^-,v^+,q) \in H_{\Gamma_N}(\operatorname{div},\Omega)^2 \times H^1_{\Gamma_D}(\Omega)$ with constants $\bar{\alpha} \ge \underline{\alpha} > 0$, which are independent of τ.

The statement of Lemma 4.1 also can be formulated as the existence of a $\gamma \in [0,1)$ such that

$$\mathcal{B}(v_h^-,v_h^+,q_h;z_h^-,z_h^+,y_h) \le \gamma$$

for all $(v_h^-,v_h^+,q_h) \in V_h^2 \times Q_h$ with $|||(v_h^-,v_h^+,q_h)|||_{\mathcal{B}} = 1$ and $(z_h^-,z_h^+,y_h) \in Z_h^2 \times Y_h$ with $|||(z_h^-,z_h^+,y_h)|||_{\mathcal{B}} = 1$. Under these assumptions, the identity

$$|||(v_h^- + z_h^-, v_h^+ + z_h^+, q_h + y_h)|||_{\mathcal{B}}^2 |||(v_h^- - z_h^-, v_h^+ - z_h^+, q_h - y_h)|||_{\mathcal{B}}^2 = 4(1 - \mathcal{B}(v_h^-,v_h^+,q_h;z_h^-,z_h^+,y_h)^2)$$

holds. Therefore, we need only show that there is a positive constant β_0 such that

$$|||(v_h^- + z_h^-, v_h^+ + z_h^+, q_h + y_h)|||_{\mathcal{B}} |||(v_h^- - z_h^-, v_h^+ - z_h^+, q_h - y_h)|||_{\mathcal{B}} \ge \beta_0$$

for all $(v_h^-,v_h^+,q_h) \in V_h^2 \times Q_h$ with $|||(v_h^-,v_h^+,q_h)|||_{\mathcal{B}} = 1$ and $(z_h^-,z_h^+,y_h) \in Z_h^2 \times Y_h$ with $|||(z_h^-,z_h^+,y_h)|||_{\mathcal{B}} = 1$ or, equivalently, $\alpha_0 > 0$ such that

$$|||(v_h^- + z_h^-, v_h^+ + z_h^+, q_h + y_h)|||_{\mathcal{A}} |||(v_h^- - z_h^-, v_h^+ - z_h^+, q_h - y_h)|||_{\mathcal{A}} \ge \alpha_0$$

for all $(v_h^-,v_h^+,q_h) \in V_h^2 \times Q_h$ with $|||(v_h^-,v_h^+,q_h)|||_{\mathcal{A}} = 1$ and $(z_h^-,z_h^+,y_h) \in Z_h^2 \times Y_h$ with $|||(z_h^-,z_h^+,y_h)|||_{\mathcal{A}} = 1$. Using the same argument as above, we are left with showing that there exists a $\tilde{\gamma} \in [0,1)$ such that

$$\mathcal{A}(v_h^-,v_h^+,q_h;z_h^-,z_h^+,y_h) \le \tilde{\gamma}|||(v_h^-,v_h^+,q_h)|||_{\mathcal{A}}|||(z_h^-,z_h^+,y_h)|||_{\mathcal{A}}$$

for all $(v_h^-,v_h^+,q_h) \in V_h^2 \times Q_h$ and $(z_h^-,z_h^+,y_h) \in Z_h^2 \times Y_h$.

(ii) Our next step is to show the existence of $\tilde{\gamma} < 1$ such that

$$\mathcal{A}_K(v_h^-,v_h^+,q_h;z_h^-,z_h^+,y_h) \le \tilde{\gamma}\mathcal{A}_K(v_h^-,v_h^+,q_h;v_h^-,v_h^+,q_h)^{1/2}\mathcal{A}_K(z_h^-,z_h^+,y_h;z_h^-,z_h^+,y_h)^{1/2}, \tag{4.4}$$

where $\mathcal{A}_K(\cdot,\cdot,\cdot;\cdot,\cdot,\cdot)$ is the elementwise bilinear form, i.e.,

$$\mathcal{A}_K(u^-,u^+,p;v^-,v^+,q) = \tau(u^-,v^-)_{0,K} + \tau^2(\operatorname{div} u^-,\operatorname{div} v^-)_{0,K} + \tau(u^+,v^+)_{0,K} + \tau^2(\operatorname{div} u^+,\operatorname{div} v^+)_{0,K} + (p,q)_{0,K} + \tau(\nabla p,\nabla q)_{0,K}$$

for all $K \in \mathcal{T}_h$. It is well known that

$$\begin{aligned}(q_h,y_h)_{0,K} &\le \tilde{\gamma}_0\|q_h\|_{0,K}\|y_h\|_{0,K},\\ (\nabla q_h,\nabla y_h)_{0,K} &\le \tilde{\gamma}_1\|\nabla q_h\|_{0,K}\|\nabla y_h\|_{0,K}\end{aligned} \tag{4.5}$$

holds for all $q_h \in Q_h$ and $y_h \in Y_h$ with $\tilde{\gamma}_0,\tilde{\gamma}_1 < 1$ (see [2]).

