A POSTERIORI ERROR ESTIMATION AND ADAPTIVITY FOR DEGENERATE PARABOLIC PROBLEMS

R.H. NOCHETTO, A. SCHMIDT, AND C. VERDI

ABSTRACT. Two explicit error representation formulas are derived for degenerate parabolic PDEs, which are based on evaluating a parabolic residual in negative norms. The resulting upper bounds are valid for any numerical method, and rely on regularity properties of solutions of a dual parabolic problem in nondivergence form with vanishing diffusion coefficient. They are applied to a practical space-time discretization consisting of C^0 piecewise linear finite elements over highly graded unstructured meshes, and backward finite differences with varying time-steps. Two rigorous a posteriori error estimates are derived for this scheme, and used in designing an efficient adaptive algorithm, which equidistributes space and time discretization errors via refinement/coarsening. A simulation finally compares the behavior of the rigorous a posteriori error estimators with a heuristic approach, and hints at the potentials and reliability of the proposed method.

1. INTRODUCTION

A posteriori error estimates are a fundamental component in the design of reliable and efficient adaptive algorithms for the numerical solution of PDEs. Even though rigorous results are available for linear and mildly nonlinear parabolic PDEs [5], [6], [7], the theory is much less satisfactory for strongly nonlinear PDEs. There are no results applicable to degenerate parabolic PDEs, which in turn exhibit lack of regularity across interfaces and corresponding numerical pollution effects. The use of highly graded meshes and varying time-steps is thus motivated by the nonlinear structure of the PDE, as opposed to domain geometry, and is a vehicle for resolving small scale features with optimal computational complexity.

In this paper we introduce a rigorous theory of a posteriori error estimation for degenerate parabolic problems of the form

$$\partial_t u - \Delta\beta(u) = f \quad \text{in } Q := \Omega \times (0,T), \tag{1.1}$$

where β is nondecreasing and Lipschitz. A typical example of industrial interest is that of solidification (classical Stefan problem), for which

$$\beta(s) := \beta_- \min(s,0) + \beta_+ \max(s-L,0). \tag{1.2}$$

1991 *Mathematics Subject Classification.* 65N15, 65N30, 65N50, 80A22, 35K65, 35R35.

Key words and phrases. degenerate parabolic equations, Stefan problem, finite elements, parabolic duality, a posteriori estimates, adaptivity.

This work was partially supported by NSF Grants DMS-9305935 and DMS-9623394, EU Grant HCM "Phase Transitions and Surface Tension", MURST, and CNR Contract 95.00735.01.

This corresponds to an ideal material with constant thermal coefficients $\beta_-, \beta_+ > 0$ and latent heat L. Any approximation U of u satisfies

$$\partial_t U - \Delta\beta(U) = f - \mathcal{R} \quad \text{in } Q, \tag{1.3}$$

where $\mathcal{R}$, an oscillatory distribution of singular character, is the so-called *parabolic residual.* In spite of the simple parabolic structure of (1.1), (1.2), its degenerate nature makes the theory of [5] fail in that it exploits the regularizing effect of a linear parabolic dual problem. The corresponding dual PDE in this context is the nonstrictly parabolic equation in *nondivergence* form

$$\partial_t \zeta + b\Delta\zeta = \eta \quad \text{in } Q, \tag{1.4}$$

with vanishing and rough diffusion coefficient $0 \le b \le \max(\beta_-, \beta_+)$ [8], [9], [10]. Such a PDE does not exhibit a regularizing mechanism, and is not computable in that b is discontinuous and depends on both u and U. Problem (1.4) measures the error accumulation in time, and is thus crucial. Our objective is to prove global regularity properties of ζ in §2, and use them in §3 to derive two representation formulas for the errors $u - U$ and $\beta(u) - \beta(U)$ in energy norms. These formulas are valid for *any* numerical method, evaluate $\mathcal{R}$ in two distinct negative norms, and lead to rigorous a posteriori upper error bounds. Since negative norms entail averaging, they are appropriate to quantify the oscillatory character of $\mathcal{R}$.

We next apply these ideas to a practical scheme consisting of C^0 piecewise linear finite elements over highly graded unstructured meshes and backward finite differences with varying time-steps. The method uses mass lumping and evaluates $\beta(U)$ solely at the nodes, which makes it easy to implement and solve iteratively. We discuss the method in §4 and derive in §5 two rigorous a posteriori error estimates for it of the form (Approaches I and II):

$$\|u - U\|_{L^\infty(0,T;H^{-1}(\Omega))} + \|\beta(u) - \beta(U)\|_{L^2(Q)} \le \mathcal{E}(u_0, f, T, \Omega; U, h, \tau). \tag{1.5}$$

The estimator $\mathcal{E}$ is computable in terms of data $u_0 = u(\cdot, 0), f, T, \Omega$, computed solution U, meshsize h, and time step τ, but entails L^1 or L^2 norms in time. This is impractical in that the entire evolution history would be needed to control $\mathcal{E}$.

We thus resort to an L^∞ norm in time, and introduce an adaptive algorithm which equidistributes space discretization errors for a uniform error distribution in time. Such errors are estimated via local a posteriori error indicators, and further equidistributed via a refinement/coarsening strategy based on bisection. This yields *compatible* consecutive meshes. These issues are fully discussed in §6.

We conclude in §7 with simulations illustrating the viability of our Approaches I and II, as well as a heuristic Approach III based on using local regularity of ζ in (1.4) and heat estimators away from discrete interfaces. We clearly show that they are all able to detect the presence of interfaces, and refine accordingly, and that Approaches II and III perform best. There is no need to compute the interface explicitly for mesh design, which is a major improvement with respect to [12].

Further simulations, comparisons of several nonlinear solvers, and a detailed description of the adaptive algorithm will be presented elsewhere [14]. Even though our error estimates (1.5) are rigorous, they do not necessarily imply $\mathcal{E}(u_0, f, T, \Omega; U, h, \tau) \to 0$ as $h, \tau \to 0$ because $\mathcal{E}$ depends on discrete quantities that change with h and τ. Stability and convergence will be assessed in [13].

2. Setting and Parabolic Duality

Let $\Omega \subset \mathbb{R}^d$ $(d > 1)$ be a bounded convex polyhedral domain, let $T > 0$ be the final time, and set $Q := \Omega \times (0, T)$. Let $\beta \in W^{1,\infty}(\mathbb{R})$ satisfy $\beta(s) = 0$ for all $s \in (0, L)$ and $0 < a \le \beta'(s) \le A$ for a.e. $s \notin (0, L)$. Let u_0 indicate the initial enthalpy, let $\theta_0 := \beta(u_0)$ denote the initial temperature, and let $F_0 := \{x \in \Omega : \theta_0(x) = 0\}$ be the initial interface. They satisfy

$$\theta_0 \in W_0^{1,\infty}(\Omega), \quad F_0 \text{ is a Lipschitz curve.}$$

Therefore, $u_0 \in W^{1,\infty}(\Omega \backslash F_0)$ and u_0 has a jump discontinuity across F_0. Finally let f be sufficiently smooth. The continuous problem then reads as follows.

Continuous problem. *Find u and θ such that*

$$\theta \in L^2(0, T; H_0^1(\Omega)), \quad u \in L^\infty(0, T; L^2(\Omega)) \cap H^1(0, T; H^{-1}(\Omega)),$$
$$\theta(x, t) = \beta(u(x, t)) \quad a.e.\ (x, t) \in Q,$$
$$u(\cdot, 0) = u_0,$$

and for a.e. $t \in (0, T)$ and all $\eta \in H_0^1(\Omega)$ the following equation holds

$$\langle \partial_t u, \eta \rangle + \langle \nabla\theta, \nabla\eta \rangle = \langle f, \eta \rangle. \tag{2.1}$$

Hereafter, $\langle \cdot, \cdot \rangle$ stands for either the inner product in $L^2(\Omega)$ or the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$. It is to be observed that the vanishing Dirichlet boundary condition on θ is assumed only for simplicity and so that the interface $F(t) := \{x \in \Omega : \theta(x, t) = 0\}$ does not include $\partial\Omega$. Existence and uniqueness for this problem are known [8], [9].

To motivate the dual problem (1.4), already studied in [8], [9], [10], we subtract (1.3) from (1.1) and integrate by parts over Q against a smooth test function ζ vanishing on $\partial\Omega \times (0, T)$. The error $e_u := u - U$ satisfies

$$\langle e_u, \zeta \rangle_{|t=T} = \langle e_u, \zeta \rangle_{|t=0} + \int_0^T \langle e_u, \partial_t\zeta + b\Delta\zeta \rangle + \mathcal{R}(\zeta), \tag{2.2}$$

where $0 \le b(x, t) \le A$ is the *discontinuous* function

$$b(x, t) := \begin{cases} \frac{\beta(u(x,t)) - \beta(U(x,t))}{u(x,t) - U(x,t)} & \text{if } u(x, t) \ne U(x, t) \\ A & \text{otherwise.} \end{cases}$$

We could thus represent norms of $e_u(\cdot, T)$ or their integrals over Q by making judicious choices of $\zeta(\cdot, T)$ and $\partial_t\zeta + b\Delta\zeta$. Evaluation of $\mathcal{R}(\zeta)$ depends on regularity of ζ, which we now investigate. Given a regularization parameter $\delta > 0$ to be chosen later, we consider two backward parabolic problems

$$\mathcal{J}(\psi) = -b^{1/2}\chi \quad \text{in } Q, \quad \psi(\cdot, T) = 0 \quad \text{in } \Omega, \tag{2.3}$$

$$\mathcal{J}(\phi) = 0 \quad \text{in } Q, \quad \phi(\cdot, T) = \rho \quad \text{in } \Omega, \tag{2.4}$$

with operator $\mathcal{J}$ in *nondivergence* form

$$\mathcal{J}(\zeta) := \partial_t\zeta + (b + \delta)\Delta\zeta, \tag{2.5}$$

$\psi, \phi = 0$ on $\partial\Omega \times (0, T)$, and $\chi \in L^2(Q)$, $\rho \in H_0^1(\Omega)$. Existence of unique solutions $\psi, \phi \in H^{2,1}(Q)$ follows directly from the theory of nonlinear strictly parabolic problems [9].

Lemma 2.1. *The following a priori bounds are valid for all* $0 \le t \le T$

$$2\|\nabla\psi(\cdot,t)\|^2_{L^2(\Omega)},\ \tfrac{1}{A+\delta}\|\partial_t\psi\|^2_{L^2(Q)},\ 4\delta\|\Delta\psi\|^2_{L^2(Q)} \le \|\chi\|^2_{L^2(Q)}. \tag{2.6}$$

Proof. We multiply (2.3) by $\Delta\psi$ and integrate by parts in $\Omega \times (t,T)$ to obtain

$$\begin{aligned}\tfrac12\int_\Omega |\nabla\psi(\cdot,t)|^2 + \int_\Omega\int_t^T (b+\delta)|\Delta\psi|^2 &= -\int_\Omega\int_t^T b^{1/2}\chi\,\Delta\psi \\ &\le \int_\Omega\int_t^T b|\Delta\psi|^2 + \tfrac14\int_\Omega\int_t^T |\chi|^2.\end{aligned}$$

This yields the a priori bounds for $\nabla\psi$ and $\Delta\psi$. Finally, we multiply (2.3) by $\partial_t\psi/(b+\delta)$ and integrate by parts in $\Omega\times(t,T)$ to get

$$\begin{aligned}\int_\Omega\int_t^T \tfrac{1}{b+\delta}|\partial_s\psi|^2 + \tfrac12\int_\Omega|\nabla\psi(\cdot,t)|^2 &= -\int_\Omega\int_t^T \tfrac{b^{1/2}}{b+\delta}\chi\,\partial_s\psi \\ &\le \tfrac12\int_\Omega\int_t^T \tfrac{1}{b+\delta}|\partial_s\psi|^2 + \tfrac12\int_\Omega\int_t^T|\chi|^2,\end{aligned}$$

because $\delta > 0$. This implies the a priori bounds for $\partial_t\psi$ because $b \le A$. □

Lemma 2.2. *The following a priori bounds are valid for all* $0 \le t \le T$

$$\|\nabla\phi(\cdot,t)\|^2_{L^2(\Omega)},\ \tfrac{2}{A+\delta}\|\partial_t\phi\|^2_{L^2(Q)},\ 2\delta\|\Delta\phi\|^2_{L^2(Q)} \le \|\nabla\rho\|^2_{L^2(\Omega)}. \tag{2.7}$$

Proof. We multiply (2.4) by $\Delta\phi$ and integrate by parts in $\Omega\times(t,T)$ to get

$$\tfrac12\int_\Omega|\nabla\phi(\cdot,t)|^2 + \int_\Omega\int_t^T(b+\delta)|\Delta\phi|^2 = \tfrac12\int_\Omega|\nabla\rho|^2.$$

The a priori bound for $\Delta\phi$ thus follows from $b \ge 0$. In view of (2.4) we further have

$$\|\partial_t\phi\|^2_{L^2(Q)} = \|(b+\delta)\Delta\phi\|^2_{L^2(Q)} \le (A+\delta)\|(b+\delta)^{1/2}\Delta\phi\|^2_{L^2(Q)} \le \tfrac{A+\delta}{2}\|\nabla\rho\|^2_{L^2(\Omega)},$$

where we have used that $b \le A$. This completes the proof. □

Corollary 2.1. *The following* $L^2_tH^2_x$ *a priori bounds are valid*

$$\int_0^T |\psi|^2_{H^2(\Omega)} \le \tfrac{1}{4\delta}\|\chi\|^2_{L^2(Q)}, \qquad \int_0^T|\phi|^2_{H^2(\Omega)} \le \tfrac{1}{2\delta}\|\nabla\rho\|^2_{L^2(\Omega)}. \tag{2.8}$$

Proof. It suffices to invoke the well-known estimate [5], [9] for convex Ω

$$|\eta|_{H^2(\Omega)} \le \|\Delta\eta\|_{L^2(\Omega)} \quad \forall\ \eta\in H^1_0(\Omega)\cap H^2(\Omega),$$

in conjunction with (2.6) and (2.7). □

The above dual problems will be instrumental in this paper. To see why, we point out that taking $\zeta = \psi$ in (2.2) yields an estimate for $\|\beta(u) - \beta(U)\|_{L^2(Q)}$, whereas taking $\zeta = \phi$ in (2.2) gives rise to an estimate for $\|u - U\|_{L^\infty(0,T;H^{-1}(\Omega))}$; see §3. We will use such an idea in §5 and present two rigorous a posteriori error estimates for problem (2.1) in these natural energy norms (Approaches I and II).

