

# SHARPLY LOCAL POINTWISE A POSTERIORI ERROR ESTIMATES FOR PARABOLIC PROBLEMS

ALAN DEMLOW AND CHARALAMBOS MAKRIDAKIS

ABSTRACT. We prove pointwise a posteriori error estimates for semi- and fully-discrete finite element methods for approximating the solution  $u$  to a parabolic model problem. Our estimates may be used to bound the finite element error  $\|u - u_h\|_{L^\infty(D)}$ , where  $D$  is an arbitrary subset of the space-time domain of definition of the given PDE. In contrast to standard global error estimates, these estimators de-emphasize spatial error contributions from space-time regions removed from  $D$ . Our results are valid on arbitrary shape-regular simplicial meshes which may change in time, and also provide insight into the contribution of mesh change to local errors. When implemented in an adaptive method, these estimates require only enough spatial mesh refinement away from  $D$  in order to ensure that local solution quality is not polluted by global effects.

## 1. INTRODUCTION

We consider finite element approximations to the model problem

$$(1.1) \quad \begin{aligned} u_t - \Delta u + u &= f \text{ in } \Omega \times (0, T], \\ \frac{\partial u}{\partial n} &= 0 \text{ on } \partial\Omega \times [0, T], \\ u(x, 0) &= u_0(x). \end{aligned}$$

Here  $\Omega \subset \mathbb{R}^N$ ,  $N \geq 2$ , is a bounded domain with smooth boundary  $\partial\Omega$ , and  $f$  and  $u_0$  are assumed to be sufficiently smooth.

Our goal in this work is to prove *sharply local* (sometimes also called *localized* in the literature) pointwise a posteriori error estimates for finite element approximations  $u_h$  to  $u$ . Many applications only require knowledge of  $u$  on some subset  $D$  of  $\Omega \times (0, T]$ . As a simple physical example, we may consider a thermal evolution problem in which one desires to monitor the temperature evolution at a single point (i.e., on  $D = x_0 \times [0, T]$  for a given  $x_0 \in \Omega$ ) or to calculate the temperature distribution accurately only at the final time ( $D = \Omega \times \{T\}$ ). More broadly, there has been much recent interest in so-called goal-oriented error estimation and adaptivity in which a posteriori error estimates and adaptive finite element methods are designed to compute physically relevant “quantities of interest” which take the form of functionals  $J(u)$  of  $u$ . Most examples of such functionals given in the literature are only locally dependent (cf. the survey articles [BR01] and [GS02]).

---

2000 *Mathematics Subject Classification.* 65N30.

*Key words and phrases.* parabolic partial differential equations, finite element methods, adaptive methods, a posteriori error estimates, pointwise error estimates, maximum norm error estimates, localized error estimates, local error estimates.

The first author was supported in part by National Science Foundation grant DMS-0713770.

When the desired information from a computation depends on  $u$  only in some subset  $D$ , it is desirable to place just enough computational resources (mesh elements and time steps) in regions away from  $D$  to ensure that solution quality in these regions does not *pollute* solution quality in the target set  $D$ . In the context of adaptive methods, achieving this goal requires a posteriori error estimates which control but do not overemphasize pollution effects. Broadly speaking, there are two main options for estimating and controlling pollution effects a posteriori. The first is to approximately compute a “generalized Green’s function” which solves a dual problem encoding information about the quantity of interest; we refer again to [BR01] and [GS02] for examples and details. Note that if in the current context  $D$  is a single point  $(x_0, t_0)$ , then we have  $J(u) = u(x_0, t_0)$ , and the “generalized Green’s function” is the *actual* Green’s function. Computational experience shows that this “dual-weighted residual” approach provides quite accurate information about the error in many situations, though such bounds are also often not reliable on coarse meshes. In addition, dual-weighted-residual methods are most effective when  $J$  is a linear functional, which excludes the case of rigorous local norm error estimation considered here. Finally, the added expense of computing a dual solution is reasonable in the context of elliptic problems, but may become less so in the the context of parabolic problems because of the added computational overhead of the time discretization.

The second main option for bounding pollution errors is to prove explicit a posteriori error estimates for a local norm of the error (or for the error at a single point). In order to make clear the intuition behind the two main types of such explicit local estimates, we briefly state two types of local (pointwise) a priori estimates for elliptic problems. Let then  $v$  and  $v_h \in S_k$  be the continuous and finite element solutions to a second-order elliptic problem. Here  $S_k$  is a standard Lagrange finite element space of degree  $k$  defined on a quasi-uniform mesh of diameter  $h$ . Following the classical local energy estimates of [NS74], Schatz and Wahlbin proved the following local pointwise bound in [SW95]. Let  $x_0 \in \Omega$ . If appropriate assumptions hold, then for any  $d \geq Ch$  we have

$$(1.2) \quad \begin{aligned} |(v - v_h)(x_0)| &\leq C\ell_h \left( \min_{\chi \in S_k} \|v - \chi\|_{L_\infty(B_d(x_0))} + d^{-k+1} \|v - v_h\|_{W_\infty^{-k+1}(\Omega)} \right) \\ &\leq C\ell_h (h^{k+1} |v|_{W_\infty^{k+1}(B_d(x_0))} + d^{-s} h^{2k} |v|_{W_\infty^{k+1}(\Omega)}). \end{aligned}$$

Here  $\|\cdot\|_{W_\infty^{-k+1}(\omega)}$  is a negative norm and  $\ell_h$  is a logarithmic factor. Thus the pointwise error is bounded by a “local approximation term”  $\min_{\chi \in S_k} \|v - \chi\|_{L_\infty(B_d(x_0))}$  of order  $h^{k+1}$  and a “global pollution term”  $d^{-k+1} \|v - v_h\|_{W_\infty^{-k+1}(\Omega)}$  which under ideal conditions has maximum order  $d^{-k+1} h^{2k}$  and which measures the effects of global solution properties upon the the quality of the approximation  $v_h$  to  $v$  at the point  $x_0$ . In [Sch98], Schatz introduced a form of sharply local pointwise a priori estimates for elliptic problems. Given  $x_0 \in \Omega$ , let  $\sigma_{x_0}(y) = \frac{h}{h+|x_0-y|}$ . Schatz proved in [Sch98] that

$$(1.3) \quad \begin{aligned} |(v - v_h)(x_0)| &\leq C\ell_h \min_{\chi \in S_k} \|\sigma_{x_0}^{k-1}(v - \chi)\|_{L_\infty(\Omega)} \\ &\leq Ch^{k+1} \ell_h \|\sigma_{x_0}^{k-1} D^{k+1} v\|_{L_\infty(\Omega)}. \end{aligned}$$

If we let  $d$  be as in (1.2) and then note that  $\sigma_{x_0}(y) \leq 1$  for  $|y - x_0| \leq d$  and  $\sigma_{x_0}(y) \leq \frac{h}{d}$  for  $|y - x_0| \geq d$ , we immediately obtain from (1.3) that  $|(v - v_h)(x_0)| \leq$

$C\ell_h(h^{k+1}\|v\|_{W_\infty^{k+1}(B_d(x_0))} + d^{-s}h^{k+1+s}\|v\|_{W_\infty^s(\Omega)})$ , precisely as in (1.2). Note, however, that  $\sigma_{x_0}(y)$  decreases smoothly with respect to the distance from  $y$  to  $x_0$ , whereas in (1.2) the distance to  $x_0$  is only taken into account by the fixed parameter  $d$ . Thus (1.3) and (1.2) both measure the degree to which global properties of  $v$  affect solution quality at  $x_0$ , but (1.3) measures more sharply the way in which the *distance* that solution features of  $v$  lie from  $x_0$  affects solution quality at  $x_0$ .

Local a posteriori estimates which imitate local a priori estimates such as (1.2) by splitting the local error explicitly into a local residual term that is of the same order as the local error norm being bounded, and a global pollution term that is heuristically of higher order and must be bounded separately have also appeared in the literature in the context of elliptic problems. The first such estimates to our knowledge appeared in [XZ00], which included both local energy and local  $L_\infty$  estimates; see also [LN03] for local energy estimates which treat effects arising from corner singularities on polygonal domains and [Dem07] for local  $W_\infty^1$  estimates analogous to (1.2). Sharply local pointwise a posteriori gradient bounds of residual type were proved for elliptic problems in [Dem06]. These estimates are valid on general shape-regular simplicial meshes (which may be highly graded) and employ a version of the weight  $\sigma$  in which the mesh parameter  $h$  reflects the local mesh size. In the parabolic context, sharply local pointwise a priori estimates were proved in [Ley04a]. Assume now that  $u_h \in S_k$  is the semidiscrete finite element approximation to the solution  $u$  to (1.1) with discrete initial data  $u_{h,0}$ . Given a fixed point  $(x_0, t_0) \in \Omega \times (0, T]$ , let  $\sigma_{x_0, t_0}(x, t) = \frac{h}{h + |x - x_0| + \sqrt{t_0 - t}}$ . It was proved in [Ley04a] that if  $0 \leq s \leq k - 1$ ,

$$(1.4) \quad |(u - u_h)(x_0, t_0)| \leq C\ell_{h,s} \min_{\chi \in C([0, t_0]; S_k)} \|\sigma_{x_0, t_0}^s(u - \chi)\|_{L_\infty(\Omega \times [0, T])},$$

where  $\ell_{h,s}$  is a logarithmic factor.

In this work we prove sharply local pointwise a posteriori error estimates which may be viewed as a posteriori counterparts to the sharply local a priori error estimates proved in [Ley04a], and also as parabolic counterparts to the similar sharply local pointwise a posteriori gradient estimates proved for elliptic problems in [Dem06]. In the absence of a time discretization, the local behavior of finite element methods for parabolic problems is similar to the local behavior of FEM for elliptic problems (see for example the local energy estimates in [Ley04a] and [STW98]). The effects of time discretizations and mesh change on local error behavior is less well understood, however, and is a major focus of the current work. Here we study the backward Euler time discretization as a simple model case. As we show more precisely below, the time discretization has little effect upon the localization of the error that is present in spatial semidiscretizations. The effects of changing spatial meshes are more subtle, but as we show below, these mesh change effects also possess a localization property similar to that seen in the spatial error. We also note here that [LW08] employs sharply local a priori results similar to (1.4) to prove asymptotic results concerning averaging estimators for fully discrete schemes for parabolic problems. In these results the time-stepping error must be strictly smaller than the space discretization error in order to obtain the desired error estimate. Here we seek instead to *adaptively balance* time and sharply local space contributions to the overall error.

We next briefly describe our results for semidiscrete finite element approximations. Let  $\mathcal{T}_h$  be a shape-regular simplicial decomposition of  $\Omega$ , let  $h_K$  be the

diameter of  $K \in \mathcal{T}_h$ , and let  $S_k$  be a standard Lagrange finite element space of polynomial degree  $k$  on  $\mathcal{T}_h$ . In addition, let  $u_h$  be a semidiscrete finite element approximation to  $u$  with initial data  $u_h(0) = P_h u_0$ , where  $P_h$  is the  $L_2$  projection onto  $S_k$ . Then for a point  $(x_0, t_0) \in \Omega \times (0, T]$  with  $x_0 \in K_0 \in \mathcal{T}_h$ ,

$$(1.5) \quad |(u - u_h)(x_0, t_0)| \leq C \left[ \|\sigma_{x_0, t_0}^{k+1}(0)(u_0 - P_h u_0)\|_{L_\infty(\Omega)} + (1 + \ell_h) \sup_{0 < t < t_0} \max_{K \in \mathcal{T}_h} (\sigma_{x_0, t_0}^{k-1}(K, t) \eta_{\infty, t}(K)) \right].$$

Here  $\eta_{\infty, t}(K)$  is an  $L_\infty$ -type residual parabolic error indicator defined on the element  $K$ ,  $\ell_h$  is a logarithmic factor which we define more precisely later, and

$$(1.6) \quad \sigma_{x_0, t_0}(K, t) = \frac{h_K}{h_K + \text{dist}(K_0, K) + \sqrt{t_0 - t}}.$$

Note that if  $(K, t)$  is removed from  $(K_0, t_0)$  in either space or time, then  $\sigma_{x_0, t_0}(K, t)$  is equivalent to the local mesh size. Thus initial data approximation effects in the term  $\|\sigma_{x_0, t_0}^{k+1}(0)(u_0 - P_h u_0)\|_{L_\infty(\Omega)}$  above are de-emphasized by a factor of approximately  $h^{k+1}$  so long as  $t_0$  is not close to 0. The weight  $\sigma$  in the second line of (1.5) also de-emphasizes error contributions from regions away from  $(K_0, t_0)$  by positive powers of  $h$  if  $k-1 > 0$ , i.e., so long as we are using quadratic or higher-order finite element spaces. More precisely, if the ‘‘parabolic distance’’  $\text{dist}(x, K) + \sqrt{t_0 - t}$  from the target point  $(x_0, t_0)$  to some element  $K$  at time  $t$  is of size  $d$ , then the residual contribution from the element  $K$  at  $t$  to the error at  $(x_0, t_0)$  will be given by  $\sigma_{x_0, t_0}^{k-1} \eta_{\infty, t}(K) \leq d^{-k+1} h_K^{k-1} \eta_{\infty, t}(K)$ . Heuristically,  $h_K^{k-1} \eta_{\infty, t}(K)$  can be thought of as a  $W_\infty^{-k+1}$ -type residual indicator that is bounded by  $\|u - u_h\|_{W_\infty^{-k+1}(K)}$ . Thus (1.5) computably bounds the error at the target point  $(x_0, t_0)$ , but without requiring that the  $L_\infty$  error be resolved globally (unless  $k = 1$ ). Note that no factor  $\sigma$  is available in the second line of (1.5) in the practically important case of piecewise linear finite element spaces, as expected; cf. [Dem04] for a counterexample in the context of a priori estimates for elliptic problems. Finally, it is trivial to take the supremum of both sides of (1.5) in order to bound  $\|u - u_h\|_{L_\infty(D)}$  for arbitrary subsets  $D$  of  $\Omega \times (0, T]$ ; cf. §3.3.