Using the affine transformation of K to a reference element, we see that

$$
\begin{aligned}
(v_h, z_h)_{0,K} &= \frac{|K|}{|K_{\mathrm{ref}}|}(v^{\mathrm{ref}}, z^{\mathrm{ref}})_{0,K_{\mathrm{ref}}} \\
&\le \tilde{\gamma}_2 \frac{|K|}{|K_{\mathrm{ref}}|}\|v^{\mathrm{ref}}\|_{0,K_{\mathrm{ref}}}\|z^{\mathrm{ref}}\|_{0,K_{\mathrm{ref}}} = \tilde{\gamma}_2\|v_h\|_{0,K}\|z_h\|_{0,K}
\end{aligned} \tag{4.6}
$$

for all $v_h \in V_h$ and $z_h \in Z_h$. The inequality

$$
(v^{\mathrm{ref}}, z^{\mathrm{ref}})_{0,K_{\mathrm{ref}}} \le \tilde{\gamma}_2\|v^{\mathrm{ref}}\|_{0,K_{\mathrm{ref}}}\|z^{\mathrm{ref}}\|_{0,K_{\mathrm{ref}}}
$$

for all $v^{\mathrm{ref}} \in V^{\mathrm{ref}}$ and $z^{\mathrm{ref}} \in Z^{\mathrm{ref}}$ with a constant $\tilde{\gamma}_2 < 1$ follows from the fact that the reference spaces V^{ref} and Z^{ref} are finite-dimensional (of dimension three and six, respectively; see Figure 4.1) and that $V^{\mathrm{ref}} \cap Z^{\mathrm{ref}} = \{0\}$.

Finally, we have $\operatorname{div} v_h$ constant on K for $v_h \in V_h$ (since v_h is linear) and

$$
\int_K \operatorname{div} z_h \, dx = \int_{\partial K} n \cdot z_h \, ds = 0 \quad \text{for all } z_h \in Z_h
$$

since for the basis functions either $\operatorname{div} z_h = 0$ on K (middle row in Figure 4.1) or $n \cdot z_h$ on ∂K (bottom row in Figure 4.1). Clearly, this implies

$$
(\operatorname{div} v_h, \operatorname{div} z_h)_{0,K} = 0 \quad \text{for all } v_h \in V_h \text{ and } z_h \in Z_h\,. \tag{4.7}
$$

Combining (4.5), (4.6), and (4.7) with the Cauchy–Schwarz inequality (for sums) leads to (4.4) with $\tilde{\gamma} = \max\{\tilde{\gamma}_0, \tilde{\gamma}_1, \tilde{\gamma}_2\}$.

(iii) Using again the Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
&\mathcal{A}(v_h^-, v_h^+, q_h; z_h^-, z_h^+, y_h) = \sum_{K\in\mathcal{T}_h} \mathcal{A}_K(v_h^-, v_h^+, q_h; z_h^-, z_h^+, y_h) \\
&\le \tilde{\gamma} \sum_{K\in\mathcal{T}_h} \mathcal{A}_K(v_h^-, v_h^+, q_h; v_h^-, v_h^+, q_h)^{1/2}\mathcal{A}_K(z_h^-, z_h^+, y_h; z_h^-, z_h^+, y_h)^{1/2} \\
&\le \tilde{\gamma} \left(\sum_{K\in\mathcal{T}_h} \mathcal{A}_K(v_h^-, v_h^+, q_h; v_h^-, v_h^+, q_h)\right)^{1/2} \left(\sum_{K\in\mathcal{T}_h} \mathcal{A}_K(z_h^-, z_h^+, y_h; z_h^-, z_h^+, y_h)\right)^{1/2} \\
&= \tilde{\gamma}|||(v_h^-, v_h^+, q_h)|||_{\mathcal{A}}|||(z_h^-, z_h^+, y_h)|||_{\mathcal{A}}
\end{aligned}
$$

for all $(v_h^-, v_h^+, q_h) \in V_h^2 \times Q_h$ and $(z_h^-, z_h^+, y_h) \in Z_h^2 \times Y_h$, which completes the proof. □

The hierarchical error estimator is defined in the usual way by

$$
\eta_K = \mathcal{B}_K(d_h^-, d_h^+, e_h; d_h^-, d_h^+, e_h)^{1/2}
$$

(see [15, section 1.4]). We are now ready to establish the equivalence of the estimator to the error under the usual saturation assumption.

THEOREM 4.2. *Assume that the saturation condition*

$$
\begin{aligned}
&\inf_{(v^-,v^+,q)\in V_{h/2}^2\times Q_{h/2}} |||(u_\tau^- - v^-, u_\tau^+ - v^+, p_\tau^+ - q)|||_{\mathcal{B}} \\
&\qquad \le \beta \inf_{(v^-,v^+,q)\in V_h^2\times Q_h} |||(u_\tau^- - v^-, u_\tau^+ - v^+, p_\tau^+ - q)|||_{\mathcal{B}}
\end{aligned} \tag{4.8}
$$

is satisfied with $\beta < 1$ uniformly in h and τ. Then, we have

$$\sum_{K\in\mathcal{T}_h} \eta_T^2 \eqsim |||(u_\tau^- - u_{\tau,h}^-, u_\tau^+ - u_{\tau,h}^+, p_\tau^+ - p_{\tau,h}^+)|||_{\mathcal{B}}^2 ; \tag{4.9}$$

i.e., these two quantities are equivalent with constants independently of h and τ.

The proof is analogous to the one for the standard hierarchical basis (see [15, section 1.4]). Its main ingredients are the strengthened Cauchy–Schwarz inequality of Lemma 4.1 and the saturation condition (4.8).