In contrast to the heat equation, problems (2.3) and (2.4) do not exhibit any regularizing effect; in fact as $\delta \to 0$ the information on second derivatives is lost (e.g., if $b = 0$ in Q then, in the limit as $\delta \to 0$, $\phi(\cdot, t) = \rho(\cdot) \in H_0^1(\Omega)$ for all $0 \le t \le T$). Since the regularity of the dual problem dictates the weights (powers of meshsize and time step) of the a posteriori error estimators, this is an early indication of the striking difference between degenerate and strictly parabolic problems.

A simple calculation shows, however, that $b \ge a_\sigma := a\sigma/(L + \sigma)$ in the open set

$$Q_\sigma := \{(x,t) \in Q : |U(x,t) - \tfrac{L}{2}| > \tfrac{L}{2} + \sigma\},$$

for any $\sigma > 0$. The proofs of Lemma 2.1, with splitting $2|b^{1/2}\chi\Delta\psi| \le b|\Delta\psi|^2 + |\chi|^2$, and Lemma 2.2 yield

$$\int_{Q_\sigma} |\Delta\psi|^2 \le \tfrac{1}{a_\sigma}\|\chi\|^2_{L^2(Q)}, \quad \int_{Q_\sigma} |\Delta\phi|^2 \le \tfrac{1}{2a_\sigma}\|\nabla\rho\|^2_{L^2(\Omega)}. \tag{2.9}$$

Then $L_t^2H_x^2$ seminorms of ψ and ϕ in any compact subset Q_σ^* of Q_σ depend on the inverse of the parabolic distance between the parabolic boundaries of Q_σ^* and Q_σ. This lack of uniform regularity in $L_t^2H_x^2$ prevents the construction of rigorous error indicators based on the heat equation away from *discrete* interfaces. However, in §§5 and 6 we present an empirical estimator (Approach III), which utilizes heat estimators in Q_0^*, and compare its performance with Approaches I and II in §7.

3. Error Representation Formulas

The purpose of this derivation is to obtain formulas for the errors

$$e_{\beta(u)} := \beta(u) - \beta(U), \quad e_u := u - U,$$

where u is the true solution and $U \in L^2(Q)$ is any other function. We do not assume that U is any specific approximation of u, and so the resulting formulas are quite general. In particular we do not require stability of U and $\beta(U)$ in the energy norms $L_t^\infty L_x^2$ and $L_t^2H_x^1$, respectively. As a by-product we rederive the usual $O(\sqrt{\varepsilon})$ rate of convergence for a vanishing viscosity approximation U of u.

For any function $\zeta \in C^0([0,T]; H^{-1}(\Omega))$, we denote $\zeta^t(\cdot) := \zeta(\cdot, t)$ for all $t \in [0,T]$. We multiply the PDE operator $\mathcal{J}(\zeta)$ in (2.5) by $e_u(\cdot, t)$, which is in $L^2(\Omega)$ for a.e. $t \in (0,T)$, and use property $be_u = e_{\beta(u)}$ to write

$$\int_Q e_u \mathcal{J}(\zeta) = \int_Q \left(u\partial_t\zeta + \beta(u)\Delta\zeta\right) - \int_Q \left(U\partial_t\zeta + \beta(U)\Delta\zeta\right) + \delta\int_Q e_u\Delta\zeta.$$

Let $U^0, U^T \in H^{-1}(\Omega)$ be given, and set $e_u^0 := u_0 - U^0$ and $e_u^T := u^T - U^T$. At this stage, both U^0 and U^T are arbitrary, but they will later be the initial value $U(\cdot, 0)$

and final value $U(\cdot,T)$ of U which make no sense in the present context. We then integrate by parts in space and time the first term on the right-hand side, employ (2.1), and add and subtract $\langle U^0,\zeta^0\rangle - \langle U^T,\zeta^T\rangle$, to arrive at

$$\langle e_u^T,\zeta^T\rangle - \int_Q e_u \mathcal{J}(\zeta) = \langle e_u^0,\zeta^0\rangle + \mathcal{R}(\zeta) - \delta\int_Q e_u\Delta\zeta, \tag{3.1}$$

where $\mathcal{R}(\zeta)$, the *parabolic residual*, is the distribution

$$\mathcal{R}(\zeta) := \langle U^0,\zeta^0\rangle - \langle U^T,\zeta^T\rangle + \int_Q \big(f\zeta + U\partial_t\zeta + \beta(U)\Delta\zeta\big). \tag{3.2}$$

Together with the initial error $\langle e_u^0,\zeta^0\rangle$, $\mathcal{R}(\zeta)$ is a measure of the amount by which U misses to be a solution of (2.1) and must be evaluated in negative norms.

We show first an $L_t^2L_x^2$ error estimate for $\beta(u)$. To this end, let $\zeta=\psi$ be the solution of the dual problem (2.3) and note that $\psi^T=0$ and

$$|e_u|b^{1/2} = (e_u e_{\beta(u)})^{1/2} \ge \tfrac{1}{A^{1/2}}|e_{\beta(u)}|.$$

From (3.1) and (3.2), we easily get

$$\begin{aligned}
\|e_{\beta(u)}\|_{L^2(Q)} &= \sup_{\chi\in L^2(Q)} \frac{\int_Q e_{\beta(u)}\chi}{\|\chi\|_{L^2(Q)}} \le A^{1/2}\sup_{\chi\in L^2(Q)}\frac{\int_Q e_u b^{1/2}\chi}{\|\chi\|_{L^2(Q)}}\\
&\le A^{1/2}\sup_{\chi\in L^2(Q)}\frac{\|e_u^0\|_{H^{-1}(\Omega)}\|\nabla\psi(\cdot,0)\|_{L^2(\Omega)}}{\|\chi\|_{L^2(Q)}}\\
&\quad + A^{1/2}\sup_{\chi\in L^2(Q)}\frac{|\mathcal{R}(\psi)|}{\|\chi\|_{L^2(Q)}}\\
&\quad + \delta^{1/2}A^{1/2}\sup_{\chi\in L^2(Q)}\frac{\|e_u\|_{L^2(Q)}\|\delta^{1/2}\Delta\psi\|_{L^2(Q)}}{\|\chi\|_{L^2(Q)}}.
\end{aligned} \tag{3.3}$$

It is thus apparent from (2.6) with $\delta\to 0$ that the representation formula (3.3) hinges on a negative norm of the residual $\mathcal{R}(\psi)$ involving first derivatives of ψ,

$$\Psi_{-1} := \sup_{\chi\in L^2(Q)}\frac{|\mathcal{R}(\psi)|}{\|\nabla\psi\|_{L^\infty(0,T;L^2(\Omega))}},$$

and leads to the following a posteriori error estimates for $\beta(u)$.

Lemma 3.1. *Let ψ be the solution of (2.3) with arbitrary $\chi\in L^2(Q)$. Then*

$$\|e_{\beta(u)}\|_{L^2(Q)} \le \left(\tfrac{A}{2}\right)^{1/2}\big(\|e_u^0\|_{H^{-1}(\Omega)} + \Psi_{-1}\big).$$

Proof. It remains to deal with the last term in (3.3). Since

$$|e_u| \le L + \tfrac{1}{a}|e_{\beta(u)}|,$$

combining (2.6) with (3.3) implies

$$(3.4)\quad \left(1-\tfrac{A^{1/2}}{2a}\delta^{1/2}\right)\|e_{\beta(u)}\|_{L^2(Q)} \le \left(\tfrac{A}{2}\right)^{1/2}\left(\|e_u^0\|_{H^{-1}(\Omega)} + \Psi_{-1} + \tfrac{1}{\sqrt{2}}L|Q|^{1/2}\delta^{1/2}\right).$$

The assertion then follows upon taking $\delta \to 0$. $\square$

To derive an $L_t^\infty H_x^{-1}$ error estimate for u, let $\zeta = \phi$ be the solution of the dual problem (2.4). Therefore, according to (3.1) and (3.2), since $e_u\mathcal{J}(\phi) = 0$ we get

$$(3.5)\quad \begin{aligned} \|e_u^T\|_{H^{-1}(\Omega)} &= \sup_{\rho\in H_0^1(\Omega)} \frac{\int_\Omega e_u^T \rho}{\|\nabla\rho\|_{L^2(\Omega)}} \\ &\le \sup_{\rho\in H_0^1(\Omega)} \frac{\|e_u^0\|_{H^{-1}(\Omega)}\|\nabla\phi(\cdot,0)\|_{L^2(\Omega)}}{\|\nabla\rho\|_{L^2(\Omega)}} \\ &\quad + \sup_{\rho\in H_0^1(\Omega)} \frac{|\mathcal{R}(\phi)|}{\|\nabla\rho\|_{L^2(\Omega)}} \\ &\quad + \delta^{1/2} \sup_{\rho\in H_0^1(\Omega)} \frac{\|e_u\|_{L^2(Q)}\|\delta^{1/2}\Delta\phi\|_{L^2(Q)}}{\|\nabla\rho\|_{L^2(\Omega)}}. \end{aligned}$$

Again, in view of (2.7) with $\delta \to 0$, this representation formula hinges on a negative norm of the residual $\mathcal{R}(\phi)$ involving first derivatives of ϕ, that is

$$\Phi_{-1} := \sup_{\rho\in H_0^1(\Omega)} \frac{|\mathcal{R}(\phi)|}{\|\nabla\phi\|_{L^\infty(0,T;L^2(\Omega))}}.$$

Lemma 3.2. *Let ϕ be the solution of (2.4) with an arbitrary $\rho \in H_0^1(\Omega)$. Then*

$$\|e_u(\cdot,T)\|_{H^{-1}(\Omega)} \le \|e_u^0\|_{H^{-1}(\Omega)} + \Phi_{-1}.$$

Proof. We make use of (2.7) and argue as in Lemma 3.1 to derive from (3.5)

$$(3.6)\quad \|e_u^T\|_{H^{-1}(\Omega)} \le \|e_u^0\|_{H^{-1}(\Omega)} + \Phi_{-1} + \tfrac{1}{\sqrt{2}a}\delta^{1/2}\|e_{\beta(u)}\|_{L^2(Q)} + \tfrac{1}{\sqrt{2}}L|Q|^{1/2}\delta^{1/2}.$$

The assertion then follows upon taking $\delta \to 0$. $\square$

Lemmas 3.1 and 3.2 lead to *Approach I* below, and are pessimistic in that they deal with the worst scenario in terms of regularity of ψ and ϕ, namely one space derivative. An alternative and fruitful avenue consists of keeping $\delta > 0$, thereby allowing H^2 space regularity of ψ and ϕ, and optimizing δ later without sending it to 0; this yields *Approach II* below and works best. To this end, we set

$$\delta_0 := \frac{R_{-2}}{C_0 a L|Q|^{1/2}}, \quad C_0 := \tfrac{1}{a}\left(A^{1/2} + \tfrac{1}{\sqrt{2}}\right),$$

where $R_{-2} := A^{1/2}\Psi_{-2} + \Phi_{-2}/\sqrt{2}$ involves the following negative norms of the residuals $\mathcal{R}(\psi)$ and $\mathcal{R}(\phi)$ with two space derivatives of ψ and ϕ

$$\Psi_{-2} := \sup_{\chi\in L^2(Q)} \frac{|\mathcal{R}(\psi)|}{\|\Delta\psi\|_{L^2(Q)}}, \quad \Phi_{-2} := \sup_{\rho\in H_0^1(\Omega)} \frac{|\mathcal{R}(\phi)|}{\|\Delta\phi\|_{L^2(Q)}}.$$

We expect δ_0 to be small because it involves $\mathcal{R}(\psi)$ and $\mathcal{R}(\phi)$. However this cannot be guaranteed a priori and is reflected in the statement of next result.