In §4 we prove an estimate similar to (1.5) for backward Euler finite element approximations to (1.1). Though their statement is somewhat more complicated, our results for fully discrete approximations differ from (1.5) mainly in that in addition to the *initial data estimator* and *spatial estimator* already present in (1.5), they include *mesh change and mesh coarsening estimators* as well as a *time discretization estimator*. Let  $\{u_h^n\}$  be a sequence of backward Euler-finite element approximations to  $u$  defined on a set of time nodes  $\{t_n\}$ , meshes  $\mathcal{T}_n$ , and degree- $k$  finite element spaces  $S^n$ , where  $0 \leq i \leq M$ . Also,  $P_h^n$  is the  $L_2$  projection onto  $S^n$ . Though its final expression is somewhat more complicated, the time estimator has principal part

$$\max_{1 \leq i \leq M} [\|u_h^n - P_h^n u_h^{n-1}\|_{L_\infty(\Omega)} + \|f - f^n\|_{L_\infty(\Omega \times [t_{n-1}, t_n])}],$$

which is of optimal order since  $u_h^n \rightarrow P_h^n u_h^{n-1}$  with order 1 as  $t_n \rightarrow t_{n-1}$  even if  $\mathcal{T}_n \neq \mathcal{T}_{n-1}$ . We emphasize that the time estimator above is *only* of optimal order in time for the backward Euler method, with no decay in error dependence as one moves away from the target time (cf. [Ley04b] for a counterexample in the context of a priori error estimates). However, the spatial estimator continues

to de-emphasize error contributions away from the target region. In addition to the spatial, time, and initial data estimators, mesh change estimators are present which are nonzero only when the mesh is refined or coarsened in passing between time steps. These mesh change estimators (whose form is given precisely later) also include weights whose form is similar to (1.6), so that the effects of mesh changes occurring away from the target region  $D$  are also de-emphasized. For problems with smooth data, this indicates that essentially uniform timesteps are necessary in order to bound the maximum error on a target region  $D$ . The *spatial* meshes, however, may be refined in regions which are close to  $D$  in the sense of the parabolic distance  $\min_{(x_0, t_0) \in D} |x - x_0| + \sqrt{|t_0 - t|}$ , but coarser in regions removed from the target point so long as  $k \geq 2$ . Numerical experiments in §5 confirm this heuristic.

We next comment on our assumptions. First, we assume that  $\partial\Omega$  is smooth, whereas previous a posteriori pointwise estimates for elliptic problems have generally been proved on polyhedral domains. We avoid polyhedral domains here because we wish to focus on the interaction of the spatial localization implied by (1.5) with the time discretization, and corner singularities arising on polyhedral domains add a great deal of technical overhead that would lengthen and cloud our presentation (cf. [Noc95], [Bom00]). Polyhedral domains are of definite practical interest, however, and global pointwise estimates on such domains are studied in the companion work [DLM09] using somewhat different techniques that largely avoid direct engagement with technicalities surround corner singularities (see Remark 4.2 below). It also is important to emphasize the generality of our assumptions concerning the grids used in both space and time discretizations here. In the context of fully discrete problems, we essentially only assume uniform shape regularity of the simplicial grids used at each time step as well as a weak compatibility assumption between grids. (Curved boundary elements must satisfy approximability conditions fulfilled by shape-regular simplicial elements; we give a brief discussion of procedures for ensuring that approximation theory results hold on curved domains below.) Thus highly graded and unstructured grids that change in an arbitrary way from time step to time step may be employed, though severe mesh changes are penalized in our estimators. In addition, no restriction whatsoever is placed on the time nodes. Modulo issues arising from the curved boundary, our results are thus truly a posteriori in that computable estimators rather than assumptions are used to control time step sizes and changes in the mesh.

The paper is organized as follows. In §2 we state some common preliminary results and definitions. In §3 we state and prove estimates for semidiscrete finite element approximations, and in §4 we state and prove analogous results for fully discrete methods. In §5 we present numerical experiments that illustrate properties of the fully discrete estimators and also confirm that pollution occurs if the mesh is not refined sufficiently away from the target set  $D$ .

## 2. PRELIMINARIES

In this section we provide a number of preliminaries.

**2.1. Analytical preliminaries.** In order to simplify notation, we first define the bilinear form

$$A(u, v) = \int_{\Omega} \nabla u \nabla v \, dx + \int_{\Omega} uv \, dx.$$

We also denote by  $(u, v)$  the  $L_2$  inner product of  $u$  and  $v$  over  $\Omega$ . We shall depress the dependence of functions upon  $t$  when doing so should cause no confusion.

Our analysis relies heavily on properties of the heat kernel for the problem (1.1). We sum up the necessary results in the following lemma.

**Lemma 2.1.** *There exists a Green's function  $G(x, t; y, s)$  for the problem (1.1). That is, there exists a kernel  $G$  such that for  $(x_0, t_0) \in \Omega \times (0, T]$ ,*

$$(2.1) \quad u(x_0, t_0) = \int_{\Omega} G(x_0, t_0; y, 0) u_0(y) \, dy + \int_0^{t_0} \int_{\Omega} G(x_0, t_0; y, s) f(y, s) \, dy \, ds$$

is a weak solution of (1.1). In addition, for  $s < t$  and  $j \geq 0$ , the spatial derivatives of  $G$  satisfy the pointwise bound

$$(2.2) \quad |D_x^j G(x, t; y, s)| \leq C(|x - y| + \sqrt{t - s})^{-N-j} e^{-c \frac{|x-y|^2}{t-s}}.$$

Here  $C$  depends on  $T$  and  $\Omega$ .

*Proof:* See [EI70]. □

As a consequence of Lemma 2.1, we also have the relationship

$$(2.3) \quad \begin{aligned} v(x_0, t_0) &= \int_{\Omega} G(x_0, t_0; y, 0) v(y, 0) \, dy \\ &+ \int_0^{t_0} [(v_t, G(x_0, t_0)) + A(v, G(x_0, t_0))] \, dt, \end{aligned}$$

where we use the abbreviation  $G(x_0, t_0) = G(x_0, t_0, \cdot, \cdot)$ . Note that (2.3) is valid for any  $v$  possessing sufficient smoothness, whereas (2.1) requires that  $u$  solves (1.1).

**2.2. Finite element preliminaries.** In this subsection we collect a few preliminaries concerning Lagrange finite element spaces defined on simplicial grids. Further properties of such spaces will be discussed in subsequent sections because some notation and technicalities differ depending upon whether a single fixed grid is employed throughout a calculation as in Section 3 (concerning semidiscrete discretizations) or a time-varying family of spaces is used as in Section 4 (concerning backward Euler discretizations).

Assume that  $\mathcal{T}_h$  is a simplicial decomposition of  $\Omega$  that is shape-regular (that is, all elements have uniformly bounded aspect ratio), and also is face-to-face in the sense that the intersection of any two simplices must be a complete edge, face, or vertex. Elements bordering  $\partial\Omega$  may be curved. Let  $S_k$  then be a standard Lagrange finite element space consisting of the continuous functions that are piecewise polynomials of degree  $k$  on  $\mathcal{T}_h$ ; for curved elements abutting  $\partial\Omega$  we also allow parametric finite element basis functions (that is, basis functions that are obtained by a non-affine mapping of polynomials from a reference domain). We denote by  $h_K$  the diameter of the element  $K \in \mathcal{T}_h$ .

We assume the existence of an interpolant  $I_h : L_1(\Omega) \rightarrow S_k$  with the following properties. Given  $K \in \mathcal{T}_h$ , let  $\omega_K$  be the patch of elements sharing a vertex with  $K$ . We then require that  $I_h$  satisfy for  $K \in \mathcal{T}_h$

$$(2.4) \quad \|I_h u\|_{L_1(K)} \leq C \|u\|_{L_1(\omega_K)},$$

and

$$(2.5) \quad \|I_h u\|_{W_1^1(K)} \leq C \|u\|_{W_1^1(\omega_K)}.$$

In addition, we require that for  $1 \leq j \leq k + 1$ ,

$$(2.6) \quad \|u - I_h u\|_{L_1(K)} \leq Ch_K^j |u|_{W_1^j(\omega_K)}$$

and

$$(2.7) \quad \|\nabla(u - I_h u)\|_{L_1(K)} \leq Ch_K^{j-1} |u|_{W_1^j(\omega_K)}.$$

On shape-regular simplicial grids having straight sides, the above are standard or easily-derived properties of for example a Scott-Zhang type interpolant that is modified to be  $L_1$ -stable (cf. the comments on page 491 of [SZ90]) or an interpolant of Clément type (cf. [Clé75]).

The construction of appropriate finite element spaces on curved domains is somewhat more complex, and we briefly discuss two options. The first is to employ parametric finite element spaces along with an appropriately defined interpolant. Such spaces and a corresponding interpolant  $I_h$  of Clément type were constructed for both two- and three-dimensional domains with smooth boundary in [Ber89]. For this interpolant, (2.4) and (2.5) may be trivially obtained from Theorem 4.1 of [Ber89] using the triangle inequality. The same theorem contains (2.6) and (2.7), but with full norms on the right hand side instead of seminorms. The presence of full norms in these inequalities would add a slight amount of technical overhead to our proofs but make no difference in our final results so we do not consider this situation explicitly. [Ber89] also reviews techniques of [Len86] (two and three dimensions) and [Zlá73] and [Sco73] (two dimensions) for constructing appropriate curved meshes and associated parametric mappings. Once meshes are constructed using one of these procedures, one must check that the associated parametric mappings do not differ too much from affine mappings (cf. (2.2) in [Ber89]). As noted in Corollary 6.1 of [Ber89], this condition is met if the mesh resolves the boundary sufficiently. The necessary conditions (which will ensure that (2.4)-(2.7) hold) can easily be checked automatically in software.

An alternative method for constructing finite element spaces and an appropriate interpolant  $I_h$  are contained in [Dem]. The finite element spaces  $S_k$  in the latter work have strictly piecewise polynomial (not parametric) basis functions. In two space dimensions, ensuring that (2.4) through (2.7) hold reduces to checking a set of simple geometric assumptions that can easily be done automatically in software. This framework has the advantage of allowing for the construction of nested discrete spaces, which is not possible when parametric spaces are used. However, the results of [Dem] are for practical purposes largely limited to space dimension two and do not easily extend from the natural boundary conditions that we consider here to essential boundary conditions.

We also note that the mesh construction approach of [Len86] and the two-dimensional meshing procedure of [Dem] easily lend themselves to adaptive mesh refinement via bisection. Both employ a standard (straight-sided) simplicial decomposition  $\tilde{\mathcal{T}}_0$  of a polyhedral approximation  $\Omega_{0,h}$  to  $\Omega$  in order to construct a curved simplicial decomposition  $\mathcal{T}_0$  of  $\Omega$ . Refinement can be carried out by bisecting elements in  $\tilde{\mathcal{T}}_0$ , projecting newly created nodes on  $\partial\Omega_{0,h}$  onto  $\partial\Omega$  in order to create a new triangulation  $\tilde{\mathcal{T}}_1$  of a polyhedral approximation  $\Omega_{1,h}$  to  $\Omega$ , and then repeating the procedures of [Len86] or [Dem] to obtain a new curved triangulation  $\mathcal{T}_1$  of  $\Omega$ . The sufficient conditions to guarantee that (2.4) through (2.7) hold cited in the above two paragraphs can then be rechecked on  $\mathcal{T}_1$ . Thus using either [Ber89] or [Dem] yields an adaptive mesh refinement procedure with checkable conditions

for adaptive boundary resolution. We also refer to [DR98] for a different treatment of a posteriori control of boundary resolution.

In addition to defining the patch  $\omega_K$  as above, we also let  $\omega'_K$  be the patch of elements sharing a vertex with any element in  $\omega_K$ , and  $\omega''_K$  be the patch of elements sharing a vertex with any element in  $\omega'_K$ . The shape-regularity of  $\mathcal{T}_h$  implies that the diameters of all elements in  $\omega''_K$  are equivalent to the diameter of  $K$ .

### 3. THE SEMIDISCRETE CASE

**3.1. Finite element approximation.** We assume that  $\mathcal{T}_h$  is a simplicial decomposition of  $\Omega$ , that  $S_k$  is a Lagrange finite element space of degree  $k$  on  $\mathcal{T}_h$ , and that an interpolant  $I_h$  satisfying (2.4) through (2.7) exists, as in §2.2. The semidiscrete approximation  $u_h \in C([0, T], S)$  of  $u$  then satisfies

$$\begin{aligned} (u_{h,t}, v_h) + A(u_h, v_h) &= (f, v_h), \quad v_h \in S_k \text{ and } t \in (0, T], \\ u_h(0) &= P_h u_0. \end{aligned}$$

Here  $P_h : L_2(\Omega) \rightarrow S_k$  is the  $L_2$  projection onto  $S_k$ . In the sequel we shall use the fact that  $u$  and  $u_h$  are Galerkin orthogonal, that is,

$$(3.1) \quad (u_t - u_{h,t}, v_h) + A(u - u_h, v_h) = 0, \quad v_h \in S_k, 0 < t \leq T.$$

**3.2. Sharply local a posteriori error estimates for the semidiscrete problem.** Given a fixed point  $x_0 \in \Omega$  and time  $t_0 \in [0, T]$ , we define the weight

$$\sigma_{x_0, t_0}(K, t) = \frac{h_K}{h_K + \text{dist}(x_0, K) + \sqrt{t_0 - t}},$$

where  $K \in \mathcal{T}_h$  and  $0 \leq t \leq t_0$ . We shall also sometimes view  $\sigma$  as a function of  $(x, t) \in \Omega \times [0, T]$  instead of  $(K, t) \in \mathcal{T}_h \times [0, T]$ , that is,

$$\sigma_{x_0, t_0}(x, t) = \frac{h_K}{h_K + \text{dist}(x_0, K) + \sqrt{t_0 - t}}, \quad x \in K \in \mathcal{T}_h.$$

For fixed  $t$ ,  $\sigma_{x_0, t_0}(\cdot, t)$  is thus piecewise constant on the mesh. For  $K \in \mathcal{T}_h$  we also define the  $L_\infty$ -type parabolic residual indicator

$$\eta_{\infty, t}(K) = h_K^2 \| (f + \Delta u_h - u_h - u_{h,t})(t) \|_{L_\infty(K)} + h_K \| \llbracket \nabla u_h(t) \rrbracket \|_{L_\infty(\partial K)}.$$

Here  $\llbracket \nabla u_h \rrbracket$  denotes the jump in normal derivative on  $\partial K \setminus \partial\Omega$  and the normal derivative itself on  $\partial K \cap \partial\Omega$ .

**Theorem 3.1.** *Let  $x_0 \in \Omega$ ,  $0 < t_0 \leq T$ , and  $\underline{h} = \min_{K \in \mathcal{T}_h} h_K$ . Assume that  $\partial\Omega$  is smooth and that the assumptions of the previous section concerning  $\mathcal{T}_h$ ,  $S_k$ , and the interpolant  $I_h$  are met. Then*

$$(3.2) \quad \begin{aligned} |(u - u_h)(x_0, t_0)| &\leq C [ \| \sigma_{x_0, t_0}(\cdot, 0)^{k+1} (u_0 - P_h u_0) \|_{L_\infty(\Omega)} \\ &\quad + (1 + \ell_{\underline{h}, t_0}) \sup_{0 < t < t_0} \max_{K \in \mathcal{T}_h} \sigma_{x_0, t_0}(K, t)^{k-1} \eta_{\infty, t}(K) ]. \end{aligned}$$

Here  $\ell_{\underline{h}, t_0} = \max(\ln \frac{t_0}{\underline{h}^2}, 0)$ , and  $C$  depends on  $T$ ,  $\Omega$ , and the shape regularity of  $\mathcal{T}_h$ .

**Remark 3.2.** It is not clear from [ËI70] precisely how  $C$  above depends upon  $T$ . For global pointwise estimates, the time dependence of constants is considered more carefully in [DLM09].