THEOREM 4.3. *The matrix associated with the variational problem* (4.2)*, using nodal bases for Z_h and Y_h, is equivalent to its diagonal, uniformly in h and τ.*

Proof. Denote the basis representations of $z_h \in Z_h$ and $y_h \in Y_h$, respectively, by

$$z_h = \sum_\mu z_h^{(\mu)} \Psi_h^{(\mu)} \text{ and } y_h = \sum_\mu y_h^{(\mu)} \Phi_h^{(\mu)} .$$

We may again use the auxiliary bilinear form $\mathcal{A}(\cdot,\cdot,\cdot;\cdot,\cdot,\cdot)$ introduced in the proof of Lemma 4.1. Due to the equivalence of the bilinear forms, it is sufficient to show that

$$\mathcal{A}(z_h^-, z_h^+, y_h; z_h^-, z_h^+, y_h) \eqsim \sum_\mu \left(z_h^{-,(\mu)}\right)^2 \mathcal{A}\left(\Psi_h^{(\mu)}, 0, 0; \Psi_h^{(\mu)}, 0, 0\right)$$
$$+ \sum_\mu \left(z_h^{+,(\mu)}\right)^2 \mathcal{A}\left(0, \Psi_h^{(\mu)}, 0; 0, \Psi_h^{(\mu)}, 0\right) + \sum_\mu \left(y_h^{(\mu)}\right)^2 \mathcal{A}\left(0, 0, \Phi_h^{(\mu)}; 0, 0, \Phi_h^{(\mu)}\right) .$$

The element matrices

$$\left[\mathcal{A}_K\left(\Psi_h^{(\mu)}, 0, 0; \Psi_h^{(\nu)}, 0, 0\right)\right]_{1\le\mu,\nu\le 6} = \left[\mathcal{A}_K\left(0, \Psi_h^{(\mu)}, 0; 0, \Psi_h^{(\nu)}, 0\right)\right]_{1\le\mu,\nu\le 6}$$

on the reference element (with corners $(-\sqrt{3}/2, -1/2), (\sqrt{3}/2, -1/2), (0, 1)$ as shown in Figure 4.1) have the form

$$\frac{\sqrt{3}}{48}\tau \begin{pmatrix} 36 & -12 & -12 & 0 & 12 & -12 \\ -12 & 36 & -12 & -12 & 0 & 12 \\ -12 & -12 & 36 & 12 & -12 & 0 \\ 0 & -12 & 12 & 10 & -1 & -1 \\ 12 & 0 & -12 & -1 & 10 & -1 \\ -12 & 12 & 0 & -1 & -1 & 10 \end{pmatrix} + \frac{\sqrt{3}}{3}\tau^2 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 4 & 4 \\ 0 & 0 & 0 & 4 & 8 & 4 \\ 0 & 0 & 0 & 4 & 4 & 8 \end{pmatrix} .$$

That this matrix is equivalent to its diagonal, uniformly in τ, is the consequence of a simple eigenvalue computation. The analogous result for

$$\left[\mathcal{A}_K\left(0, 0, \Phi_h^{(\mu)}; 0, 0, \Phi_h^{(\nu)}\right)\right]_{1\le\mu,\nu\le 3}$$

for the standard hierarchical basis can be found in [2]. ∎

Theorem 4.3 implies that we can use the conjugate gradient method with Jacobi preconditioning to solve (4.2) with a number of iterations that is bounded in τ and h.

5. Time and space adaptive algorithm. We recall from [12] the definition of the norm

$$|||(v,q)|||_{\tau_j} := \left(\int_{t_{j-1}}^{t_j} \left(\|v(\sigma)\|_{0,\Omega}^2 + \tau_j \|\operatorname{div} v(\sigma)\|_{0,\Omega}^2 \right.\right.$$
$$\left.\left. + \frac{1}{\tau_j}\|q(\sigma)\|_{0,\Omega}^2 + \|\nabla q(\sigma)\|_{0,\Omega}^2 \right) d\sigma \right)^{1/2} \tag{5.1}$$

(with time-step-size $\tau_j = t_j - t_{j-1}$) in which we measure the discretization error. If our aim is to achieve $|||(u-u_\tau, p-p_\tau)|||_{\tau_j} \leq \texttt{tol}$ in each time-step for some prescribed tolerance tol, then we should require

$$|||(u-u_\tau, p-p_\tau)|||^2_{\tau_j} \eqsim \mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f) \leq (\texttt{tol})^2 .$$

If this requirement is fulfilled, we accept the approximation and have to suggest the next time-step. Otherwise, a new time-step has to be chosen to repeat the step. In what follows, both of the above cases will be considered. We use the notation τ_{old} and τ_{new} to avoid confusion and set $\text{Error}_{\text{old}} := \mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f)$. In both cases (accepted or repeated step), our aim is to achieve $\text{Error}_{\text{new}} \leq \texttt{tol}$. In order to minimize the computational effort, we also want to achieve $\text{Error}_{\text{new}} \approx \texttt{tol}$. On the other hand, [12, Thm. 5.4] gives $\text{Error}_{\text{old}} \approx C_1 \tau_{\text{old}}^{3/2}$ and $\text{Error}_{\text{new}} \approx C_2 \tau_{\text{new}}^{3/2}$. Dividing the first of these relations by the second and assuming $C_1 \approx C_2$, we obtain

$$\left(\frac{\tau_{\text{new}}}{\tau_{\text{old}}}\right)^{3/2} \approx \frac{\text{Error}_{\text{new}}}{\text{Error}_{\text{old}}}. \tag{5.2}$$

Keeping in mind our error criterion and introducing a safety factor in order to avoid infinite loops leads to

$$\tau_{\text{new}} = \left(\frac{\texttt{tol}}{\text{Error}_{\text{old}}}\right)^{2/3} \tau_{\text{old}} \delta \,, \tag{5.3}$$

with some suitable $\delta \in (0,1)$. Furthermore, the time-step size should be restricted from above and below, as we will see later in the algorithm.