Lemma 3.3. *Let ψ be the solution of* (2.3) *with an arbitrary $\chi \in L^2(Q)$ and let ϕ be the solution of* (2.4) *with an arbitrary $\rho \in H_0^1(\Omega)$. Then*

$$\|e_{\beta(u)}\|_{L^2(Q)} + \|e_u(\cdot,T)\|_{H^{-1}(\Omega)} \le \sqrt{2}C_0 a\|e_u^0\|_{H^{-1}(\Omega)} + \begin{cases} 2\big(C_0 aL|Q|^{1/2}R_{-2}\big)^{1/2} & \text{if } \delta_0 \le \frac{1}{C_0^2} \\ 2C_0R_{-2} & \text{if } \delta_0 > \frac{1}{C_0^2}. \end{cases}$$

Proof. We add twice (3.3) to (3.5), and argue as with (3.4) and (3.6). Since

$$\frac{|\mathcal{R}(\psi)|}{\|\chi\|_{L^2(Q)}} \le \frac{1}{2\delta^{1/2}}\frac{|\mathcal{R}(\psi)|}{\|\Delta\psi\|_{L^2(Q)}}, \qquad \frac{|\mathcal{R}(\phi)|}{\|\nabla\rho\|_{L^2(\Omega)}} \le \frac{1}{\sqrt{2}\delta^{1/2}}\frac{|\mathcal{R}(\phi)|}{\|\Delta\phi\|_{L^2(Q)}},$$

we readily get

$$(2 - C_0\delta^{1/2})\|e_{\beta(u)}\|_{L^2(Q)} + \|e_u(\cdot,T)\|_{H^{-1}(\Omega)} \le \sqrt{2}C_0 a\|e_u^0\|_{H^{-1}(\Omega)} + q(\delta), \tag{3.7}$$

where $q(\delta) := q_-\delta^{-1/2} + q_+\delta^{1/2}$ with $q_- := R_{-2}$ and $q_+ := C_0aL|Q|^{1/2}$. We now observe that $\delta = \delta_0$ minimizes $q(\delta)$. If $C_0\delta_0^{1/2} \le 1$, then the first assertion follows trivially from $q(\delta_0) = 2(q_-q_+)^{1/2}$. Otherwise, if $C_0\delta_0^{1/2} > 1$, then $q_+ < C_0^2 q_-$ and $q(1/C_0^2) < 2C_0q_-$. This concludes the argument. □

If U is a finite element solution, then the last term in Lemmas 3.1–3.3 (parabolic residual) can be further evaluated via Galerkin orthogonality; this is accomplished in §5. We stress that U need not be a discrete solution as the following application of Lemma 3.3 illustrates. Let U be the solution of (2.1) with $\beta_\varepsilon(s) := \beta(s) + \varepsilon s$ in place of $\beta(s)$, namely U is the usual vanishing viscosity approximation of u. The proof below is different from the original one in [10], easily extends to other perturbations of β such as that in [10], and is valid under *minimal regularity* of both u_0 and f which precludes compactness.

Corollary 3.1. *The following perturbation estimate is valid for ε small*

$$\|e_{\beta(u)}\|_{L^2(Q)} + \|e_u\|_{L^\infty(0,T;H^{-1}(\Omega))} \le C\varepsilon^{1/2},$$

where constant C depends on $\|u_0\|_{H^{-1}(\Omega)}, |\Omega|, T, L, a, A,$ and $\|f\|_{L^1(0,T;H^{-1}(\Omega))}$.

Proof. Since $\mathcal{R}(\zeta) = -\varepsilon\int_Q U\Delta\zeta$, the assertion results from Lemma 3.3 upon realizing that $R_{-2} \le aC_0\varepsilon\|U\|_{L^2(Q)} \le C\varepsilon$. To derive a bound on $\|U\|_{L^2(Q)}$ under minimal regularity, we consider (2.1) with β_ε instead of β and strong data $U^0 \in L^2(\Omega)$ and $f \in L^2(0,T;H^{-1}(\Omega))$. We take $\eta = GU(\cdot,t) \in H_0^1(\Omega)$, where the Green's operator G is the inverse of $-\Delta$ in $H_0^1(\Omega)$, and obtain

$$\tfrac{1}{2}d_t\langle GU,U\rangle + \langle\beta_\varepsilon(U),U\rangle = \langle Gf,U\rangle. \tag{3.8}$$

We then integrate (3.8) in time for $0 \le t \le s \le T$, the maximum of $\|U(\cdot,t)\|_{H^{-1}(\Omega)}$ being attained at $t = s$. We easily deduce

$$\begin{aligned} &\tfrac{1}{2}\|U(\cdot,s)\|^2_{H^{-1}(\Omega)} + \tfrac{1}{A}\int_0^s \|\beta_\varepsilon(U)(\cdot,t)\|^2_{L^2(\Omega)}dt \\ &\quad\le \tfrac{1}{2}\|U^0\|^2_{H^{-1}(\Omega)} + \|U(\cdot,s)\|_{H^{-1}(\Omega)}\int_0^s \|Gf(\cdot,t)\|_{H_0^1(\Omega)}dt, \end{aligned} \tag{3.9}$$

which yields a bound for $\|U(\cdot,s)\|_{H^{-1}(\Omega)}$ without use of Gronwall's inequality. We insert this bound in (3.9), now for arbitrary s, and use that $\beta'_\varepsilon(s) \geq \beta'(s) \geq a > 0$ and Ω is bounded to arrive at the desired estimate for $\|U\|_{L^2(Q)}$ for strong data. We finally observe that the mapping $(U^0, f) \to (U, \beta_\varepsilon(U))$ from $H^{-1}(\Omega) \times L^1(0,T;H^{-1}(\Omega)) \to L^\infty(0,T;H^{-1}(\Omega)) \times L^2(Q)$ is Lipschitz, which is easily seen with the technique leading to (3.9). This allows us to regularize the weak data $U^0 \in H^{-1}(\Omega)$ and $f \in L^1(0,T;H^{-1}(\Omega))$ via $U^0_\sigma \in L^2(\Omega)$ and $f_\sigma \in L^2(0,T;H^{-1}(\Omega))$ and eventually pass to the limit in the regularization parameter σ. □

4. Finite Element Discretization

We now introduce the fully discrete problem, which combines continuous piecewise linear finite elements in space with backward differences in time.

We denote by τ_n the time step at the n-th step and set $t^n := \sum_{i=1}^n \tau_i$. Let N be the total number of time steps, that is $t^N = T$. For any function $\zeta \in C^0((t^{n-1}, t^n]; H^{-1}(\Omega))$, we denote $\zeta^n(\cdot) := \zeta(\cdot, t^n)$.

We denote by $\mathcal{M}^n$ a uniformly regular partition of Ω into simplices [3]. Mesh $\mathcal{M}^n$ is obtained from $\mathcal{M}^{n-1}$ by refining/coarsening, and thus $\mathcal{M}^n$ and $\mathcal{M}^{n-1}$ are *compatible*. Given a triangle $S \in \mathcal{M}^n$, h_S stands for its diameter and ρ_S for its sphericity and they satisfy $h_S \leq 2\rho_S/\sin(\alpha_S/2)$, where α_S is the minimum angle of S. Uniform regularity of the family of triangulations is equivalent to $\alpha_S \geq \alpha > 0$, with α independent of n. We also denote by $\mathcal{B}^n$ the collection of interior interelement boundaries e of $\mathcal{M}^n$ in Ω; h_e stands for the size of $e \in \mathcal{B}^n$.

Let $\mathbb{V}^n \subset H^1_0(\Omega)$ indicate the usual space of piecewise linear finite elements over $\mathcal{M}^n$. Let $\{x^n_k\}_{k=1}^{K^n}$ denote the interior nodes of $\mathcal{M}^n$. Let $I^n : C^0_0(\bar\Omega) \to \mathbb{V}^n$ be the usual Lagrange interpolation operator, namely $(I^n\eta)(x^n_k) = \eta(x^n_k)$ for all $1 \leq k \leq K^n$. Finally, let the discrete inner product $\langle\cdot,\cdot\rangle^n$ be defined by

$$\langle\varphi,\eta\rangle^n := \int_\Omega I^n(\varphi\eta)dx = \tfrac{1}{d+1}\sum_{S\in\mathcal{M}^n}|S|\sum_{x^n_k\in S}\varphi(x^n_k)\eta(x^n_k) \quad \forall\, \varphi,\eta\in\mathbb{V}^n.$$

This corresponds to the vertex quadrature rule, which can be easily evaluated element by element and leads to mass lumping [3].

We observe the validity of the local estimates

$$\begin{aligned} &\|\varphi\eta - I^n(\varphi\eta)\|_{L^\infty(S)} \leq \tfrac18 h_S^2|\nabla\varphi||\nabla\eta| \quad \forall\, \varphi,\eta\in\mathbb{V}^n,\\ &\|\nabla\varphi\|_{L^2(S)} \leq \tfrac13\rho_S^{-1}\|\varphi\|_{L^2(S)} \quad \forall\,\varphi\in\mathbb{V}^n. \end{aligned} \tag{4.1}$$

Consequently, the element inner product $\langle\varphi,\eta\rangle^n_S := \int_S I^n(\varphi\eta)$ satisfies [15]

$$\|\varphi\|_{L^2(S)} \leq (\langle\varphi,\varphi\rangle^n_S)^{1/2} =: \|\varphi\|^n_S \leq C\|\varphi\|_{L^2(S)} \quad \forall\,\varphi\in\mathbb{V}^n,$$

and the discrepancy between $\langle\varphi,\eta\rangle_S := \int_S\varphi\eta$ and $\langle\varphi,\eta\rangle^n_S$ can be bounded by

$$\begin{aligned} |\langle\varphi,\eta\rangle_S - \langle\varphi,\eta\rangle^n_S| &\leq \tfrac18 h_S^2\|\nabla\varphi\|_{L^2(S)}\|\nabla\eta\|_{L^2(S)}\\ &\leq \tfrac{1}{24}h_S^2\rho_S^{-1}\|\varphi\|_{L^2(S)}\|\nabla\eta\|_{L^2(S)} \quad \forall\,\varphi,\eta\in\mathbb{V}^n. \end{aligned} \tag{4.2}$$

The discrete initial enthalpy $U^0\in\mathbb{V}^0$ is defined at nodes x^0_k of $\mathcal{M}^0 = \mathcal{M}^1$ to be

$$U^0(x^0_k) := u_0(x^0_k) \quad \forall\, x^0_k\in\Omega\backslash F_0, \quad U^0(x^0_k) := 0 \quad \forall\, x^0_k\in F_0. \tag{4.3}$$

Hence, U^0 is easy to evaluate in practice. Then we set $\Theta^0 := I^0\beta(U^0)$.

Discrete problem. *Given $U^{n-1}, \Theta^{n-1} \in \mathbb{V}^{n-1}$, then $\mathcal{M}^{n-1}$ and τ_{n-1} are modified as described below to get $\mathcal{M}^n$ and τ_n and thereafter $U^n, \Theta^n \in \mathbb{V}^n$ computed according to $\Theta^n = I^n\beta(U^n)$ and*

$$\frac{1}{\tau_n}\langle U^n - I^n U^{n-1}, \varphi\rangle^n + \langle \nabla\Theta^n, \nabla\varphi\rangle = \langle I^n f^n, \varphi\rangle^n \quad \forall\, \varphi \in \mathbb{V}^n. \tag{4.4}$$

In view of the constitutive relation $\Theta^n = I^n\beta(U^n)$ being enforced only at the nodes, and the use of mass lumping, (4.4) is easy to implement and yields a monotone operator in $\mathbb{R}^{K^n}$. This problem is solved below via an optimized nonlinear SOR [12]. However, these computational tricks introduce further consistency errors that are apparent from (3.2). Whether these devices preserve optimal accuracy is still to be explored.

We finally introduce further notation. The following sets will be used later

$$\begin{aligned}
\mathcal{F}^n &:= \cup\{S \in \mathcal{M}^n : \Theta^n(x) = 0 \text{ for some } x \in S\} && \textit{discrete free boundary},\\
\mathcal{T}^n &:= \cup\{S \in \mathcal{M}^n : \Theta^n(x)\Theta^{n-1}(x) \le 0 \text{ for some } x \in S\} && \textit{transition region},\\
\mathcal{C}^n &:= \cup\{S \in \mathcal{M}^n : S \text{ is coarsened from } \mathcal{M}^{n-1}\} && \textit{coarsening set}.
\end{aligned}$$

We point out that the compatibility of $\mathcal{M}^n$ and $\mathcal{M}^{n-1}$ yields $I^nU^{n-1} \ne U^{n-1}$ only in $\mathcal{C}^n$; in fact $(\mathbb{V}^{n-1} \cap \mathbb{V}^n)_{|S} = \mathbb{V}^{n-1}_{|S}$ except for elements S in $\mathcal{C}^n$.

Let the *jump* J^n_e of $\nabla\Theta^n$ across $e \in \mathcal{B}^n$ be

$$J^n_e := [\![\nabla\Theta^n]\!]_e \cdot \nu_e = (\nabla\Theta^n_{|S_1} - \nabla\Theta^n_{|S_2}) \cdot \nu_e.$$

If the unit normal vector ν_e to e always points from S_2 to S_1, then J^n_e is well defined. For any element $S \in \mathcal{M}^n$, J^n_S stands for the jumps of $\nabla\Theta^n$ across $\partial S \backslash \partial\Omega$.