*Proof:* Let  $x_0 \in K_0$ , and also let  $Q_{t_0} = \Omega \times (0, t_0)$ . We shall use the abbreviations  $G(x, t) = G(x_0, t_0; x, t)$  and  $G(t) = G(\cdot, t)$ . Using (2.3), the Galerkin orthogonality relationship (3.1), the scaled trace inequality  $\|v\|_{L_1(\partial T)} \leq C(h_T^{-1}\|v\|_{L_1(T)} + \|\nabla v\|_{L_1(T)})$  for  $T \in \mathcal{T}_h$ , and standard techniques for proving residual-type bounds (cf. [Dem06]), we then calculate that

$$\begin{aligned}
(u - u_h)(x_0, t_0) &= (u_0 - P_h u_0, G(0)) + \int_0^{t_0} ((u - u_h)_t, G) + A(u - u_h, G) \\
&= (u_0 - P_h u_0, G(0) - I_h G(0)) \\
&\quad + \int_0^{t_0} [((u - u_h)_t, G - I_h G) + A(u - u_h, G - I_h G)] dt \\
(3.3) \quad &\leq \|\sigma_{x_0, t_0}^{k+1} (u_0 - P_h u_0)\|_{L_\infty(\Omega)} \|\sigma_{x_0, t_0}^{-k-1} (G(0) - I_h G(0))\|_{L_1(\Omega)} \\
&\quad + (\|\sigma_{x_0, t_0}^{-k+1} h^{-2} (G - I_h G)\|_{L_1(Q_{t_0})} + \|\sigma_{x_0, t_0}^{-k+1} h^{-1} \nabla (G - I_h G)\|_{L_1(Q_{t_0})}) \\
&\quad \cdot \sup_{0 < t < t_0} \max_{K \in \mathcal{T}_h} \sigma_{x_0, t_0}(K, t)^{k-1} \eta_{\infty, t}(K).
\end{aligned}$$

Thus we must establish that

$$(3.4) \quad \|\sigma_{x_0, t_0}^{-k-1} (G(0) - I_h G(0))\|_{L_1(\Omega)} \leq C,$$

$$(3.5) \quad \|\sigma_{x_0, t_0}^{-k+1} (h^{-2} |G - I_h G| + h^{-1} |\nabla (G - I_h G)|)\|_{L_1(Q_{t_0})} \leq C(1 + \ell_{\underline{h}, t_0}).$$

In order to prove (3.4), we first consider the case  $\sqrt{t_0} \leq h(x_0)$ . Employing the properties (2.4) through (2.7) of  $I_h$ , noting that  $\sigma_{x_0, t_0}$  is equivalent to 1 on  $\omega'_{K_0}$  and is equivalent on any two neighboring elements, and noting that  $h_K \leq C \text{dist}(K, K_0)$  for  $K \subset \Omega \setminus \omega_{K_0}$  (cf. [Dem06], Proposition 2.1), we find that

$$\begin{aligned}
\|\sigma_{x_0, t_0}^{-k-1} (G(0) - I_h G(0))\|_{L_1(\Omega)} &\leq C \|G(0) - I_h G(0)\|_{L_1(\omega'_{K_0})} \\
&\quad + \|\sigma_{x_0, t_0}^{-k-1} (G(0) - I_h G(0))\|_{L_1(\Omega \setminus \omega'_{K_0})}.
\end{aligned}$$

Using the  $L_1$  stability of  $I_h$  and (2.2), we find that

$$C \|G(0) - I_h G(0)\|_{L_1(\omega'_{K_0})} \leq C \|G(0)\|_{L_1(\Omega)} \leq C.$$

Using (2.2) and (2.6), noting that  $h(x) \leq C|x - x_0|$  for  $x \in \Omega \setminus \omega_{K_0}$ , and recalling that  $\sigma$  is equivalent on neighboring elements, we next compute that

$$\begin{aligned}
&\|\sigma_{x_0, t_0}^{-k-1} (G(0) - I_h G(0))\|_{L_1(\Omega \setminus \omega'_{K_0})} \\
&\leq C \|(h + |\cdot - x_0| + \sqrt{t_0})^{k+1} D^{k+1} G\|_{L_1(\Omega \setminus \omega_{K_0})} \\
&\leq C \int_{\Omega} (|x - x_0| + \sqrt{t_0})^{k+1} (|x - x_0| + \sqrt{t_0})^{-N-k-1} e^{-\frac{c|x-x_0|^2}{t_0}} dx \\
&\leq C \int_0^\infty \frac{r^{N-1}}{(r+1)^N} e^{-cr^2} dr \\
&\leq C.
\end{aligned}$$

Here we have transformed into polar coordinates with  $r = \frac{|x-x_0|}{\sqrt{t_0}}$ .

If  $\sqrt{t_0} \geq h(x_0)$ , then  $h(x) + |x - x_0| + \sqrt{t_0} \leq C(|x - x_0| + \sqrt{t_0})$  for all  $x \in \Omega$ . Using (2.2) and (2.6), we then compute that

$$\begin{aligned}
& \|\sigma_{x_0, t_0}^{-k-1}(G(0) - I_h G(0))\|_{L_1(\Omega)} \\
& \leq \|(|\cdot - x_0| + h + \sqrt{t_0})^{k+1} h^{-k-1} (G - I_h G)\|_{L_1(\Omega)} \\
& \leq \|(|\cdot - x_0| + \sqrt{t_0})^{k+1} D^{k+1} G\|_{L_1(\Omega)} \\
& \leq C \|(|\cdot - x_0| + \sqrt{t_0})^{-N} e^{-\frac{c|\cdot - x_0|^2}{t_0}}\|_{L_1(\Omega)} \\
& \leq C.
\end{aligned}$$

Next we prove (3.5). Let  $K_0 \in \mathcal{T}_h$  be an element containing  $x_0$ , and define  $Q_0 = \omega_{K_0} \times (t_0 - h(x_0)^2, t_0)$ ,  $Q_1 = \omega'_{K_0} \times (t_0 - h(x_0)^2, t_0)$ , and  $Q_2 = \omega''_{K_0} \times (t_0 - h(x_0)^2, t_0)$ . Noting that  $\sigma_{x_0, t_0}$  is equivalent to 1 on  $Q_1$  and employing the interpolation properties (2.4) through (2.7), we find that

$$\begin{aligned}
& \int_{Q_{t_0}} \sigma_{x_0, t_0}^{-k+1} (h^{-2} |G - I_h G| + h^{-1} |\nabla(G - I_h G)|) \, dx \, dt \\
& = \int_{Q_1} \sigma_{x_0, t_0}^{-k+1} (h^{-2} |G - I_h G| + h^{-1} |\nabla(G - I_h G)|) \, dx \, dt \\
(3.6) \quad & + \int_{Q_{t_0} \setminus Q_1} \sigma_{x_0, t_0}^{-k+1} (h^{-2} |G - I_h G| + h^{-1} |\nabla(G - I_h G)|) \, dx \, dt \\
& \leq Ch(x_0)^{-1} \int_{Q_2} |\nabla G| \, dx \, dt \\
& + \int_{Q_{t_0} \setminus Q_0} (h + |x - x_0| + \sqrt{t_0 - t})^{-k+1} |D^{k+1} G| \, dx \, dt.
\end{aligned}$$

In order to bound the term  $h(x_0)^{-1} \int_{Q_2} |\nabla G| \, dx \, dt$ , we set  $r = \frac{|x - x_0|}{\sqrt{t}}$  and compute

$$\begin{aligned}
& h(x_0)^{-1} \int_{Q_2} |\nabla G| \, dx \, dt \leq h(x_0)^{-1} \int_{t_0 - h(x_0)^2}^{t_0} \int_{|x - x_0| \leq Ch(x_0)} |\nabla G| \, dx \, dt \\
& \leq Ch(x_0)^{-1} \int_0^{h(x_0)^2} \int_{|x - x_0| \leq Ch(x_0)} (|x - x_0| + \sqrt{t})^{-N-1} e^{-c \frac{|x - x_0|^2}{t}} \, dx \, dt \\
& \leq Ch(x_0)^{-1} \int_0^{h(x_0)^2} \frac{1}{\sqrt{t}} \int_0^\infty \frac{r^{N-1}}{(r+1)^{N+1}} e^{-cr^2} \, dr \, dt \\
& \leq Ch(x_0)^{-1} \sqrt{t} \Big|_0^{h(x_0)^2} \\
& \leq C.
\end{aligned}$$

In order to bound the last term in (3.6), we again use (2.2) and the fact that  $h \leq C|x - x_0|$  for  $(x, t) \in Q_{t_0} \setminus Q_0$  to find

$$\begin{aligned}
& \int_{Q_{t_0} \setminus Q_0} (h + |x - x_0| + \sqrt{t_0 - t})^{-k-1} |D^{k+1}G| \, dx \, dt \\
& \leq C \int_{Q_{t_0} \setminus Q_0} (|x - x_0| + \sqrt{t_0 - t})^{-N-2} e^{-c|x-x_0|^2/(t_0-t)} \, dx \, dt \\
(3.7) \quad & \leq C \int_0^{t_0} \int_{|x-x_0| \geq ch(x_0)} (|x - x_0| + \sqrt{t_0 - t})^{-N-2} \, dx \, dt \\
& \quad + C \int_0^{t_0-h(x_0)^2} \int_{|x-x_0| \geq 0} (|x - x_0| + \sqrt{t_0 - t})^{-N-2} \, dx \, dt,
\end{aligned}$$

where the last integral is not present if  $t_0 \leq h(x_0)^2$ . Using the transformation  $r = \frac{|x-x_0|}{\sqrt{t_0-t}}$ , we calculate that

$$\begin{aligned}
& \int_0^{t_0} \int_{|x-x_0| \geq ch(x_0)} (|x - x_0| + \sqrt{t_0 - t})^{-N-2} \, dx \, dt \\
& \leq \int_0^{h(x_0)^2} \frac{1}{t} \int_{h(x_0)/\sqrt{t}}^{\infty} r^{-3} \, dr \, dt \\
& \leq \int_0^{h(x_0)^2} \frac{1}{t} \frac{t}{h(x_0)^2} \, dt \\
& \leq C.
\end{aligned}$$

We finally compute that if  $t(0) \geq h(x_0)^2$ , then

$$\begin{aligned}
(3.8) \quad C \int_0^{t_0-h(x_0)^2} \int_{|x-x_0| \geq 0} (|x - x_0| + \sqrt{t_0 - t})^{-N-2} \, dx \, dt & \leq C \int_{h(x_0)^2}^{t_0} \frac{1}{t} \, dt \\
& \leq C \ln \frac{t_0}{h(x_0)^2}.
\end{aligned}$$

Collecting the preceding estimates into (3.3) completes the proof of Theorem 3.1.  $\square$

**3.3. Sharply local estimates for arbitrary subsets.** Practically speaking, it is sometimes of interest to bound the maximum error over some arbitrary subset  $D$  of  $\Omega \times (0, T]$  and not just at a single point. This can be trivially accomplished by taking the maximum of (3.2) over  $D$ .

**Corollary 3.3.** *Let  $D \subset \Omega \times (0, T]$ , and let*

$$\sigma_D(x, t) = \sup_{(x_0, t_0) \in D} \sigma_{x_0, t_0}(x, t).$$

*Let also  $t_D = \sup_{(x, t) \in D} t$ . Then under the conditions of Theorem 3.1,*

$$\begin{aligned}
(3.9) \quad \|u - u_h\|_{L^\infty(D)} & \leq C [\|\sigma_D(\cdot, 0)^{k+1} (u_0 - P_h u_0)\|_{L^\infty(\Omega)} \\
& \quad + (1 + \ell_{\underline{h}, t_D}) \sup_{0 < t < t_D} \max_{K \in \mathcal{T}_h} \sigma_D(K, t)^{k-1} \eta_{\infty, t}(K)].
\end{aligned}$$

*Here  $\ell_{\underline{h}, t_D} = \max(\ln \frac{t_D}{\underline{h}^2}, 0)$ , and  $C$  depends on  $T$ ,  $\Omega$ , and the shape regularity of  $\mathcal{T}_h$ .*

**Remark 3.4.** Given a set  $D$ , let the parabolic distance  $\rho_D(x, t)$  to  $D$  be given by  $\rho_D(x, t) = \min_{(x_0, t_0) \in D} |x - x_0| + \sqrt{t - t_0}$ , and for  $K \subset \Omega$ , let  $\rho_D(K, t) = \min_{x \in K} \rho(x, t)$ . Then

$$(3.10) \quad \sigma_D(K, t) = \frac{h_K}{h_K + \rho_D(K, t)}.$$

If  $D$  has simple geometry, it is thus not difficult to compute  $\sigma_D$ .

**Remark 3.5.** We now comment on the effects of the weight  $\sigma_D$  in (3.9). First, if we assume that  $\inf_{(x_0, t_0) \in D} t_0 = \tilde{t} > 0$ , then we have from the definition of  $\rho_D$  above that  $\sigma_D(K, 0) \leq \frac{h_K}{\sqrt{\tilde{t}}}$ . Thus if  $u_0 \in C^{k+1}(\Omega)$  and the mesh is quasi-uniform with diameter  $h$ , then the initial data term satisfies

$$(3.11) \quad \|\sigma_D(\cdot, 0)^{k+1}(u_0 - P_h u_0)\|_{L_\infty(\Omega)} \leq Ch^{2k+2}\tilde{t}^{-(k+1)/2}.$$

That is, the effects of the error in approximating the initial data upon the finite element error at a later time are of higher order even when  $k = 1$ .

Next we recall the heuristic that multiplying a residual error indicator by a factor of  $h_K$  has the effect of lowering the derivative index of the Sobolev norm being used to measure the error by one. For example, one obtains an  $L_2$ -type residual error indicator by multiplying the  $H^1$ -type residual indicator by  $h_K$ . Note that  $\sigma_D(K, t)^{k-1} \eta_{\infty, t}(K) \leq \frac{1}{\rho_D(K, t)^{k-1}} h_T^{k-1} \eta_{\infty, t}(K)$ . Using the heuristic above,  $h_T^{k-1} \eta_{\infty, t}(K)$  is a  $W_\infty^{-k+1}$ -type residual error indicator. Thus the estimate (3.9) roughly speaking measures the pollution error in regions removed from  $D$  in a  $W_\infty^{-k+1}$  norm, which is weaker than the  $L_\infty$  norm when  $k \geq 2$  (that is, when quadratic and higher-order elements are used). Note also that in contrast to local estimates of the form (1.2), the weight  $\sigma_D$  gives a smooth transition from regions in which  $L_\infty$ -type residual indicators  $\eta_{\infty, t}(K)$  measure elementwise error contributions to regions in which  $W_\infty^{-k+1}$  residual indicators  $h_K^{-k+1} \eta_{\infty, t}(K)$  measure elementwise error contributions.

