Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Adaptive Space-Time Finite Element Methods for Parabolic Optimization Problems

D. Meidner, B. Vexler

RICAM-Report 2006-08

ADAPTIVE SPACE-TIME FINITE ELEMENT METHODS FOR PARABOLIC OPTIMIZATION PROBLEMS

DOMINIK MEIDNER[†] AND BORIS VEXLER[‡]

Abstract. In this paper we derive a posteriori error estimates for space-time finite element discretization of parabolic optimization problems. The provided error estimates assess the discretization error with respect to a given quantity of interest and separate the influence of different parts of the discretization (time, space, and control discretization). This allows to set up an efficient adaptive algorithm which successively improves the accuracy of the computed solution by construction of locally refined meshes for time and space discretizations.

Key words. parabolic equations, optimal control, parameter identification, a posteriori error estimation, mesh refinement

AMS subject classifications. 65N30, 49K20, 65M50, 35K55

1. Introduction. In this paper we develop an adaptive algorithm for efficient solution of time-dependent optimization problems governed by parabolic partial differential equations. The optimization problems are formulated in a general setting including optimal control as well as parameter identification problems. Both, time and space discretization of the state equation are based on the finite element method as proposed e.g. in [10, 11]. In [2] we have shown that this type of discretization allows for a natural translation of the optimality conditions from the continuous to the discrete level. This gives rise to exact computation of the derivatives required in the optimization algorithms on the discrete level.

The main goal of this paper is to derive a posteriori error estimates which assess the error between the solution of the continuous and the discrete optimization problem with respect to a given quantity of interest. This quantity of interest may coincide with the cost functional or expresses another goal for the computation. In order to set up an efficient adaptive algorithm we will separate the influence of the time and space discretizations on the error in the quantity of interest. This allows to ballance different types of error and successively to improve the accuracy by construction of locally refined meshes for time and space discretizations.

The use of adaptive techniques based on a posteriori error estimation is well accepted in the context of finite element discretization of partial differential equations, see e.g. [9, 26, 3]. In the last years the application of these techniques is also investigated for optimization problems governed by partial differential equations. Energy-type error estimators for the error in the state, control and the adjoint variable are developed in [19, 20] in the context of distributed elliptic optimal control problems subject to pointwise control constraints. Recently, these techniques are also applied in the context of optimal control problems governed by linear parabolic equations, see [18]. In a recent preprint [23] an anisotropic error estimate is derived for the error due to the space discretization of an optimal control problem governed by linear heat equation.

However, in many applications, the error in global norms does not provide useful error bounds for the error in the quantity of physical interest. In [1, 3] a general

[†]Institut für Angewandte Mathematik, Ruprecht-Karls-Universität Heidelberg, INF 294, 69120 Heidelberg, Germany, dominik.meidner@iwr.uni-heidelberg.de

[‡]Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenberger Straße 69, 4040 Linz, Austria, boris.vexler@oeaw.ac.at

concept for a posteriori estimation of the discretization error with respect to the cost functional in the context of optimal control problems is presented. In papers [4, 5], this approach is extended to the estimation of the discretization error with respect to an arbitrary functional depending on both the control and the state variable, i.e. with respect to a quantity of interest. This allows, among other things, an efficient treatment of parameter identification and model calibration problems. The main contribution of this paper is the extension of these approaches to optimization problems governed by parabolic partial differential equations.

In this paper, we consider optimization problems under constraints of (nonlinear) parabolic differential equations

$$\begin{aligned} \partial_t u + A(q,u) &= f \\ u(0) &= u_0(q). \end{aligned} \tag{1.1}$$

Here, the state variable is denoted by u and the control variable by q. Both, the differential operator A and the initial condition u_0 may depend on q. This allows a simultaneous treatment of both, optimal control and parameter identification problems. For optimal control problems, the operator A is typically given by

$$A(q,u) = \bar{A}(u) - B(q),$$

with a (nonlinear) operator $\bar{A}$ and a (usually linear) control operator B. In parameter identification problems, the variable q denotes the unknown parameters to be determined and may enter the operator A in a nonlinear way. The case of initial control is included via the q-dependent initial condition $u_0(q)$.

The target of the optimization is to minimize a given cost functional $J(q,u)$ subject to the state equation (1.1).

For the numerical solution of this optimization problem the state variable has to be discretized in space and in time. Moreover, if the control (parameter) space is infinite dimensional, it has to be discretized, too. For fixed time, space, and control discretizations this leads to a finite dimensional optimization problem. We introduce σ as a general discretization parameter including the space, time, and the control discretization and denote the solution of the discrete problem by (q_σ, u_σ). For this discrete solution we derive an a posteriori error estimate with respect to the cost functional J of the following form:

$$J(q,u) - J(q_\sigma, u_\sigma) \approx \eta_k^J + \eta_h^J + \eta_d^J \tag{1.2}$$

Here, η_k^J, η_h^J, and η_d^J denote the error estimators, which can be evaluated from the computed discrete solution: η_k^J assess the error due to the time discretization, η_h^J due to the space discretization, and η_d^J due to the discretization of the control space. The structure of the error estimate (1.2) allows for equilibration of different discretization errors within an adaptive refinement algorithm to be described in the sequel.

For many optimization problems the quantity of physical interest coincide with the cost functional, which explains the choice of the error measure (1.2). However, in the case of parameter identification or model calibration problems, the cost functional is only an instrument for the estimation of the unknown parameters. Therefore, the value of the cost functional in the optimum and the corresponding discretization error are of secondary importance. This motivates error estimation with respect to a given functional I depending on the state and the control (parameter) variable. In this

paper we extend the corresponding results from [4, 5, 27] to parabolic problems and derive an a posteriori error estimator of the form

$$I(q,u) - I(q_\sigma, u_\sigma) \approx \eta_k^I + \eta_h^I + \eta_d^I,$$

where again η_k^I and η_h^I estimate the temporal and spatial discretization errors and η_d^I estimates the discretization error due to the discretization of the control space.

In Section 5.2 we will describe an adaptive algorithm based on these error estimators. Within this algorithm the time, space, and control discretizations are separately refined for efficient reduction of the total error equilibrating different types of the error. This local refinement relies on the computable representation of the error estimators as a sum of local contributions (error indicators), see the discussion in Section 5.1.

To the authors knowledge, this is the first paper describing the a posteriori error estimation for optimization problems governed by parabolic differential equations including the separation of different types of the discretization error.

The outline of the paper is as follows: In the next section we describe necessary optimality conditions for the problem under consideration and sketch the Newton-type optimization algorithm on the continuous level. This algorithm will be applied on the discrete level for fixed discretizations within an adaptive refinement procedure. In Section 3 we present the space time finite element discretization of the optimization problem. Section 4 is devoted to the derivation of the error estimators in a general setting. In Section 5 we discuss numerical evaluation of these error estimators and the adaptive algorithm in details. In the last section we present two numerical examples illustrating the behavior of the proposed methods. The first example deals with boundary control of the heat equation, whereas the second one is concerned with the identification of Arrhenius parameters in a simplified gaseous combustion model by means of point measurements of the concentrations.

2. Optimization. The optimization problems considered in this paper are formulated in the following abstract setting: Let Q be a Hilbert space for the controls (parameters) with scalar product $(\cdot,\cdot)_Q$. Moreover, let V and H be Hilbert spaces, which build together with the dual space V^* of V a Gelfand triple $V \hookrightarrow H \hookrightarrow V^*$. The duality pairing between the Hilbert spaces V and its dual V^* is denoted by $\langle\cdot,\cdot\rangle_{V^*\times V}$ and the scalar product in H by $(\cdot,\cdot)_H$. A typical choice for these space could be

$$V = \left\{ v \in H^1(\Omega) \;\middle|\; v\big|_{\partial\Omega_{\mathrm{D}}} = 0 \right\} \text{ and } H = L^2(\Omega), \tag{2.1}$$

where $\partial\Omega_{\mathrm{D}}$ denotes the part of the boundary of Ω with prescribed Dirichlet boundary conditions.

For a time interval $I = (0,T)$ we introduce the Hilbert space $X := W(0,T)$ defined as

$$W(0,T) = \left\{ v \;\middle|\; v \in L^2(I,V) \text{ and } \partial_t v \in L^2(I,V^*) \right\}. \tag{2.2}$$

It is well known that the space X is continuously embedded in $C(\bar{I},H)$, see e.g. [8]. Furthermore, we use the inner product of $L^2(I,H)$ given by

$$(u,v) := (u,v)_{L^2(I,H)} = \int_0^T (u(t),v(t))_H \, dt \tag{2.3}$$

for setting up the weak formulation of the state equation. This is possible since due to the properties of the Gelfand's triple the inner product on H is an equivalent representation of the duality pairing of V and V^*.

By means of the spatial semi-linear form $\bar{a}\colon Q \times V \times V \to \mathbb{R}$ defined for a differential operator $A\colon Q \times V \to V^*$ by

$$\bar{a}(q,\bar{u})(\bar{\varphi}) := \langle A(q,\bar{u}), \bar{\varphi}\rangle_{V^*\times V},$$

we can define the semi-linear form $a(\cdot,\cdot)(\cdot)$ on $Q \times X \times X$ as

$$a(q,u)(\varphi) := \int_0^T \bar{a}(q,u(t))(\varphi(t))\,dt$$

which is assumed to be three times continuously differentiable and linear in the third argument.

REMARK 2.1. *If the control variable q depends on time, this has to be incorporated by an obvious modification of the definitions of the semi-linear forms.*

After these preliminaries, we pose the *state equation* in a weak form: Find for given control $q \in Q$ the *state variable* $u \in X$ such that

$$\begin{aligned}(\partial_t u, \varphi) + a(q,u)(\varphi) &= (f,\varphi) \quad \forall \varphi \in X,\\ u(0) &= u_0(q),\end{aligned} \tag{2.4}$$

where $f \in L^2(0,T;V^*)$ represents the right hand side of the state equation and $u_0\colon Q \to H$ denotes a three times continuously differentiable mapping describing parameter-dependent initial conditions.

The cost functional $J\colon Q \times X \to \mathbb{R}$ is defined using two three times continuously differentiable functionals $J_1\colon V \to \mathbb{R}$ and $J_2\colon H \to \mathbb{R}$ by

$$J(q,u) = \int_0^T J_1(u)\,dt + J_2(u(T)) + \frac{\alpha}{2}\|q-\bar{q}\|_Q^2, \tag{2.5}$$

where the regularization (or cost) term is added which involves $\alpha \geq 0$ and a reference parameter $\bar{q} \in Q$.

The corresponding optimization problem is formulated as follows:

$$\text{Minimize } J(q,u) \text{ subject to the state equation (2.4), } (q,u) \in Q \times X. \tag{2.6}$$

The question of existence and uniqueness of solutions to such optimization problems is discussed e.g. in [17, 12, 25]. Throughout the paper, we assume problem (2.6) to admit a (locally) unique solution.

Provided the existence of a solution operator $S\colon Q \supset Q_0 \to X$ on an open subset Q_0 containing the optimal solution, we can define the reduced cost functional $j\colon Q_0 \to \mathbb{R}$ by $j(q) = J(q,S(q))$. This definition allows to reformulate problem (2.6) as an unconstrained optimization problem:

$$\text{Minimize } j(q), \quad q \in Q_0. \tag{2.7}$$

For the reduced optimization problem (2.7) we apply Newton's method to reach a control q which satisfies the first order necessary optimality condition

$$j'(q)(\tau q) = 0, \qquad \forall \tau q \in Q.$$

Starting with an initial guess q^0, the next Newton iterate is obtained by $q^{i+1} = q^i + \delta q$, where the update $\delta q \in Q$ is the solution of the linear problem:

$$j''(q)(\delta q, \tau q) = -j'(q)(\tau q), \qquad \forall \tau q \in Q. \tag{2.8}$$

Thus, we need suitable expressions for the first and second derivatives of the reduced cost functional j. To this end, we introduce the Lagrangian $\mathcal{L}\colon Q \times X \times X \to \mathbb{R}$, defined as

$$\mathcal{L}(q,u,z) = J(q,u) + (f - \partial_t u, z) - a(q,u)(z) - (u(0) - u_0(q), z(0))_H. \tag{2.9}$$

With its aid, we obtain the following standard representation of the first derivative $j'(q)(\tau q)$:

THEOREM 2.1.