For the adaptive refinement of the triangulation we use the fact that, in addition to the functional which measures the combined error in time and space, we have an estimator for the space discretization error. Starting from an initial triangulation $\mathcal{T}_0$, we construct a sequence of adaptively refined triangulations $\mathcal{T}_l$ by evaluating the hierarchical estimator with respect to the corresponding spaces V_l and Q_l. The hierarchical surplus spaces on $\mathcal{T}_l$ are denoted by Z_l and Y_l. This is done as long as the value of the estimator for the global discretization error in space is not much smaller than the value of the least-squares functional. Furthermore, a limit on the maximum number of levels $l_{\max}$ is implemented.

With the user-defined parameters tol, $\gamma, \delta \in (0,1)$, $\tau_{\max}$, and $\tau_{\min}$, and the initial time-step τ_0, our adaptive algorithm reads as follows:

t $= 0$; $\tau = \tau_0$; $p^{\text{old}} = p_0$;
while $t < T$,
  compute $(u_{\tau,h}^-, u_{\tau,h}^+, p_{\tau,h}^+) \in V_0^2 \times Q_0$ such that it solves (3.1);
  compute $(d_h^-, d_h^+, e_h) \in Z_0^2 \times Y_0$ such that it solves (4.2);
  compute $\eta_K = \mathcal{B}_K(d_h^-, d_h^+, e_h; d_h^-, d_h^+, e_h)^{1/2}$ for each $K \in \mathcal{T}_0$; $\eta_0^2 = \sum_{K \in \mathcal{T}_0} \eta_K^2$;
  $l = 0$;
  while $l \leq l_{\max}$ **and not** $[\eta_l \ll \mathcal{F}(u_{\tau,h}^-, u_{\tau,h}^+, p_{\tau,h}^+; p^{\text{old}}, f)]$,
    $l = l + 1$;
    $\mathcal{T}_l =$ refine $\hat{K} \in \mathcal{T}_{l-1}$ with $\eta_{\hat{K}} > \gamma \max_{K \in \mathcal{T}_{l-1}} \eta_K$;

compute $(u^-_{\tau,h}, u^+_{\tau,h}, p^+_{\tau,h}) \in V_l^2 \times Q_l$ such that it solves (3.1);

compute $(d^-_h, d^+_h, e_h) \in Z_l^2 \times Y_l$ such that it solves (4.2);

compute $\eta_K = \mathcal{B}_K(d^-_h, d^+_h, e_h; d^-_h, d^+_h, e_h)^{1/2}$ for each $K \in \mathcal{T}_l$; $\eta_l^2 = \sum_{K \in \mathcal{T}_l} \eta_K^2$;

end

$\text{Error}_{\text{est}} = \mathcal{F}(u^-_{\tau,h}, u^+_{\tau,h}, p^+_{\tau,h}; p^{\text{old}}, f)^{1/2}$;

if $\text{Error}_{\text{est}} \le \texttt{tol}$ %% accept this step

$$t = t + \tau \; ; \; \tau = \min\left\{ \tau_{\max}, \left(\frac{\texttt{tol}}{\text{Error}_{\text{est}}}\right)^{2/3} \tau\delta \right\} \; ; \; p^{\text{old}} = p^+_{\tau,h}$$

else %% the computation of this step is not acceptable

$$\tau = \max\left(\tau_{\min}, \left(\frac{\texttt{tol}}{\text{Error}_{\text{est}}}\right)^{2/3} \tau\delta \right) \; ;$$

end

end

The time-stepping part of the above adaptive scheme is quite standard and may be found in many textbooks on numerical methods for systems of ordinary differential equations (see, for example, [6, section 5.1] or [8, section II.4]). A fully adaptive algorithm based on a combination of linear-implicit discretization in time with a hierarchical error estimator in space is also described in [10]. The least-squares approach with our specific choice of piecewise linear functions in time leads to a method that is third-order in the classical sense for the scalar variable p. Other approaches of comparable accuracy include the discontinuous Galerkin method with piecewise linear functions (see [7] and [14, Chap. 12] for related aspects of a posteriori error estimation). Another well-known approach to adaptive time-stepping is given by embedded Runge–Kutta methods. Implicit Runge–Kutta methods of Radau IIA type, for example, also lead to third-order accurate methods which can be augmented by an error estimator (see [9, section IV.5]). Those error estimators are tailored to measure the maximal error over time of the primary variable, for example, in $L^\infty((0,T), L^2(\Omega))$ for the discontinuous Galerkin method in [7]. In contrast, our least-squares approach estimates the error in $|||(\cdot,\cdot)|||_{\tau_j}$, which is an integral norm over time for both the primary (scalar) and the dual (flux) variables.