#### 4. THE FULLY DISCRETE CASE

**4.1. Fully discrete finite element approximation.** Let  $0 = t_0 < t_1 < \dots < t_J = T$ ,  $I_n = (t_{n-1}, t_n)$ , and  $\tau_n = |I_n|$ . For each  $0 \leq n \leq J$ , let  $\mathcal{T}_n$  be a shape-regular simplicial decomposition of  $\Omega$  as in §2.2. Let also  $S^n$  be a space of continuous piecewise polynomials of degree  $k$ ,  $k \geq 1$ , on  $\mathcal{T}_n$  for each  $0 \leq n \leq J$ . We assume that the meshes  $\mathcal{T}_n$  are uniformly shape-regular in the sense that the aspect ratios of all elements are uniformly bounded, and additionally assume the existence of interpolants  $I_h^n : L_1(\Omega) \rightarrow S^n$  that satisfy the approximation properties (2.4) through (2.7) with uniform constants on curved boundary elements as well as interior elements. Thus our estimates allow for spatial meshes that are highly graded and unstructured. They may also change in an arbitrary fashion from time step to time step subject to a certain compatibility constraint described in the next paragraph, though the estimators that we develop may penalize severe mesh changes. In addition, no relationship is assumed either between time steps or between the size of time steps and the mesh.

Next we state a *compatibility* assumption on the meshes  $\{\mathcal{T}_n\}_{n=0, \dots, J}$ ; cf. Appendix A of [LM06] for more details. Two simplicial decompositions  $\mathcal{T}_i$  and  $\mathcal{T}_j$  of  $\Omega$  are said to be compatible if they are derived from the same macro triangulation  $\mathcal{M}$  by an *admissible* refinement procedure which preserves shape regularity and assures

that for any elements  $K \in \mathcal{T}_i$  and  $K' \in \mathcal{T}_j$ , either  $K \cap K' = \emptyset$ ,  $K \subset K'$ , or  $K' \subset K$ . The bisection-based refinement procedure used for example in the ALBERTA finite element toolbox (cf. [SS05]) is known to be admissible.

There is a natural partial ordering of compatible triangulations, with  $\mathcal{T}_i \leq \mathcal{T}_j$  if  $\mathcal{T}_j$  is a refinement of  $\mathcal{T}_i$ . The *finest common coarsening*  $\mathcal{T}_i \wedge \mathcal{T}_j$  of  $\mathcal{T}_i$  and  $\mathcal{T}_j$  is defined in a natural way, and  $\hat{h}_{ij} = \max(h_i, h_j)$ , where  $\hat{h}_{ij}$  is the local mesh size function for  $\mathcal{T}_i \wedge \mathcal{T}_j$ . Finally, let  $S^i$  and  $S^j$  be finite element spaces of degree  $k$  on  $\mathcal{T}_i$  and  $\mathcal{T}_j$ .  $\hat{S}^{ij} = S^i \cap S^j$  is then the corresponding space of degree  $k$  on  $\mathcal{T}_i \wedge \mathcal{T}_j$ . We also let  $\hat{I}_h^{ij} : L_1(\Omega) \rightarrow \hat{S}^{ij}$  be an interpolant satisfying (2.4) through (2.7). We shall generally be concerned with finest common coarsenings of meshes  $\mathcal{T}_{n-1}$  and  $\mathcal{T}_n$ . Thus we write  $\hat{\mathcal{T}}_n = \mathcal{T}_{n-1} \wedge \mathcal{T}_n$ ,  $\hat{h}_n = \hat{h}_{n-1, n}$ ,  $\hat{S}^n = \hat{S}^{n-1, n}$ , and  $\hat{I}_h^n = \hat{I}_h^{n-1, n}$ .

Letting  $v^n(x) = v(t_n, x)$  for  $v \in C^0(\Omega \times [0, T])$ , we discretize the weak form of (1.1) using the finite element spaces  $S^n$  for the spatial discretization and the backward Euler method for the time discretization as follows. Let  $u_h^0 \in S^0$  be the  $L_2$  projection of  $u_0$  onto  $S^0$ .  $u_h^n \in S^n$ ,  $1 \leq n \leq J$ , is then defined via the recursion

$$(4.1) \quad \left( \frac{u_h^n - u_h^{n-1}}{\tau_n}, \phi_n \right) + A(u_h^n, \phi_n) = (f^n, \phi_n) \text{ for all } \phi_n \in S^n.$$

In order to obtain a time-continuous approximation to  $u$ , we first define the ‘‘hat’’ functions  $\phi_n(t)$  to be the functions that are continuous, piecewise linear with respect to the time mesh, and which satisfy  $\phi_n(t_n) = 1$ ,  $\phi_n(t_i) = 0$  for  $i \neq n$ . Then we interpolate the functions  $u_h^n$  linearly between  $t_{n-1}$  and  $t_n$ :

$$(4.2) \quad u_h(x, t) = \sum_{n=0}^J \phi_n(t) u_h^n(x).$$

Let then  $u_{h,t}^n = \frac{d}{dt} u_h|_{I_n}$ , that is,

$$u_{h,t}^n(x) = \frac{u_h^n(x) - u_h^{n-1}(x)}{\tau_n}, \quad n \geq 1.$$

Next we define a set of discrete differential operators  $A_h^n$ . For  $u \in H^1(\Omega)$ , let  $A_h^n u \in S^n$  be given by

$$(4.3) \quad A(u, v_h) = (A_h^n u, v_h), \quad v_h \in S^n.$$

Combining (4.1) and (4.3), we see that

$$(4.4) \quad f^n - u_{h,t}^n = [A_h^n u_h^n] + [f^n - P_h^n f^n] + \left[ \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n} \right],$$

where  $P_h^n$  is the  $L_2$  projection onto  $S^n$ . Rewriting (4.4) yields

$$(4.5) \quad u_h^n = (I + \tau_n A_h^n)^{-1} (P_h^n u_h^{n-1} + \tau_n P_h^n f^n).$$

Thus  $u_h^n \rightarrow P_h^n u_h^{n-1}$  with optimal order  $\tau_n$  as  $\tau_n \rightarrow 0$ . We do not expect that  $u_h^n \rightarrow u_h^{n-1}$  when  $\mathcal{T}_n$  is coarser than  $\mathcal{T}_{n-1}$ .

**4.2. Elliptic reconstruction.** It will be convenient to employ the *elliptic reconstruction* defined for semidiscrete problems in [MN03] and for fully discrete schemes in [LM06]. Analogous to our definition of  $u_h$ , we first define the reconstruction at time nodes and then interpolate linearly between them. For  $0 \leq n \leq J$ , let  $\mathcal{R}^n u_h^n$  solve

$$(4.6) \quad A(\mathcal{R}^n u_h^n, v) = (g^n, v), \quad v \in H^1(\Omega),$$

where

$$(4.7) \quad g^n = \begin{cases} A_h^0 u_h^0 + f^0 - P_h^0 f^0, & n = 0, \\ f^n - u_{h,t}^n, & n \geq 1. \end{cases}$$

Thus  $\mathcal{R}^n u_h^n$  solves a *continuous* elliptic problem with homogeneous natural boundary conditions. Note also from (4.4) that the definitions of  $g^0$  and  $g^n$ ,  $n \geq 1$ , differ only in that  $g^0$  does not include a term of the form  $(u_h^{n-1} - P_h^n u_h^{n-1})/\tau_n$ , which is not defined for  $n = 0$  and which is in addition  $L_2$ -orthogonal to  $S^n$ . We finally define the time-continuous reconstruction

$$(4.8) \quad \mathcal{R}u_h(x, t) = \sum_{n=0}^J \phi_n(t) \mathcal{R}^n u_h^n(x).$$

Error relationships between  $\mathcal{R}^n u_h^n$  and  $u_h^n$  and between  $\mathcal{R}u_h$  and  $u$  will play a fundamental role in our analysis. First note that (4.3), (4.4), (4.6), and (4.7) together yield the Galerkin orthogonality relationship

$$(4.9) \quad A(u_h^n - \mathcal{R}^n u_h^n, v_h) = 0, \quad v_h \in S^n, 0 \leq n \leq J.$$

Let  $t \in I_n$  and  $v \in H^1(\Omega)$ . Using the fact that  $\phi_{n-1}(t) + \phi_n(t) = 1$  along with (4.6) through (4.9), we compute

$$(4.10) \quad \begin{aligned} & ((u - \mathcal{R}u_h)_t, v) + A(u - \mathcal{R}u_h, v) \\ &= (f, v) - ((\mathcal{R}u_h)_t, v) - \phi_{n-1}(t)A(\mathcal{R}^{n-1}u_h^{n-1}, v) - \phi_n(t)A(\mathcal{R}^n u_h^n, v) \\ &= (f, v) - ((\mathcal{R}u_h)_t, v) - \phi^{n-1}(t)(g^n - g^{n-1}, v) \\ &\quad - (\phi_n(t) + \phi_{n-1}(t))A(\mathcal{R}^n u_h^n, v) \\ &= (f, v) - ((\mathcal{R}u_h)_t, v) - \phi^{n-1}(t)(g^n - g^{n-1}, v) - (f^n - u_{h,t}^n, v) \\ &= ((u_h - \mathcal{R}u_h)_t, v) + (f - f^n, v) + \phi_{n-1}(t)(g^n - g^{n-1}, v). \end{aligned}$$

**Remark 4.1.** Our definition of the elliptic reconstruction differs somewhat from that used in [LM06], where  $\mathcal{R}^n u_h^n$  is instead taken to solve  $A(\mathcal{R}^n u_h^n, v) = (A_h^n u_h^n, v)$ ,  $v \in H^1(\Omega)$ . Comparing (4.4) and (4.7), we see that the right-hand-side data used in our definition and that in [LM06] differ by  $[f^n - P_h^n f^n] + [(u_h^{n-1} - P_h^n u_h^{n-1})/\tau_n]$  except when  $n = 0$ , where the two definitions differ only by  $f^0 - P_h^0 f_0$ . This difference is orthogonal to  $S^n$ , so both definitions contain the same ‘‘principal’’ part.

**Remark 4.2.** The elliptic reconstruction is an efficient bookkeeping device for handling the time discretization error in our development, but not absolutely essential to our proofs. Thus we chose not to employ the reconstruction in the previous subsection concerning semidiscrete approximations. This situation stands in contrast to that in the companion work [DLM09], where the elliptic reconstruction plays an essential role. In the latter work, a posteriori error estimates are proved in  $L_\infty(\Omega \times [0, T])$ , and local error behavior is not considered. In particular, it is shown in [DLM09] that using the elliptic reconstruction technique, one can combine stability and strong stability of the *continuous parabolic* solution operator with a posteriori estimates in  $L_\infty$  for the corresponding *elliptic* finite element method in order to obtain a posteriori estimates in  $L_\infty$  for parabolic finite element methods. In other words, [DLM09] employs properties of the continuous parabolic problem and properties of the discrete elliptic problem in order to obtain  $L_\infty$  estimates,

whereas here we directly employ properties of both the discrete and continuous parabolic problems but make no use of existing elliptic estimates.

**4.3. A posteriori estimate for the fully discrete problem.** In this subsection we develop an a posteriori estimator for the fully discrete case. We begin by adapting our definition of the weight  $\sigma$  and error indicator  $\eta_{\infty,t}$  to the time-discrete case. Let  $h_n(x)$  be the local mesh size on  $\mathcal{T}_n$  and let  $h_{n,K}$  be the diameter of  $K \in \mathcal{T}_n$ . Also let  $(\tilde{x}, \tilde{t})$  be a fixed point lying in  $\Omega \times (0, T]$ . We then define the weight

$$\sigma_{n,\tilde{x},\tilde{t}}(K, t) = \frac{h_{n,K}}{h_{n,K} + \text{dist}(K, \tilde{x}) + \sqrt{\tilde{t} - t}},$$

where  $K \in \mathcal{T}_n$  and  $0 \leq t \leq \tilde{t}$ . We shall also require a similar weight defined with respect to a finest common coarsening  $\hat{\mathcal{T}}_n$  and thus define for  $\hat{K} \in \hat{\mathcal{T}}_n$

$$\hat{\sigma}_{n,\tilde{x},\tilde{t}}(\hat{K}, t) = \frac{\hat{h}_{n,\hat{K}}}{\hat{h}_{n,\hat{K}} + \text{dist}(\hat{K}, \tilde{x}) + \sqrt{\tilde{t} - t}}.$$

We occasionally write  $\sigma_n = \sigma_{n,\tilde{x},\tilde{t}}$  when the dependence is clear from the context.

Next we define several types of error indicator. We first consider a *spatial* or *elliptic* indicator. For  $K \in \mathcal{T}_n$  we define the  $L_\infty$ -type residual indicator

$$\eta_{\infty,n}(K) = h_{n,K}^2 \|g^n + \Delta u_h^n - u_h^n\|_{L_\infty(K)} + h_{n,K} \|[\nabla u_h^n]\|_{L_\infty(\partial K)}.$$

Next we define *time* indicators

$$\eta_{time,data,n} = \|f - f^n\|_{L_\infty[L_\infty(\Omega); I_n]}$$

and

$$(4.11) \quad \begin{aligned} \eta_{time,n;(\tilde{x},\tilde{t})} &= \|u_h^n - P_h^n u_h^{n-1}\|_{L_\infty(\Omega)} \\ &+ \max_{K \in \mathcal{T}_n} \sigma_n(K, t_n)^s h_{n,K}^2 \|(I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(K)}. \end{aligned}$$

Thirdly, for  $K \in \mathcal{T}_n$ ,  $1 \leq i \leq M$ , we define *coarsening* indicators

$$\begin{aligned} \eta_{coarse,n}(K) &= h_{n,K}^2 \left( \left\| \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n} \right\|_{L_\infty(K)} + \|A_h^n(P_h^n u_h^{n-1} - u_h^{n-1})\|_{L_\infty(K)} \right) \\ &+ h_{n,K} \|\nabla(P_h^n u_h^{n-1} - u_h^{n-1})\|_{L_\infty(K)} + \|P_h^n u_h^{n-1} - u_h^{n-1}\|_{L_\infty(K)}. \end{aligned}$$

Finally, for  $\hat{K} \in \hat{\mathcal{T}}_n$  and  $1 \leq i \leq M$  we define *mesh change* indicators

$$(4.12) \quad \eta_{mesh,n}(\hat{K}) = \hat{h}_{n,\hat{K}}^2 \|(A_h^n - A_h^{n-1})u_h^{n-1} - (P_h^n - P_h^{n-1})f^{n-1}\|_{L_\infty(\hat{K})}.$$

We briefly note that the spatial and the two time indicators  $\eta_{\infty,n}$  and  $\eta_{time,data,n}$  and  $\eta_{time,n}$  are of optimal order in space and time, respectively. The coarsening and mesh change indicators  $\eta_{coarse,n}$  and  $\eta_{mesh,n}$  also scale optimally in space. The coarsening indicator  $\eta_{coarse,n}$  disappears unless the mesh is coarsened in proceeding from  $\mathcal{T}_{n-1}$  to  $\mathcal{T}_n$ , and the mesh change indicator  $\eta_{mesh,n}$  also disappears unless the mesh changes (is coarsened, refined, or both) in proceeding from  $\mathcal{T}_{n-1}$  to  $\mathcal{T}_n$ . The properties of these estimators are discussed more thoroughly in remarks below.