- *If for given $q \in Q$ the state $u \in X$ fulfills the state equation*

$$\mathcal{L}'_z(q,u,z)(\varphi) = 0, \quad \forall \varphi \in X,$$

- *and if additionally $z \in X$ is chosen as solution of the adjoint state equation*

$$\mathcal{L}'_u(q,u,z)(\varphi) = 0, \quad \forall \varphi \in X,$$

then the following expression of first derivative of the reduced cost functional holds:

$$\begin{aligned} j'(q)(\tau q) &= \mathcal{L}'_q(q,u,z)(\tau q) \\ &= \alpha(q - \bar{q}, \tau q)_Q - a'_q(q,u)(\tau q) + (u'_0(q)(\tau q), z(0))_H. \end{aligned}$$

REMARK 2.2. *The optimality system of the considered optimization problem* (2.6) *is given by the derivatives of the Lagrangian used in Theorem 2.1 above:*

$$\begin{aligned} &\mathcal{L}'_z(q,u,z)(\varphi) = 0, \quad \forall \varphi \in X \qquad \textit{(State equation)}, \\ &\mathcal{L}'_u(q,u,z)(\varphi) = 0, \quad \forall \varphi \in X \qquad \textit{(Adjoint state equation)}, \\ &\mathcal{L}'_q(q,u,z)(\psi) = 0, \quad \forall \psi \in Q \qquad \textit{(Gradient equation)}. \end{aligned} \tag{2.10}$$

For the explicit formulation of the dual equation in this setting see e.g. [2].

In the same manner one can gain representations of the second derivatives of j in terms of the Lagrangian, see e.g. [2] where two different kinds of expressions are discussed: Either one can build up the whole Hessian and solve the system (2.8) by an arbitrary linear solver, or one just computes matrix-vector products of the Hessian times a given vector and uses this to solve (2.8) by the conjugate gradient method.

The presented Newton's method will be used to solve discrete optimization problems arising from discretizing the states and the controls as e.g. shown in the following section. In practical realizations, Newton's method has to be combined with some globalizations techniques as line search or trust region to enlarge its area of convergence, see e.g. [22, 7].

REMARK 2.3. *The solution u of the underlying state equation is typically required in the whole time interval for the computation of the adjoint solution z. If all data are stored, the storage grows linearly with respect to the number of time intervals in the time discretization. For reducing the required memory one can apply checkpointing techniques, see e.g. [13, 14]. In [2] we analyze such a strategy in the context of space-time finite element discretization of parabolic optimization problems.*

3. Discretization. In this section, we discuss the discretization of the optimization problem (2.6). To this end, we use Galerkin finite element methods in space and time to discretize the state equation. This allows us to give a natural computable representation of the discrete gradient and Hessian in the same manner as shown in Section 2 for the continuous problem. The use of exact discrete derivatives is important for the convergence of the optimization algorithms. Moreover, our systematic approach to a posteriori error estimation relies on using the Galerkin-type discretizations.

The first of the following subsection is devoted to semi-discretization in time by *continuous Galerkin (cG)* and *discontinuous Galerkin (dG)* methods. Subsection 3.2 deals with the space discretization of the semi-discrete problems arising from time discretization. For the numerical analysis of these schemes we refer to [10].

The discretization of the control space Q is kept rather abstract by choosing an finite dimensional subspace $Q_d \subset Q$. A possible concretion of this choice is shown in the numerical examples in Section 6. For the variational discretization concept, where the control variable is not discretized explicitly, we refer to [15], for a superconvergence based discretization of the control variable see [21].

3.1. Time Discretization of the States. To define a semi-discretization in time, let us partition the time interval $\bar{I} = [0, T]$ as

$$\bar{I} = \{0\} \cup I_1 \cup I_2 \cup \cdots \cup I_M$$

with subintervals $I_m = (t_{m-1}, t_m]$ of size k_m and time points

$$0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = T.$$

We define the discretization parameter k as a piecewise constant function by setting $k\big|_{I_m} = k_m$ for $m = 1, \ldots, M$.

By means of the subintervals I_m, we define for $r \in \mathbb{N}_0$ two semi-discrete spaces X_k^r and $\widetilde{X}_k^r$:

$$\begin{aligned} X_k^r &= \left\{ v_k \in C(\bar{I}, V) \;\middle|\; v_k\big|_{I_m} \in \mathcal{P}^r(I_m, V) \right\} \subset X \\ \widetilde{X}_k^r &= \left\{ v_k \in L^2(I, V) \;\middle|\; v_k\big|_{I_m} \in \mathcal{P}^r(I_m, V) \text{ and } v_k(0) \in H \right\} \end{aligned}$$

Here, $\mathcal{P}^r(I_m, V)$ denotes the space of polynomials up to order r defined on I_m with values in V. Thus, X_k^r consist of piecewise polynomials which are continuous in time and will be used as trial space in the continuous Galerkin method whereas the functions in $\widetilde{X}_k^r$ may have discontinuities at the edges of the subintervals I_m. This space will be used in the sequel as test space in the continuous Galerkin method and as trial and test space in the discontinuous Galerkin method.

3.1.1. Continuous Galerkin (cG) Methods. Using the semi-discrete spaces defined above, the cG(r) formulation of the state equation can directly stated as: Find for given control $q_k \in Q$ a state $u_k \in X_k^r$ such that

$$\begin{aligned} (\partial_t u_k, \varphi) + a(q_k, u_k)(\varphi) = (f, \varphi) \quad \forall \varphi \in \widetilde{X}_k^{r-1}, \\ u_k(0) = u_0(q_k). \end{aligned} \tag{3.1}$$

Here, the inner product on X has to be extended on $\widetilde{X}_k^r$ via its definition (2.3).

The corresponding semi-discretized optimization problem reads:

$$\text{Minimize } J(q_k, u_k) \text{ subject to the state equation (3.1), } (q_k, u_k) \in Q \times X_k^r. \quad (3.2)$$

Since the state equation semi-discretized by the cG(r) method has the same form as in the continuous setting, the corresponding Lagrangian is analogically defined on $Q \times X_k^r \times \widetilde{X}_k^{r-1}$ as

$$\mathcal{L}(q_k, u_k, z_k) = J(q_k, u_k) + (f - \partial_t u_k, z_k) - a(q_k, u_k)(z_k) - (u_k(0) - u_0(q_k), z_k(0))_H.$$

3.1.2. Discontinuous Galerkin (dG) Methods. To define the dG(r) discretization we employ the following definition for functions $v_k \in \widetilde{X}_k^r$:

$$v_{k,m}^+ := \lim_{t \to 0^+} v_k(t_m + t), \quad v_{k,m}^- := \lim_{t \to 0^+} v_k(t_m - t) = v_k(t_m), \quad [v_k]_m := v_{k,m}^+ - v_{k,m}^-$$

Then, the dG(r) semi-discretization of the state equation (2.4) reads: Find for given control $q_k \in Q$ a state $u_k \in \widetilde{X}_k^r$ such that

$$\begin{aligned} \sum_{m=1}^{M} \int_{I_m} (\partial_t u_k, \varphi)_H \, dt + a(q_k, u_k)(\varphi) + \sum_{m=0}^{M-1} ([u_k]_m, \varphi_m^+)_H = (f, \varphi), \quad \forall \varphi \in \widetilde{X}_k^r, \\ u_{k,0}^- = u_0(q_k). \end{aligned} \quad (3.3)$$

The semi-discrete optimization problem for the dG(r) time discretization has the form:

$$\text{Minimize } J(q_k, u_k) \text{ subject to the state equation (3.3), } (q_k, u_k) \in Q \times \widetilde{X}_k^r. \quad (3.4)$$

Then we pose the Lagrange functional $\widetilde{\mathcal{L}}\colon Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r \to \mathbb{R}$ associated with the dG(r) time discretization for the state equation as

$$\begin{aligned} \widetilde{\mathcal{L}}(q_k, u_k, z_k) = J(q_k, u_k) + (f, z_k) - \sum_{m=1}^{M} \int_{I_m} (\partial_t u_k, z_k)_H \, dt \\ - a(q_k, u_k)(z_k) - \sum_{m=0}^{M-1} ([u_k]_m, z_{k,m}^+)_H - (u_{k,0}^- - u_0(q_k), z_{k,0}^-)_H. \end{aligned}$$

3.2. Space Discretization of the States. In this subsection, we first describe the finite element discretization in space. To this end, we consider two or three dimensional shape-regular meshes, see e.g. [6]. A mesh consists of quadrilateral or hexahedral cells K, which constitute a non-overlapping cover of the computational domain $\Omega \subset \mathbb{R}^n$, $n \in \{2, 3\}$. The corresponding mesh is denoted by $\mathcal{T}_h = \{K\}$, where we define the discretization parameter h as a cellwise constant function by setting $h\big|_K = h_K$ with the diameter h_K of the cell K.

On the mesh $\mathcal{T}_h$ we construct a conform finite element space $V_h \subset V$ in a standard way:

$$V_h^s = \left\{ v \in V \;\middle|\; v\big|_K \in \mathcal{Q}^s(K) \text{ for } K \in \mathcal{T}_h \right\}$$

Here, $\mathcal{Q}^s(K)$ consists of shape functions obtained via bi- or tri-linear transformations of polynomials in $\widehat{\mathcal{Q}^s}(\widehat{K})$ defined on the reference cell $\widehat{K} = (0, 1)^n$.

To obtain the fully discretized versions of the time discretized state equations (3.1) and (3.3), we utilize the space-time finite element spaces

$$X_{k,h}^{r,s} = \left\{ v_{kh} \in C(\bar{I}, V_h^s) \;\middle|\; v_{kh}\big|_{I_m} \in \mathcal{P}^r(I_m, V_h^s) \right\} \subset X_k^r$$

and

$$\widetilde{X}_{k,h}^{r,s} = \left\{ v_{kh} \in L^2(I, V_h^s) \;\middle|\; v_{kh}\big|_{I_m} \in \mathcal{P}^r(I_m, V_h^s) \text{ and } v_{kh}(0) \in V_h^s \right\} \subset \widetilde{X}_k^r.$$

REMARK 3.1. *By the above definition of the discrete spaces $X_{k,h}^{r,s}$ and $\widetilde{X}_{k,h}^{r,s}$, we have assumed that the spatial discretization is fixed for all time intervals. However, in many application problems the use of different meshes $\mathcal{T}_h^m$ for each of the subintervals I_m will lead to more efficient adaptive discretizations. The consideration of such dynamically changing meshes can be included in the formulation of the dG(r) schemes in a natural way. The corresponding formulation of the cG(r) method is more involved due to the continuity requirement in the trial space. The treatment of dynamic meshes for parabolic optimization problems within an adaptive algorithm will be analyzed in a forthcoming paper.*

Then, the so called cG(s)cG(r) discretization of the state equation (2.4) can be stated as: Find for given control $q_{kh} \in Q$ a state $u_{kh} \in X_{k,h}^{r,s}$ such that

$$\begin{aligned} (\partial_t u_{kh}, \varphi) + a(q_{kh}, u_{kh})(\varphi) &= (f, \varphi) \quad \forall \varphi \in \widetilde{X}_{k,h}^{r-1,s}, \\ u_{kh}(0) &= u_0(q_{kh}), \end{aligned} \tag{3.5}$$

and the cG(s)dG(r) discretization has the form: Find for given control $q_{kh} \in Q$ a state $u_{kh} \in \widetilde{X}_{k,h}^{r,s}$ such that

$$\sum_{m=1}^{M} \int_{I_m} (\partial_t u_{kh}, \varphi)_H \, dt + a(q_{kh}, u_{kh})(\varphi) + \sum_{m=0}^{M-1} ([u_{kh}]_m, \varphi_m^+)_H = (f, \varphi), \quad \forall \varphi \in \widetilde{X}_{k,h}^{r,s},$$
$$u_{kh,0}^- = u_0(q_{kh}). \tag{3.6}$$

Thus, the optimization problems with fully discretized states are given by

$$\text{Minimize } J(q_{kh}, u_{kh}) \text{ subject to the state equation (3.5), } (q_{kh}, u_{kh}) \in Q \times X_{k,h}^{r,s} \tag{3.7}$$

for the cG(s)cG(r) discretization and by

$$\text{Minimize } J(q_{kh}, u_{kh}) \text{ subject to the state equation (3.6), } (q_{kh}, u_{kh}) \in Q \times \widetilde{X}_{k,h}^{r,s} \tag{3.8}$$

for the cG(s)dG(r) discretization of the state space.