6. Computational results. In all the following examples, $a \equiv c \equiv 1$.

Example 1. On $\Omega = [-1,1]^2$ we solve the system (2.1) with right-hand side $f \equiv 0$. The prescribed boundary conditions are

$$p|_{[-1,1]\times\{1,-1\}} = 0 \, , \; \langle n, u\rangle|_{\{1,-1\}\times[-1,1]} = 0,$$

i.e., Dirichlet conditions on the top and bottom boundary segments and Neumann conditions on the left and right. Initial conditions are

$$p(0,x) = \cos\left(\frac{\pi}{2}(x_1 - 1)\right) \sin\left(\frac{\pi}{2}(x_2 - 1)\right) \, .$$

The exact solution of this problem is given by

$$p(t,x) = \exp\left(-\frac{\pi^2}{2}\, t\right) \cos\left(\frac{\pi}{2}(x_1 - 1)\right) \sin\left(\frac{\pi}{2}(x_2 - 1)\right),$$

TABLE 6.1
Consistency error measured in terms of the functional $\mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f)$.

h	1/8	1/16	1/32	1/64
$\tau = 0.2$	$3.22 \cdot 10^{-2}$	$2.10 \cdot 10^{-2}$	$1.82 \cdot 10^{-2}$	$1.75 \cdot 10^{-2}$
$\tau = 0.1$	$1.15 \cdot 10^{-2}$	$4.48 \cdot 10^{-3}$	$2.70 \cdot 10^{-3}$	$2.25 \cdot 10^{-3}$
$\tau = 0.05$	$5.54 \cdot 10^{-3}$	$1.55 \cdot 10^{-3}$	$5.36 \cdot 10^{-4}$	$2.81 \cdot 10^{-4}$
$\tau = 0.025$	$2.89 \cdot 10^{-3}$	$7.44 \cdot 10^{-4}$	$1.98 \cdot 10^{-4}$	$6.10 \cdot 10^{-5}$
$\tau = 0.0125$	$1.49 \cdot 10^{-3}$	$3.80 \cdot 10^{-4}$	$9.61 \cdot 10^{-5}$	$2.49 \cdot 10^{-5}$

TABLE 6.2
Error at time $t = 1$.

h	$\|p - p_{\tau,h}\|_{1,\Omega}$	$\|u - u_{\tau,h}\|_{\mathrm{div},\Omega}$
$\tau = 0.2$	$2.30 \cdot 10^{-3}$	$8.03 \cdot 10^{-3}$
$\tau = 0.1$	$3.45 \cdot 10^{-4}$	$3.00 \cdot 10^{-3}$
$\tau = 0.05$	$4.79 \cdot 10^{-5}$	$9.30 \cdot 10^{-4}$
$\tau = 0.025$	$6.42 \cdot 10^{-6}$	$2.88 \cdot 10^{-4}$
$\tau = 0.0125$	$1.31 \cdot 10^{-6}$	$1.55 \cdot 10^{-4}$

and its approximation is rather unproblematic. The purpose of this example is to illustrate the theoretically predicted approximation order. We use a uniform time-step τ and uniformly refined triangulations in order to illustrate the convergence of our method.

Table 6.1 shows the consistency error at the first time-step measured in terms of the functional. We see the expected reduction of the functional proportionally to τ^3 (equivalent to the consistency error norm $|||(\eta_u, \eta_p)|||_\tau^2$ by [12, Thm. 4.3]). Note that for smaller values of τ, the approximation order seems to be getting lower, which is due to the fact that we are not able to sufficiently reduce the approximation error in space. Table 6.2 shows the error at time $t = 1$ for different τ (for fixed $h = 1/64$). We see the order 3 for the approximation of p expected from our analysis in [12]. The approximation for u is of lower order.

Example 2. Here again, $\Omega = [-1, 1]^2$ and $f \equiv 0$. Homogeneous Dirichlet boundary conditions are prescribed on $\partial\Omega$, and the initial condition is

$$p(0, x) = \min\{1 - x_1, 1 + x_1, 1 - x_2, 1 + x_2\}\,.$$

Due to the nonsmoothness of the initial condition and the resulting insufficient spatial accuracy, we see a lower order of the method during the first time-steps until the order predicted by the theory is reached. Figure 6.1 shows the order α in $\mathcal{F}(u_\tau^-, u_\tau^+, p_\tau^+; p_\tau(t), f)^{1/2} = \mathcal{O}(\tau^\alpha)$ observed in the computations over time. Obviously, the expected order of 1.5 is reached after some short initial period.

Example 3. The following example is a very challenging test for our one-step method (cf. [7, Ex. 9.2]; see also [3, Ex. 8.3]). On $\Omega = [-2, 2]^2$, we solve the system (2.1) with right-hand side $f \equiv 0$ and boundary condition $p = 0$ on $\partial\Omega$. The initial condition

$$p(0, x) = 250 \exp\left(-250 \|x\|^2\right)$$

can be viewed as an approximation to a point source or δ-function. The exact solution of this problem is given by

$$p(t, x) = \frac{1}{4t + 1/250} \exp\left(-\frac{\|x\|^2}{4t + 1/250}\right)\,.$$