**Theorem 4.3.** *Let  $0 < \tilde{t} \leq T$ , let  $M$  be such that  $1 \leq M \leq J$  and  $t_{M-1} < \tilde{t} \leq t_M$ , and let  $\tilde{x} \in \Omega$ . Also, let  $\underline{h}_M = \min_{0 \leq n \leq M} \min_{K \in \mathcal{T}_n} h_K$  and  $\hat{\underline{h}}_M = \min_{0 \leq n \leq M} \min_{\hat{K} \in \hat{\mathcal{T}}_n} \hat{h}_K$ . Assume that  $\partial\Omega$  is smooth and that the assumptions of*

§4.1 concerning the meshes  $\mathcal{T}_n$ , the finite element spaces  $S^n$ , and the interpolants  $I_h^n$  are met. Then for  $0 \leq s \leq k-1$ ,

$$\begin{aligned}
|(u - u_h)(\tilde{x}, \tilde{t})| &\leq C[\|\sigma_{0,\tilde{x},\tilde{t}}(\cdot, 0)^{k+1}(u_0 - P_h u_0)\|_{L^\infty(\Omega)} \\
&\quad + (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{0 \leq n \leq M} \max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}(K, t_{n+1})^s \eta_{\infty,n}(K) \\
(4.13) \quad &\quad + \max_{1 \leq n \leq M} [\tilde{t} \cdot \eta_{time,data,n} + (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \eta_{time,n;(\tilde{x}, \tilde{t})}] \\
&\quad + (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}(K, t_{n+1})^s \eta_{coarse,n}(K) \\
&\quad + (1 + \hat{\ell}_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \max_{\hat{K} \in \hat{\mathcal{T}}_n} \hat{\sigma}_{n,\tilde{x},\tilde{t}}(\hat{K}, t_n)^s \eta_{mesh,n}(\hat{K}),
\end{aligned}$$

where  $\ell_{\underline{h}_M, \tilde{t}} = \max(0, \ln(\tilde{t}/\underline{h}_M^2))$ ,  $\hat{\ell}_{\underline{h}_M, \tilde{t}} = \max(0, \ln(\tilde{t}/\hat{h}_M^2))$ ,  $\ell_{\tau, \tilde{t}} = 0$  if  $M \leq 2$  and  $\ell_{\tau, \tilde{t}} = \ln \frac{\tilde{t}}{\tau_{M-1}}$  otherwise, and we use the convention  $t_M = t_{M+1} = \tilde{t}$  when evaluating  $\sigma_{n,\tilde{x},\tilde{t}}$  and  $\hat{\sigma}_{n,\tilde{x},\tilde{t}}$ . If in addition  $0 \leq s \leq 2$ , then

$$\begin{aligned}
|(u - u_h)(\tilde{x}, \tilde{t})| &\leq C[\|\sigma_{0,\tilde{x},\tilde{t}}(\cdot, 0)^{k+1}(u_0 - P_h u_0)\|_{L^\infty(\Omega)} \\
&\quad + (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{0 \leq n \leq M} \max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}(K, t_n)^s \eta_{\infty,n}(K) \\
(4.14) \quad &\quad + \max_{1 \leq n \leq M} [\tilde{t} \cdot \eta_{time,data,n} + (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \eta_{time,n;(\tilde{x}, \tilde{t})}] \\
&\quad + (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}(K, t_n)^s \eta_{coarse,n}(K) \\
&\quad + (1 + \hat{\ell}_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \max_{\hat{K} \in \hat{\mathcal{T}}_n} \hat{\sigma}_{n,\tilde{x},\tilde{t}}(\hat{K}, t_n)^s \eta_{mesh,n}(\hat{K}).
\end{aligned}$$

In the above estimates,  $C$  depends on  $\Omega$ ,  $T$ , the degree  $k$  of the finite element spaces, and the shape regularity of the elements in the meshes  $\mathcal{T}_n$ ,  $0 \leq n \leq J$ .

Before proving Theorem 4.3 we make a series of remarks concerning its properties.

**Remark 4.4.** The only difference between (4.13) and (4.14) is that the weight  $\sigma_{n,\tilde{x},\tilde{t}}$  on the second and fourth lines of the estimate is evaluated at  $t_{n+1}$  in (4.13), but at  $t_n$  in (4.14). The latter estimate is more convenient in practice because the estimator  $\max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}(K, t_n)^s \eta_{\infty,n}(K)$  can be computed without knowing what the next node  $t_{n+1}$  is, while computing  $\max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}(K, t_{n+1})^s \eta_{\infty,n}(K)$  requires knowing  $t_{n+1}$ . We have only been able to establish (4.14) for  $s \leq 2$ . This restriction does allow for the best possible weighting in the piecewise quadratic and cubic cases.

**Remark 4.5.** The right hand sides of (4.13) and (4.14) split easily into *initial data errors* (the first line of each estimate), *spatial estimators* (second line), *time estimator* (third line), and *coarsening* and *mesh change* estimators (fourth and fifth lines, respectively). The main two terms in the time indicator are  $\eta_{time,data,n}$  and the term  $\|u^n - P_h^n u_h^{n-1}\|_{L^\infty(\Omega)}$  from the time indicator  $\eta_{time,n;(\tilde{x}, \tilde{t})}$  in (4.11). Recalling the comment following (4.5), we see that these terms are both of optimal order in time, but their influence upon the error at the target point  $(\tilde{x}, \tilde{t})$  does not decay as one moves away from  $(\tilde{x}, \tilde{t})$  in either time or space. However, the influence of the initial data, spatial, coarsening, and mesh change error estimators upon the error at  $(\tilde{x}, \tilde{t})$  does indeed decay as one moves away in time and space so long as  $s > 0$  (which is always possible if  $k \geq 2$ , i.e., quadratic or higher-order elements are used). The last term  $\max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}(K, t_n)^s (h_{n,K}^2 \|(I - P_h^n)(f^n - f^{n-1}) +$

$A_h^n(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(K)}$  in the time estimator  $\eta_{time,n;(\tilde{x},\tilde{t})}$  is of mixed character.  $h^2 A_h^n(u_h^n - P_h^n u_h^{n-1})$  scales heuristically as  $u_h^n - P_h^{n-1} u_h^{n-1}$ , though one cannot use inverse estimates to prove this because of the global nature of  $A_h^n$ . Thus it is of optimal order in time, but also has a local character when  $s > 0$  due to the weight  $\sigma$ . Taken together, these observations suggest that if  $u$  is smooth, then in order to ensure a small error at some given time  $\tilde{t}$  it is necessary to take essentially uniform (small) time steps throughout the calculation, but spatial meshes at time nodes  $t_n < \tilde{t}$  may be substantially coarser than those that will be required at the final time.

**Remark 4.6.** We next compare the estimate (4.13) for fully discrete approximations with the estimate (3.2) for semidiscrete approximations. The first terms (the initial data errors) are essentially the same in both estimates. The second terms (the spatial errors) differ mainly in that the time accumulation in (3.2) is taken as supremum over a continuous time interval, whereas in (4.13) the time accumulation occurs as a maximum over the discrete time intervals. In addition, mesh changes and the time discretization are accounted for in (4.13) but not in (3.2).

**Remark 4.7.** Next we further discuss the time indicators. In [DLM09], a single time estimator  $\sum_{n=1}^M \int_{I_n} \|f - f^n + \phi_{n-1}(g^n - g^{n-1})\|_{L_\infty(\Omega)} dt$  is employed. Note

$$(4.15) \quad \begin{aligned} g^n - g^{n-1} = & [A_h^n u_h^n - A_h^{n-1} u_h^{n-1}] + [(f^n - P_h^n f^n) - (f^{n-1} - P_h^{n-1} f^{n-1})] \\ & + \left[ \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n} - \frac{u_h^{n-2} - P_h^{n-1} u_h^{n-2}}{\tau_{n-1}} \right] \end{aligned}$$

for  $n \geq 2$  (the expression for  $n = 1$  excludes the last half of the last term above). We instead split  $g^n - g^{n-1}$  in order to separate time discretization and mesh change effects and then manipulate the resulting time terms so that the remaining “second-derivative” factors involving  $A_h^n$  are multiplied by  $h^2$ .  $\sum_{n=1}^M \int_{I_n} \|f - f^n + \phi_{n-1}(g^n - g^{n-1})\|_{L_\infty(\Omega)} dt$  is of optimal order in time when no mesh change is allowed, but the estimators we present here are necessary to properly account for mesh change effects and have correct spatial scaling as well.

One disadvantage of the time estimator presented in Theorem 4.3 is that it appears requires the computation of linear operators that are spatially globally dependent ( $A_h^n$  and  $P_h^n$ ) at each time step. Note however from (4.15) that if  $\mathcal{T}_n = \mathcal{T}_{n-1}$ ,  $(I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1}) = g^n - g^{n-1} = (f^n - u_{h,t}^n) - (f^{n-1} - u_{h,t}^{n-1})$ . Thus we must only compute  $A_h^n$  and  $P_h^n$  when the mesh changes and can otherwise avoid this added overhead.

**Remark 4.8.** It is common in practice to modify the backward Euler scheme in (4.1) by replacing  $P_h^n u_h^{n-1}$  with  $L_h^n u_h^{n-1}$ , where  $L_h^n$  is the standard nodal Lagrange interpolant. This is for example the case in the ALBERTA adaptive library (cf. [SS05]). Note that this definition does not coincide with ours when mesh coarsening is allowed, and does not admit all of the spatial orthogonalities that we employ to prove that mesh change effects, and in particular coarsening effects, have a local character.

**Remark 4.9.** We finally consider the possibility of using the estimators in Theorem 4.3 in order to carry out fully (space and time) adaptive computations. In (4.14), it is easy to see that the initial data estimator (first line) can be easily reduced to any given error tolerance by adapting the initial mesh. In addition, one expects that

the space estimator (second line) and time estimator (third line) can be controlled at each time step  $t_n$  by adjusting the mesh  $\mathcal{T}_n$  and the time step  $\tau_n$ . It is clear that the coarsening estimator (fourth line) may also be reduced or eliminated at a given time step by refining  $\mathcal{T}_n$ . In practice, a limited number of spatial coarsening steps are typically carried out at each time step, and one might thus adapt the viewpoint that the coarsening estimator is an “on-off switch” for coarsening: Coarsening is allowed if it is small enough, but not if it isn’t. Note that disallowing coarsening is unlikely to have a negative effect on the other error terms.

Control of the final mesh-change term (fifth line) is somewhat more complex. The term  $(P_h^n - P_h^{n-1})f^{n-1}$  is a data approximation term, and may be controlled at each time step by splitting it as  $(P_h^n - I)f^{n-1} + (I - P_h^{n-1})f^{n-1}$  and controlling coarsening. The difficulty lies in the term  $(A_h^n - A_h^{n-1})u_h^{n-1}$ . Assume momentarily that  $\mathcal{T}_{n-1}$  is quasi-uniform of size  $h_{n-1}$  and that  $\mathcal{T}_n$  is obtained by refining  $\mathcal{T}_{n-1}$  uniformly so that  $\mathcal{T}_n$  is quasi-uniform of size  $h_n$ . If  $\|u_h^{n-1} - u^{n-1}\|_{L^\infty(\Omega)} \leq Ch_{n-1}^{k+1}$  and  $u^{n-1}$  is smooth, it is not hard to show that  $\|(A_h^n - A_h^{n-1})u_h^{n-1}\|_{L^\infty(\Omega)} \leq C(1 + \frac{h_{n-1}}{h_n})h_{n-1}^{k-1}$ . Scaling by  $\hat{h}_n^2 = h_{n-1}^2$  as in (4.12) and assuming that  $(A_h^n - A_h^{n-1})u_h^{n-1}$  dominates the indicator  $\eta_{mesh,n}$  yields  $\eta_{mesh,n} \leq C(1 + \frac{h_{n-1}}{h_n})h_{n-1}^{k+1}$ . That is, we expect that  $\eta_{mesh,n} \rightarrow 0$  as  $\hat{h}_n \rightarrow 0$  if the refinement depth between  $\mathcal{T}_{n-1}$  and  $\mathcal{T}_n$  is bounded, but  $\eta_{mesh,n}$  may in fact blow up if too many refinements are carried out in passing from  $\mathcal{T}_{n-1}$  to  $\mathcal{T}_n$  relative to the mesh size. Simple numerical experiments confirm this behavior, which is expected since  $A_h^n \rightarrow (-\Delta + I)$  as  $h_n \rightarrow 0$ , at least for  $k \geq 1$ , and  $u_h^{n-1}$  is not in  $W_\infty^2$ . Adaptive codes typically refine at most a few times between successive mesh levels, so severe blowup is not generally expected in practical situations. Thus practically speaking,  $\eta_{mesh,n}$  is of optimal order (and also multiplied by the localizing weight  $\hat{\sigma}$  in the final a posteriori estimate), but adaptive control of the mesh change estimator remains a nontrivial task.

In [Dup82] Dupont gave an example for which

$$\max_n \sup_\Omega h_n \rightarrow 0, \quad \max_n \sup_\Omega \hat{h}_n \not\rightarrow 0,$$

and for which the discrete solution does not converge to the exact solution. Thus it is not unexpected that one must control mesh change as well as mesh size in order to obtain a rigorous a posteriori upper bound. In [LM06], the elliptic reconstruction is used to obtain optimal-order estimators in various norms ( $L_\infty(L_2)$ ,  $L_\infty(H^1)$ , and  $H^1(L_2)$ ) for backward Euler schemes. The estimators presented there also involve the maximum mesh size  $\hat{h}_n$  at each time step in an error term that measures mesh change, though in a slightly different fashion than here. Automatically controlling the effects of mesh change in fully adaptive codes employing backward Euler time-stepping schemes is thus not an entirely solved problem, especially in the context of non-standard norms such as we employ here.

**4.4. Proof of Theorem 4.3.** *Proof:* Let  $\tilde{x} \in \Omega$ , let  $G(y, t) = G(\tilde{x}, \tilde{t}; y, t)$ , and let  $G(t) = G(\cdot, t)$ . Using (2.3), we calculate

$$(4.16) \quad (u - u_h)(\tilde{x}, \tilde{t}) = (u_0 - P_h^0 u_0, G(0)) + \int_0^{\tilde{t}} [((u - u_h)_t, G) + A(u - u_h, G)] dt.$$

The term  $(u_0 - P_h^0 u_0, G(0))$  above may be handled exactly as in (3.3) and following, so we must only concentrate on the time integral term in (4.16).