The definition of the Lagrangians $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ for fully discretized states can directly be transfered from the formulations for semi-discretization in time just by restriction of the state spaces X_k^r and $\widetilde{X}_k^r$ to the subspaces $X_{k,h}^{r,s}$ and $\widetilde{X}_{k,h}^{r,s}$, respectively. With the aid of these Lagrangians, the derivatives of the reduced functionals $j_k(q_k) = J(q_k, S_k(u_k))$ and $j_{kh}(q_{kh}) = J(q_{kh}, S_{kh}(u_{kh}))$ on the different discretization levels can be expressed in the same manner as described on the continuous level in Theorem 2.1. Thus, we obtain exact derivatives of the reduced cost functional on the discrete level, see [2] for details.

REMARK 3.2. *The dG(r) and cG(r) schemes are known to be time discretization schemes of order $r+1$. The cG(r) schemes lead to a A-stable discretization whereas the dG(r) schemes are even strongly A-stable.*

REMARK 3.3. *The lower order methods dG(0) and cG(1) can be reinterpreted as time stepping schemes using numerical integration. Thereby, the dG(0) discretization leads to variations of the backward Euler scheme depending on the chosen quadrature rule for the righthandside, and the cG(1) discretization results in variants of the Crank-Nicolson scheme. The exact computation of the derivatives on the discrete level mentioned above is not disturbed even by the numerical integration. This can be shown using a duality argument with respect to the inner product based on the underlying quadrature rule.*

3.3. Discretization of the Controls. As proposed in the beginning of the current section, the discretization of the control space Q is kept rather abstract. It is done by choosing a finite dimensional subspace $Q_d \subset Q$. Then, the formulation of the state equation, the optimization problems and the Lagrangians defined on the fully discretized state space can directly be transfered to the level with fully discretized control and state spaces by replacing Q by Q_d. The full discrete solutions will be indicated by the subscript σ which collects the discretization indices k, h and d.

4. Derivation of the A Posteriori Error Estimator. In this section, we will establish a posteriori error estimators for the error arising due to the discretization of the control and state spaces in terms of the cost functional J and an arbitrary quantity of interest I.

For this, we first recall an abstract result from [3] which we will later use to establish the desired a posteriori error estimators:

PROPOSITION 4.1. *Let Y be a function space and L a differentiable functional on Y. We seek a stationary point y of L on Y, that is*

$$L'(y)(\hat{y}) = 0 \quad \forall \hat{y} \in Y. \tag{4.1}$$

This equation is approximated by a Galerkin method using a finite dimensional subspace $Y_0 \subset Y$. The discrete problem seeks y_0 satisfying

$$L'(y_0)(\hat{y}_0) = 0 \quad \forall \hat{y}_0 \in Y_0. \tag{4.2}$$

Then we have for arbitrary $\hat{y}_0 \in Y_0$ the error representation

$$L(y) - L(y_0) = \frac{1}{2} L'(y_0)(y - \hat{y}_0) + \mathcal{R}, \tag{4.3}$$

where the remainder term $\mathcal{R}$ is given with $e := y - y_0$ by

$$\mathcal{R} = \frac{1}{2} \int_0^1 L'''(y_0 + se)(e,e,e) \cdot s \cdot (s-1)\, ds.$$

In the sequel, we present the derivation of an error estimator for the fully discrete optimization problem in the case of discontinuous Galerkin (dG) time discretization only. The continuous Galerkin (cG) time discretization can be treated in a similar way.

4.1. Error Estimator for the Cost Functional. In the sequel, we use the abstract result of Proposition 4.1 for derivation of error estimators in terms of the cost functional J:

$$J(q,u) - J(q_\sigma, u_\sigma)$$

Here, $(q,u) \in Q \times X$ denotes the continuous optimal solution of (2.6) and $(q_\sigma, u_\sigma) = (q_{khd}, u_{khd}) \in Q_d \times \widetilde{X}_{k,h}^{r,s}$ is the optimal solution of the full discretized problem.

To separate the influences of the different discretizations on the discretization error we are interested in, we split

$$\begin{aligned} J(q,u) - J(q_\sigma, u_\sigma) = \quad & J(q,u) - J(q_k, u_k) \\ & + J(q_k, u_k) - J(q_{kh}, u_{kh}) \\ & + J(q_{kh}, u_{kh}) - J(q_\sigma, u_\sigma), \end{aligned}$$

where $(q_k, u_k) \in Q \times \widetilde{X}_k^r$ is the solution of the time discretized problem (3.4) and $(q_{kh}, u_{kh}) \in Q \times \widetilde{X}_{k,h}^{r,s}$ is the solution of the time and space discretized problem (3.8) with still undiscretized control space Q.

THEOREM 4.2. *Let (q,u,z), (q_k, u_k, z_k), (q_{kh}, u_{kh}, z_{kh}), and $(q_\sigma, u_\sigma, z_\sigma)$ be stationary points of $\mathcal{L}$ resp. $\widetilde{\mathcal{L}}$ on the different levels of discretization, i.e.*

$$\begin{aligned} \mathcal{L}'(q,u,z)(\hat{q},\hat{u},\hat{z}) = \widetilde{\mathcal{L}}'(q,u,z)(\hat{q},\hat{u},\hat{z}) = 0, \quad & \forall (\hat{q},\hat{u},\hat{z}) \in X \times X \times Q, \\ \widetilde{\mathcal{L}}'(q_k,u_k,z_k)(\hat{q}_k,\hat{u}_k,\hat{z}_k) = 0, \quad & \forall (\hat{q}_k,\hat{u}_k,\hat{z}_k) \in \widetilde{X}_k^r \times \widetilde{X}_k^r \times Q, \\ \widetilde{\mathcal{L}}'(q_{kh},u_{kh},z_{kh})(\hat{q}_{kh},\hat{u}_{kh},\hat{z}_{kh}) = 0, \quad & \forall (\hat{q}_{kh},\hat{u}_{kh},\hat{z}_{kh}) \in \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s} \times Q, \\ \widetilde{\mathcal{L}}'(q_\sigma,u_\sigma,z_\sigma)(\hat{q}_\sigma,\hat{u}_\sigma,\hat{z}_\sigma) = 0, \quad & \forall (\hat{q}_\sigma,\hat{u}_\sigma,\hat{z}_\sigma) \in \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s} \times Q_d. \end{aligned}$$

Then there holds for the errors with respect to the cost functional due to the time, space, and control discretizations:

$$\begin{aligned} J(q,u) - J(q_k,u_k) &= \frac{1}{2}\widetilde{\mathcal{L}}'(q_k,u_k,z_k)(q-\hat{q}_k, u-\hat{u}_k, z-\hat{z}_k) + \mathcal{R}_k \\ J(q_k,u_k) - J(q_{kh},u_{kh}) &= \frac{1}{2}\widetilde{\mathcal{L}}'(q_{kh},u_{kh},z_{kh})(q_k-\hat{q}_{kh}, u_k-\hat{u}_{kh}, z_k-\hat{z}_{kh}) + \mathcal{R}_h \\ J(q_{kh},u_{kh}) - J(q_\sigma,u_\sigma) &= \frac{1}{2}\widetilde{\mathcal{L}}'(q_\sigma,u_\sigma,z_\sigma)(q_{kh}-\hat{q}_\sigma, u_{kh}-\hat{u}_\sigma, z_{kh}-\hat{z}_\sigma) + \mathcal{R}_d. \end{aligned}$$

Here, $(\hat{q}_k,\hat{u}_k,\hat{z}_k) \in \widetilde{X}_k^r \times \widetilde{X}_k^r \times Q$, $(\hat{q}_{kh},\hat{u}_{kh},\hat{z}_{kh}) \in \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s} \times Q$, and $(\hat{q}_\sigma,\hat{u}_\sigma,\hat{z}_\sigma) \in \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s} \times Q_d$ can be chosen arbitrary and the remainder terms $\mathcal{R}_k$, $\mathcal{R}_h$, and $\mathcal{R}_d$ have the same form as given in Proposition 4.1 for $L = \widetilde{\mathcal{L}}$.

Proof. Since all the used solution pairs are optimal solutions of the optimization problem on different discretizations levels, we obtain for arbitrary $z \in X$, $z_k \in \widetilde{X}_k^r$, and $z_{kh}, z_\sigma \in \widetilde{X}_{k,h}^{r,s}$

$$J(q,u) - J(q_k,u_k) = \widetilde{\mathcal{L}}(q,u,z) - \widetilde{\mathcal{L}}(q_k,u_k,z_k) \tag{4.4a}$$

$$J(q_k,u_k) - J(q_{kh},u_{kh}) = \widetilde{\mathcal{L}}(q_k,u_k,z_k) - \widetilde{\mathcal{L}}(q_{kh},u_{kh},z_{kh}) \tag{4.4b}$$

$$J(q_{kh},u_{kh}) - J(q_\sigma,u_\sigma) = \widetilde{\mathcal{L}}(q_{kh},u_{kh},z_{kh}) - \widetilde{\mathcal{L}}(q_\sigma,u_\sigma,z_\sigma), \tag{4.4c}$$

whereas the identity

$$J(q,u) = \mathcal{L}(q,u,z) = \widetilde{\mathcal{L}}(q,u,z)$$

follows from the fact that the $u \in X$ is continuous and thus the additional jump terms in $\widetilde{\mathcal{L}}$ compared to $\mathcal{L}$ vanish.