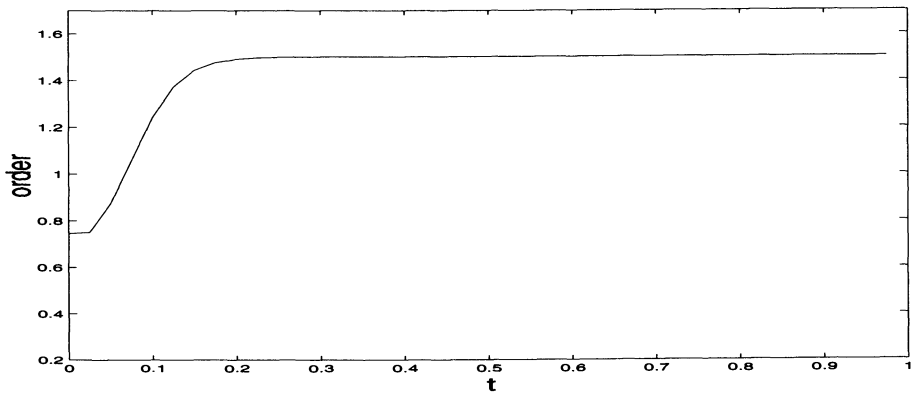

FIG. 6.1. *Example* 2: *Observed consistency order.*

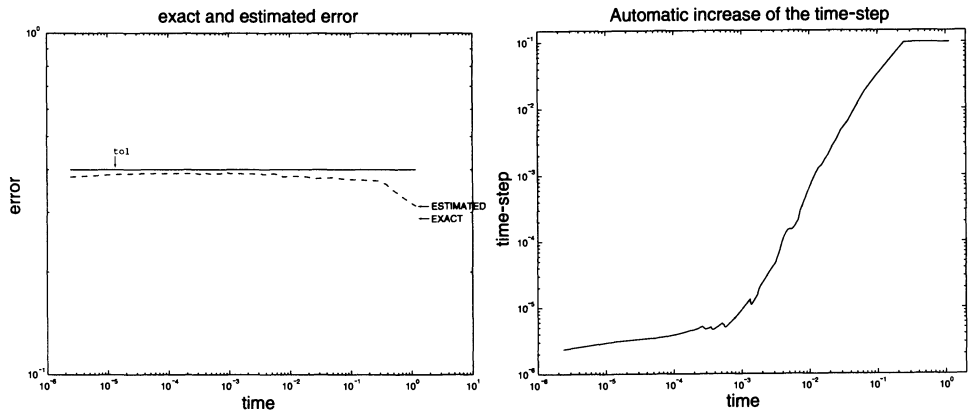

FIG. 6.2. *Example* 3: *Quality of the error estimator and time-steps.*

For $t \in (0, 1]$, which is the time interval for our computations, the homogeneous Dirichlet boundary conditions are very closely fulfilled. The program was started with a time-step of $\tau_0 = 1.78 \cdot 10^{-6}$ with a required tolerance of $\texttt{tol} = 0.4$. The discretization in space uses a triangulation which is the result of five adaptive refinement steps. We can notice by the right graph in Figure 6.2 the good performance of our error estimator compared to the exact error over time. The curve labeled "exact" indicates the norm

$$|||(e_u, e_p)|||_\tau = \left(\int_0^\tau \left(\|e_u(t+\sigma)\|_{0,\Omega}^2 + \tau \, \|\mathrm{div}\, e_u(t+\sigma)\|_{0,\Omega}^2 + \frac{1}{\tau} \|e_p(t+\sigma)\|_{0,\Omega}^2 + \|\nabla e_p(t+\sigma)\|_{0,\Omega}^2 \right) d\sigma \right)^{1/2}$$

(cf. [12, Thm. 4.3]) of the error $(e_u, e_p) = (u - u_{\tau,h}, p - p_{\tau,h})$. Note that this norm was computed only approximately using the trapezoidal rule for the time integration. The left part of Figure 6.2 shows the time-steps depending on time, showing that the estimator works efficiently. Notice that we bound the maximal time-step by 0.1.

In Figure 6.3 we demonstrate the solution and the corresponding triangulation at the beginning of the algorithm. Finally, Figure 6.4 shows the solution and the

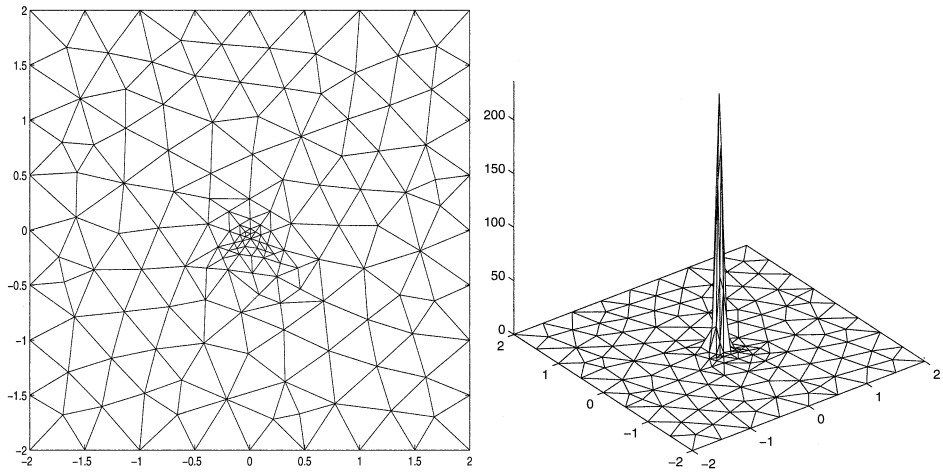

FIG. 6.3. *Example* 3: *Triangulation and solution at* $t = 2 \cdot 10^{-6}$ *on level* 3.