Employing (4.10) while noting that  $G(t) \in H^1(\Omega)$ , we find that

$$\begin{aligned}
(4.17) \quad & \int_0^{\tilde{t}} [((u - u_h)_t, G) + A(u - u_h, G)] dt \\
&= \int_0^{\tilde{t}} [((u - \mathcal{R}u_h)_t, G) + A(u - \mathcal{R}u_h, G) \\
&\quad + ((\mathcal{R}u_h - u_h)_t, G) + A(\mathcal{R}u_h - u_h, G)] dt \\
&= \int_0^{\tilde{t}} A(\mathcal{R}u_h - u_h, G) dt + \sum_{n=1}^M \int_{I_n} [(f - f^n + \phi_{n-1}(g^n - g^{n-1}), G)] dt.
\end{aligned}$$

In order to bound the term  $\int_0^{\tilde{t}} A(\mathcal{R}u_h - u_h, G) dt$  from (4.17), we recall the definitions (4.2) and (4.8) and calculate

$$(4.18) \quad \int_0^{\tilde{t}} A(\mathcal{R}u_h - u_h, G) = \sum_{n=0}^M \int_{t_{n-1}}^{t_{n+1}} \phi_n A(\mathcal{R}^n u_h^n - u_h^n, G) dt.$$

In what follows we use the convention  $t_n = 0$  for  $n < 0$  and  $t_M = t_{M+1} = \tilde{t}$ . Let  $I_h^n : L_1(\Omega) \rightarrow S^n$  satisfy (2.4) through (2.7). Employing (4.9) and standard techniques for proving residual-type estimates, we find that for  $0 \leq n \leq M$ ,

$$\begin{aligned}
(4.19) \quad & \sum_{n=0}^M \int_{t_{n-1}}^{t_{n+1}} \phi_n A(\mathcal{R}^n u_h^n - u_h^n, G) dt \\
&= \sum_{n=0}^M \int_{t_{n-1}}^{t_{n+1}} \phi_n A(\mathcal{R}^n u_h^n - u_h^n, G - I_h^n G) dt \\
&\leq C \max_{0 \leq n \leq M} \max_{t_{n-1} \leq t \leq t_{n+1}} \max_{K \in \mathcal{T}_n} \phi_n(t) \sigma_{n, \tilde{x}, \tilde{t}}^s(K, t) \eta_{\infty, n}(K) \\
&\quad \cdot \sum_{n=0}^M \int_{t_{n-1}}^{t_{n+1}} \int_{\Omega} \sigma_{n, \tilde{x}, \tilde{t}}^{-s} (h_n^{-2} |G - I_h^n G| + h_n^{-1} |\nabla(G - I_h^n G)|) dx dt.
\end{aligned}$$

We next bound each term in the last sum of (4.19). If  $\tilde{t} - t_{n+1} < h_n(\tilde{x})^2$ , we let  $K_0 \in \mathcal{T}_n$  be an element containing  $\tilde{x}$ . Also, we let  $\tilde{t}_n = \tilde{t} - h_n(\tilde{x})^2$  if  $\tilde{t} - t_{n-1} > h_n(\tilde{x})^2$  and  $\tilde{t}_n = t_{n-1}$  otherwise. Finally let  $Q_0 = \omega_{K_0} \times (\tilde{t}_n, t_{n+1})$ ,  $Q_1 = \omega'_{K_0} \times (\tilde{t}_n, t_{n+1})$ , and  $Q_2 = \omega''_{K_0} \times (\tilde{t}_n, t_{n+1})$  if  $\tilde{t} - t_{n+1} < h_n(\tilde{x})^2$  and  $Q_0 = Q_1 = Q_2 = \emptyset$  otherwise. Noting that  $\sigma_{n, \tilde{x}, \tilde{t}}$  is equivalent to 1 on  $Q_2$  and recalling the fact that  $s \leq k - 1$ , we use the properties of  $I_h^n$  to compute

$$\begin{aligned}
(4.20) \quad & \int_{t_{n-1}}^{t_{n+1}} \int_{\Omega} \sigma_{n, \tilde{x}, \tilde{t}}^{-s} (h_n^{-2} |G - I_h^n G| + h_n^{-1} |\nabla(G - I_h^n G)|) dx dt \\
&\leq C \int_{Q_1} (h_n^{-2} |G - I_h^n G| + h_n^{-1} |\nabla(G - I_h^n G)|) dx dt \\
&\quad + \int_{\Omega \times (t_{n-1}, t_{n+1}) \setminus Q_1} \sigma_{n, \tilde{x}, \tilde{t}}^{-k+1} (h_n^{-2} |G - I_h^n G| + h_n^{-1} |\nabla(G - I_h^n G)|) dx dt \\
&\leq C h_n(\tilde{x})^{-1} \int_{Q_2} |\nabla G| dx dt \\
&\quad + \int_{\Omega \times (t_{n-1}, t_{n+1}) \setminus Q_0} (|x - \tilde{x}| + \sqrt{\tilde{t} - t})^{k-1} |D^{k+1} G| dx dt.
\end{aligned}$$

Using (2.2), noting that  $h_n(\tilde{x})^{-1} \leq \frac{1}{\sqrt{\tilde{t}-t}}$  for  $t \in [\tilde{t}_n, t_{n+1}]$ , and using the transformation  $r = \frac{|x-\tilde{x}|}{\sqrt{\tilde{t}-t}}$ , we next find that

$$\begin{aligned}
Ch_n(\tilde{x})^{-1} \int_{Q_2} |\nabla G| dx dt &\leq Ch_n(\tilde{x})^{-1} \int_{\tilde{t}_n}^{t_{n+1}} \int_{|x-\tilde{x}| \leq Ch_n(\tilde{x})} (|x-\tilde{x}| + \sqrt{\tilde{t}-t})^{-N-1} dx dt \\
&\leq Ch_n(\tilde{x})^{-1} \int_{\tilde{t}_n}^{t_{n+1}} \frac{1}{\sqrt{\tilde{t}-t}} \int_0^\infty \frac{r^{N-1}}{(r+1)^{N+1}} dr dt \\
&\leq \begin{cases} Ch_n(\tilde{x})^{-1} \int_{\tilde{t}-h_n(\tilde{x})^2}^{\tilde{t}} \frac{1}{\sqrt{\tilde{t}-t}} dt, & n \geq M-2, \\ C \int_{\tilde{t}_n}^{t_{n+1}} \frac{1}{\tilde{t}-t} dt, & n < M-2. \end{cases}
\end{aligned}$$

Recalling that  $t_{n-1} \leq \tilde{t}_n$ , we thus have

$$(4.21) \quad Ch_n(\tilde{x})^{-1} \int_{Q_2} |\nabla G| dx dt \leq \begin{cases} C, & n \geq M-2, \\ C \int_{t_{n-1}}^{t_{n+1}} \frac{1}{\tilde{t}-t} dt, & n < M-2. \end{cases}$$

Inserting (4.21) into (4.20), defining  $Q_{min} = B_{\underline{h}_M}(\tilde{x}) \times [\tilde{t} - \underline{h}_M^2, \tilde{t}]$ , recalling the definition of  $Q_0$ , and computing as in (3.7) through (3.8), we find that

$$\begin{aligned}
(4.22) \quad &\sum_{n=0}^M \int_{t_{n-1}}^{t_{n+1}} \int_{\Omega} \sigma_{n,\tilde{x},\tilde{t}}^{-s} (h_n^{-2} |G - I_h^n G| + h_n^{-1} |\nabla(G - I_h^n G)|) dx dt \\
&\leq C(1 + \int_{Q_{\tilde{t}} \setminus Q_{min}} (|x-\tilde{x}| + \sqrt{\tilde{t}-t})^{k-1} |D^{k+1} G| dx dt \\
&\quad + \sum_{n=0}^{M-3} \int_{t_{n-1}}^{t_{n+1}} \frac{1}{\tilde{t}-t} dt) \\
&\leq C(1 + \ell_{\underline{h}_M, \tilde{t}} + \int_0^{t_{M-2}} \frac{1}{\tilde{t}-t} dt) \\
&\leq C(1 + \ell_{\underline{h}_M, \tilde{t}} + \ln \frac{\tilde{t}}{\tilde{t}-t_{M-2}}) \\
&\leq C(1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}).
\end{aligned}$$

We now proceed to bound the term  $\sum_{n=1}^M \int_{I_n} [(f - f^n + \phi_{n-1}(g^n - g^{n-1}), G)] dt$  from (4.17). Employing (4.15), we first expand this term as follows:

$$\begin{aligned}
(4.23) \quad & \sum_{n=1}^M \int_{I_n} [(f - f^n + \phi_{n-1}(g^n - g^{n-1}), G)] dt = \left[ \sum_{n=1}^M \int_{I_n} (f - f^n, G) \right. \\
& + \left[ \sum_{n=1}^M \int_{I_n} \phi_{n-1}((I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1}), G) dt \right] \\
& + \left[ \sum_{n=1}^M \int_{I_n} \phi_{n-1}(A_h^n(P_h^n u_h^{n-1} - u_h^{n-1}) \right. \\
& \quad \left. + \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n} - \frac{u_h^{n-2} - P_h^{n-1} u_h^{n-2}}{\tau_{n-1}}, G) dt \right] \\
& \left. + \left[ \sum_{n=1}^M \int_{I_n} \phi_{n-1}((A_h^n - A_h^{n-1})u_h^{n-1} - (P_h^n - P_h^{n-1})f^{n-1}, G) dt \right] \right] \\
& \equiv I + II + III + IV.
\end{aligned}$$

To bound  $I$ , we note from (2.2) that  $\|G\|_{L_1[L_1(\Omega);(0,\tilde{t})]} \leq C\tilde{t}$ . Thus

$$I \leq C\tilde{t} \max_{1 \leq n \leq J} \|f - f^n\|_{L_\infty[L_\infty(\Omega);I_n]}.$$

In order to bound term  $II$  from (4.23), we first use the definition of  $A_h^n$  and integration by parts to compute

$$\begin{aligned}
(4.24) \quad & ((I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1}), G) \\
& = ((I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1}), G - I_h^n G) \\
& \quad + A(u_h^n - P_h^n u_h^{n-1}, I_h^n G) \\
& = ((I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1}), G - I_h^n G) \\
& \quad + A(u_h^n - P_h^n u_h^{n-1}, I_h^n G - G) + A(u_h^n - P_h^n u_h^{n-1}, G).
\end{aligned}$$

Computing as in (4.22) and then employing the inverse inequality  $h_{n,K} \|\nabla(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(K)} \leq \|u_h^n - P_h^n u_h^{n-1}\|_{L_\infty(K)}$  while recalling that  $|\sigma_{n,\tilde{x},\tilde{t}}| \leq 1$ , we next find that

$$\begin{aligned}
(4.25) \quad & \sum_{n=1}^M \int_{I_n} \phi_{n-1} [((I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1}), G - I_h^n G) \\
& \quad + A(u_h^n - P_h^n u_h^{n-1}, I_h^n G - G)] dt \\
& \leq \sum_{n=1}^M \left( \int_{\Omega} \sigma_{n,\tilde{x},\tilde{t}}(K, t)^{-s} (h_n^{-2} |G - I_h^n G| + h_n^{-1} |\nabla(G - I_h^n G)|) dx dt \right) \\
& \quad \cdot \max_{K \in \mathcal{T}_n} \sigma_{n,\tilde{x},\tilde{t}}^s (h_{n,K}^2 \|(I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(K)} \\
& \quad + h_{n,K} \|\nabla(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(K)}) \\
& \leq C(1 + \ell_{h_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \eta_{time,n;(\tilde{x},\tilde{t})}.
\end{aligned}$$

We next mimic the definitions given in the paragraph preceding (4.20) as follows. If  $\tilde{t} - t_n \leq h_n(\tilde{x})^2$ , we let  $K_0 \in \mathcal{T}_n$  be an element containing  $\tilde{x}$ . Also, let  $\tilde{t}_n = \tilde{t} - h_n(\tilde{x})^2$  if  $\tilde{t} - t_{n-1} > h_n(\tilde{x})^2$  and  $\tilde{t} = t_n$  otherwise, and with this definition of  $\tilde{t}_n$ , let  $Q_0$ ,

$Q_1$ , and  $Q_2$  be precisely as previously defined. Finally let  $\psi$  be a cutoff function that is piecewise linear on  $\mathcal{T}_n$  and satisfies  $\psi \equiv 1$  on  $\omega'_{K_0}$  and  $\psi \equiv 0$  on  $\Omega \setminus \omega''_{K_0}$  if  $t \in (\tilde{t} - h_n(\tilde{x})^2, \tilde{t})$ , and  $\psi \equiv 0$  otherwise. Also note that  $|\nabla\psi(x)| \leq Ch_n(\tilde{x})^{-1}$  if  $x \in \omega''_{K_0} \setminus \omega'_{K_0}$  and  $t \in (\tilde{t} - h_n(\tilde{x})^2, \tilde{t})$ , and  $\nabla\psi(x) = 0$  otherwise. Noting that  $A(u_h^n - P_h^n u_h^{n-1}, G) = (\nabla(u_h^n - P_h^n u_h^{n-1}), \nabla G) + (u_h^n - P_h^n u_h^{n-1}, G)$ , and then integrating by parts, we compute

$$\begin{aligned} & (\nabla(u_h^n - P_h^n u_h^{n-1}), \nabla G) \\ &= (\nabla(u_h^n - P_h^n u_h^{n-1}), \psi \nabla G) + (\nabla(u_h^n - P_h^n u_h^{n-1}), (1 - \psi) \nabla G) \\ &\leq h_n(x_0) \|\nabla(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(\text{supp}(\psi))} h_n(x_0)^{-1} \|\nabla G\|_{L_1(\text{supp}(\psi))} \\ &\quad + \|u_h^n - P_h^n u_h^{n-1}\|_\Omega (\|\nabla\psi \nabla G\|_{L_1(\omega''_{K_0})} + \|(1 - \psi) \Delta G\|_{L_1(\Omega)}). \end{aligned}$$

Employing an inverse inequality on the term  $h_n(x_0) \|\nabla(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(\text{supp}(\psi))}$ , recalling (4.21), defining  $Q_{min}$  as in the paragraph preceding (4.22), and computing as in (3.7) through (3.8), we have

$$\begin{aligned} & \sum_{n=1}^M \int_{I_n} A(u_h^n - P_h^n u_h^{n-1}, G) dt \\ (4.26) \quad & \leq C \max_{1 \leq n \leq M} \|u_h^n - P_h^n u_h^{n-1}\|_{L_\infty(\Omega)} \\ & \quad \cdot (1 + \|G + \Delta G\|_{L_1([\Omega \times (0, \tilde{t}]) \setminus Q_{min})} + \int_0^{\tilde{t} - \tau_{M-1}} \frac{1}{\tilde{t} - t} dt) \\ & \leq C(1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \|u_h^n - P_h^n u_h^{n-1}\|_{L_\infty(\Omega)}. \end{aligned}$$

Inserting (4.26) and (4.25) into (4.24) completes the bound of term *II*.