Let U be the piecewise constant extension of $\{U^n\}$ defined by $U(\cdot, 0) = U^0(\cdot)$ and $U(\cdot, t) := U^n(\cdot)$ for all $t^{n-1} < t \le t^n$ with $n \ge 1$. Let

$$U_t(\cdot, t) := \frac{U^n(\cdot) - I^nU^{n-1}(\cdot)}{\tau_n} \quad \forall\, t^{n-1} < t \le t^n,\ n \ge 1, \tag{4.5}$$

and let the *interior residual* R^n be

$$R^n := I^n f^n - U_t(\cdot, t^n).$$

5. A Posteriori Error Analysis

We now state and prove two rigorous a posteriori error estimates. Theorem 5.1 is based on Lemmas 3.1 and 3.2, that is on H^1 regularity of the solutions ψ and ϕ of the dual problems (2.3) and (2.4), and leads to Approach I. Exploiting H^2 space regularity of ψ and ϕ in the spirit of Lemma 3.3 yields Theorem 5.2 and corresponding Approach II. The derivation below parallels, and in fact extends, that in §3 but exploits Galerkin orthogonality to express negative norms of the residuals $\mathcal{R}(\psi)$ and $\mathcal{R}(\phi)$ in terms of computable quantities.

Theorem 5.1 (Approach I). *There exist constants $C_i > 0$ depending on the minimum angle of $\mathcal{M}^n$ and the space dimension d, and $C_* := A^{1/2}/\sqrt{2} + 1$ and $C_\sharp := A^{1/2} C_0 a$, such that the following a posteriori error estimates holds*

$$\|e_{\beta(u)}\|_{L^2(Q)} + \|e_u(\cdot, T)\|_{H^{-1}(\Omega)} \leq \mathcal{E}^{\mathrm{I}}(u_0, f, T, \Omega; U, h, \tau) := C_* \sum_{i \neq 4} \mathcal{E}_i + C_\sharp \mathcal{E}_4,$$

where the error indicators $\mathcal{E}_i = \mathcal{E}_i^{\mathrm{I}}$ for $0 \leq i \leq 8$ are given by

$$\begin{aligned}
\mathcal{E}_0 &:= \|u_0 - U^0\|_{H^{-1}(\Omega)} && \textit{initial error,} \\
\mathcal{E}_1^{\mathrm{I}} &:= C_1 \sum_{n=1}^N \tau_n \Big(\sum_{e \in \mathcal{B}^n} h_e \|J_e^n\|_{L^2(e)}^2 \Big)^{1/2} && \textit{jump residual,} \\
\mathcal{E}_2^{\mathrm{I}} &:= C_2 \sum_{n=1}^N \tau_n \Big(\sum_{S \in \mathcal{M}^n} h_S^2 \|R^n\|_{L^2(S)}^2 \Big)^{1/2} && \textit{interior residual,} \\
\mathcal{E}_3^{\mathrm{I}} &:= \sum_{n=1}^N \tau_n \|\nabla \beta(U^n) - \nabla I^n \beta(U^n)\|_{L^2(\Omega)} && \textit{constitutive relation,} \\
\mathcal{E}_4 &:= \Big(\sum_{n=1}^N \tau_n \|U^n - I^n U^{n-1}\|_{L^2(\Omega)}^2 \Big)^{1/2} && \textit{time residual,} \\
\mathcal{E}_5 &:= \sum_{n=1}^N \|I^n U^{n-1} - U^{n-1}\|_{H^{-1}(\Omega)} && \textit{coarsening,} \\
\mathcal{E}_6 &:= C_6 \sum_{n=1}^N \tau_n \Big(\sum_{S \in \mathcal{M}^n} h_S^4 \|\nabla R^n\|_{L^2(S)}^2 \Big)^{1/2} && \textit{quadrature,} \\
\mathcal{E}_7 &:= \sum_{n=1}^N \tau_n \|f^n - I^n f^n\|_{H^{-1}(\Omega)} && \textit{interpolation,} \\
\mathcal{E}_8 &:= \sum_{n=1}^N \int_{t^{n-1}}^{t^n} \|f - f^n\|_{H^{-1}(\Omega)} && \textit{time discretization.}
\end{aligned}$$

All indicators $\mathcal{E}_i$ can be evaluated explicitly in terms of the computed solution U, initial datum u_0, and source term f. Indicators $\mathcal{E}_0, \mathcal{E}_1^{\mathrm{I}}, \mathcal{E}_2^{\mathrm{I}}, \mathcal{E}_4$, and $\mathcal{E}_5$ are essential, and are also present for the heat equation but with different weights and cumulative effect in time [5]. The error accumulation is measured here in L^1 for $\mathcal{E}_1^{\mathrm{I}}, \mathcal{E}_2^{\mathrm{I}}, \mathcal{E}_5$ and in L^2 for $\mathcal{E}_4$, whereas it is in L^∞ for the heat equation; the latter exhibits a weaker dependence on T for $T \geq 1$. The powers of meshsize in $\mathcal{E}_1^{\mathrm{I}}$ and $\mathcal{E}_2^{\mathrm{I}}$ are smaller than those for the heat equation, namely h_e^3 and h_S^4, respectively, thereby reflecting the degenerate nature of (1.1), or equivalently the lack of H^2 space regularity of ψ and ϕ.

The remaining indicators $\mathcal{E}_3^{\mathrm{I}}, \mathcal{E}_6, \mathcal{E}_7, \mathcal{E}_8$ are not essential and could in principle be removed at the expense of complicating the implementation of (4.4). In particular, we note that $\beta(U^n) \neq I^n \beta(U^n)$ only for $S \subset \mathcal{F}^n$ and $\Theta^n{}_{|S} \neq 0$, provided β is piecewise linear, and that $\mathcal{E}_6 \leq C \mathcal{E}_2^{\mathrm{I}}$ as results from (4.1).

The following theorem yields the same weights for $\mathcal{E}_1^{\mathrm{I}}$ and $\mathcal{E}_2^{\mathrm{I}}$ as the heat equation but yet with a worse error accumulation in time. Note that this improvement comes at the expense of a smaller outermost power, namely $1/4$ instead of $1/2$.

Theorem 5.2 (Approach II). *Let* $C_\dagger := (A + 1/C_0^2)^{1/2}(2A^{1/2} + 1/\sqrt{2}), C_\ddagger := \sqrt{2}C_0 a$,

$$\delta_1 := \frac{\mathcal{E}_1^{\mathrm{II}} + \mathcal{E}_2^{\mathrm{II}} + \mathcal{E}_3^{\mathrm{II}}}{L|Q|^{1/2}},$$

and

$$\mathcal{E}_1^{\mathrm{II}} := C_1\Big(\sum_{n=1}^N \tau_n \sum_{e\in\mathcal{B}^n} h_e^3 \|J_e^n\|_{L^2(e)}^2\Big)^{1/2} \qquad \textit{jump residual,}$$

$$\mathcal{E}_2^{\mathrm{II}} := C_2\Big(\sum_{n=1}^N \tau_n \sum_{S\in\mathcal{M}^n} h_S^4 \|R^n\|_{L^2(S)}^2\Big)^{1/2} \qquad \textit{interior residual,}$$

$$\mathcal{E}_3^{\mathrm{II}} := \Big(\sum_{n=1}^N \tau_n \|\beta(U^n) - I^n\beta(U^n)\|_{L^2(\Omega)}^2\Big)^{1/2} \qquad \textit{constitutive relation.}$$

The following a posteriori error estimate holds

$$\|e_{\beta(u)}\|_{L^2(Q)} + \|e_u(\cdot,T)\|_{H^{-1}(\Omega)} \le \mathcal{E}^{\mathrm{II}}(u_0, f, T, \Omega; U, h, \tau) := C_\ddagger\Big(\mathcal{E}_0 + \sum_{i=5}^8 \mathcal{E}_i\Big)$$
$$+ C_\dagger\mathcal{E}_4 + \sqrt{2}C_\ddagger \begin{cases} L^{1/2}|Q|^{1/4}\big(\mathcal{E}_1^{\mathrm{II}} + \mathcal{E}_2^{\mathrm{II}} + \mathcal{E}_3^{\mathrm{II}}\big)^{1/2} & \textit{if } \delta_1 \le \frac{1}{C_0^2} \\ C_0\big(\mathcal{E}_1^{\mathrm{II}} + \mathcal{E}_2^{\mathrm{II}} + \mathcal{E}_3^{\mathrm{II}}\big) & \textit{if } \delta_1 > \frac{1}{C_0^2}. \end{cases}$$

5.1. Residuals. We first express the residual $\mathcal{R}(\zeta)$, for a generic function $\zeta \in H^{2,1}(Q)$, in terms of computable quantities. We notice that, since U is piecewise constant in time, summation by parts yields

$$\int_0^T \langle U, \partial_t\zeta\rangle = \sum_{n=1}^N \langle U^n, \zeta^n - \zeta^{n-1}\rangle = \langle U^N, \zeta^N\rangle - \langle U^0, \zeta^0\rangle - \sum_{n=1}^N \langle U^n - U^{n-1}, \zeta^{n-1}\rangle.$$

Moreover, in light of (4.5), the last term can be written equivalently as

$$\sum_{n=1}^N \langle U^n - U^{n-1}, \zeta^{n-1}\rangle = \int_0^T \langle U_t, \zeta\rangle$$
$$+ \sum_{n=1}^N \int_{t^{n-1}}^{t^n} \langle U_t, \zeta^{n-1} - \zeta\rangle + \sum_{n=1}^N \langle I^n U^{n-1} - U^{n-1}, \zeta^{n-1}\rangle.$$

Rearranging terms, the residual $\mathcal{R}(\zeta)$ becomes

$$\begin{aligned}\mathcal{R}(\zeta) = &\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\big(\langle R^n,\zeta\rangle - \langle\nabla I^n\beta(U^n),\nabla\zeta\rangle\big)\\ &-\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\langle U_t,\zeta^{n-1}-\zeta\rangle - \sum_{n=1}^{N}\langle I^nU^{n-1}-U^{n-1},\zeta^{n-1}\rangle\\ &+\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\langle\nabla I^n\beta(U^n)-\nabla\beta(U^n),\nabla\zeta\rangle + \sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\langle f-I^nf^n,\zeta\rangle.\end{aligned} \tag{5.1}$$

We next rewrite the discrete problem (4.4) as follows

$$\langle R^n,\varphi\rangle - \langle\nabla I^n\beta(U^n),\nabla\varphi\rangle = \langle R^n,\varphi\rangle - \langle R^n,\varphi\rangle^n \quad \forall\,\varphi\in\mathbb{V}^n,$$

and subtract this expression from the right hand side of the residual $\mathcal{R}(\zeta)$ given by (5.1). This crucial step is usually referred to as *Galerkin orthogonality*. We finally decompose the integral $\langle\nabla\Theta^n,\nabla(\zeta-\varphi)\rangle$ over all elements $S\in\mathcal{M}^n$ and next integrate by parts to obtain the equivalent expression

$$-\langle\nabla I^n\beta(U^n),\nabla(\zeta-\varphi)\rangle = \sum_{e\in\mathcal{B}^n}\langle\!\langle J_e^n,\zeta-\varphi\rangle\!\rangle_e \quad \forall\,\varphi\in\mathbb{V}^n,$$

where $\langle\!\langle\cdot,\cdot\rangle\!\rangle_e$ denotes the L^2-scalar product on $e\in\mathcal{B}^n$. Thus we easily arrive at

$$\begin{aligned}\mathcal{R}(\zeta) = &\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\sum_{e\in\mathcal{B}^n}\langle\!\langle J_e^n,\zeta-\varphi\rangle\!\rangle_e + \sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\langle R^n,\zeta-\varphi\rangle\\ &+\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\langle\nabla I^n\beta(U^n)-\nabla\beta(U^n),\nabla\zeta\rangle - \sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\langle U_t,\zeta^{n-1}-\zeta\rangle\\ &-\sum_{n=1}^{N}\langle I^nU^{n-1}-U^{n-1},\zeta^{n-1}\rangle + \sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\big(\langle R^n,\varphi\rangle - \langle R^n,\varphi\rangle^n\big)\\ &+\sum_{n=1}^{N}\int_{t^{n-1}}^{t^n}\langle f-I^nf^n,\zeta\rangle =: \mathrm{I}+\cdots+\mathrm{VII}.\end{aligned}$$

Since we need to approximate ζ under minimal regularity assumptions, we resort to the Clément interpolation operator $\Pi^n : L^2(\Omega)\to\mathbb{V}^n$, which satisfies the following local approximation properties [4], for all $\eta\in H^k(\Omega)$ and $k=1,2$,

$$\begin{aligned}\|\eta-\Pi^n\eta\|_{L^2(S)} + h_S\|\nabla(\eta-\Pi^n\eta)\|_{L^2(S)} &\le \tilde{C}h_S^k|\eta|_{H^k(\tilde{S})},\\ \|\eta-\Pi^n\eta\|_{L^2(e)} &\le \tilde{C}h_e^{k-1/2}|\eta|_{H^k(\tilde{S})},\end{aligned} \tag{5.2}$$

where $\tilde{S}$ is the union of all elements surrounding $S\in\mathcal{M}^n$ or $e\in\mathcal{B}^n$. Constants $\tilde{C}$ depend solely on the minimum angle of the mesh $\mathcal{M}^n$. An important by-product of uniform mesh regularity is that the number of adjacent simplices to a given element is bounded by a constant M independent of n, meshsizes, and time steps. Hence

$$\sum_{S\in\mathcal{M}^n} \|\eta\|^2_{L^2(\tilde S)} \le M\|\eta\|^2_{L^2(\Omega)} \quad \forall\, \eta\in L^2(\Omega). \tag{5.3}$$

This, in conjunction with (5.2) for $k=1$, yields

$$\|\nabla\Pi^n\eta\|_{L^2(\Omega)} \le (1+\tilde C M^{1/2})\|\nabla\eta\|_{L^2(\Omega)} \quad \forall\, \eta\in H^1_0(\Omega). \tag{5.4}$$

We now estimate each term I to VII separately to derive Theorems 5.1 and 5.2. We argue with $\zeta=\psi$, solution of (2.3) with $\chi\in L^2(Q)$. Similar estimates are valid for $\zeta=\phi$, solution of (2.4) with $\rho\in H^1_0(\Omega)$. We make extensive use of the a priori estimates of Lemma 2.1 (and Lemma 2.2).