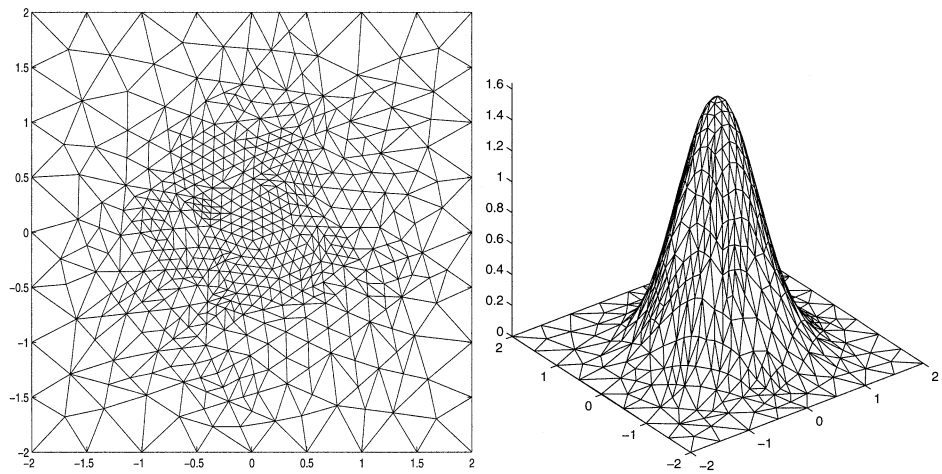

FIG. 6.4. *Example* 3: *Triangulation and solution at* $t = 0.1$ *on level* 3.

corresponding triangulation after a time period of 0.1.

Example 4. The next example is a test for the adaptive refinement strategy for the space discretization, which is a modified version of [1, Ex. 2] (cf. also [3, Ex. 8.4]). Roughly speaking, this constitutes the rotation of p in Example 3 on a radius of $r = 1/8$. Note that the right-hand side f depends on t in Example 3 and that our theory from [12] does not cover this case. However, the method still can be applied by using $f(t), f(t+\tau/2)$, and $f(t+\tau)$ for an approximate evaluation of the functional (2.3). The domain is $\Omega = [-1,1]^2$. Again, homogeneous Dirichlet boundary conditions are prescribed on $\partial\Omega$ and the initial condition is

$$p(0,x) = 0.8 \exp\left(-80 \left\| x - \begin{pmatrix} 1/2 \\ 0 \end{pmatrix} \right\|^2\right).$$

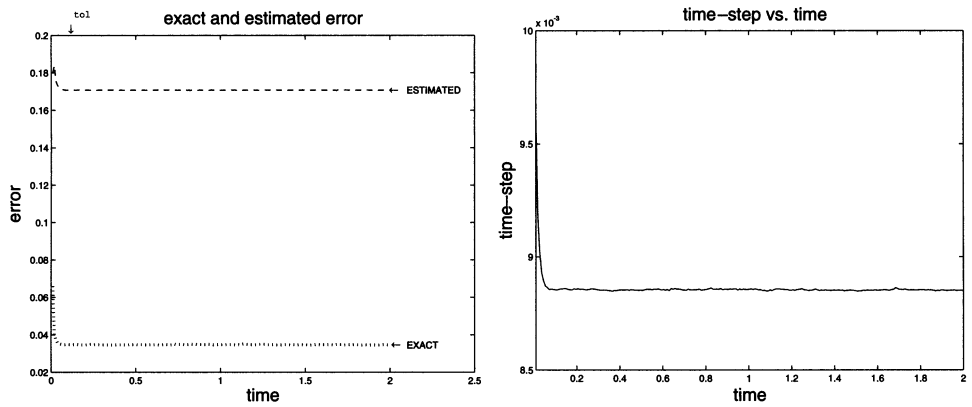

FIG. 6.5. *Example* 4: *Quality of the error estimator and time steps for* `tol` $= 0.2$.

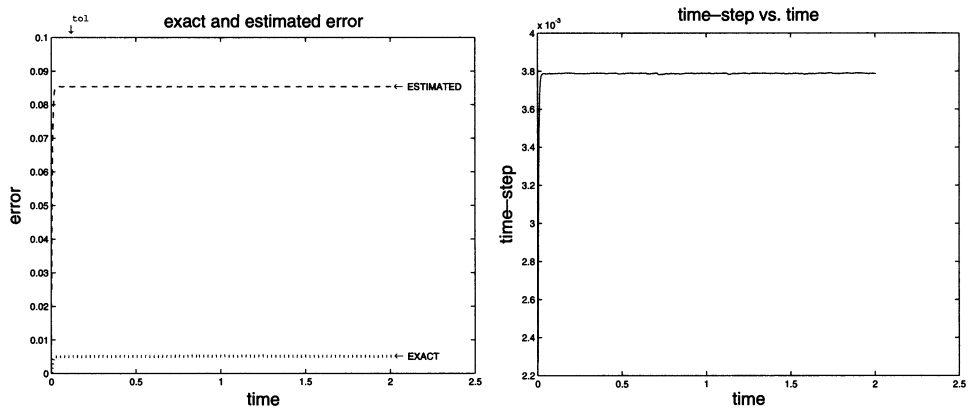

FIG. 6.6. *Example* 4: *Quality of the error estimator and time steps for* `tol` $= 0.1$.

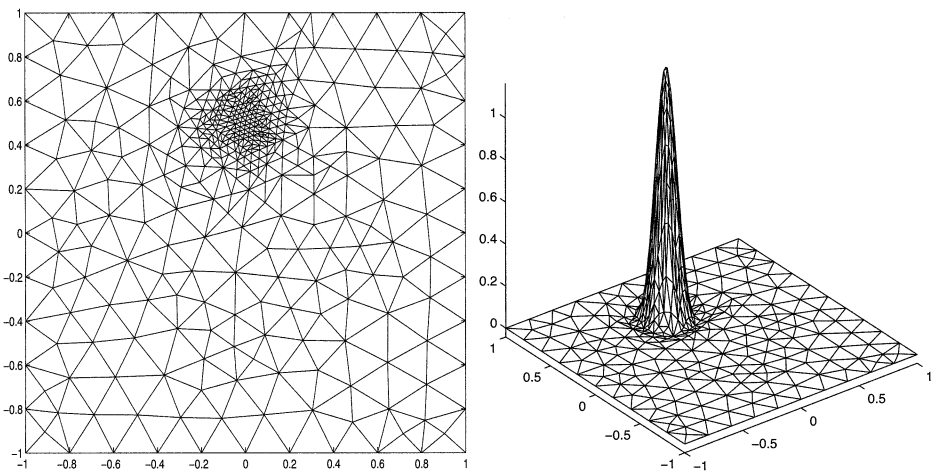

FIG. 6.7. *Example* 4: *Triangulation and solution at* $t = 0.505$ *on level* 3.