To apply the abstract error identity (4.3) on the three righthandsides in (4.4), we choose the spaces Y and Y_0 of Proposition 4.1 as

$$\begin{aligned}
&\text{for (4.4a)}: \quad && Y = Q \times (X \cup \widetilde{X}_k^r) \times (X \cup \widetilde{X}_k^r) && Y_0 = Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r \\
&\text{for (4.4b)}: \quad && Y = Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r && Y_0 = Q \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s} \\
&\text{for (4.4c)}: \quad && Y = Q \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s} && Y_0 = Q_d \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s}.
\end{aligned}$$

Hence, the choice of the second and third pairing of $Y_0 \subset Y$ is obvious, since we have $\widetilde{X}_{k,h}^{r,s} \subset \widetilde{X}_k^r$ and $Q_d \subset Q$. For the choice of the spaces for (4.4a), we have to take into account the fact that $\widetilde{X}_k^r \not\subset X$. Thus, to fulfill the prerequisites of Theorem 4.2, we have chosen the state space in Y as the union of X and $\widetilde{X}_k^r$. The validity of (4.1) for this choice is shown by a density argument. □

By means of the residuals of the three equations building the optimality system (2.10)

$$\begin{aligned}
\tilde{\rho}^u(q,u)(\varphi) &:= \widetilde{\mathcal{L}}'_z(q,u,z)(\varphi), \\
\tilde{\rho}^z(q,u,z)(\varphi) &:= \widetilde{\mathcal{L}}'_u(q,u,z)(\varphi), \\
\tilde{\rho}^q(q,u,z)(\varphi) &:= \widetilde{\mathcal{L}}'_q(q,u,z)(\varphi),
\end{aligned}$$

the statement of Theorem 4.2 can be rewritten as

$$J(q,u) - J(q_k,u_k) \approx \frac{1}{2}\Big(\tilde{\rho}^u(q_k,u_k)(z-\hat{z}_k) + \tilde{\rho}^z(q_k,u_k,z_k)(u-\hat{u}_k)\Big) \tag{4.5a}$$

$$J(q_k,u_k) - J(q_{kh},u_{kh}) \approx \frac{1}{2}\Big(\tilde{\rho}^u(q_{kh},u_{kh})(z_k-\hat{z}_{kh}) + \tilde{\rho}^z(q_{kh},u_{kh},z_{kh})(u_k-\hat{u}_{kh})\Big) \tag{4.5b}$$

$$J(q_{kh},u_{kh}) - J(q_\sigma,u_\sigma) \approx \frac{1}{2}\tilde{\rho}^q(q_\sigma,u_\sigma,z_\sigma)(q_{kh}-\hat{q}_\sigma). \tag{4.5c}$$

Here, we employed the fact, that the terms

$$\begin{aligned}
&\tilde{\rho}^q(q_k,u_k,z_k)(q-\hat{q}_k), && \tilde{\rho}^q(q_{kh},u_{kh},z_{kh})(q_k-\hat{q}_{kh}), \\
&\tilde{\rho}^u(q_\sigma,u_\sigma)(z_{kh}-\hat{z}_\sigma), && \tilde{\rho}^z(q_\sigma,u_\sigma,z_\sigma)(u_{kh}-\hat{u}_\sigma)
\end{aligned}$$

are zero for the choice

$$\begin{aligned}
\hat{q}_k &= q \in Q, && \hat{q}_{kh} = q_k \in Q, \\
\hat{z}_\sigma &= z_{kh} \in \widetilde{X}_{k,h}^{r,s}, && \hat{u}_\sigma = u_{kh} \in \widetilde{X}_{k,h}^{r,s}.
\end{aligned}$$

This is possible since for the errors $J(q,u)-J(q_k,u_k)$ and $J(q_k,u_k)-J(q_{kh},u_{kh})$ only the state space is discretized and for $J(q_{kh},u_{kh})-J(q_\sigma,u_\sigma)$ we keep the discrete state space while discretizing the control space Q.

4.2. Error Estimator for an Arbitrary Functional. We now tend toward an error estimation of the different types of discretization errors in terms of a given functional $I\colon Q \times X \to \mathbb{R}$ describing the quantity of interest.

To this end, we define exterior Lagrangians $\mathcal{M}\colon (Q\times X\times X)^2 \to \mathbb{R}$ and $\widetilde{\mathcal{M}}\colon (Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r)^2 \to \mathbb{R}$ as

$$\mathcal{M}(\xi,\chi) = I(q,u) + \mathcal{L}'(\xi)(\chi)$$

with $\xi = (q, u, z), \chi = (p, v, y)$ and

$$\widetilde{\mathcal{M}}(\xi_k, \chi_k) = I(q_k, u_k) + \widetilde{\mathcal{L}}'(\xi_k)(\chi_k)$$

with $\xi_k = (q_k, u_k, z_k), \chi_k = (p_k, v_k, y_k)$.

Now we are in a similar setting as in the subsection before: We split the total discretization error with respect to I as

$$\begin{aligned} I(q,u) - I(q_\sigma, u_\sigma) = \quad & I(q,u) - I(q_k, u_k) \\ & + I(q_k, u_k) - I(q_{kh}, u_{kh}) \\ & + I(q_{kh}, u_{kh}) - I(q_\sigma, u_\sigma) \end{aligned}$$

and obtain the following theorem:

THEOREM 4.3. *Let (ξ, χ), (ξ_k, χ_k), (ξ_{kh}, χ_{kh}), and $(\xi_\sigma, \chi_\sigma)$ be stationary points of $\mathcal{M}$ resp. $\widetilde{\mathcal{M}}$ on the different levels of discretization, i.e.*

$$\begin{aligned} \mathcal{M}'(\xi,\chi)(\hat\xi,\hat\chi) = \widetilde{\mathcal{M}}'(\xi,\chi)(\hat\xi,\hat\chi) = 0, &\quad \forall (\hat\xi,\hat\chi) \in (Q \times X \times X)^2, \\ \widetilde{\mathcal{M}}'(\xi_k,\chi_k)(\hat\xi_k,\hat\chi_k) = 0, &\quad \forall (\hat\xi_k,\hat\chi_k) \in (Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r)^2, \\ \widetilde{\mathcal{M}}'(\xi_{kh},\chi_{kh})(\hat\xi_{kh},\hat\chi_{kh}) = 0, &\quad \forall (\hat\xi_{kh},\hat\chi_{kh}) \in (Q \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s})^2, \\ \widetilde{\mathcal{M}}'(\xi_\sigma,\chi_\sigma)(\hat\xi_\sigma,\hat\chi_\sigma) = 0, &\quad \forall (\hat\xi_\sigma,\hat\chi_\sigma) \in (Q_d \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s})^2. \end{aligned}$$

Then there holds for the errors with respect to the quantity of interest due to the time, space, and control discretizations:

$$\begin{aligned} I(q,u) - I(q_k,u_k) &= \frac{1}{2}\widetilde{\mathcal{M}}'(\xi_k,\chi_k)(\xi - \hat\xi_k, \chi - \hat\chi_k) + \mathcal{R}_k, \\ I(q_k,u_k) - I(q_{kh},u_{kh}) &= \frac{1}{2}\widetilde{\mathcal{M}}'(\xi_{kh},\chi_{kh})(\xi_k - \hat\xi_{kh}, \chi_k - \hat\chi_{kh}) + \mathcal{R}_h, \\ I(q_{kh},u_{kh}) - I(q_\sigma,u_\sigma) &= \frac{1}{2}\widetilde{\mathcal{M}}'(\xi_\sigma,\chi_\sigma)(\xi_{kh} - \hat\xi_\sigma, \chi_{kh} - \hat\chi_\sigma) + \mathcal{R}_d. \end{aligned}$$

Here, $(\hat\xi_k, \hat\chi_k) \in (Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r)^2$, $(\hat\xi_{kh}, \hat\chi_{kh}) \in (Q \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s})^2$, and $(\hat\xi_\sigma, \hat\chi_\sigma) \in (Q_d \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s})^2$ can be chosen arbitrary and the remainder terms $\mathcal{R}_k$, $\mathcal{R}_h$, and $\mathcal{R}_d$ have the same form as given in Proposition 4.1 for $L = \widetilde{\mathcal{M}}$.

Proof. Due to the optimality of the solution pairings on the different discretization levels we have the representations

$$I(q,u) - I(q_k,u_k) = \widetilde{\mathcal{M}}(\xi,\chi) - \widetilde{\mathcal{M}}(\xi_k,\chi_k) \tag{4.6a}$$

$$I(q_k,u_k) - I(q_{kh},u_{kh}) = \widetilde{\mathcal{M}}(\xi_k,\chi_k) - \widetilde{\mathcal{M}}(\xi_{kh},\chi_{kh}) \tag{4.6b}$$

$$I(q_{kh},u_{kh}) - I(q_\sigma,u_\sigma) = \widetilde{\mathcal{M}}(\xi_{kh},\chi_{kh}) - \widetilde{\mathcal{M}}(\xi_\sigma,\chi_\sigma), \tag{4.6c}$$

where the identity

$$I(q,u) = \mathcal{M}(\xi,\chi) = \widetilde{\mathcal{M}}(\xi,\chi)$$

again follows from the fact that the $u \in X$ is continuous and thus the additional jump terms in $\widetilde{\mathcal{M}}$ compared to $\mathcal{M}$ vanish.

Similar to the proof of Theorem 4.2, we choose the spaces Y and Y_0 for application of Proposition 4.1 as

$$\begin{aligned}
&\text{for (4.6a)}: && Y = (Q \times (X \cup \widetilde{X}_k^r) \times (X \cup \widetilde{X}_k^r))^2 && Y_0 = (Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r)^2\\
&\text{for (4.6b)}: && Y = (Q \times \widetilde{X}_k^r \times \widetilde{X}_k^r)^2 && Y_0 = (Q \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s})^2\\
&\text{for (4.6c)}: && Y = (Q \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s})^2 && Y_0 = (Q_d \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s})^2
\end{aligned}$$

and end up with the stated error representations. □

To apply Theorem 4.3 for instance to $I(q_{kh}, u_{kh}) - I(q_\sigma, u_\sigma)$, we have to require that

$$\widetilde{\mathcal{M}}'(\xi_\sigma, \chi_\sigma)(\hat{\xi}_\sigma, \hat{\chi}_\sigma) = 0, \quad \forall (\hat{\xi}_\sigma, \hat{\chi}_\sigma) \in (\widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s} \times Q_d)^2.$$

For solving this system, we have consider the concrete form of $\widetilde{\mathcal{M}}'$:

$$\begin{aligned}
&\widetilde{\mathcal{M}}'(\xi_\sigma, \chi_\sigma)(\delta\xi_\sigma, \delta\chi_\sigma) =\\
&\qquad I_q'(q_\sigma, u_\sigma)(\delta q_\sigma) + I_u'(q_\sigma, u_\sigma)(\delta u_\sigma) + \widetilde{\mathcal{L}}'(\xi_\sigma)(\delta\chi_\sigma) + \widetilde{\mathcal{L}}''(\xi_\sigma)(\chi_\sigma, \delta\xi_\sigma)
\end{aligned}$$

Since $\xi_\sigma = (q_\sigma, u_\sigma, z_\sigma)$ is the solution of the discrete optimization problem, it fulfills already $\widetilde{\mathcal{L}}'(\xi_\sigma)(\delta\chi_\sigma) = 0$. Thus, the solution triple $\chi_\sigma = (p_\sigma, v_\sigma, y_\sigma) \in Q_d \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s}$ has to fulfill

$$\begin{aligned}
&\widetilde{\mathcal{L}}''(\xi_\sigma)(\chi_\sigma, \delta\xi_\sigma) =\\
&\qquad - I_q'(q_\sigma, u_\sigma)(\delta q_\sigma) - I_u'(q_\sigma, u_\sigma)(\delta u_\sigma), \quad \forall \delta\xi_\sigma \in Q_d \times \widetilde{X}_{k,h}^{r,s} \times \widetilde{X}_{k,h}^{r,s}. \quad (4.7)
\end{aligned}$$

Solving this system of equations is apart from a different righthandside equivalent to the execution of one step of a (reduced) SQP-type method.