Selecting $\varphi(\cdot,t)=\Pi^n\psi(\cdot,t)$ for $t^{n-1}<t\le t^n$ and using (2.6), (5.2), and (5.3), terms I and II can be bounded as follows

$$|\mathrm{I}| \le C_1\sum_{n=1}^N\int_{t^{n-1}}^{t^n}\Big(\sum_{e\in\mathcal{B}^n} h_e\|J^n_e\|^2_{L^2(e)}\Big)^{1/2}\|\nabla\psi(\cdot,t)\|_{L^2(\Omega)} \le \tfrac{1}{\sqrt2}\mathcal{E}^{\mathrm{I}}_1\|\chi\|_{L^2(Q)},$$

$$|\mathrm{II}| \le C_2\sum_{n=1}^N\int_{t^{n-1}}^{t^n}\Big(\sum_{S\in\mathcal{M}^n} h^2_S\|R^n\|^2_{L^2(S)}\Big)^{1/2}\|\nabla\psi(\cdot,t)\|_{L^2(\Omega)} \le \tfrac{1}{\sqrt2}\mathcal{E}^{\mathrm{I}}_2\|\chi\|_{L^2(Q)}.$$

Moreover, (2.6) also yields

$$|\mathrm{III}| \le \sum_{n=1}^N\int_{t^{n-1}}^{t^n}\|\nabla\beta(U^n)-\nabla I^n\beta(U^n)\|_{L^2(\Omega)}\|\nabla\psi(\cdot,t)\|_{L^2(\Omega)} \le \tfrac{1}{\sqrt2}\mathcal{E}^{\mathrm{I}}_3\|\chi\|_{L^2(Q)}.$$

Estimators $\mathcal{E}^{\mathrm{I}}_1, \mathcal{E}^{\mathrm{I}}_2$, and $\mathcal{E}^{\mathrm{I}}_3$ are those defined in Theorem 5.1; constants C_1 and C_2 depend upon $\tilde C$ in (5.2) and M in (5.3). Alternatively, if $\mathcal{E}^{\mathrm{II}}_1, \mathcal{E}^{\mathrm{II}}_2$, and $\mathcal{E}^{\mathrm{II}}_3$ are the estimators in Theorem 5.2, in light of (2.8) we can also write

$$|\mathrm{I}| \le C_1\sum_{n=1}^N\int_{t^{n-1}}^{t^n}\Big(\sum_{e\in\mathcal{B}^n} h^3_e\|J^n_e\|^2_{L^2(e)}\Big)^{1/2}|\psi(\cdot,t)|_{H^2(\Omega)} \le \tfrac12\delta^{-1/2}\mathcal{E}^{\mathrm{II}}_1\|\chi\|_{L^2(Q)},$$

$$|\mathrm{II}| \le C_2\sum_{n=1}^N\int_{t^{n-1}}^{t^n}\Big(\sum_{S\in\mathcal{M}^n} h^4_S\|R^n\|^2_{L^2(S)}\Big)^{1/2}|\psi(\cdot,t)|_{H^2(\Omega)} \le \tfrac12\delta^{-1/2}\mathcal{E}^{\mathrm{II}}_2\|\chi\|_{L^2(Q)},$$

and, upon integration by parts,

$$|\mathrm{III}| \le \sum_{n=1}^N\int_{t^{n-1}}^{t^n}\|\beta(U^n)-I^n\beta(U^n)\|_{L^2(\Omega)}\|\Delta\psi(\cdot,t)\|_{L^2(\Omega)} \le \tfrac12\delta^{-1/2}\mathcal{E}^{\mathrm{II}}_3\|\chi\|_{L^2(Q)}.$$

With the aid of (2.6) and the fact that $\psi-\psi^{n-1}=\int_{t^{n-1}}^t\partial_s\psi$, we readily obtain

$$\begin{aligned}|\mathrm{IV}| &\le \Big(\sum_{n=1}^N\tau_n\|U^n-I^nU^{n-1}\|^2_{L^2(\Omega)}\Big)^{1/2}\Big(\int_0^T\|\partial_t\psi\|^2_{L^2(\Omega)}\Big)^{1/2}\\ &\le (A+\delta)^{1/2}\mathcal{E}_4\|\chi\|_{L^2(Q)}.\end{aligned}$$

Using (2.6) again, now combined with (4.2) and (5.4), we easily deduce

$$|\mathrm{VI}| \le C_6 \sum_{n=1}^{N} \int_{t^{n-1}}^{t^n} \Big(\sum_{S \in \mathcal{M}^n} h_S^4 \|\nabla R^n\|_{L^2(S)}^2 \Big)^{1/2} \|\nabla \psi(\cdot,t)\|_{L^2(\Omega)} \le \tfrac{1}{\sqrt{2}} \mathcal{E}_6 \|\chi\|_{L^2(Q)},$$

where C_6 depends on $\tilde{C}$ and M. The remaining bounds follow directly from (2.6)

$$|\mathrm{V}| \le \sum_{n=1}^{N} \|I^n U^{n-1} - U^{n-1}\|_{H^{-1}(\Omega)} \|\nabla \psi^{n-1}\|_{L^2(\Omega)} \le \tfrac{1}{\sqrt{2}} \mathcal{E}_5 \|\chi\|_{L^2(Q)},$$

$$|\mathrm{VII}| \le \sum_{n=1}^{N} \int_{t^{n-1}}^{t^n} \|f - I^n f^n\|_{H^{-1}(\Omega)} \|\nabla \psi(\cdot,t)\|_{L^2(\Omega)} \le \tfrac{1}{\sqrt{2}} (\mathcal{E}_7 + \mathcal{E}_8) \|\chi\|_{L^2(Q)}.$$

Collecting all the previous estimates we get the following bound for $\mathcal{R}(\psi)$

$$\frac{|\mathcal{R}(\psi)|}{\|\chi\|_{L^2(Q)}} \le \tfrac{1}{\sqrt{2}} \sum_{i=5}^{8} \mathcal{E}_i + (A+\delta)^{1/2} \mathcal{E}_4 + \begin{cases} \tfrac{1}{\sqrt{2}} (\mathcal{E}_1^{\mathrm{I}} + \mathcal{E}_2^{\mathrm{I}} + \mathcal{E}_3^{\mathrm{I}}) \\ \tfrac{1}{2} \delta^{-1/2} (\mathcal{E}_1^{\mathrm{II}} + \mathcal{E}_2^{\mathrm{II}} + \mathcal{E}_3^{\mathrm{II}}). \end{cases} \tag{5.5}$$

Similarly, using (2.7) we obtain

$$\frac{|\mathcal{R}(\phi)|}{\|\nabla \rho\|_{L^2(\Omega)}} \le \sum_{i=5}^{8} \mathcal{E}_i + \tfrac{1}{\sqrt{2}} (A+\delta)^{1/2} \mathcal{E}_4 + \begin{cases} \mathcal{E}_1^{\mathrm{I}} + \mathcal{E}_2^{\mathrm{I}} + \mathcal{E}_3^{\mathrm{I}} \\ \tfrac{1}{\sqrt{2}} \delta^{-1/2} (\mathcal{E}_1^{\mathrm{II}} + \mathcal{E}_2^{\mathrm{II}} + \mathcal{E}_3^{\mathrm{II}}). \end{cases} \tag{5.6}$$

5.2. Proofs of Theorems 5.1 and 5.2. Upon adding $A^{1/2}$ times (5.5) to (5.6) and taking $\delta \to 0$, Theorem 5.1 is an easy consequence of (3.3) and (3.5).

In order to prove Theorem 5.2 we proceed as in Lemma 3.3. We take $\delta \le 1/C_0^2$ and resort to (3.7), which in the present context becomes

$$\|e_{\beta(u)}\|_{L^2(Q)} + \|e_u(\cdot,T)\|_{H^{-1}(\Omega)} \le C_\ddagger \Big(\mathcal{E}_0 + \sum_{i=5}^{8} \mathcal{E}_i \Big) + C_\dagger \mathcal{E}_4 + \tfrac{1}{\sqrt{2}} C_\ddagger q(\delta, T) \tag{5.7}$$

with

$$q(\delta,T) := q_-(T)\delta^{-1/2} + q_+(T)\delta^{1/2}, \quad q_-(T) := \mathcal{E}_1^{\mathrm{II}} + \mathcal{E}_2^{\mathrm{II}} + \mathcal{E}_3^{\mathrm{II}}, \quad q_+(T) := L|Q|^{1/2}.$$

We then realize that $\delta = \delta_1$ optimizes $q(\delta, T)$, and argue as in Lemma 3.3.

To derive the $L^\infty(0,T;H^{-1}(\Omega))$ error bound of (1.5) we finally allow T to be arbitrary and note that $\mathcal{E}_i(T)$, as well as $q_-(T)$ and $q_+(T)$, are nondecreasing in T. We thus obtain (1.5) with $\mathcal{E}(u_0, f, T, \Omega; U, h, \tau) = 2\mathcal{E}^{\mathrm{I}}$ or alternatively $2\mathcal{E}^{\mathrm{II}}$.

5.3. Approach III. The local treatment of terms I, II, and III above, in conjunction with (2.9), suggests removing the factor $\delta^{-1/2}$ away from discrete interfaces. This idea, heuristic in the sense that $L_t^2 H_x^2$ regularity of ψ and ϕ is not necessarily uniform, yields a method somewhat in between Approaches I and II. We set the estimators $\mathcal{E}_1^{\mathrm{III}}, \mathcal{E}_2^{\mathrm{III}}, \mathcal{E}_3^{\mathrm{III}}$ equal to those of Approach I near discrete interfaces and equal to those of Approach II (heat estimators), but with factor D/C_* instead of 1 away from discrete interfaces, where $D := a^{1/2}C_0$, and obtain

$$\|e_{\beta(u)}\|_{L^2(Q)} + \|e_u(\cdot,T)\|_{H^{-1}(\Omega)} \le \mathcal{E}^{\mathrm{III}}(u_0,f,T,\Omega;U,h,\tau)$$
$$:= C_*\Big(\mathcal{E}_0 + \sum_{i=5}^{8}\mathcal{E}_i\Big) + C_\sharp\mathcal{E}_4 + C_*(\mathcal{E}_1^{\mathrm{III}} + \mathcal{E}_2^{\mathrm{III}} + \mathcal{E}_3^{\mathrm{III}}).$$

To this end, we add $A^{1/2}$ times (5.5) to (5.6), let $\delta \to 0$, replace $\|D^2\zeta\|_{L^2(Q_0)}$ by $\|\Delta\zeta\|_{L^2(Q_0)}$ for $\zeta = \psi, \phi$ and use (2.9), and ultimately pretend that the lower bound of b is a instead of $a_0 = 0$. If ψ and ϕ were computable, we could keep these H_x^2 norms and evaluate them locally, in which case they could be viewed as weights.

6. Adaptive Algorithm

The error estimators $\mathcal{E}(u_0,f,T,\Omega;U,h,\tau)$ of §5 entail an L^1 or L^2 norm in time, which is impractical in that the entire evolution history would be needed to control $\mathcal{E}$. We resort here to an L^∞ norm in time, and explain our error equidistribution strategy, which is based upon minimizing the spatial degrees of freedom for a uniform error distribution in time. A similar strategy is derived in [1] for a linear elliptic problem. We then fully discuss element error indicators, and all tests necessary for mesh and time step admissibility.