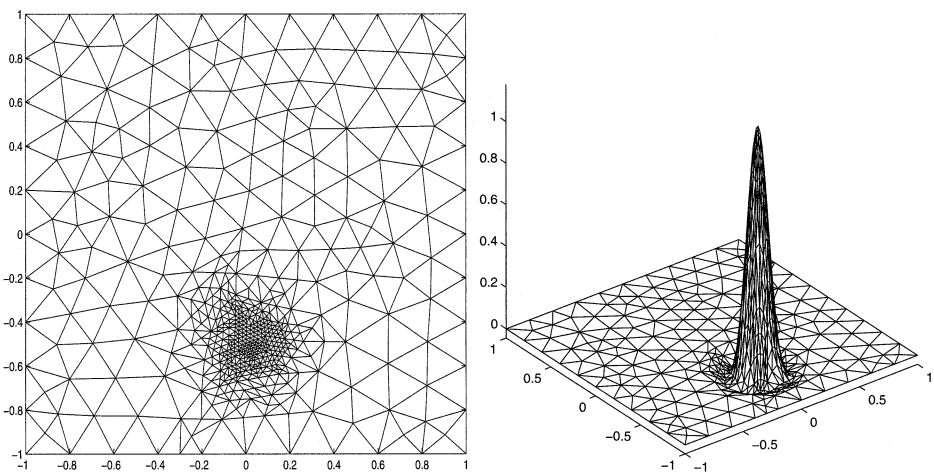

FIG. 6.8. *Example* 4: *Triangulation and solution at* $t = 1.508$ *on level* 3.

The exact solution of this problem is given by

$$p(t,x) = 0.8 \exp\left(-80 \left\| x - \frac{1}{2} \begin{pmatrix} \cos(\pi t) \\ \sin(\pi t) \end{pmatrix} \right\|^2\right)$$

and the right-hand side $f(t,x) := \partial_t p - \Delta p$. Again, for $t \in (0,2]$ the boundary condition is approximately fulfilled. The program was started with a time-step of $\tau_0 = 0.01$. We compute until $T = 2$, when a full circulation is reached. Figure 6.5 shows the results with $\texttt{tol} = 0.2$, while the tolerance was set to $\texttt{tol} = 0.1$ in Figure 6.6. In Figures 6.7 and 6.8, the computed solution is shown at two different times.

Acknowledgments. We are thankful to the anonymous referees for helpful comments and suggestions.

REFERENCES

[1] S. ADJERID AND J. E. FLAHERTY, *A local refinement finite-element method for two-dimensional parabolic systems*, SIAM J. Sci. Stat. Comput., 9 (1988), pp. 792–811.
[2] R. E. BANK, *Hierarchical bases and the finite element method*, Acta Numer., 5 (1996), pp. 1–43.
[3] F. A. BORNEMANN, *An Adaptive Multilevel Approach to Parabolic Equations in Two Space Dimensions*, Technical report TR 91-7, Konrad-Zuse-Zentrum für Informationstechnik, Berlin, 1991.
[4] D. BRAESS, *Finite Elements*, Cambridge University Press, Cambridge, 1997.
[5] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer, New York, 1994.
[6] P. DEUFLHARD AND F. BORNEMANN, *Numerische Mathematik* II, De Gruyter, Berlin, 1994.
[7] K. ERIKSSON AND C. JOHNSON, *Adaptive finite element methods for parabolic problems* I: *A linear model problem*, SIAM J. Numer. Anal., 28 (1991), pp. 43–77.
[8] E. HAIRER, S. P. NORSETT, AND G. WANNER, *Solving Ordinary Differential Equations* I, 2nd ed., Springer, Berlin, 1993.
[9] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations* II, 2nd ed., Springer, Berlin, 1996.
[10] J. LANG, *Adaptive Multilevel Solution of Nonlinear Parabolic PDE Systems*, Springer, Berlin, 2000.

[11] M. MAJIDI, *Least-Squares-Galerkin-Verfahren für nichtlineare parabolische Anfangs-Randwertprobleme*, Ph.D. thesis, Institüt für Angewandte Mathematik, Universität Hannover, Germany, 2000.

[12] M. MAJIDI AND G. STARKE, *Least-squares Galerkin methods for parabolic problems* I: *Semi-discretization in time*, SIAM J. Numer. Anal., 39 (2001), pp. 1302–1323.

[13] G. STARKE, *Least-squares mixed finite element solution of variably saturated subsurface flow problems*, SIAM J. Sci. Comput., 21 (2000), pp. 1869–1885.

[14] V. THOMÉE, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin, 1997.

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

[16] B. I. WOHLMUTH AND R. H. W. HOPPE, *A comparison of a posteriori error estimators for mixed finite element discretizations by Raviart-Thomas elements*, Math. Comp., 68 (1999), pp. 1347–1378.