After splitting $y_\sigma = y_\sigma^{(0)} + y_\sigma^{(1)}$, where $y_\sigma^{(0)} \in \widetilde{X}_{k,h}^{r,s}$ is the solution of

$$\widetilde{\mathcal{L}}_{zu}''(\xi_\sigma)(y_\sigma^{(0)}, \varphi) = -I_u'(q_\sigma, u_\sigma)(\varphi), \quad \forall \varphi \in \widetilde{X}_{k,h}^{r,s},$$

we can rewrite system (4.7) in terms of the full discrete reduced Hessian $j_\sigma''(q)$ as

$$j_\sigma''(q_\sigma)(p_\sigma, \delta q_\sigma) = -I_q'(q_\sigma, u_\sigma)(\delta q_\sigma) - \mathcal{L}_{zq}''(\xi_\sigma)(y_\sigma^{(0)}, \delta q_\sigma), \quad \forall \delta q_\sigma \in Q_d,$$

where $j_\sigma''(q_\sigma)(p_\sigma, \delta q_\sigma)$ can be expressed as

$$\widetilde{\mathcal{L}}_{qq}''(\xi_\sigma)(p_\sigma, \delta q_\sigma) + \widetilde{\mathcal{L}}_{uq}''(\xi_\sigma)(v_\sigma, \delta q_\sigma) + \widetilde{\mathcal{L}}_{zq}''(\xi_\sigma)(y_\sigma^{(1)}, \delta q_\sigma).$$

The computation of $j_\sigma''(q_\sigma)(p_\sigma, \cdot)$ requires here the solution of the two auxiliary equations for $v_\sigma \in \widetilde{X}_{k,h}^{r,s}$ and $y_\sigma^{(1)} \in \widetilde{X}_{k,h}^{r,s}$:

$$\begin{aligned}
\widetilde{\mathcal{L}}_{uz}''(\xi_\sigma)(v_\sigma, \varphi) &= -\widetilde{\mathcal{L}}_{qz}''(\xi_\sigma)(p_\sigma, \varphi), \quad \forall \varphi \in \widetilde{X}_{k,h}^{r,s}\\
\widetilde{\mathcal{L}}_{zu}''(\xi_\sigma)(y_\sigma^{(1)}, \varphi) &= -\widetilde{\mathcal{L}}_{qu}''(\xi_\sigma)(p_\sigma, \varphi) - \widetilde{\mathcal{L}}_{uu}''(\xi_\sigma)(v_\sigma, \varphi), \quad \forall \varphi \in \widetilde{X}_{k,h}^{r,s}
\end{aligned}$$

By means of the residuals of the presented equations for p, v and y, i.e.

$$\begin{aligned}
\tilde{\rho}^v(\xi, p, v)(\varphi) &:= \widetilde{\mathcal{L}}_{uz}''(\xi)(v, \varphi) + \widetilde{\mathcal{L}}_{qz}''(\xi)(p, \varphi)\\
\tilde{\rho}^y(\xi, p, v, y)(\varphi) &:= \widetilde{\mathcal{L}}_{zu}''(\xi)(y, \varphi) + \widetilde{\mathcal{L}}_{qu}''(\xi)(p, \varphi) + \widetilde{\mathcal{L}}_{uu}''(\xi)(v, \varphi) + I_u'(q, u)(\varphi)\\
\tilde{\rho}^p(\xi, p, v, y)(\varphi) &:= \widetilde{\mathcal{L}}_{qq}''(\xi)(p, \varphi) + \widetilde{\mathcal{L}}_{uq}''(\xi)(v, \varphi) + \widetilde{\mathcal{L}}_{zq}''(\xi)(y, \varphi) + I_q'(q, u)(\varphi),
\end{aligned}$$

and the already defined residuals $\tilde{\rho}^u$, $\tilde{\rho}^z$ and $\tilde{\rho}^q$ the result of Theorem 4.3 can be expressed as

$$\begin{aligned}
I(q,u) - I(q_k,u_k) \approx \frac{1}{2}\Big(&\tilde{\rho}^u(q_k,u_k)(y-\hat{y}_k) + \tilde{\rho}^z(q_k,u_k,z_k)(v-\hat{v}_k) \\
&+ \tilde{\rho}^v(\xi_k,p_k,v_k)(z-\hat{z}_k) + \tilde{\rho}^y(\xi_k,p_k,v_k,y_k)(u-\hat{u}_k)\Big) \\
I(q_k,u_k) - I(q_{kh},u_{kh}) \approx \frac{1}{2}\Big(&\tilde{\rho}^u(q_{kh},u_{kh})(y_k-\hat{y}_{kh}) + \tilde{\rho}^z(q_{kh},u_{kh},z_{kh})(v_k-\hat{v}_{kh}) \\
&+ \tilde{\rho}^v(\xi_{kh},p_{kh},v_{kh})(z_k-\hat{z}_{kh}) \\
&+ \tilde{\rho}^y(\xi_{kh},p_{kh},v_{kh},y_{kh})(u_k-\hat{u}_{kh})\Big) \\
I(q_{kh},u_{kh}) - I(q_\sigma,u_\sigma) \approx \frac{1}{2}\Big(&\tilde{\rho}^q(q_\sigma,u_\sigma,z_\sigma)(p_{kh}-\hat{p}_\sigma) + \tilde{\rho}^p(\xi_\sigma,p_\sigma,v_\sigma,y_\sigma)(q_{kh}-\hat{q}_\sigma)\Big).
\end{aligned}$$

As for the estimator for the error in the cost functional, we employed here the fact, that the terms

$$\begin{aligned}
&\tilde{\rho}^q(q_k,u_k,z_k)(p-\hat{p}_k), && \tilde{\rho}^p(\xi_k,p_k,v_k,y_k)(q-\hat{q}_k), \\
&\tilde{\rho}^q(q_{kh},u_{kh},z_{kh})(p_k-\hat{p}_{kh}), && \tilde{\rho}^p(\xi_{kh},p_{kh},v_{kh},y_{kh})(q_k-\hat{q}_{kh}), \\
&\tilde{\rho}^u(q_\sigma,u_\sigma)(y_{kh}-\hat{y}_\sigma), && \tilde{\rho}^z(q_\sigma,u_\sigma,z_\sigma)(v_{kh}-\hat{v}_\sigma), \\
&\tilde{\rho}^v(\xi_\sigma,p_\sigma,v_\sigma)(z_{kh}-\hat{z}_\sigma), && \tilde{\rho}^y(\xi_\sigma,p_\sigma,v_\sigma,y_\sigma)(u_{kh}-\hat{u}_\sigma)
\end{aligned}$$

vanish if $\hat{p}_k$, $\hat{q}_k$, $\hat{p}_{kh}$, $\hat{q}_{kh}$, $\hat{y}_\sigma$, $\hat{v}_\sigma$, $\hat{z}_\sigma$, $\hat{u}_\sigma$ are chosen appropriately.

REMARK 4.1. *As already mentioned in the introduction of this section, we obtain almost identical results for the time discretization by the continuous Galerkin method as presented here. The difference simply consists in the tilde on the variables. The arguments of the proofs keep exactly the same.*

REMARK 4.2. *For the error estimation with respect to the cost function no additional equations have to be solved. The error estimation with respect to a given quantity of interest requires the computation of the auxiliary variables p_σ, v_σ, y_σ. The additional numerical effort is similar to the execution of one step of the SQP or Newton's method.*

5. Numerical Realization.

5.1. Evaluation of the Error Estimators. In this subsection, we concretize the a posteriori error estimator developed in the previous section for the cG(1)cG(1) and cG(1)dG(0) space-time discretizations on quadrilateral meshes in two space dimensions. That is, we consider the combination of cG(1) or dG(0) time discretization with piecewise bi-linear finite elements for the space discretization. As in the previous section, we will only present the concrete expressions for the dG time discretization, the cG discretization can be treated in exactly the same manner.

The error estimates presented in the previous section involve interpolation errors of the time, space, and the control discretizations. We approximate these errors using interpolations in higher order finite element spaces. To this end, we introduce linear operators Π_h, Π_k, and Π_d, which will map the computed solutions to the

approximations of the interpolation errors:

$$\begin{aligned} z - \hat{z}_k &\approx \Pi_k z_k & u - \hat{u}_k &\approx \Pi_k u_k \\ z_k - \hat{z}_{kh} &\approx \Pi_h z_{kh} & u_k - \hat{u}_{kh} &\approx \Pi_h u_{kh} \\ q_{kh} - \hat{q}_\sigma &\approx \Pi_d q_\sigma \end{aligned}$$

$$\begin{aligned} y - \hat{y}_k &\approx \Pi_k y_k & v - \hat{v}_k &\approx \Pi_k v_k \\ y_k - \hat{y}_{kh} &\approx \Pi_h y_{kh} & v_k - \hat{v}_{kh} &\approx \Pi_h v_{kh} \\ p_{kh} - \hat{p}_\sigma &\approx \Pi_d p_\sigma \end{aligned}$$

For the here considered case of cG(1)cG(1) and cG(1)dG(0) discretizations of the state space, the operators are chosen depending on the test and trial space as

$$\begin{aligned} \Pi_k &= I_k^{(1)} - \mathrm{id} \quad \text{with} \quad I_k^{(1)} \colon \widetilde{X}_k^0 \to X_k^1, \\ \Pi_k &= I_{2k}^{(2)} - \mathrm{id} \quad \text{with} \quad I_{2k}^{(2)} \colon X_k^1 \to X_{2k}^2, \\ \Pi_h &= I_{2h}^{(2)} - \mathrm{id} \quad \text{with} \quad I_{2h}^{(2)} \colon \begin{cases} X_{k,h}^{1,1} \to X_{k,2h}^{1,2} \\ \widetilde{X}_{k,h}^{0,1} \to \widetilde{X}_{k,2h}^{0,2}. \end{cases} \end{aligned}$$

The action of the piecewise linear and piecewise quadratic interpolation operators $I_k^{(1)}$ and $I_{2k}^{(2)}$ in time is depicted in Figures 5.1 and 5.2. The piecewise bi-quadratic spatial interpolation $I_{2h}^{(2)}$ can be easily computed if the underlying mesh provides a patch structure. That is, one can always combine four (eight) adjacent cells to a macro cell on which the bi-quadratic interpolation can be defined. An example of such an patched mesh is shown in Figure 5.3.

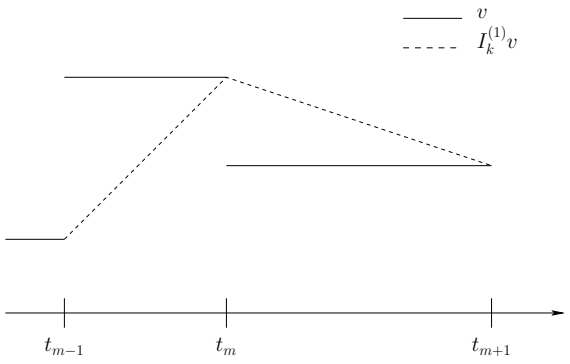

FIG. 5.1. *Piecewise Linear Interpolation of a Piecewise Constant Function*

The choice of Π_d depends on the discretization of the control space Q. If the finite dimensional subspaces Q_d are constructed similar to the discrete state spaces, one can directly choose for Π_d a modification of the operators Π_k and Π_h defined above. If for example the controls q only depend on time and the discretization is done with piecewise constant polynomials, we can choose $\Pi_d = I_d^{(1)} - \mathrm{id}$. If the control space Q is already finite dimensional, which is usually the case in the context of parameter estimation, it is possible to choose $\Pi_d = 0$ and thus, the estimator for the error $J(q_{kh}, u_{kh}) - J(q_\sigma, u_\sigma)$ is zero—as well as this discretization error itself.

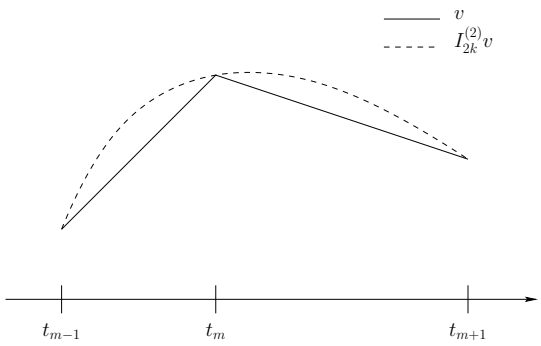

FIG. 5.2. *Piecewise Quadratic Interpolation of a Piecewise Linear Function*

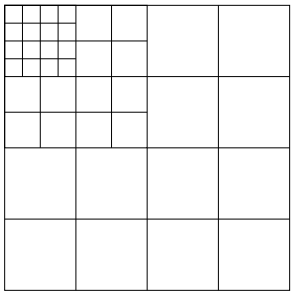

FIG. 5.3. *Patched Mesh*

In order to make the error representations from the previous section computable, we replace the residuals linearized on the solution of semi-discretized problems by the linearization at full discrete solutions.