6.1. Mesh Design. Let $\mathcal{M}(t)$ be a time-dependent mesh with variable meshsize $h(x,t)$ and let $\tau(t)$ be the underlying variable time-step. Since $h^d\tau$ is proportional to the volume of a generic space-time finite element, then the computational complexity of (4.4) can be accounted for by the total number of degrees of freedom

$$M^* := \int_Q h(x,t)^{-d}\tau(t)^{-1}\sigma(x,t)\,dxdt = \int_0^T \tau(t)^{-1}M(t)\,dt, \tag{6.1}$$

where $0 < \sigma_- \le \sigma(x,t) \le \sigma_+$ is a local measure of element shape regularity and

$$M(t) := \int_\Omega h(x,t)^{-d}\sigma(x,t)\,dx \tag{6.2}$$

stands for the cardinality of $\mathcal{M}(t)$. Let $\mathcal{E}$ be a generic a priori error of the form

$$\mathcal{E} = \int_0^T \mathcal{E}(t)dt = \int_Q \big(h(x,t)^\alpha + \tau(t)^\beta\big)E(x,t)\,dxdt. \tag{6.3}$$

Given an error tolerance ε, we then pose the following question: optimize h and τ for an error distribution

$$\mathcal{E} = \varepsilon. \tag{6.4}$$

We have to minimize (6.1) subject to the constraint (6.4). This constrained optimization problem is equivalent to seeking a saddle point of the functional

$$\mathcal{L}(h,\tau,\lambda) := \int_Q h^{-d}\tau^{-1}\sigma - \lambda\Big(\varepsilon - \int_Q (h^\alpha + \tau^\beta)E\Big),$$

with Lagrange multiplier λ. Differentiation with respect to h and τ yields

$$0 = \int_Q \zeta\, \partial_h \mathcal{L}(h,\tau,\lambda) = \int_0^T \int_\Omega (-\sigma d h^{-(d+1)}\tau^{-1} + \lambda\alpha h^{\alpha-1}E)\zeta,$$

$$0 = \int_Q \eta\, \partial_\tau \mathcal{L}(h,\tau,\lambda) = \int_0^T \eta \int_\Omega (-\sigma h^{-d}\tau^{-2} + \lambda\beta\tau^{\beta-1}E),$$

where $\zeta = \zeta(x,t)$ and $\eta = \eta(t)$ are smooth functions. Hence

$$h^\alpha E = \frac{d}{\lambda\alpha} h^{-d}\tau^{-1}\sigma, \quad \tau^\beta \int_\Omega E = \frac{1}{\lambda\beta}\tau^{-1}M(t). \tag{6.5}$$

These relations, in conjunction with (6.1) and (6.3), yield

$$\varepsilon = \frac{d}{\lambda\alpha}\int_Q h^{-d}\tau^{-1}\sigma + \frac{1}{\lambda\beta}\int_0^T \tau^{-1}M(t) = \frac{d\beta+\alpha}{\lambda\alpha\beta} M^*.$$

We thus deduce the following expression for the Lagrange multiplier

$$\lambda = \frac{d\beta+\alpha}{\alpha\beta}\,\frac{M^*}{\varepsilon}.$$

We conclude from (6.5) that the optimization procedure consists of equidistributing the local space-time errors according to the recipes

$$M^*\sigma^{-1}h^{d+\alpha}\tau E = \frac{d\beta}{\alpha+d\beta}\,\varepsilon, \quad \frac{M^*}{M(t)}\tau^{\beta+1}\int_\Omega E = \frac{\alpha}{\alpha+d\beta}\,\varepsilon. \tag{6.6}$$

A serious difficulty of (6.6) is that M^* is never available for all times so as to make the optimization feasible.

A different and more stringent objective would be to equidistribute pointwise discretization errors in (6.3). This leads to the local requirements

$$T|\Omega|h^\alpha E = \tfrac{1}{2}\varepsilon, \quad T\tau^\beta \int_\Omega E = \tfrac{1}{2}\varepsilon, \tag{6.7}$$

which corresponds to having a constant integrand in (6.3) and yields (6.4). This method does not require M^*, and can thus be implemented. If we further impose $\tau = T/N$ to be constant, we end up with the method chosen in [11] for a priori mesh design.

Upon combining (6.6) and (6.7), we get a third strategy that can still be implemented: we optimize the spatial degrees of freedom for a uniform error distribution $\mathcal{E}(t)$ in time. Minimizing (6.2) for each $0 < t \le T$, subject to the constraints

$$T\int_\Omega h(x,t)^\alpha E(x,t)\,dx = \tfrac{1}{2}\varepsilon, \quad T\tau(t)^\beta \int_\Omega E(x,t)\,dx = \tfrac{1}{2}\varepsilon, \tag{6.8}$$

results in the following restrictions on h and τ, which are intermediate between (6.6) and (6.7),

$$TM(t)\sigma(x,t)^{-1}h(x,t)^{d+\alpha}E(x,t) = \tfrac{1}{2}\varepsilon, \quad T\tau(t)^\beta \int_\Omega E(x,t)\,dx = \tfrac{1}{2}\varepsilon. \tag{6.9}$$

We adopt such a viewpoint here, and discuss its implementation next. Note first that E would depend on the discrete solution, and so implicitly on h and τ, if $\mathcal{E}$ in (6.3) were an a posteriori error estimator. So the above analysis is a priori.

6.2. Equidistribution Strategy. The *equidistribution strategy* (ES) is an iterative procedure that in the kth step improves upon a mesh density $h_k(x,t)$ by means of the following two opposite operations, based on the first constraint in (6.9). Let $0 < \Lambda < 1$ be a refinement factor and $M_k = \int_\Omega h_k^{-d}\sigma$ from (6.2) with $h = h_k$.

(a) *Refinement.* Set $h_{k+1} = \Lambda h_k$ for all elements satisfying

$$\rho_k := 2TM_k\sigma^{-1}h_k^{d+\alpha}E > \varepsilon. \tag{6.10}$$

(b) *Coarsening.* Set $h_{k+1} = \Lambda^{-1}h_k$ for all elements verifying

$$\rho_k < \Lambda^{2d+\alpha}\varepsilon. \tag{6.11}$$

In practice, more elements will be refined and fewer will be coarsened to preserve mesh conformity. In fact, every local mesh modification involves also adjacent elements. Since

$$M_{k+1} = \int_\Omega h_{k+1}^{-d}\sigma \le \Lambda^{-d}\int_\Omega h_k^{-d}\sigma = \Lambda^{-d}M_k,$$

we see that $\rho_{k+1} \le \Lambda^\alpha\rho_k$ for refined elements whereas $\rho_{k+1} \le \Lambda^{-2d-\alpha}\rho_k < \varepsilon$ for coarsened elements. We conclude that coarsened elements will not be candidates for refinement in the $(k+1)$th step, because $\rho_{k+1} < \varepsilon$ and so test (a) fails, and that

$$\rho_{k+1} \le \min\big(\varepsilon, \Lambda^\alpha\rho_k\big).$$

The latter guarantees convergence of ES in a finite number of steps.

6.3. Element Indicators. We now introduce error indicators of the form (6.8) with density $E(x,t)$ computable element by element and denoted by $E^n(S)$ for $S \in \mathcal{M}^n$ and $t = t^n$. We first replace the global norm $H^{-1}(\Omega)$ in $\mathcal{E}_0, \mathcal{E}_5, \mathcal{E}_7, \mathcal{E}_8$ by the $L^2(\Omega)$ norm scaled with the smallest diameter d_Ω of Ω, which is the Poincaré constant. We next show how to determine computable global indicators E_0 (initial error), $E_{\tau,1}, E_{\tau,2}$ (temporal errors) and $E_{h,1}, E_{h,2}^{\mathrm{k}}$ (spatial errors) satisfying

$$\begin{gathered}\mathcal{E}_0 \le E_0,\\ \mathcal{E}_4 \le E_{\tau,1}, \quad \mathcal{E}_8 \le E_{\tau,2},\\ \mathcal{E}_5 + \mathcal{E}_6 + \mathcal{E}_7 \le E_{h,1}, \quad \mathcal{E}_1^{\mathrm{k}} + \mathcal{E}_2^{\mathrm{k}} + \mathcal{E}_3^{\mathrm{k}} \le E_{h,2}^{\mathrm{k}},\end{gathered} \tag{6.12}$$

where $k = \mathrm{I}, \mathrm{II}$ stands for the approach. Each of these errors E_p is obtained from *element indicators* $E_p^n(S)$ via

$$E_p = \max_{1\le n\le N}\Big(\sum_{S\in\mathcal{M}^n} E_p^n(S)\Big)^{1/2},$$

where the max is superfluous for $p=0$. The element indicators are given by

$$
\begin{aligned}
E_0(S) &:= d_\Omega^2 |S| \|u_0 - U^0\|_{L^\infty(S)}^2 && \textit{local initial error},\\
E_{\tau,1}^n(S) &:= T\|U^n - I^n U^{n-1}\|_{L^2(S)}^2 && \textit{local time residual},\\
E_{\tau,2}^n(S) &:= T^2 d_\Omega^2 \|f - f^n\|_{L^\infty(t^{n-1},t^n;L^2(S))}^2 && \textit{local time discretization},\\
E_{h,1}^n(S) &:= 3T^2 \Big(d_\Omega^2 \tau_n^{-2} \|I^n U^{n-1} - U^{n-1}\|_{L^2(S)}^2 && \textit{local coarsening}\\
&\qquad + C_6^2 h_S^4 \|\nabla R^n\|_{L^2(S)}^2 && \textit{local quadrature}\\
&\qquad + d_\Omega^2 \|f^n - I^n f^n\|_{L^2(S)}^2 \Big) && \textit{local interpolation},\\
E_{h,2}^{\mathrm{I},n}(S) &:= 3T^2 \Big(\tfrac{1}{2} C_1^2 h_S \|J_S^n\|_{L^2(\partial S)}^2 && \textit{local jump residual } \mathrm{I}\\
&\qquad + C_2^2 h_S^2 \|R^n\|_{L^2(S)}^2 && \textit{local interior residual } \mathrm{I}\\
&\qquad + \|\nabla\beta(U^n) - \nabla I^n \beta(U^n)\|_{L^2(S)}^2 \Big) && \textit{local constitutive relation } \mathrm{I},\\
E_{h,2}^{\mathrm{II},n}(S) &:= 3T \Big(\tfrac{1}{2} C_1^2 h_S^3 \|J_S^n\|_{L^2(\partial S)}^2 && \textit{local jump residual } \mathrm{II}\\
&\qquad + C_2^2 h_S^4 \|R^n\|_{L^2(S)}^2 && \textit{local interior residual } \mathrm{II}\\
&\qquad + \|\beta(U^n) - I^n \beta(U^n)\|_{L^2(S)}^2 \Big) && \textit{local constitutive relation } \mathrm{II}.
\end{aligned}
$$

Since $u_0 \in W^{1,\infty}(\Omega\backslash F_0)$, interpolation gives $\|u_0 - U^0\|_{L^\infty(S)} \le C h_S \|\nabla u_0\|_{L^\infty(S)} = O(h_S)$ away from F_0. If S intersects F_0, instead, then u_0 has a jump of size L which yields $\|u_0 - U^0\|_{L^\infty(S)} = O(L)$. In both instances, the ensuing orders match that of the local jump residual I, which indicates that $\mathcal{M}_0$ will not be drastically modified. If in addition $u_0 \in W^{2,\infty}(\Omega\backslash F_0)$, then $\|u_0 - U^0\|_{L^\infty(S)} \le C h_S^2 \|D^2 u_0\|_{L^\infty(S)} = O(h_S^2)$ away from F_0, which yields the order $O(h_S^{4+d})$ of the local jump residual II.

According to §5.3, we introduce the local indicators for heuristic Approach III which is based on combining $E_{h,2}^{\mathrm{I},n}(S)$ and $E_{h,2}^{\mathrm{II},n}(S)$. With $D = a^{1/2} C_0$ we set

$$
E_{h,2}^{\mathrm{III},n}(S) := \begin{cases} E_{h,2}^{\mathrm{I},n}(S) & \text{if } S \subset \mathcal{T}^n \\ \frac{D^2}{C_*^2} E_{h,2}^{\mathrm{II},n}(S) & \text{if } S \not\subset \mathcal{T}^n. \end{cases}
$$

A simple but tedious calculation shows that (6.12) is fulfilled, and the following inequality is valid for Approaches k = I and III (heuristic for Approach III)

$$
\mathcal{E}^{\mathrm{k}}(u_0, f, T, \Omega; U, h, \tau) \le E^{\mathrm{k}} := C_*(E_0 + E_{\tau,2} + E_{h,1} + E_{h,2}^{\mathrm{k}}) + C_\sharp E_{\tau,1}. \tag{6.13}
$$

On the other hand, if $\delta \le 1/C_0^2$ and $q(\delta) := E_{h,2}^{\mathrm{II}} \delta^{-1/2} + L T^{1/2} |\Omega|^{1/2} \delta^{1/2}$, then (5.7) and (6.12) lead to

$$
\mathcal{E}^{\mathrm{II}}(u_0, f, T, \Omega; U, h, \tau) \le E^{\mathrm{II}} := C_\ddagger(E_0 + E_{\tau,2} + E_{h,1}) + C_\dagger E_{\tau,1} + \tfrac{1}{\sqrt{2}} C_\ddagger q(\delta). \tag{6.14}
$$

6.4. Adaptive Strategy. Given an error tolerance ε, the objective is to adaptively select time steps τ and mesh densities h so that the spacial degrees of freedom are optimized for a uniform error distribution in time and

$$E^{\mathrm{k}}(u_0, f, T, \Omega; U, h, \tau) \le \varepsilon. \tag{6.15}$$

To this end we equidistribute the local estimators of §6.3 according to (6.9) via ES, and use *bisection* to perform refinement/coarsening operations (6.10) and (6.11). Bisection creates *compatible* consecutive meshes, extends naturally from 2D to 3D, and is handy for combined refinement/coarsening operations; we refer to [2].