Our arguments for bounding term *III* from (4.23) are similar to those for term *II*. Recalling the definition (4.3) of  $A_h^n$ , we first compute

$$\begin{aligned} & (A_h^n(P_h^n u_h^{n-1} - u_h^{n-1}), G) = (A_h^n(P_h^n u_h^{n-1} - u_h^{n-1}), G - I_h^n G) \\ (4.27) \quad & \quad + A(P_h^n u_h^{n-1} - u_h^{n-1}, I_h^n G - G) + A(P_h^n u_h^{n-1} - u_h^{n-1}, G). \\ & = (A_h^n(P_h^n u_h^{n-1} - u_h^{n-1}), G - I_h^n G) + A(P_h^n u_h^{n-1} - u_h^{n-1}, I_h^n G - G) \\ & \quad + (P_h^n u_h^{n-1} - u_h^{n-1}, (-\Delta G + G) - I_h^n(-\Delta G + G)). \end{aligned}$$

Similarly,

$$(4.28) \quad \left( \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n}, G \right) = \left( \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n}, G - I_h^n G \right)$$

Arguing as previously, we find that

$$\begin{aligned} (4.29) \quad & \sum_{n=1}^M \int_{I_n} \int_\Omega |\sigma_{n, \tilde{x}, \tilde{t}}^{-s} [h_n^{-2} |I_h^n G - G| + h_n^{-1} |\nabla(I_h^n G - G)|] dx dt \\ & \leq C(1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}). \end{aligned}$$

Noting that  $A(P_h^n u_h^{n-1} - u_h^{n-1}, G) = (\nabla(P_h^n u_h^{n-1} - u_h^{n-1}), \nabla G) + (P_h^n u_h^{n-1} - u_h^{n-1}, G)$ , we compute

$$\begin{aligned} & (\nabla(P_h^n u_h^{n-1} - u_h^{n-1}), \nabla G) \\ &= (\nabla(P_h^n u_h^{n-1} - u_h^{n-1}), \psi \nabla G) + (\nabla(P_h^n u_h^{n-1} - u_h^{n-1}), (1 - \psi) \nabla G) \\ &\leq \|\nabla(P_h^n u_h^{n-1} - u_h^{n-1})\|_{L_\infty(\text{supp}(\psi))} \|\nabla G\|_{L_1(\text{supp}(\psi))} \\ &\quad + (P_h^n u_h^{n-1} - u_h^{n-1}, \nabla \cdot [(1 - \psi) \nabla G] - I_h^n \nabla \cdot [(1 - \psi) \nabla G]). \end{aligned}$$

This expression may be bounded precisely as in the preceding paragraph after noting that  $I_h^n$  may be defined so that the support of  $I_h^n \nabla \cdot [(1 - \psi) \nabla G]$  is the same as that of  $\nabla \cdot [(1 - \psi) \nabla G]$ . Combining (4.27), (4.28), and (4.29), one finds that for  $0 \leq s \leq k - 1$ ,

$$\begin{aligned} (4.30) \quad III &\leq C(1 + \ell_{\underline{h}_M, \bar{t}} + \ell_{\tau, \bar{t}}) \left[ \max_{1 \leq n \leq M} \max_{t_{n-1} \leq t \leq t_n} \max_{K \in \mathcal{T}_n} \sigma_{n, \bar{x}, \bar{t}}(K, t) \right. \\ &\quad \cdot (h_{n, K}^2 \left\| \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n} \right\|_{L_\infty(K)} + h_{n, K}^2 \|A_h^n(P_h^n u_h^{n-1} - u_h^{n-1})\|_{L_\infty(K)} \\ &\quad + h_{n, K} \|\nabla(P_h^n u_h^{n-1} - u_h^{n-1})\|_{L_\infty(K)} + \|P_h^n u_h^{n-1} - u_h^{n-1}\|_{L_\infty(K)}) \\ &\quad \left. + \max_{1 \leq n \leq M} \max_{t_n \leq t \leq t_{n+1}} \max_{K \in \hat{\mathcal{T}}_n} \phi_n(t) \sigma_{n, \bar{x}, \bar{t}}(K, t) h_{n, K}^2 \left\| \frac{u_h^{n-1} - P_h^n u_h^{n-1}}{\tau_n} \right\|_{L_\infty(K)} \right]. \end{aligned}$$

Next we bound the term  $IV$  from (4.23). Noting that  $((A_h^n - A_h^{n-1})u_h^{n-1} - (P_h^n - P_h^{n-1})f^{n-1}, v) = 0$  for any  $v \in \hat{S}^n$  and proceeding as in (4.19) and following but now with the finest common coarsenings  $\hat{\mathcal{T}}_n$  replacing the original mesh  $\mathcal{T}_n$ , we find

$$\begin{aligned} (4.31) \quad IV &= \sum_{n=1}^J \int_{I_n} \phi_{n-1}((A_h^n - A_h^{n-1})u_h^{n-1} - (P_h^n - P_h^{n-1})f^{n-1}, G - \hat{I}_h^n G) \, dt \\ &\leq C(1 + \hat{\ell}_{\underline{h}_M, \bar{t}} + \ell_{\tau, \bar{t}}) \max_{1 \leq n \leq M} \max_{t_{n-1} \leq t \leq t_n} \max_{\hat{K} \in \hat{\mathcal{T}}_n} \phi_{n-1}(t) \hat{\sigma}_{n, \bar{x}, \bar{t}}(\hat{K}, t) \hat{h}_{n, \hat{K}}^2 \\ &\quad \cdot \|(A_h^n - A_h^{n-1})u_h^{n-1} - (P_h^n - P_h^{n-1})f^{n-1}\|_{L_\infty(\hat{K})}. \end{aligned}$$

Finally we consider the term  $\max_{t_{n-1} \leq t \leq t_{n+1}} \phi_n(t) \sigma_{n, \bar{x}, \bar{t}}^s$  from (4.19) and other similar terms from (4.30) and (4.31). Note that  $\phi_n$  and  $\sigma_{n, \bar{x}, \bar{t}}$  are the only terms in these upper bounds that depend on  $t$ . Observe also that  $\phi_n \sigma_n^s$  is strictly increasing in  $t$  for  $t_{n-1} \leq t \leq t_n$ . Thus  $\max_{t_{n-1} \leq t \leq t_{n+1}} \phi_n(t) \sigma_{n, \bar{x}, \bar{t}}^s = \max_{t_n \leq t \leq t_{n+1}} \phi_n(t) \sigma_{n, \bar{x}, \bar{t}}^s$ . For any nonnegative  $s$ , we also have

$$(4.32) \quad \max_{t_n \leq t \leq t_{n+1}} \phi_n(t) \sigma_{n, \bar{x}, \bar{t}}^s \leq \sigma_{n, \bar{x}, \bar{t}}^s(\cdot, t_{n+1}).$$

Collecting (4.32) and (4.22) into (4.19) and then collecting (4.19) and (4.23) into (4.18) and then (4.17) yields the first three terms in (4.13). Similar arguments may be used to obtain the last two (coarsening and mesh change) terms in (4.13) from (4.30) and (4.31), thus completing the proof of (4.13). In addition, it is not hard to compute that for  $t_n \leq t \leq t_{n+1}$  and  $0 \leq s \leq 2$ ,

$$\frac{d}{dt} \phi_n(t) \sigma_{n, \bar{x}, \bar{t}}^s(K, t) \leq \sigma_{n, \bar{x}, \bar{t}}^s \frac{1}{\tau_{n+1}} \left( \frac{s(t_{n+1} - t)}{2} - 1 \right) \leq 0.$$

Thus if  $s \leq 2$ ,  $\max_{t_n \leq t \leq t_{n+1}} \phi_n(t) \sigma_{n, \bar{x}, \bar{t}}^s(\cdot, t)^s$  is taken on at  $t = t_n$ , and we obtain (4.14).  $\square$

**4.5. Fully discrete sharply local estimates for arbitrary subsets.** As in §3.3, we may trivially obtain estimates that are valid over arbitrary subsets of  $\Omega \times (0, T]$ .

**Corollary 4.10.** *Let  $D \subset \Omega \times (0, T]$ , and let*

$$\sigma_{n,D}(x, t) = \sup_{(x_0, t_0) \in D} \sigma_{n, x_0, t_0}(x, t).$$

$\hat{\sigma}_{n,D}$  is defined similarly. In addition, define

$$\begin{aligned} \eta_{time, n; D} &= \|u_h^n - P_h^n u_h^{n-1}\|_{L_\infty(\Omega)} \\ &+ \max_{K \in \mathcal{T}_n} \sigma_{n,D}(K, t_n)^s h_{n,K}^2 \|(I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(K)}. \end{aligned}$$

Let also  $t_D = \sup_{(x,t) \in D} t$ , let  $t_{M-1} < t_D \leq t_M$ , and assume  $0 \leq s \leq \min(k-1, 2)$ . Then under the conditions of Theorem 4.3,

$$\begin{aligned} \|u - u_h\|_{L_\infty(D)} &\leq C[\|\sigma_{0,D}(\cdot, 0)\|_{L_\infty(\Omega)}]^{k+1} \|u_0 - P_h u_0\|_{L_\infty(\Omega)} \\ &+ (1 + \ell_{\underline{h}_M, t_D} + \ell_{\tau, t_D}) \max_{0 \leq n \leq M} \max_{K \in \mathcal{T}_n} \sigma_{n,D}(K, t_n)^s \eta_{\infty, n}(K) \\ &+ \max_{1 \leq n \leq M} [\tilde{t} \cdot \eta_{time, data, n} + (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \eta_{time, n; D}] \\ &+ (1 + \ell_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \max_{K \in \mathcal{T}_n} \sigma_{n,D}(K, t_n)^s \eta_{coarse, n}(K) \\ &+ (1 + \hat{\ell}_{\underline{h}_M, \tilde{t}} + \ell_{\tau, \tilde{t}}) \max_{1 \leq n \leq M} \max_{\hat{K} \in \hat{\mathcal{T}}_n} \hat{\sigma}_{n,D}(\hat{K}, t_n)^s \eta_{mesh, n}(\hat{K}). \end{aligned}$$

In the above estimates,  $C$  depends on  $\Omega$ ,  $T$ , the degree  $k$  of the finite element spaces, and the shape regularity of the elements in the meshes  $\mathcal{T}_n$ ,  $0 \leq n \leq M$ .

## 5. NUMERICAL EXPERIMENTS

In this section we describe some numerical experiments in one space dimension which illustrate the properties of our estimators and confirm their utility in constructing adaptive codes. Note that while the theoretical results of our paper and the bulk of practical applications concern space dimension  $N \geq 2$ , the main issues addressed in this paper concerning interaction of time and space errors can be expected to arise even for space dimension  $N = 1$ . Thus we carry out our computations in this simple model situation.

**5.1. Basic parameters.** In our computations we take  $\Omega = (0, 1) \subset \mathbb{R}$  and  $\tilde{t} = 0.1$ . For Experiments 1 and 2 below, our test solution is

$$u(x, t) = \arctan(\cos(\pi x^{100t+1})).$$

Note that  $u$  has an interior layer whose sharpness and position depend upon  $t$  and that  $\frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(1, t) = 0$  (see Figure 3 for a graph of  $u(x, 0.1)$ ). Based on numerical tests using several test solutions and values of  $\tilde{t}$ , we employed space and time estimators with the following constants:

$$\begin{aligned} est_{space} &= 0.02 \max_{0 \leq n \leq M} \max_{K \in \mathcal{T}_n} \sigma_{n,D}(K, t_n)^s \eta_{\infty, n}(K), \\ est_{time} &= \max_{1 \leq n \leq M} (0.002 \cdot \tilde{t} \cdot \eta_{time, data, n} + 0.2 \|u_h^n - P_h^n u_h^{n-1}\|_{L_\infty(\Omega)} \\ &+ 0.02 \max_{K \in \mathcal{T}_n} \sigma_{n,D}(K, t_n)^s h_{n,K}^2 \|(I - P_h^n)(f^n - f^{n-1}) + A_h^n(u_h^n - P_h^n u_h^{n-1})\|_{L_\infty(K)}) \end{aligned}$$

The coarsening, initial data, and mesh change indicators were similarly scaled by 0.02, since the constants in these indicators arise in a manner similar to those in the

spatial estimator. Our main focus in choosing the above constants was to balance the relative contributions of the various parts of our overall error estimator in order to obtain effective error indicators; we do not attempt to justify them theoretically. Also note that adaptivity (choosing time steps and refining and coarsening spatial meshes) was carried out entirely on the basis of the initial data, spatial, and time indicators; mesh change and coarsening indicators were not taken into account. However, the results we present below indicate that these contributions are generally similar to the spatial indicators. No logarithmic factors were taken into account.

Given an error tolerance  $\epsilon$ , our adaptive code first refined the initial mesh so as to ensure that the initial data and initial spatial estimators were bounded by  $\frac{\epsilon}{2}$ . Subsequent time steps were also chosen so that the local time indicators (scaled as in the preceding paragraph) were bounded above by  $\frac{\epsilon}{2}$  and below by  $\frac{\epsilon}{4}$ . Spatial mesh refinement was carried out if the spatial estimator at any time step exceeded  $\frac{\epsilon}{2}$ , and coarsening was carried out if the spatial estimators for a given pair of compatible elements (i.e., ones having the same parent element) were below  $\frac{\epsilon}{32}$ .

Finally, quadratic elements ( $k = 2$ ) were used in all computations. We thus took  $s = 0$  in order to control the global maximum error and  $s = 1$  in order to carry out computations reflecting the local character of the error.

**5.2. Experiment 1: Orders of convergence.** In the first experiment, adaptive calculations were carried out based on global refinement in both space and time, that is,  $s = 0$ . This enables us to confirm that our basic estimators are of optimal order and also provides a benchmark with which to compare adaptive calculations based on sharply local estimators. We took  $\epsilon$  to be 0.1, 0.05, 0.025, 0.01, 0.005, 0.0025, 0.001, and 0.0001.

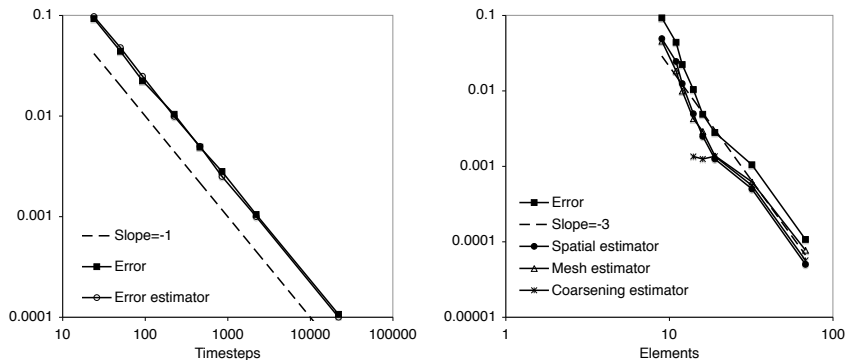


FIGURE 1. Global error  $\|u - u_h\|_{L_\infty((0,1) \times (0,1))}$  and estimators vs. number of time steps (left) and number of elements (right) in an adaptive algorithm for controlling  $\|u - u_h\|_{L_\infty}$  globally.

In Figure 1 we confirm that adaptivity based on our time estimators yields optimal-order decrease of the error. The left plot in Figure 1 confirms that the overall error decreases with the inverse of the number of time steps, and in addition that  $\frac{est_{space} + est_{time}}{\|u - u_h\|_{L_\infty((0,1) \times (0,1))}} \approx 1$  (that is, the effectivity indices are very close to 1). In the right plot of Figure 1, we see that the error and global spatial estimator similarly decrease with optimal order  $\frac{1}{DOF^3}$ , where  $DOF$  is the maximum number of spatial elements appearing in the computation. Figure 1 also indicates that the

coarsening and mesh change estimators consistently have values similar to or less than the spatial estimator whenever they are nonzero.