We finally obtain the following computable a posteriori error estimator for the cost functional J

$$J(q,u) - J(q_\sigma, u_\sigma) \approx \eta_k^J + \eta_h^J + \eta_d^J$$

with

$$\begin{aligned}
\eta_k^J &:= \frac{1}{2}\Big(\tilde{\rho}^u(q_\sigma, u_\sigma)(\Pi_k z_\sigma) + \tilde{\rho}^z(q_\sigma, u_\sigma, z_\sigma)(\Pi_k u_\sigma)\Big)\\
\eta_h^J &:= \frac{1}{2}\Big(\tilde{\rho}^u(q_\sigma, u_\sigma)(\Pi_h z_\sigma) + \tilde{\rho}^z(q_\sigma, u_\sigma, z_\sigma)(\Pi_h u_\sigma)\Big)\\
\eta_d^J &:= \frac{1}{2}\tilde{\rho}^q(q_\sigma, u_\sigma, z_\sigma)(\Pi_d q_\sigma).
\end{aligned}$$

For the quantity of interest I the error estimator is given by:

$$I(q,u) - I(q_\sigma, u_\sigma) \approx \eta_k^I + \eta_h^I + \eta_d^I$$

with

$$\begin{aligned}
\eta_k^I &:= \frac{1}{2}\Big(\tilde{\rho}^u(q_\sigma,u_\sigma)(\Pi_k y_\sigma) + \tilde{\rho}^z(q_\sigma,u_\sigma,z_\sigma)(\Pi_k v_\sigma) \\
&\quad + \tilde{\rho}^v(\xi_\sigma,v_\sigma,p_\sigma)(\Pi_k z_\sigma) + \tilde{\rho}^y(\xi_\sigma,v_\sigma,y_\sigma,p_\sigma)(\Pi_k u_\sigma)\Big) \\
\eta_h^I &:= \frac{1}{2}\Big(\tilde{\rho}^u(q_\sigma,u_\sigma)(\Pi_h y_\sigma) + \tilde{\rho}^z(q_\sigma,u_\sigma,z_\sigma)(\Pi_h v_\sigma) \\
&\quad + \tilde{\rho}^v(\xi_\sigma,v_\sigma,p_\sigma)(\Pi_h z_\sigma) + \tilde{\rho}^y(\xi_\sigma,v_\sigma,y_\sigma,p_\sigma)(\Pi_h u_\sigma)\Big) \\
\eta_d^I &:= \frac{1}{2}\Big(\tilde{\rho}^q(q_\sigma,u_\sigma,z_\sigma)(\Pi_d p_\sigma) + \tilde{\rho}^p(\xi_\sigma,v_\sigma,y_\sigma,p_\sigma)(\Pi_d q_\sigma)\Big).
\end{aligned}$$

To give an impression of the terms that have to be evaluated for the error estimators, we present for the cG(1)dG(0) discretization the explicit form of state residuals $\tilde{\rho}^u(q_\sigma,u_\sigma)(\Pi_k z_\sigma)$ and $\tilde{\rho}^u(q_\sigma,u_\sigma)(\Pi_h z_\sigma)$ and the adjoint residuals $\tilde{\rho}^z(q_\sigma,u_\sigma,z_\sigma)(\Pi_k u_\sigma)$ and $\tilde{\rho}^z(q_\sigma,u_\sigma,z_\sigma)(\Pi_h u_\sigma)$. For simplicity of notation, we assume here q to be independent on time. Since we evaluate the arising integrals over time for the residuals weighted with z_σ or u_σ by the right endpoint rule and for the residuals weighted with $I_k^{(1)} z_\sigma$ or $I_k^{(1)} u_\sigma$ by the trapezoidal rule, we have to ensure the righthandside f to be continuous in time, i.e. $f \in C(\bar{I}, H)$. Then we obtain with the abbreviations $U_m = u_\sigma\big|_{I_m}$ and $Z_m = z_\sigma\big|_{I_m}$ the following parts of the error estimators:

$$\begin{aligned}
\tilde{\rho}^u(q_\sigma,u_\sigma)(\Pi_k z_\sigma) = \sum_{m=1}^{M}\Big\{ &(U_m - U_{m-1}, Z_m - Z_{m-1})_H \\
&+ \frac{k_m}{2}\bar{a}(q_\sigma,U_m)(Z_m - Z_{m-1}) \\
&+ \frac{k_m}{2}(f(t_{m-1}),Z_{m-1})_H - \frac{k_m}{2}(f(t_m),Z_m)_H\Big\}
\end{aligned}$$

$$\begin{aligned}
\tilde{\rho}^z(q_\sigma,u_\sigma,z_\sigma)(\Pi_k u_\sigma) = \sum_{m=1}^{M}\Big\{ &\frac{k_m}{2}\bar{a}'_u(q_\sigma,U_m)(U_m,Z_m) \\
&- \frac{k_m}{2}\bar{a}'_u(q_\sigma,U_{m-1})(U_{m-1},Z_m) \\
&+ \frac{k_m}{2}J_1'(U_{m-1})(U_{m-1}) - \frac{k_m}{2}J_1'(U_m)(U_m)\Big\}
\end{aligned}$$

$$\begin{aligned}
\tilde{\rho}^u(q_\sigma,u_\sigma)(\Pi_h z_\sigma) = \sum_{m=1}^{M}\Big\{ &k_m(f(t_m), I_{2h}^{(2)} Z_m - Z_m)_H \\
&- k_m\bar{a}(q_\sigma,U_m)(I_{2h}^{(2)} Z_m - Z_m) \\
&- (U_m - U_{m-1}, I_{2h}^{(2)} Z_m - Z_m)_H\Big\} \\
&- (U_0 - u_0(q_\sigma), I_{2h}^{(2)} Z_0 - Z_0)_H
\end{aligned}$$

$$\begin{aligned}\tilde{\rho}^z(q_\sigma, u_\sigma, z_\sigma)(\Pi_h u_\sigma) = \sum_{m=1}^{M} \Big\{ & k_m J_1'(U_m)(I_{2h}^{(2)} U_m - U_m) \\ & - k_m \bar{a}_u'(q_\sigma, U_m)(I_{2h}^{(2)} U_m - U_m, Z_m) \\ & + (I_{2h}^{(2)} U_{m-1} - U_{m-1}, Z_m - Z_{m-1})_H \Big\} \\ & + J_2'(U_M)(I_{2h}^{(2)} U_M - U_M) - (I_{2h}^{(2)} U_M - U_M, Z_M)_H\end{aligned}$$

For the cG(1)cG(1) discretization the terms that have to be evaluated are very similar and the evaluation can be treated as presented here for the cG(1)dG(0) discretization.

The presented a posteriori error estimators are directed towards two aims: assessment of the discretization error and improvement of the accuracy by local refinement. For the second aim the information provided by the error estimator have to be localized to cellwise or nodewise contributions (local error indicators). For details of the localization procedure we refer e.g. to [3].

5.2. Adaptive Algorithm. Goal of the adaption of the different types of discretizations has to be the equilibrated reduction of the corresponding discretization errors. If a given tolerance TOL has to be reached, this can be done by refining each discretization as long as the value of this part of the error estimator is greater than 1/3TOL. We want to present here a strategy which will equilibrate the different discretization errors even if no tolerance is given.

Aim of the equilibration algorithm presented in the sequel is to obtain discretization such that

$$|\eta_k| \approx |\eta_h| \approx |\eta_d|$$

and to keep this property during the further refinement. Here, the estimators η_i denote the estimators η_i^J for the cost functional J or η_i^I for the quantity of interest I.

For doing this equilibration, we choose an "equilibration factor" $e \approx 1-5$ and propose the following strategy: We compute a permutation (a, b, c) of the discretization indices (k, h, d) such that

$$|\eta_a| \geq |\eta_b| \geq |\eta_c|,$$

and define the relations

$$\gamma_{ab} := \left| \frac{\eta_a}{\eta_b} \right| \geq 1, \qquad \gamma_{bc} := \left| \frac{\eta_b}{\eta_c} \right| \geq 1.$$

Then we decide by means of Table 5.1 in every repetition of the adaptive refinement algorithm given by Algorithm 5.1 which discretization shall be refined. For every discretization to be adapted we select by means of the local error indicators the cells for refinement. For this purpose there are several strategies available, see e.g. [3].

ALGORITHM 5.1 (Adaptive Refinement Algorithm).

1: *Choose an initial triple of discretizations $\mathcal{T}_{\sigma_0}$, $\sigma_0 = (k_0, h_0, d_0)$ for the space-time discretization of the states and an appropriate discretization of the controls and set $n = 0$.*
2: ***loop***
3: *Compute the optimal solution pair $(q_{\sigma_n}, u_{\sigma_n})$*
4: *Evaluate the a posteriori error estimators η_{k_n}, η_{h_n} and η_{d_n}.*

5: ***if*** $\eta_{k_n} + \eta_{h_n} + \eta_{d_n} \leq TOL$ ***then***
6: ***break***
7: ***else***
8: *Determine the discretization(s) to be refined by means of Table 5.1.*
9: ***end if***
10: *Refine* $\mathcal{T}_{\sigma_n} \to \mathcal{T}_{\sigma_{n+1}}$ *depending on the size of* η_{k_n}, η_{h_n}, *and* η_{d_n} *to equilibrate the three discretization errors.*
11: *Increment* n.
12: ***end loop***

TABLE 5.1
Equilibration Strategy

Relation between the estimators	Discretizations to be refined
$\gamma_{ab} \leq e$ and $\gamma_{bc} \leq e$	a, b, and c
$\gamma_{bc} > e$	a and b
else ($\gamma_{ab} > e$ and $\gamma_{bc} \leq e$)	a

6. Numerical Examples. This section is devoted to the numerical validation of the theoretical results presented in the previous sections. This will be done by means of an optimal control problem with time-dependent boundary control (cf. Subsection 6.1) and a parameter estimation problem (cf. Subsection 6.2).

6.1. Example 1: Neumann Boundary Control Problem. We consider the linear parabolic state equation on the two-dimensional unit square $\Omega := (0,1)^2$ with final time $T = 1$ given by

$$\begin{aligned} \partial_t u - \nu\Delta u + u &= f && \text{in } \Omega \times I, \\ \partial_n u(x,t) &= 0 && \text{on } \Gamma_0 \times I, \\ \partial_n u(x,t) &= q^{(i)}(t) && \text{on } \Gamma_i \times I, i = 1,2 \\ u(x,0) &= 0 && \text{on } \Omega. \end{aligned} \tag{6.1}$$

The control $q = (q^{(1)}, q^{(2)})$ acts as only time-depended boundary control of Neumann type on two parts of the boundary denoted by Γ_1 and Γ_2. Thus, the control space Q is chosen as $[L^2(I)]^2$ and the spaces V and H used in the definition of the state space X are set to $V = H^1(\Omega)$ and $H = L^2(\Omega)$.

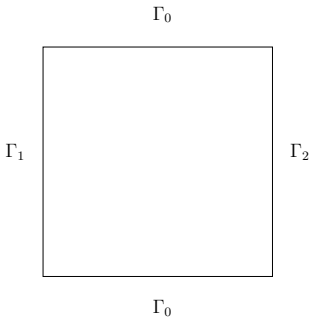

FIG. 6.1. *Example 1: Computational Domain* Ω

As cost functional J to be minimized subject to the state equation we choose the functional

$$J(q,u) := \frac{1}{2}\int_0^T \int_\Omega (u(x,t)-1)^2\,dx\,dt + \frac{\alpha}{2}\int_0^T \{q_1^2(t)+q_2^2(t)\}\,dt$$

of tracking type endowed with a $L^2(I)$-regularization.

For the computations, the righthandside f is chosen as

$$f(x,t) = 10t\exp\left(1-\frac{1}{1-100\|x-\tilde{x}\|^2}\right), \quad \tilde{x} = \left(\frac{2}{3},\frac{1}{2}\right)$$

and the parameters α and ν are set to

$$\alpha = 0.1, \quad \nu = 0.1.$$

The discretization of the state space is done here via the cG(1)cG(1) space-time Galerkin method which is a variant of the Crank-Nicolson scheme. Consequently, the state is discretized in time by piecewise linear and the adjoint state by piecewise constant polynomials. The controls are discretized using piecewise constant polynomials on a partition of the time interval I which has to be at most as fine as the time discretization of the states.