Given refinement parameters $\Gamma > 0$ and coarsening parameters $\gamma > 0$ satisfying

$$\Gamma_0 + \Gamma_\tau + \Gamma_h \le 1, \quad \gamma_\tau < \Gamma_\tau, \quad \gamma_h < \Gamma_h,$$

then time steps and mesh densities for Approaches k = I and III are reduced until

$$\begin{aligned} C_* E_0 &\le \Gamma_0\,\varepsilon, \\ \tfrac12 \gamma_\tau\,\varepsilon \le C_\sharp E_{\tau,1}, C_* E_{\tau,2} &\le \tfrac12 \Gamma_\tau\,\varepsilon, \\ \tfrac12 \gamma_h\,\varepsilon \le C_* E_{h,1}, C_* E^{\mathrm{k}}_{h,2} &\le \tfrac12 \Gamma_h\,\varepsilon. \end{aligned} \tag{6.16}$$

In fact, from (6.13) we readily obtain (6.15). Achieving (6.15) for Approach II is more problematic in that the optimal choice of δ in (6.14) turns out to be

$$\delta_2 := \frac{E^{\mathrm{II}}_{h,2}}{LT^{1/2}|\Omega|^{1/2}},$$

which is not a computable quantity as it involves a maximum over all time steps. To overcome this difficulty we restrict the meshsize and time step as follows

$$\begin{aligned} C_\ddagger E_0 &\le \Gamma_0\,\varepsilon, \\ \tfrac12 \gamma_\tau\,\varepsilon \le C_\dagger E_{\tau,1}, C_\ddagger E_{\tau,2} &\le \tfrac12 \Gamma_\tau\,\varepsilon, \\ \tfrac12 \gamma_h\,\varepsilon \le C_\ddagger E_{h,1} &\le \tfrac12 \Gamma_h\,\varepsilon, \\ \tfrac14 G\gamma_h\,\varepsilon \le C_0^2 a E^{\mathrm{II}}_{h,2} &\le \tfrac14 G\Gamma_h\,\varepsilon, \end{aligned} \tag{6.17}$$

with

$$G := \min\Big(1, \frac{\Gamma_h\,\varepsilon}{4aLT^{1/2}|\Omega|^{1/2}}\Big).$$

We claim this yields (6.15). In fact, if $G < 1$ then

$$\delta_2 = \frac{E^{\mathrm{II}}_{h,2}}{LT^{1/2}|\Omega|^{1/2}} \le \frac{\Gamma_h^2\,\varepsilon^2}{16 C_0^2 a^2 L^2 T |\Omega|} = \tfrac{G^2}{C_0^2} < \tfrac{1}{C_0^2},$$

whence

$$\tfrac{1}{\sqrt2} C_\ddagger q(\delta_2) = 2 C_0 a L^{1/2} T^{1/4} |\Omega|^{1/4} (E^{\mathrm{II}}_{h,2})^{1/2} \le \tfrac12 \Gamma_h\,\varepsilon.$$

On the other hand, if $G = 1$, then with $\delta = 1/C_0^2$ in (6.14) we deduce

$$\tfrac{1}{\sqrt2} C_\ddagger q\big(\tfrac{1}{C_0^2}\big) = C_0^2 a E^{\mathrm{II}}_{h,2} + aLT^{1/2}|\Omega|^{1/2} \le \tfrac14 \Gamma_h\,\varepsilon + \tfrac14 \Gamma_h\,\varepsilon = \tfrac12 \Gamma_h\,\varepsilon.$$

We now describe an algorithm based on (6.16) for Approach I (and III). The implementation of Approach II is similar, but based on (6.17) instead. Let M^n denote the cardinality of $\mathcal{M}^n$ at any step of ES.

6.4.1. Initial Mesh. Given a coarse mesh, ES bisects all $S \in \mathcal{M}^0$ such that

$$C_*^2 E_0(S) > \frac{\Gamma_0^2\, \varepsilon^2}{M^0},$$

as suggested by (6.10) without T. We interpolate u_0 by means of (4.3) to find U^0 (and Θ^0) and estimate $E_0(S)$ as $E_0(S) \approx h_S^2 |S| \|\nabla U^0\|_{L^\infty(S)}^2$.

6.4.2. Time Step Selection. Starting with $\tau_n = \tau_{n-1}$, the algorithm checks whether

$$E_\tau^n := 4 \max \Big(C_\sharp^2 \sum_{S \in \mathcal{M}^n} E_{\tau,1}^n(S), C_*^2 \sum_{S \in \mathcal{M}^n} E_{\tau,2}^n(S) \Big) > \Gamma_\tau^2\, \varepsilon^2, \quad E_\tau^n < \gamma_\tau^2\, \varepsilon^2.$$

In the first case τ_n is reduced, whereas in the second one (corresponding to τ_n being too small) τ_n is accepted but the initial guess for the next time step size is enlarged.

6.4.3. Mesh Size Selection. Starting from $\mathcal{M}^n = \mathcal{M}^{n-1}$, ES checks whether

$$E_h^{\mathrm{I},n}(S) := 4C_*^2 \max \left(E_{h,1}^n(S), E_{h,2}^{\mathrm{I},n}(S) \right) > \frac{\Gamma_h^2\, \varepsilon^2}{M^n}, \quad E_h^{\mathrm{I},n}(S) < \frac{\gamma_h^2\, \varepsilon^2}{M^n},$$

as suggested by (6.10) and (6.11). Then refinement and coarsening operations are performed accordingly, with the precaution of choosing $\gamma_h \ll \Gamma_h$ properly to prevent ES from alternating such operations over the same elements.

6.4.4. Flow Chart. We summarize Approach I in the following flow diagram:

start with $\mathcal{M}^1 = \mathcal{M}^0$, τ^1, U^0
for $n \geq 1$, while $t^{n-1} < T$
  (1) set $t^n = t^{n-1} + \tau^n$
    solve for U^n, Θ^n
    compute error estimators E_τ^n and $E_h^{\mathrm{I},n}(S)$ for $S \in \mathcal{M}^n$
    if $E_\tau^n > \Gamma_\tau^2 \varepsilon^2$, reduce τ^n and goto (1)
  (2) for every $S \in \mathcal{M}^n$
      if $E_h^{\mathrm{I},n}(S) > \Gamma_h^2 \varepsilon^2 / M^n$, refine S
      if $E_h^{\mathrm{I},n}(S) < \gamma_h^2 \varepsilon^2 / M^n$, coarsen S if possible
    if the mesh was changed
      solve again for U^n, Θ^n
      compute error estimators
      if $E_\tau^n > \Gamma_\tau^2 \varepsilon^2$, reduce τ^n and goto (1)
      if $E_h^{\mathrm{I},n} > \Gamma_h^2 \varepsilon^2$, goto (2)
    accept U^n, Θ^n and $\mathcal{M}^n$, τ^n
    set $\mathcal{M}^{n+1} = \mathcal{M}^n$, $\tau^{n+1} = \tau^n$
    if $E_\tau^n < \gamma_\tau^2 \varepsilon^2$, enlarge τ^{n+1}

After each iteration of ES, both U^n and Θ^n are recalculated on the new mesh using the new time step size. To minimize the overall computational cost, a compromise is reached between the minimization of the number of mesh elements (degrees of freedom) and iterations of ES. ES stops iterating as soon as

$$E_h^{\mathrm{I},n} := \sum_{S\in\mathcal{M}^n} E_h^{\mathrm{I},n}(S) \le \Gamma_h^2\,\varepsilon^2$$

is fulfilled, thereby allowing some elements to violate the local error tolerance. Consequently, discretization errors might not be equidistributed. As implemented, and shown in the flow diagram, at least one mesh modification per time step is performed in order to permit elements near the moving interface to be refined, even if the global error bound is already fulfilled by the old mesh.

6.5. Convergence. We elaborate here on convergence of the algorithm of §6.4; a more comprehensive study will be given elsewhere [13]. For all $1\le n\le N$, we assume the following discrete a priori estimates, which are slightly stronger than the natural bounds

$$\|\nabla\Theta^n\|_{L^2(\Omega)} = O(1), \quad |\mathcal{F}^n\backslash\mathcal{Z}^n| = o(1),$$
$$\|U^n - I^nU^{n-1}\|^2_{L^2(\Omega)} + \tau_n^{-1}\|\Theta^n - I^n\Theta^{n-1}\|^2_{L^2(\Omega)} + \sum_{S\subset\Omega\backslash\mathcal{F}^n} h_S\|J_S^n\|^2_{L^2(\partial S)} = o(1),$$

where $\mathcal{Z}^n := \cup\{S\in\mathcal{M}^n : \Theta^n_{|S} = 0\}$ is the numerical mush. We now examine a few relevant terms, using the hyperbolic and parabolic relations

$$r_H^n := \max_{S\subset\mathcal{T}^n}\frac{h_S^2}{\tau_n^2} \le C_H, \quad r_P^n := \max_{S\subset\Omega\backslash\mathcal{T}^n}\frac{h_S^2}{\tau_n} \le C_P. \tag{6.18}$$

We consider first the jump residual, for which it suffices to estimate the contribution of $\mathcal{F}^n\backslash\mathcal{Z}^n$, namely

$$\sum_{e\subset\mathcal{F}^n\backslash\mathcal{Z}^n} h_e\|J_e^n\|^2_{L^2(e)} \le C\|\nabla\Theta^n\|^2_{L^2(\mathcal{F}^n\backslash\mathcal{Z}^n)} = o(1).$$

The same calculation applies to the constitutive relation. For the interior residual, we decompose the integral over $\mathcal{T}^n$ and the complement, where we use $U^n - I^nU^{n-1} = \Theta^n - I^n\Theta^{n-1}$, to deduce

$$\begin{aligned}\sum_{S\in\mathcal{M}^n}\frac{h_S^2}{\tau_n^2}\|U^n - I^nU^{n-1}\|^2_{L^2(S)} &\le r_H^n\|U^n - I^nU^{n-1}\|^2_{L^2(\Omega)}\\ &+ \frac{r_P^n}{\tau_n}\|\Theta^n - I^n\Theta^{n-1}\|^2_{L^2(\Omega)} = o(1).\end{aligned}$$

Finally, $\|U^n - I^nU^{n-1}\|_{L^2(\Omega)} = o(1)$ also controls the time residual. This demonstrates that the a posteriori estimators of §6.4 tend to 0 as both meshsize and time-step approach 0. Therefore the goal (6.15) is achievable. On the other hand, (6.18) reflects the hyperbolic structure of the interface as well as the parabolic structure of the problem elsewhere. We refer to [11] for a similar observation.

7. Simulations

We present results of the adaptive method for the model problem from [12] with known exact solution

$$u(x,y,t) := \begin{cases} 0.75\big(6r^2 - \frac{1}{6}\big) & \text{if } r < \frac{1}{6}, \\ 1 + \big(1.5 - \alpha'(t)\frac{y-\alpha(t)}{r}\big)\big(r - \frac{1}{6}\big) & \text{if } r \geq \frac{1}{6}, \end{cases} \tag{7.1}$$

where $r := \big((x - \frac{1}{3})^2 + (y - \alpha(t))^2\big)^{1/2}$. The interface is a circle of constant radius $r = \frac{1}{6}$ centered at $(\frac{1}{3}, \alpha(t))$ with $\alpha(t) := 0.5 + 0.1\sin(12.5t)$, and thus oscillates in the vertical direction with highly varying normal velocity. Initial datum, boundary values, and heat source f are directly computed from (7.1). The domain is $\Omega = (0.1, 0.7) \times (0.2, 0.8)$ and the final time is $T = 1$.

Note that f is discontinuous across the true interface. Therefore a simple minded use of $I^n f$ leads to large errors which dominate the local indicators. However a discontinuous f is not realistic in practice, being just the cost of having a simple exact solution. To examine the essense of the proposed methodology, a special quadrature is used for those triangles crossed by the exact free boundary: their intersection is determined first and then separate quadrature used in the resulting quadrilateral and triangle. The nonhomogeneous Dirichlet boundary condition is handled in the standard fashion since the interface does not touch the fixed boundary.

Several simulations were carried out with error tolerances $\varepsilon = 20, 14, 10, 7, 5$. The following parameters were used for partitioning the total error into initial, temporal, and spatial components

$$\Gamma_0 = \Gamma_\tau = 0.2, \quad \Gamma_h = 0.6, \quad \gamma_\tau = 0.155, \quad \gamma_h = 0.2683 \text{ (I)},\ 0.1897 \text{ (II, III)}.$$

Since the expected local coarsening error is proportional to the power of meshsize, and such a power is smaller for Approach I than for II and III, more coarsening (larger γ_h) is allowed for Approach I. The results are summarized in Figures 7.1 to 7.4.