**5.3. Experiment 2: Localization.** Adaptive calculations were carried out with  $D_0 = \Omega \times (0, \tilde{t}] = (0, 1) \times (0, .1)$  (global error, as in the previous subsection), with  $D_1 = [0, .5] \times \{t = 0.1\}$ , and with  $D_2 = [.5, 1] \times \{t = 0.1\}$ .  $u$  is smooth on  $D_1$ , whereas the interior layer at  $t = 0.1$  lies in  $0.5 \leq x \leq 1$  (cf. Figure 3), so more spatial elements are required to resolve  $u$  on  $D_2$  than on  $D_1$ . An error plot showing optimal-order decrease with respect to both the time and spatial discretizations is shown in Figure 2. The impact of localization on the time discretization is considered in the right plot of Figure 2, where it can be seen that our adaptive algorithm employing sharply local estimators produces very nearly the same time discretization as the adaptive algorithm designed to control the global maximum error. This strongly supports the assertion that our time indicators largely filter out spatial error effects.

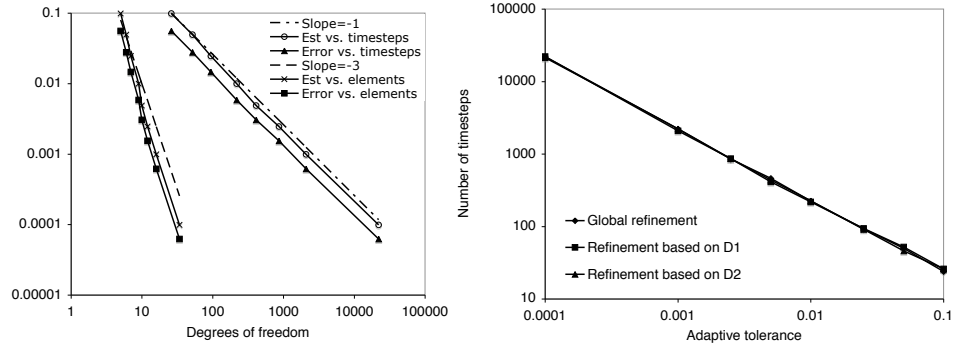


FIGURE 2. *Left:* Optimal-order decrease of  $\|u - u_h\|_{L^\infty(D_1)}$  with respect to the number mesh elements and number of timesteps. *Right:* Number of timesteps produced by adaptive algorithms with given tolerances. Considered are algorithms for controlling the global error, and for controlling the maximum errors on  $D_1$  and  $D_2$ .

Figure 3 contains a comparison of the time evolution of the numbers of spatial elements used in the adaptive code with a tolerance of  $\epsilon = 0.001$ . The code designed to control the global error predictably employs the largest number of elements. Controlling the error on  $D_1$ , where  $u$  is very smooth, requires the least number of elements. Resolving the interior layer in  $D_2$  requires a number of elements *at the final time*  $t = 0.1$  comparable to the number required at  $t = 0.1$  in the computation controlling the global error, but at previous times a relatively small number of elements is required. This confirms our assertion that an adaptive code based on a sharply local estimator should allow for spatial meshes that are relatively coarse in regions removed from the target region  $D$  but more heavily refined near  $D$ . Figure 3 contains a comparison of the meshes at the final time  $t = 0.1$  produced by the adaptive code when  $D = D_1$  and  $D = D_2$ , which shows that as expected  $D = D_2$  leads to a more heavily refined mesh on  $[0.5, 1]$  when  $t = 0.1$ . As is also expected, controlling the error on  $D_1$  or  $D_2$  does not control the error globally.

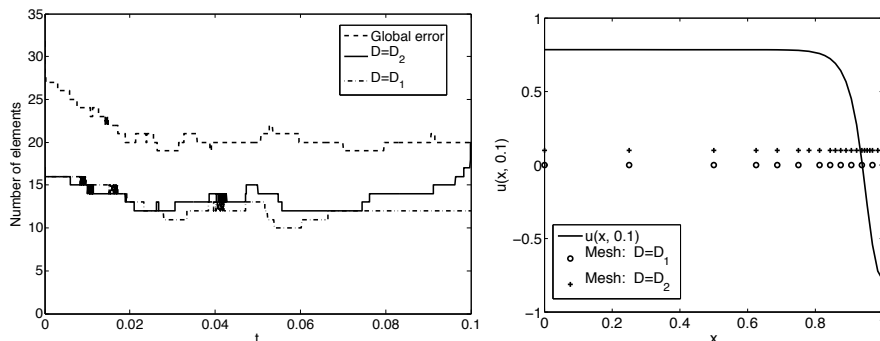


FIGURE 3. *Left:* Number of spatial elements vs. time produced by adaptive codes designed to control two different local errors (on  $D = [0, .5] \times t = .1$  and  $D = [.5, 1] \times t = .1$ ) and the global error. *Right:*  $u(x, 0.1)$  along with the meshes for  $t = 0.1$  produced by the adaptive algorithm with tolerance  $\epsilon = 10^{-3}$  for  $D = D_1$  and  $D = D_2$ .

For example,  $\epsilon = 10^{-4}$  and  $D = D_1$  yields  $\|u - u_h\|_{L_\infty(D)} = 6.23 \times 10^{-5}$ , but  $\|u - u_h\|_{L_\infty((0,1) \times [0,.1])} = 8.03 \times 10^{-4}$ .

Finally, we note that the behavior of the coarsening and mesh change estimators is essentially the same in the case of sharply local estimators as when the global error is controlled. That is, these estimators are consistently approximately the size of the spatial estimator or smaller (we do not present such data for sharply local estimators, as it is quite similar to Figure 1).

**5.4. Experiment 3: Pollution.** In this experiment we illustrate pollution effects. Here we let  $D = [0, 0.5] \times [0, 0.1]$ , that is, the target domain is the left half of the unit interval, at every time in the specified time interval. In order to demonstrate that some refinement is necessary globally in order to obtain good local approximation, we set  $s = 20$  in our code. This forces refinement in  $D$  while maintaining mesh smoothness globally, but causes very little refinement away from  $D$ . As our test solution we chose

$$(5.1) \quad u(x, t) = (1 - e^{-t}) \cos(4\pi x).$$

Adaptive calculations were carried out as above with the adaptive tolerance  $\epsilon$  taken to be 0.1, 0.05, 0.025, 0.01, 0.005, 0.0025, 0.001, 0.0001, 0.00005, 0.000025, and 0.00001. In the left graph of Figure 4, we observe that the error initially decreases as the mesh is refined, but the error decrease eventually stalls and the error plot levels out. This indicates that further *local* refinement on  $D$  will not lead to further error decrease on  $D$ , that is, lack of global resolution of the solution  $u$  pollutes the solution quality on  $D$ . The right plot of Figure 4 contains the discrete solution  $u_h|_{t=0.1}$  and mesh obtained with  $\epsilon = 0.00001$ . Here we observe the lack of global resolution.

**5.5. Conclusion.** The error estimators that we have produced here substantially isolate the time and spatial effects present in the backward Euler finite element discretization, as can be seen both by analyzing their theoretical properties and from numerical tests. This represents a substantial step forward in understanding

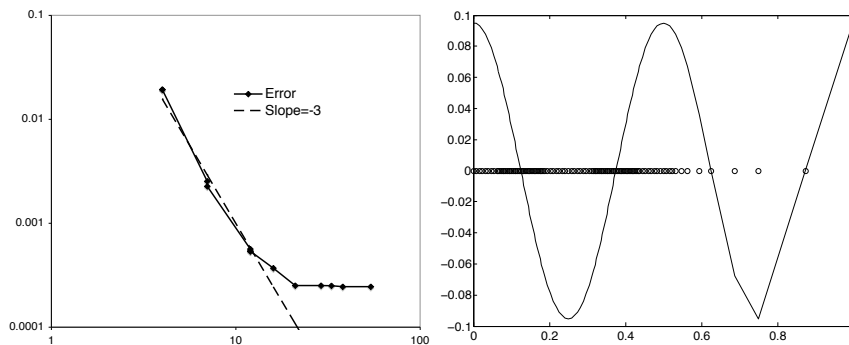


FIGURE 4. *Left:*  $\|u - u_h\|_{L^\infty(D)}$  vs. number of spatial elements; pollution is observed as the mesh is refined. *Right:*  $u_h(x, 0.1)$  along with the mesh for  $t = 0.1$  produced by the adaptive algorithm with tolerance  $\epsilon = 10^{-4}$ .

the interactions of local errors and time discretizations. However, the low order of the backward Euler time discretization is a drawback of our adaptive method, and an interesting question for future research is the interaction of higher-order time discretizations with local errors. An additional open question is the possibility of obtaining localization in the time error for some time scheme.

#### ACKNOWLEDGEMENTS

The authors appreciate insightful comments from two referees which led to improvements in the paper.

#### REFERENCES

- [Ber89] Christine Bernardi, *Optimal finite-element interpolation on curved domains*, SIAM J. Numer. Anal. **26** (1989), no. 5, 1212–1240. MR MR1014883 (91a:65228)
- [Bom00] Mats Boman, *On a posteriori error analysis in the maximum norm*, Ph.D. thesis, Chalmers University of Technology and Göteborg University, 2000.
- [BR01] Roland Becker and Rolf Rannacher, *An optimal control approach to a posteriori error estimation in finite element methods*, Acta Numer. **10** (2001), 1–102. MR 2 009 692
- [Clé75] Ph. Clément, *Approximation by finite element functions using local regularization*, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. RAIRO Analyse Numérique **9** (1975), no. R-2, 77–84. MR MR0400739 (53 #4569)
- [Dem04] Alan Demlow, *Piecewise linear finite element methods are not localized*, Math. Comp. **73** (2004), no. 247, 1195–1201 (electronic). MR MR2047084 (2005a:65129)
- [Dem06] ———, *Localized pointwise a posteriori error estimates for gradients of piecewise linear finite element approximations to second-order quasilinear elliptic problems*, SIAM J. Numer. Anal. **44** (2006), no. 2, 494–514. MR MR2218957
- [Dem07] ———, *Local a posteriori estimates for pointwise gradient errors in finite element methods for elliptic problems*, Math. Comp. **76** (2007), no. 257, 19–42 (electronic). MR MR2261010
- [Dem] Alan Demlow, *Finite element interpolation of nonsmooth functions on curved domains using polynomial basis functions*, Tech. report, In preparation.
- [DLM09] Alan Demlow, Omar Lakkis, and Charalambos Makridakis, *A posteriori error estimates in the maximum norm for parabolic problems*, SIAM J. Numer. Anal. **47** (2009), no. 3, 2157–2176.

- [DR98] W. Dörfler and M. Rumpf, *An adaptive strategy for elliptic problems including a posteriori controlled boundary approximation*, Math. Comp. **67** (1998), no. 224, 1361–1382. MR MR1489969 (99b:65141)
- [Dup82] Todd Dupont, *Mesh modification for evolution equations*, Math. Comp. **39** (1982), no. 159, 85–107. MR MR658215 (84g:65131)
- [ÈI70] S. D. Èidel'man and S. D. Ivasišen, *Investigation of the Green's matrix of a homogeneous parabolic boundary value problem*, Trans. Moscow Math. Soc. **23** (1970), 179–242. MR MR0367455 (51 #3697)
- [GS02] M.B. Giles and E. Süli, *Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality*, Acta Numer. **11** (2002), 145–236.
- [Len86] M. Lenoir, *Optimal isoparametric finite elements and error estimates for domains involving curved boundaries*, SIAM J. Numer. Anal. **23** (1986), no. 3, 562–580. MR MR842644 (87m:65163)
- [Ley04a] Dmitriy Leykekhman, *Pointwise localized error estimates for parabolic finite element equations*, Numer. Math. **96** (2004), no. 3, 583–600. MR MR2028727 (2004k:65175)
- [Ley04b] ———, *Pointwise weighted error estimates for parabolic finite element equations*, Ph.D. thesis, Cornell University, 2004.
- [LM06] Omar Lakkis and Charalambos Makridakis, *Elliptic reconstruction and a posteriori error estimates for fully discrete linear parabolic problems*, Math. Comp. **75** (2006), no. 256, 1627–1658 (electronic). MR MR2240628 (2007e:65122)
- [LN03] Xiaohai Liao and Ricardo H. Nochetto, *Local a posteriori error estimates and adaptive control of pollution effects*, Numer. Methods Partial Differential Equations **19** (2003), no. 4, 421–442. MR MR1980188 (2004c:65130)
- [LW08] D. Leykekhman and L. B. Wahlbin, *A posteriori error estimates by recovered gradients in parabolic finite element equations*, BIT **48** (2008), no. 3, 585–605. MR MR2447987
- [MN03] Charalambos Makridakis and Ricardo H. Nochetto, *Elliptic reconstruction and a posteriori error estimates for parabolic problems*, SIAM J. Numer. Anal. **41** (2003), no. 4, 1585–1594 (electronic). MR MR2034895 (2004k:65157)
- [Noc95] Ricardo H. Nochetto, *Pointwise a posteriori error estimates for elliptic problems on highly graded meshes*, Math. Comp. **64** (1995), no. 209, 1–22. MR 95c:65172
- [NS74] Joachim A. Nitsche and Alfred H. Schatz, *Interior estimates for Ritz-Galerkin methods*, Math. Comp. **28** (1974), 937–958. MR 51 #9525
- [Sch98] Alfred H. Schatz, *Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids. I. Global estimates.*, Math. Comp. **67** (1998), no. 223, 877–899. MR 98j:65082
- [Sco73] L.R. Scott, *Finite element techniques for curved boundaries*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1973.
- [SS05] Alfred Schmidt and Kunibert G. Siebert, *Design of adaptive finite element software*, Lecture Notes in Computational Science and Engineering, vol. 42, Springer-Verlag, Berlin, 2005, The finite element toolbox ALBERTA, With 1 CD-ROM (Unix/Linux). MR MR2127659 (2005i:65003)
- [STW98] A. H. Schatz, V. Thomée, and L. B. Wahlbin, *Stability, analyticity, and almost best approximation in maximum norm for parabolic finite element equations*, Comm. Pure Appl. Math. **51** (1998), no. 11-12, 1349–1385. MR MR1639143 (99h:65171)
- [SW95] Alfred H. Schatz and Lars B. Wahlbin, *Interior maximum-norm estimates for finite element methods, Part II*, Math. Comp. **64** (1995), no. 211, 907–928.
- [SZ90] L. Ridgway Scott and Shangyou Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp. **54** (1990), no. 190, 483–493. MR 90j:65021
- [XZ00] Jinchao Xu and Aihui Zhou, *Local and parallel finite element algorithms based on two-grid discretizations*, Math. Comp. **69** (2000), no. 231, 881–909. MR MR1654026 (2000j:65102)
- [Zlá73] Miloš Zlámal, *Curved elements in the finite element method. I*, SIAM J. Numer. Anal. **10** (1973), 229–240. MR MR0395263 (52 #16060)

UNIVERSITY OF KENTUCKY, DEPARTMENT OF MATHEMATICS, 715 PATTERSON OFFICE TOWER,  
LEXINGTON, KY 40506-0027

*E-mail address:* `demlow@ms.uky.edu`

DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY OF CRETE, GR- 71409 HERAKLION,  
GREECE, AND, INSTITUTE FOR APPLIED AND COMPUTATIONAL MATHEMATICS, FOUNDATION FOR  
RESEARCH AND TECHNOLOGY-HELLAS, VASILIKA VOUTON P.O.Box 1527, GR-71110 HERAKLION,  
GREECE

*E-mail address:* `makr@tem.uoc.gr`