REMARK 6.1. *If the discretization of the control is chosen such that the gradient equation*

$$\int_{\Gamma_i} z(x,t)\,dx + \alpha q^{(i)}(t) = 0, \qquad i=1,2, \quad t\in I$$

can be fulfilled pointwise on the discrete level, the residual ρ^q of this equation as well as the error due to discretization of the control space vanish, cf. (4.5c)*. Thus, it is only reasonable to discretize the controls at most as fine as the adjoint state.*

In Table 6.1 we show the development of the discretization error and the a posteriori error estimators during an adaptive run with local refinement of all three types of discretizations. Here, M denotes the number of time steps, N denotes the number of nodes in the spatial mesh, and $\dim Q_d$ is the number of degrees of freedom for the discretization of the control. The effectivity index given in the last column of this table is defined as usual by

$$I_{\text{eff}} := \frac{J(q,u)-J(q_\sigma,u_\sigma)}{\eta_k^J+\eta_h^J+\eta_q^J}.$$

The table also demonstrates the desired equilibration of the different discretization errors and the sufficient quality of the error estimators.

A comparison of the error $J(q,u)-J(q_\sigma,u_\sigma)$ for the different refinement strategies is depicted in Figure 6.2:

- "uniform": Here, we apply uniform refinement of all discretizations after each run of the optimization loop.
- "uniform equilibration": Here, we also allow only for uniform refinements but use the error estimators within the equilibration strategy (Table 5.1) to decide which discretizations have to be refined.

TABLE 6.1
Example 1: Local Refinement with Equilibration

M	N	$\dim Q_d$	η_k^J	η_h^J	η_q^J	$\eta_k^J + \eta_h^J + \eta_q^J$	$J(q,u) - J(q_\sigma, u_\sigma)$	I_{eff}
64	25	16	$-9.7 \cdot 10^{-05}$	$2.0 \cdot 10^{-03}$	$-8.5 \cdot 10^{-04}$	$1.088 \cdot 10^{-03}$	$-2.567 \cdot 10^{-04}$	-0.2360
64	81	20	$-1.1 \cdot 10^{-04}$	$-1.0 \cdot 10^{-03}$	$-3.2 \cdot 10^{-04}$	$-1.543 \cdot 10^{-03}$	$-7.818 \cdot 10^{-04}$	0.5065
64	289	20	$-1.3 \cdot 10^{-04}$	$-4.8 \cdot 10^{-04}$	$-3.2 \cdot 10^{-04}$	$-9.458 \cdot 10^{-04}$	$-8.009 \cdot 10^{-04}$	0.8468
74	813	32	$-4.7 \cdot 10^{-05}$	$-2.2 \cdot 10^{-05}$	$-1.3 \cdot 10^{-04}$	$-2.058 \cdot 10^{-04}$	$-2.116 \cdot 10^{-04}$	1.0285
74	813	48	$-4.8 \cdot 10^{-05}$	$-2.2 \cdot 10^{-05}$	$-7.7 \cdot 10^{-05}$	$-1.476 \cdot 10^{-04}$	$-1.493 \cdot 10^{-04}$	1.0109
87	2317	76	$-2.7 \cdot 10^{-05}$	$1.1 \cdot 10^{-05}$	$-2.9 \cdot 10^{-05}$	$-4.516 \cdot 10^{-05}$	$-4.559 \cdot 10^{-05}$	1.0094
104	8213	128	$-1.8 \cdot 10^{-05}$	$2.7 \cdot 10^{-06}$	$-1.3 \cdot 10^{-05}$	$-2.931 \cdot 10^{-05}$	$-2.842 \cdot 10^{-05}$	0.9696
208	8213	128	$-4.3 \cdot 10^{-06}$	$2.7 \cdot 10^{-06}$	$-1.5 \cdot 10^{-05}$	$-1.674 \cdot 10^{-05}$	$-1.661 \cdot 10^{-05}$	0.9923
208	8213	192	$-4.2 \cdot 10^{-06}$	$2.7 \cdot 10^{-06}$	$-7.0 \cdot 10^{-06}$	$-8.573 \cdot 10^{-06}$	$-8.335 \cdot 10^{-06}$	0.9722

- "local equilibration": Here, we combine local refinement of all discretizations with the proposed equilibration strategy.

It shows for example, that to reach a discretization error of $4 \cdot 10^{-5}$ the uniform refinement needs about 70 times the number of degrees of freedom the fully adaptive refinement needs.

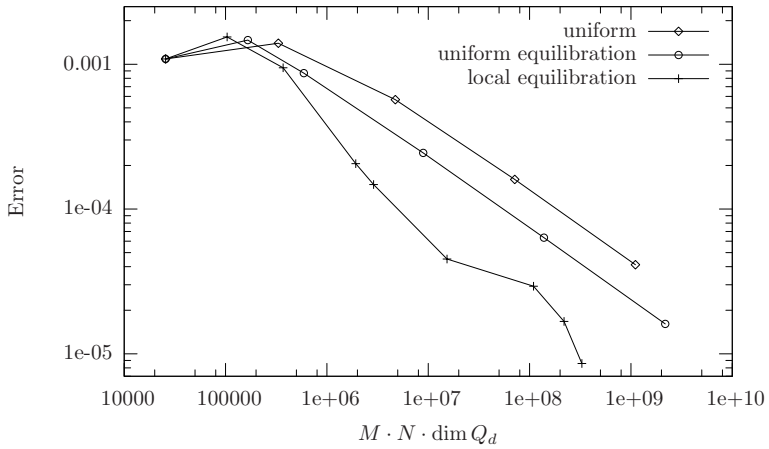

FIG. 6.2. *Example 1: Comparison of Different Refinement Strategies*

In Table 6.2 we present the numerical justification for splitting the total discretization error in three parts regarding the discretization of time, space, and control: The table demonstrates the independence of each part of the error estimator on the refinement of the other parts. This feature is especially important to reach a equilibration of the discretization errors by applying the adaptive refinement algorithm.

6.2. Example 2: Parameter Estimation. The state equation for the following example is taken from [16]. It describes the major part of gaseous combustion under the low Mach number hypothesis. Under this assumption, the motion of the fluid becomes independent from temperature and species concentration. Hence, one can solve the temperature and the species equation alone specifying any solenoidal velocity field.

Introducing the dimensionless temperature $\theta = \frac{T - T_{\text{unburnt}}}{T_{\text{burnt}} - T_{\text{unburnt}}}$, denoting by Y the

TABLE 6.2
Example 1: Independence of One Part of the Error Estimator on the Refinement of the other Parts

M	N	$\dim Q_d$	η_k^J	η_h^J	η_q^J
256	289	16		$-4.9104 \cdot 10^{-04}$	$-8.6152 \cdot 10^{-04}$
512	289	16		$-4.9110 \cdot 10^{-04}$	$-8.6232 \cdot 10^{-04}$
1024	289	16	—	$-4.9111 \cdot 10^{-04}$	$-8.6251 \cdot 10^{-04}$
2048	289	16		$-4.9111 \cdot 10^{-04}$	$-8.6256 \cdot 10^{-04}$
4096	289	16		$-4.9112 \cdot 10^{-04}$	$-8.6258 \cdot 10^{-04}$
1024	25	16	$-3.8360 \cdot 10^{-07}$		$-8.7015 \cdot 10^{-04}$
1024	81	16	$-4.3463 \cdot 10^{-07}$		$-8.5900 \cdot 10^{-04}$
1024	289	16	$-4.5039 \cdot 10^{-07}$	—	$-8.6251 \cdot 10^{-04}$
1024	1089	16	$-4.5529 \cdot 10^{-07}$		$-8.6398 \cdot 10^{-04}$
1024	4225	16	$-4.6096 \cdot 10^{-07}$		$-8.6432 \cdot 10^{-04}$
4096	289	16	$-2.8171 \cdot 10^{-08}$	$-4.9112 \cdot 10^{-04}$	
4096	289	32	$-3.0332 \cdot 10^{-08}$	$-4.8826 \cdot 10^{-04}$	
4096	289	64	$-3.1317 \cdot 10^{-08}$	$-4.8688 \cdot 10^{-04}$	—
4096	289	128	$-3.1704 \cdot 10^{-08}$	$-4.8651 \cdot 10^{-04}$	
4096	289	256	$-3.1828 \cdot 10^{-08}$	$-4.8642 \cdot 10^{-04}$	

species concentration, and assuming constant diffusion coefficients yields

$$\begin{aligned} \partial_t \theta - \Delta \theta &= \omega(Y, \theta) && \text{in } \Omega \times I, \\ \partial_t Y - \frac{1}{\mathrm{Le}} \Delta Y &= -\omega(Y, \theta) && \text{in } \Omega \times I, \end{aligned} \tag{6.2}$$

where the Lewis number Le is the ratio of diffusivity of heat and diffusivity of mass. We use a simple one-species reaction mechanism governed by an Arrhenius law

$$\omega(Y, \theta) = \frac{\beta^2}{2\mathrm{Le}} Y \mathrm{e}^{\frac{\beta(\theta-1)}{1+\alpha(\theta-1)}},$$

in which an approximation for large activation energy has been employed.

Here, we consider a freely propagating laminar flame described by (6.2) and its response to a heat absorbing obstacle, a set of cooled parallel rods with rectangular cross section (cf. Figure 6.3). Thus, the boundary condition are chosen as

$$\begin{aligned} \theta &= 1 && \text{on } \Gamma_{\mathrm{D}} \times I, \\ Y &= 0 && \text{on } \Gamma_{\mathrm{D}} \times I, \\ \partial_n \theta &= 0 && \text{on } \Gamma_{\mathrm{N}} \times I, \\ \partial_n Y &= 0 && \text{on } \Gamma_{\mathrm{N}} \times I, \\ \partial_n \theta &= -k\theta && \text{on } \Gamma_{\mathrm{R}} \times I, \\ \partial_n Y &= 0 && \text{on } \Gamma_{\mathrm{R}} \times I, \end{aligned}$$

where the heat absorption is modeled by Robin boundary conditions on Γ_{R}.

The initial condition is the analytical solution of an one-dimensional right-travel-

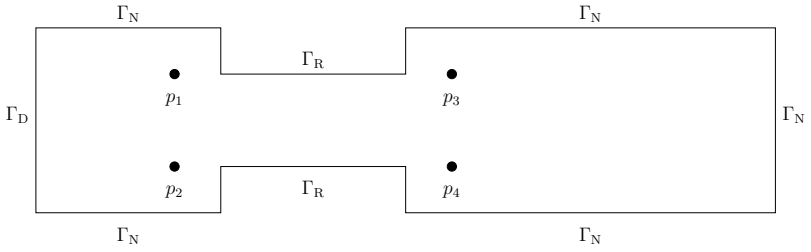

FIG. 6.3. *Example 2: Computational Domain Ω and Measurement Points p_i*

ing flame in the limit $\beta \to \infty$ located left of the obstacle:

$$\theta(0,x) = \begin{cases} 1, & \text{for } x_1 \le \tilde{x}_1 \\ \mathrm{e}^{\tilde{x}_1 - x_1}, & \text{for } x_1 > \tilde{x}_1 \end{cases} \qquad \text{on } \Omega,$$

$$Y(0,x) = \begin{cases} 0, & \text{for } x_1 \le \tilde{x}_1 \\ 1 - \mathrm{e}^{\mathrm{Le}(\tilde{x}_1 - x_1)}, & \text{for } x_1 > \tilde{x}_1 \end{cases} \qquad \text{on } \Omega.$$

For the computations, the occurring parameters are set to

$$\mathrm{Le} = 1, \qquad \beta = 10, \qquad k = 0.1, \qquad \tilde{x}_1 = 9$$

whereas the parameter α occurring in the Arrhenius law will be the objective of the parameter estimation.