Figure 7.1 shows the meshes at time $t = 0.3$ produced by Approaches I with $\varepsilon = 14\,(10)$, and II and III with $\varepsilon = 10\,(7)$. Meshes from left to right correspond to Approaches I, II, and III. The true errors for all three pictures on top and bottom are similar. Approach I leads to more triangles than II and III because of the lower power of meshsize in the estimators. Approaches II and III are comparable.

Figure 7.2 shows three meshes generated by Approach II for quite different interface velocities. In the left and middle pictures the velocity is highest, and the interface is moving upwards ($t = 0.5$) and downwards ($t = 0.765$) as reflected in the isotherms (the interface is enhanced with a bold line). The rightmost picture corresponds to vanishing velocity and so to circular isotherms ($t = 0.876$). Higher interface velocities lead to higher refinement near the free boundary, because velocity is proportional to the temperature gradient jump. Higher velocities yield also larger temperature gradients in the liquid phase and corresponding additional mesh refinement. The triangle counts are 1670, 1734, and 1366, respectively.

Figure 7.3 illustrates the effect of γ_h for Approach I with the interface moving downwards. The ratio Γ_h/γ_h cannot be close to 1 nor too small. The former situation leads to mesh oscillations due to repeated coarsening/refinement operations

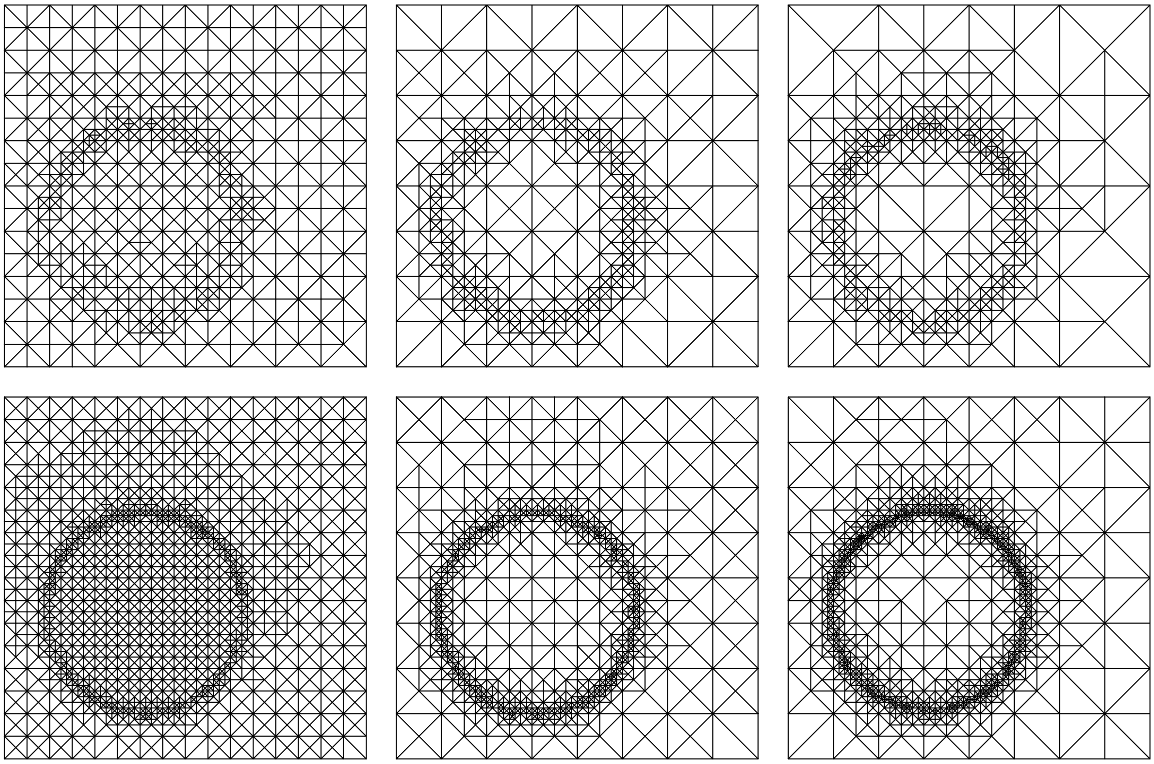

FIGURE 7.1. Approaches I with $\varepsilon = 14(10)$ and II and III with $\varepsilon = 10(7)$; Triangle count: I 1016 (2747), II 700 (1592), III 786 (2129)

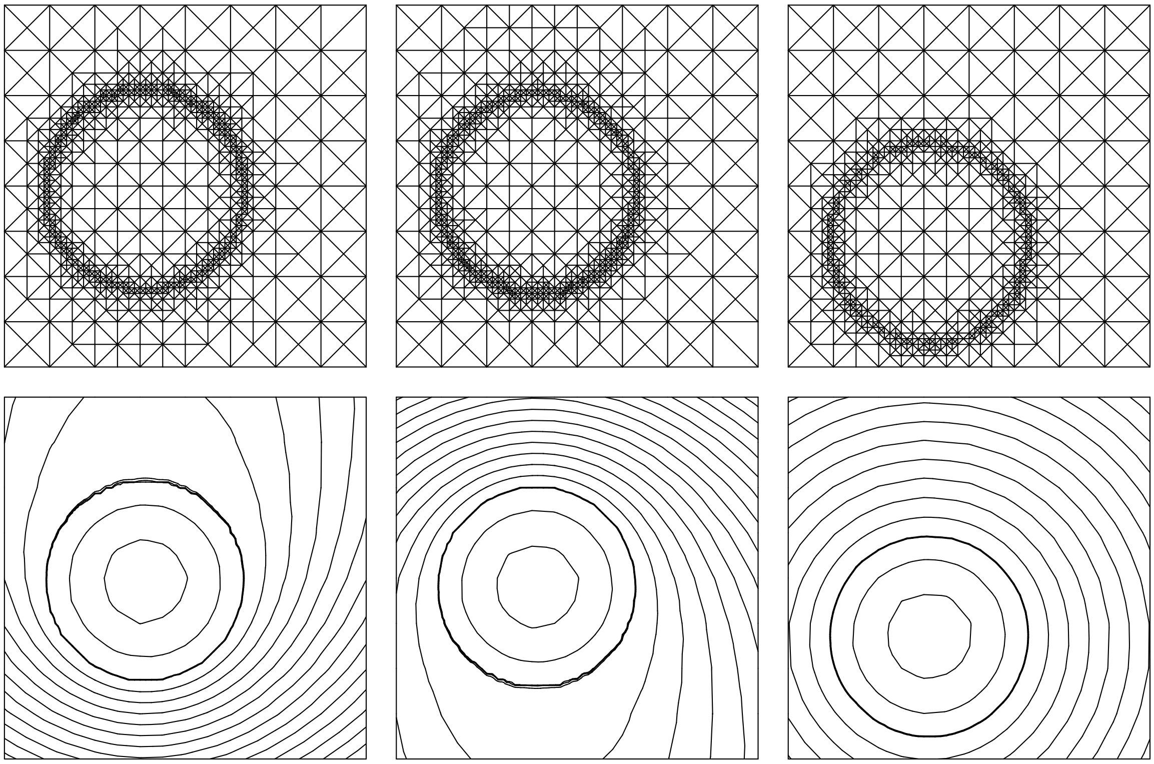

FIGURE 7.2. Meshes and isotherms for Approach II with $\varepsilon = 7$

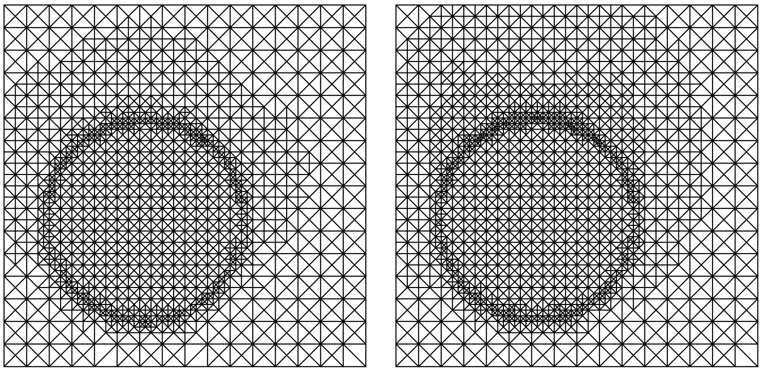

FIGURE 7.3. Effect of coarsening for Approach I and $\varepsilon = 10$, $\gamma_h/\Gamma_h = \sqrt{0.2}(\sqrt{0.1})$.

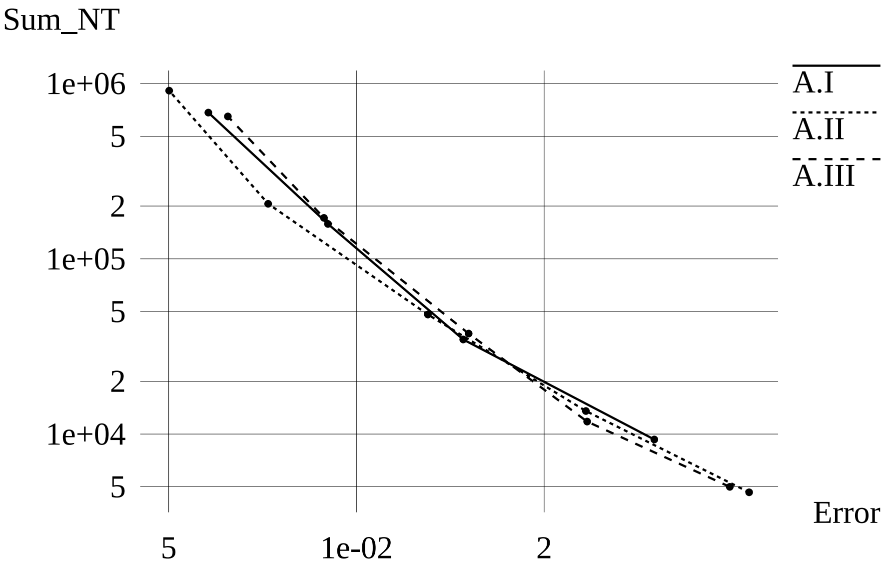

FIGURE 7.4. Total element count vs true error for Approaches I, II, and III.

over the same elements. A small ratio instead leads to meshes with unnecessary triangles, which thereby reflect the evolution history. This is an undesirable event.

In Figure 7.4 we provide a quantitative comparison: a plot of the total element count (added over all time steps), which is proportional to the total amount of work, as a function of the true error. Approach II performs slightly better in the range tested.

REFERENCES

1. I. Babuška and W.C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal. **15** (1978), 736–754.
2. E. Bänsch, *Local mesh refinement in 2 and 3 dimensions*, IMPACT Comput. Sci. Engrg. **3** (1991), 181–191.
3. P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam,

1978.
4. Ph. Clément, *Approximation by finite element functions using local regularization*, RAIRO Anal. Numér. **9** (1975), 77–84.
5. K. Eriksson and C. Johnson, *Adaptive finite element methods for parabolic problems I: a linear model problem*, SIAM J. Numer. Anal. **28** (1991), 43–77.
6. ———, *Adaptive finite element methods for parabolic problems II: optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$*, SIAM J. Numer. Anal. **32** (1995), 706–740.
7. ———, *Adaptive finite element methods for parabolic problems IV: nonlinear problems*, SIAM J. Numer. Anal. **32** (1995), 1750–1763.
8. A. Friedman, *Variational Principles and Free Boundary Problems*, Wiley, New York, 1982.
9. O.A. Ladyzenskaja, V. Solonnikov, and N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, vol. TMM 23, AMS, Providence, 1968.
10. R.H. Nochetto, *Error estimates for multidimensional singular parabolic problems*, Japan J. Indust. Appl. Math. **4** (1987), 111–138.
11. R.H. Nochetto, M. Paolini, and C. Verdi, *An adaptive finite elements method for two-phase Stefan problems in two space dimensions. Part I: stability and error estimates. Supplement*, Math. Comp. **57** (1991), 73–108, S1–S11.
12. ———, *An adaptive finite elements method for two-phase Stefan problems in two space dimensions. Part II: implementation and numerical experiments*, SIAM J. Sci. Statist. Comput. **12** (1991), 1207–1244.
13. R.H. Nochetto, A. Schmidt, and C. Verdi, *Adapting meshes and time-steps for phase change problems*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. (9) Mat. Appl. **8** (1997), 273–292.
14. ———, *Adaptive solution of phase change problems over unstructured tetrahedral meshes*, Grid Generation and Adaptive Algorithms (M. Luskin *et al.* eds.), Springer Verlag, New York (to appear).
15. P.A. Raviart, *The use of numerical integration in finite element methods for solving parabolic equations*, Topics in Numerical Analysis (J.J.H. Miller ed.), Academic Press, London, 1973, pp. 233–264.

Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA.
E-mail address: `rhn@math.umd.edu`

Institut für Angewandte Mathematik, Universität Freiburg, 79106 Freiburg, Germany.
E-mail address: `alfred@mathematik.uni-freiburg.de`

Dipartimento di Matematica, Università di Milano, 20133 Milano, Italy.
E-mail address: `verdi@paola.mat.unimi.it`