To use the same notations as in the theoretical parts of this article, we define the pair of solution components $u := (\theta, Y) \in \hat{u} + X^2$ and denote the parameter α to be estimated by $q \in Q := \mathbb{R}$. For definition of the state space X we use the spaces V and H as given by (2.1). The function $\hat{u}$ is defined to fulfill the prescribed Dirichlet data as $\hat{u}\big|_{\Gamma_D} = (1, 0)$.

The unknown parameter α is estimated here using information from pointwise measurements of θ and Y at four measurement points $p_i \in \Omega$ $(i = 1, \ldots, 4)$ at final time $T = 60$. This parameter identification problem can be formulated as a cost functional of least squares type:

$$J(q,u) = \frac{1}{2} \sum_{i=1}^{4} \Big\{ (\theta(p_i, T) - \tilde{\theta}_i)^2 + (Y(p_i, T) - \tilde{Y}_i)^2 \Big\}$$

The values of artificial measurements $\tilde{\theta}_i$ and $\tilde{Y}_i$ $(i = 1, \ldots, 4)$ are obtained from a reference solution computed with fine discretizations.

The consideration of point measurements does not fulfill the assumption on the cost functional in (2.5), since the point evaluation is not bounded as a functional on $H = L^2(\Omega)$. Therefore, the point functionals here may be understood as regularized functionals defined on $L^2(\Omega)$. For an a priori error analysis of an elliptic parameter identification problems with pointwise measurements we refer to [24].

For this type of parameter estimation problems one is usually not interested in reducing the discretization error measured in terms of the cost functional. The focus is rather on reducing the error in the parameter q to be estimated. Hence, we use the quantity of interest I given by

$$I(q,u) = q$$

and apply the techniques presented in Section 4.2 for estimating the discretization error with respect to I. Since the control space Q in this application is given as $Q = \mathbb{R}$, it is not necessary to discretize Q. Thus, there is no discretization error due to the Q-discretization and the a posteriori error estimator consists only of η_k^I and η_h^I.

The results of a computation with equilibrated adaption of the space and time discretization using cG(1)dG(0) are shown in Table 6.3. The discretization parameters M and N as well as the effectivity index I_{eff} are defined as in Example 1.

TABLE 6.3
Example 2: Local Refinement with Equilibration

M	N	η_k^I	η_h^I	$\eta_k^I + \eta_h^I$	$I(q,u) - I(q_{kh}, u_{kh})$	I_{eff}
512	269	$-8.4 \cdot 10^{-03}$	$4.3 \cdot 10^{-02}$	$3.551 \cdot 10^{-02}$	$-2.859 \cdot 10^{-02}$	-0.8051
512	685	$-9.0 \cdot 10^{-03}$	$5.2 \cdot 10^{-03}$	$-3.778 \cdot 10^{-03}$	$-4.854 \cdot 10^{-02}$	12.8480
690	1871	$-3.7 \cdot 10^{-03}$	$-1.4 \cdot 10^{-02}$	$-1.860 \cdot 10^{-02}$	$-3.028 \cdot 10^{-02}$	1.6280
968	5611	$-2.9 \cdot 10^{-03}$	$-6.3 \cdot 10^{-03}$	$-9.292 \cdot 10^{-03}$	$-1.104 \cdot 10^{-02}$	1.1885
1036	14433	$-2.7 \cdot 10^{-03}$	$-2.3 \cdot 10^{-03}$	$-5.118 \cdot 10^{-03}$	$-5.441 \cdot 10^{-03}$	1.0630
1044	43979	$-2.7 \cdot 10^{-03}$	$-8.3 \cdot 10^{-04}$	$-3.613 \cdot 10^{-03}$	$-3.588 \cdot 10^{-03}$	0.9932

Similar to Example 1, we compare in Figure 6.4 the fully adaptive refinement with equilibration and uniform refinements with and without equilibration. By local refinement of all involved discretizations we reduce the necessary degrees of freedom to reach a total error of 10^{-2} by a factor of 11 compared to a uniform refinement without equilibration.

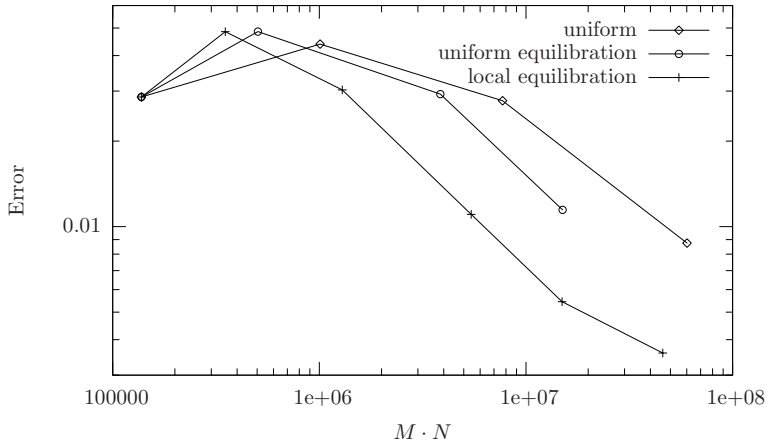

FIG. 6.4. *Example 2: Comparison of Different Refinement Strategies*

Finally, we present in the Figures 6.5 and 6.6 a typical locally refined spatial mesh and a distribution of the time step size obtained by the space-time-adaptive refinement.

REFERENCES

[1] R. BECKER, H. KAPP, AND R. RANNACHER, *Adaptive finite element methods for optimal control of partial differential equations: Basic concepts*, SIAM J. Control Optimization, 39 (2000), pp. 113–132.

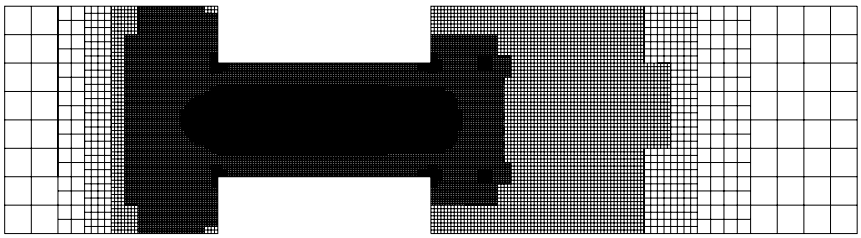

FIG. 6.5. *Example 2: Local Refined Mesh*

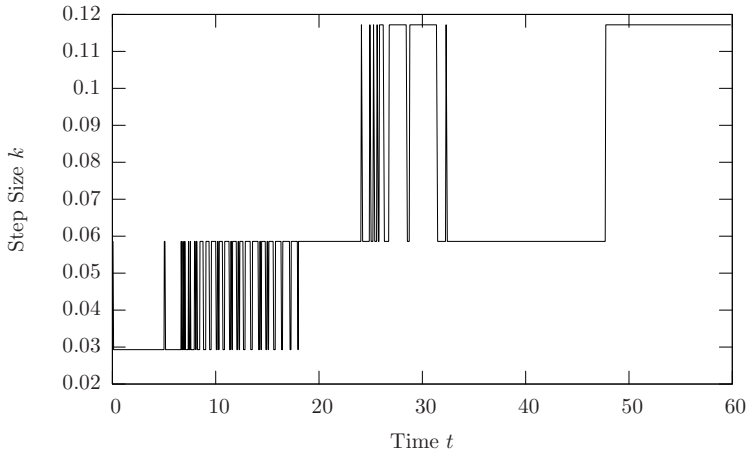

FIG. 6.6. *Example 2: Visualization of the Adaptively Determined Time Step Size k*

[2] R. BECKER, D. MEIDNER, AND B. VEXLER, *Efficient numerical solution of parabolic optimization problems by finite element methods*, submitted, (2005).
[3] R. BECKER AND R. RANNACHER, *An optimal control approach to a-posteriori error estimation*, in Acta Numerica 2001, A. Iserles, ed., Cambridge University Press, 2001, pp. 1–102.
[4] R. BECKER AND B. VEXLER, *A posteriori error estimation for finite element discretizations of parameter identification problems*, Numer. Math., 96 (2004), pp. 435–459.
[5] ———, *Mesh refinement and numerical sensitivity analysis for parameter calibration of partial differential equations*, J. Comp. Physics, 206 (2005), pp. 95–110.
[6] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, vol. 40 of Classics Appl. Math., SIAM, Philadelphia, 2002.
[7] A. CONN, N. GOULD, AND P. TOINT, *Trust-region methods*, SIAM, MPS, Philadelphia, 2000.
[8] R. DAUTRAY AND J.-L. LIONS, *Mathematical Analysis and Numerical Methods for Science and Technology: Evolution Problems I*, vol. 5, Springer-Verlag, Berlin, 1992.
[9] K. ERIKSSON, D. ESTEP, P. HANSBO, AND C. JOHNSON, *Introduction to adaptive methods for differential equations*, in Acta Numerica 1995, A. Iserles, ed., Cambridge University Press, 1995, pp. 105–158.
[10] ———, *Computational differential equations*, Cambridge University Press, Cambridge, 1996.
[11] K. ERIKSSON, C. JOHNSON, AND V. THOMÉE, *Time discretization of parabolic problems by the discontinuous Galerkin method*, RAIRO Modelisation Math. Anal. Numer., 19 (1985), pp. 611–643.
[12] A. V. FURSIKOV, *Optimal Control of Distributed Systems: Theory and Applications*, vol. 187 of Transl. Math. Monogr., AMS, Providence, 1999.
[13] A. GRIEWANK, *Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation*, Optim. Methods Softw., 1 (1992), pp. 35–54.
[14] A. GRIEWANK AND A. WALTHER, *Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation*, ACM Trans. Math. Software, 26 (2000),

pp. 19–45.
[15] M. Hinze, *A variational discretization concept in control constrained optimization: The linear-quadratic case*, Comput. Optim. Appl., 30 (2005), pp. 45–61.
[16] J. Lang, *Adaptive Multilevel Solution of Nonlinear Parabolic PDE Systems. Theory, Algorithm, and Applications*, vol. 16 of Lecture Notes in Earth Sci., Springer-Verlag, Berlin, 1999.
[17] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, vol. 170 of Grundlehren Math. Wiss., Springer-Verlag, Berlin, 1971.
[18] W. Liu, H. Ma, T. Tang, and N. Yan, *A posteriori error estimates for discontinuous galerkin time-stepping method for optimal control problems governed by parabolic equations*, SIAM J. Numer. Anal., 42 (2004), pp. 1032–1061.
[19] W. Liu and N. Yan, *A posteriori error estimates for distributed convex optimal control problems*, Adv. Comput. Math, 15 (2001), pp. 285–309.
[20] ———, *A posteriori error estimates for control problems governed by nonlinear elliptic equations*, Appl. Num. Math., 47 (2003), pp. 173–187.
[21] C. Meyer and A. Rösch, *Superconvergence properties of optimal control problems*, SIAM J. Control Optim., 43 (2004), pp. 970–985.
[22] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, 1999.
[23] M. Picasso, *Anisotropic a posteriori error estimates for an optimal control problem governed by the heat equation*, preprint, Ecole Polytechnique Fédérale de Lausanne, 2005.
[24] R. Rannacher and B. Vexler, *A priori error estimates for the finite element discretization of elliptic parameter identification problems with pointwise measurements*, SIAM J. Control Optim., 44 (2005), pp. 1844–1863.
[25] F. Tröltzsch, *Optimale Steuerung partieller Differentialgleichungen*, Friedr. Vieweg & Sohn Verlag, Wiesbaden, 2005.
[26] R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley/Teubner, New York-Stuttgart, 1996.
[27] B. Vexler, *Adaptive Finite Elements for Parameter Identification Problems*, PhD Thesis, Institut für Angewandte Mathematik, Universität Heidelberg, 2001.